LNCS 5066

Manfred Tscheligi
Marianna Obrist
Artur Lugmayr (Eds.)

Changing Television Environments

6th European Conference, EuroITV 2008
Salzburg, Austria, July 2008
Proceedings

Springer

Lecture Notes in Computer Science 5066

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Manfred Tscheligi Marianna Obrist
Artur Lugmayr (Eds.)

Changing Television Environments

6th European Conference, EuroITV 2008
Salzburg, Austria, July 3 - 4, 2008
Proceedings

Springer

Volume Editors

Manfred Tscheligi
Marianna Obrist
University of Salzburg
ICT&S Center, Sigmund-Haffner-Gasse 18
5020 Salzburg, Austria
E-mail: {manfred.tscheligi; marianna.obrist}@sbg.ac.at

Artur Lugmayr
Tampere University of Technology (TUT)
NAMU Lab., P.O. Box 553/Korkeakoulunkatu 1
33101 Tampere, Finland
E-mail: artur.lugmayr@tut.fi

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.5, H.4, H.3, I.3, I.7, J.5

LNCS Sublibrary: SL 3 – Information Systems and Application,
incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-69477-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-69477-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12323289 06/3180 5 4 3 2 1 0

Preface

Since the time when interactive television emerged as a medium for the home environment, it has been permanently evolving. Changing requirements and user behavior, e.g., the demand for being mobile and have access to information and entertainment anywhere and any time, are challenging interactive TV. New kinds of interactive services have to be conceived for the increasing mobile, ubiquitous requirements of the different user groups. In these changing environments, a better understanding of emerging contexts and their implications is essential. This gave birth to the idea for the theme of the EuroITV 2008 Conference: "Changing Television Environments." EuroITV 2008, the 6th edition of the European Conference on Interactive Television, was organized and hosted by the HCI and Usability Unit, ICT&S Center, University of Salzburg, Austria. The EuroITV Conference Series started at Brighton University in 2003 and 2004. It was followed by Aalborg University in 2005, Athens University of Economics and Business in 2006 and by CWI (Centrum Voor Wiskunde en Informatica) in Amsterdam 2007. We would like to thank all former Chairs for making this great conference series happen and for providing us with the opportunity to host EuroITV 2008.

The annual conference features work on different aspects of interactive television, IPTV, mobile TV, digital content production, entertainment computing, usability and audience research, as well as changes in technical requirements and technologies for future television. EuroITV 2008 covered significant aspects of the interactive television domain including submissions on user studies, technical challenges related to new developments as well as new kinds of formats reaching beyond iTV@Home. Mobile TV, social TV, personalization and recommender systems are of high interest within this research community. Usability and accessibility topics are also gaining an important role. These topics were covered in the 156 submissions received, out of which 5 workshops, 3 tutorials, 6 demos and 5 doctoral consortiums were accepted; 42 papers (full and short) were selected for presentation using a double-blind reviewing process.

A conference like EuroITV is a team effort, and we would like to thank all contributors. Especially, we would like to express our gratitude to the University of Salzburg for enabling this conference, as well as the ICT&S Center for its generous support. We extend our gratitude to the people who devoted time before and during the conference and to our cooperation partners and sponsors for making this conference possible.

July 2008

Manfred Tscheligi
Marianna Obrist
Artur Lugmayr

Organization

The 6th European Conference on Interactive TV, EuroITV 2008, was organized by the HCI and Usability Unit, ICT&S Center, University of Salzburg, Austria.

Organizing Committee

General Chair	Manfred Tscheligi (University of Salzburg, Austria)
Conference Chairs	Marianna Obrist (University of Salzburg, Austria)
	Artur Lugmayr (Tampere University of Technology, Finland)
Short Papers and Posters	Regina Bernhaupt (University of Salzburg, Austria)
	David Geerts (K.U.Leuven, Belgium)
Workshops	Pablo Cesar (CWI, The Netherlands)
	Hendrik Knoche (University College London, UK)
Tutorials	Judith Masthoff (University of Aberdeen, UK)
Demonstrations	Erika Reponen (Nokia Research Center, Tampere, Finland)
	Thomas Mirlacher (University of Salzburg, Austria)
Doctoral Consortium	Lyn Pemberton (University of Brighton, UK)
Industrial Case Studies	Sander A.J.P. Limonard (TNO ICT, The Netherlands)
	Owen Daly-Jones (Serco Usability Services, UK)
	Gunnar Harboe (Motorola Labs, USA)
Industrial Liaison	Michael Darnell (Microsoft TV, USA)
	Bart Hemmeryckx-Deleersnijder (Alcatel-Lucent, Belgium)
Local Organization	Elke Beck (University of Salzburg, Austria)
	Alexander Meschtscherjakov (University of Salzburg, Austria)
Local Arrangements	Martin Altmanninger
	Axel Baumgartner
	Hermann Huber
	Sara Kepplinger
	Michael Lankes
	Philipp Losbichler
	Christiane Moser
	Martin Murer
	Michael Pirker
	Wolfgang Reitberger
	Stefan Riegler

Local Organization	Daniel Schweiger Astrid Weiss David Wilfinger Daniela Wurhofer

Program Committee

S. Agamanolis, Distance Lab, UK
D. Alliez, NDS Technologies, France
L. Aroyo, Free University, The Netherlands
L. Barkhuus, University of California, USA
G. Barros, University of Sao Paulo, Brazil
A. Benjelloun Touimi, Orange FT group, France
A. Berglund, Linkoping University, Sweden
R. Bernhaupt, University of Salzburg, Austria
M. Bove, MIT Media Lab, USA
P. B. Brandtzæg, SINTEF ICT, Norway
V. Bruns, Tampere University of Technology, Finland
D. Bulterman, CWI, The Netherlands
B. Bushoff, Sagasnet, Germany
P.S. Cesar, CWI, The Netherlands
K. Chorianopoulos, Ionian University, Greece
G. Harboe, Motorola Labs, USA
C. Herrero, HUT, Finland
L. Eronen, Finland, Independent Researcher, Finland
D. Geerts, Mediacentrum, K.U.Leuven, Belgium
S. Goldenberg, Georgia Tech, USA
L. F. Gomes Soares, PUC-Rio, Brazil
G. Gonzalez-Sanchez, Mediapro R+D, Spain
R. Goularte, Sao Paulo University, Brazil
M. Hausenblas, Joanneum Research ForschungsGmbH, Austria
J. Jensen, Aalborg University, Denmark
S. Jumisko-Pyykko, Tampere University of Technology, Finland
J. Kallenbach, Helsinki University of Technology, Finland
I. Kegel, BT Group plc, UK
H. Knoche, University College London, UK
C. Klimmt, Hanover University of Music and Drama, Germany
G. Lekakos, Athens University of Economics and Business, Greece
M. Lopez-Nores, University of Vigo, Spain
A. Lugmayr, Tampere University of Technology, Finland
E. Mantzari, Athens University of Economics and Business, Greece
J. Masthoff, University of Aberdeen, UK
T. Mirlacher, University of Salzburg, Austria
M. Obrist, University of Salzburg, Austria
J.J. Pazos-Arias, University of Vigo, Spain

L. Pemberton, University of Brighton, UK
C. Peng, VTT, Finland
T. Rasmussen, Aalborg University, Denmark
M. Rauterberg, Technical University Eindhoven, The Netherlands
M. Rice, University of Dundee, UK
S. Reymann, Tampere University of Technology, Finland
M.M. Saleemi, Abo Akademi University, Finland
W. Seager, University College London, UK
D. Sloo, Microsoft Corporation, USA
C. de Salles Soares Neto, PUC-Rio, Brazil
M. Springett, Middlesex University, UK
T. Strandvall, Tobii Technology, Schweden
M. Tscheligi, University of Salzburg, Austria
A. Vrechopoulos, Athens University of Economics and Business, Greece
P. Vuorimaa, Helsinki University of Technology, Finland
A. Weiss, University of Salzburg, Austria
D. Wilfinger, University of Salzburg, Austria
Z. Yu, Kyoto University, Japan

Conference Co-organized by

IFIP, International Federation for Information Processing

Conference Sponsored by:

CITIZEN MEDIA
Telekom Austria
Nokia
Microsoft Mediaroom
Ruwido
CableLink by Salzburg AG

Conference in Cooperation with:

ICT&S Center, Advanced Studies and Research in Information and Communication Technologies and Society, University of Salzburg University of Salzburg
ACM, Association for Computing Machinery
ACM SIGCHI, ACM Special Interest Group on Computer-Human Interaction
ACM SIGMM, ACM Special Interest Group on Multimedia
ACM SIGWEB, ACM Special Interest Group on Hypertext, Hypermedia and Web
NAMU, New AMbient MUltimedia research group, Tampere, Finland
EuroITV, European Interactive Television (http://www.euroitv.org/)
UITV.INFO, Understanding Interactive Television (http://uitv.info)

Table of Contents

Interactive TV

Interactive Authoring

Personalisation and Recommender Systems

New TV Environments

iTV Architectures and Systems

User Interfaces and Interaction Design

User Studies

Accessibility

Interactive Television – A Brief Media History

Jens F. Jensen

Center for Applied Experience Economy, Aalborg University
Department of Communication, 9220 Aalborg East, Denmark
jensf@vrmedialab.dk

Abstract. The paper is an attempt at tracing the history of interactive television. The lead that is followed is the various experimental attempts with and the commercial launches of interactive television that have been conducted, as well as the results that have been achieved. This will be done by identifying the various historical phases with different characteristic technological waves, business models, services and contents concepts.

Keywords: Interactive Television, interactive media, media history.

1 Introduction: Holy Grail or Vapourware?

Gathered from the information in the contemporary press, the electronic media and even scientific journals one is easily left with the impression that interactive television is a new phenomenon, or maybe even that it belongs in the future [1, 2]. In reality, however, interactive television has a long history – actually, just as long as television itself. Back in the 1920s, when television technology was invented, interactive communication was – here in the shape of one-way video and two-way audio – actually among the formats that was seriously considered and tested.

However, this was not to be the dominant form that the television media assumed when it was eventually presented and generally accepted. Quite the opposite, television more or less became the model example for a 'push' media and one-way mass communication. Nevertheless, the idea of making television interactive has emerged time and time again throughout the development of the media.

Most often, however, this other side of the history of the television media – the flipside of the mass media and one-way communication – is overlooked or forgotten. That this historical perspective is missing is – as, e.g., John Carey draws attention to [1, 2] – unfortunate since quite an amount of knowledge, experiences and points are gathered within this part of the media's history. Knowledge which relates to technological problems and problem solutions, business models that worked or did not work, consumers' preferences, demands and reactions, etc., which are highly useful in the current development of new forms of interactive media.

During the last five decades, there have been numerous attempts with both experimental and commercial testing of interactive television concepts. However, different problems – unripe technologies, underdeveloped infrastructure, lack of contents, failing demand – have for the most part hindered an actual breakthrough. In a sense interactive television has been under way for the last 50 years.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 1–10, 2008.

At the same time, interactive television is one of the most hyped technologies of that period. Repeatedly, it has been announced that we now faced a decisive technological and commercial breakthrough for interactive television. Every other year, the birth of the technology has been proclaimed. More or less every time, however, the matter transpired to be a false alarm. Presumably, no other technology has been characterised by so many ’false starts’, so many hyped beginnings, and so many stumbling launches. Seen in this perspective, interactive television must be one of the last half centuries most failure ridden technologies. Unrivalled, it must take the price for being the ultimate ’vapourware’ of the last 50 years – a designation, which in the computer business refers to soft- or hardware that is announced and advertised for long before their launching – and which often are never realised.

There have been numerous attempts at finding the ‘killer application’, which by itself would drive the spreading of the technology. This has happened to the extent that interactive television has been named the Holy Grail of the Information Age by some commentators. For in the same manner as the medieval knights, according to the King Arthur legends, went forth to find the Holy Grail – the chalice of Christ – the most priced of all objects, which would grant eternal life and happiness to its finder; so the ‘knights’ of the media industry goes on a crusade to track down the never yet found nor seen wonder product. A product which is expected to catapult the media industry into new and profitable arenas.

Just as the birth of interactive television has been declared repeatedly, so has its death. Countless times the more sceptical part of the observers has passed the death. For example, in the mid-90s, E. Schwartz [3] wrote off the entire idea of interactive television in an articled in *Wired*, in which the title rhetorically inquired: “People Are Supposed to Pay for This Stuff?” At the end of the 90s, the *New York Times* proclaimed the death of interactive television. And as late as June 2004, A. Rise [4] wrote a commentary in *adage* entitled: “Why Interactive Television has no Future”. However, the rapports of the death of interactive television have been – to paraphrase Mark Twain – greatly exaggerated. The notion of adding interactive elements to the television media continues to stir.

This paper is an attempt at tracing the history of interactive television. The lead that is to be followed is the various experimental attempts with and the commercial launches of interactive television that have been conducted, as well as the results that have been achieved. This will be done by identifying the various historical phases with different characteristic technological waves, business models, services and contents concepts, primarily with point of departure in Europe and the United States.

Within the framework of this paper, ‘interactive television’ will be understood in the broad sense as the merger between conventional television and new interactive information and communication technologies. More specifically, interactive television is a form of television that also bases itself on actual physical interaction with the media in the form of choices, decisions, and communicative input. In this manner it becomes possible for the viewer to gain control over what is seen, when it is seen, and how it is seen, or there is opened up for the actual possibility for active participation in programmes or upload of content generated by the user [5, 6, 7, 8, 9].

2 Phase 1: Video Telephony in the 1950s and 60s

The actual beginning for the modern era of interactive television is often linked with the telephone companies' special version of interactive television: video telephony [2]. In the United States, Bell Telephone Laboratories began to experiment with sending and receiving pictures via the telephone lines as early as the 1920s; in 1956, the first prototype of a 'PicturePhone' was completed; and at the 1964 World's Fair in New York, the first model, named 'Mod 1', was introduced to the audience. Visitors at Bell Labs booth were here invited to call a similar booth in Disneyland. With this technology, which was already at that time described as 'crossing a telephone with a TV set' in advertisements, one could not only hear but also see the person with which one was telephoning. During the following years, 'The PicturePhone' was tested in cooperation with AT&T in different market tests and limited services. And by the end of the 60s, it was launched as an actual product, primarily aimed at the business world. At the launch of the commercial PicturePhone service in Pittsburgh in 1970, leading personnel at AT&T predicted that the PicturePhone's spread would exceed one million by 1980.

However, the PicturePhone never really caught on. There were several reasons for this. A. Moyers writes: "Survey results showed that most people did not like the PicturePhone. The keyboard was clumsy, and the picture was small. Moreover, most people felt uncomfortable at the idea of being seen during a telephone conversation" [10]. In addition, the picture quality was poor, the price or the service was too expensive, and the user could only communicate with the network of persons who already had a PicturePhone, which for obvious reasons had to be a limited amount. Actually, the need and demand for being able to see other persons turned out to be limited in most situations where the device was tested [2]. In 1973, AT&T took the Picture Phone off the market after having invested an estimated 130 to 500 million US dollars in the projected [11].

The picture telephone is thus another example of a technology which has experienced several 'false starts'. In principle, it has been technologically possible to produce and sell picture telephones in the last four decades. But it has never been possible to make the technology catch on. We have to go to this side of the millennium before video telephony is – albeit once again slowly – spread in connection with 3G mobile telephony.

3 Phase 2: Analogue ITV in the Late 70s and Early 80s

The first round of testing of actual television based interactive television took place in the latter half of the 70s; a period that was relatively rich in tests involving interactive television. These tests were not based on digital technologies but were all analogue. Nevertheless, they, to varying degrees, did have interactive elements.

In 1977 Warner-Amex, now called Time Warner, launched the QUBE system in Columbus; the first commercial interactive service on a larger scale. In connection with the history of interactive television, this system has reached a special status since it often is sited as an example of early launch of interactive television which failed.

QUBE was a cable television system based on 30 analogue television channels distributed among ten broadcast channels, ten pay-per-view channels, and ten channels with original, interactive services. In addition to these downstream channels, the system was equipped with a narrow-band upstream return channel, which was used by the interactive services. The QUBE clients were equipped with a decoder or set-top box with five knobs. Via this they could participate in game shows, choose sports events, order pay-tv, participate in opinion polls and voting, etc. The viewers pushed the knobs on the box, the selections were processed by a computer, and later the result was announced on the screen. From the beginning, the press coverage of the system was enormous, and many households subscribed to the service.

However, the actual use of the interactive elements was generally low with the exception of, for instance, certain gaming formats which attracted some viewers and produced an intense interactive participation. In addition to that, certain major events generated some viewer participation, as for instance when subscribers were asked to voice their opinions on a speech just given by President Carter.

In the long run, the system turned out to be expensive in terms of maintenance: the technology was expensive for the consumers as well as for the cable operator; and it was expensive to produce the interactive programmes and to maintain the return channel. "It caused problems for the interactive service as regards operational security", writes Carey and continues: "The budget for QUBE-programmes was very low compared to budgets for broadcast networks. 'Interactivity' with low production costs could not compete with the network programmes. In addition to this, those who produced the programmes had to begin from scratch: there were very few previous experiences with designing programmes" [2].

Thus, the attempt to turn the service into a commercially viable business never succeeded. And in 1984 the system was quietly closed down. One of the most direct consequences of the QUBE experiment was that the testing of the interactive formats led to the development of new program components, which later became the basis for the creation of renowned media phenomena such as MTV, Nickelodeon, QVC and The Movie Channel.

4 Phase 3: The Interactive Revolution in the 1980s

Generally, the 1980s were characterised by the breakthrough for and the spread of a broad spectrum of new interactive media and technologies that invaded the domestic setting, the workplaces and places of education and which offered a greater degree of control over the use of media: video cassette recorders, gaming consoles, videogames, personal computers, cash dispensers, microwave ovens, information stalls in public places and so on. This breakthrough had a scope and a depth that it actually makes sense to talk about a sort of 'interactive turn' in the media culture. A turn in the direction of interactive technologies and services which generally trained the consumers to interact with technological consumer products, delivered new forms of interaction with machines and contents, and generally opened up for the user to have more control over the media experience.

Within the area of interactive television, the era was generally characterised by a scaling-down of the very ambitious interactive television services to more simple technological solutions in the direction of interactive texts on television and opinion polls during television programmes via special telephone services.

Within the area of cable, Cox Cable launched the videotext service Indax in Omaha in the early 80s. A service which ran over a two-way cable and offered home banking, shopping, information services, and educational contents, solely based on text and simple graphics. And correspondingly, Time Inc. launched the service Time Teletext in Orlando and San Diego. A one-way teletext service which, however, did use an entire cable channel and hence could transmit data in an extent and at a speed that made it possible to simulate interactivity in connection with, for instance, games and quizzes. Both Indax and Time Teletext were withdrawn upon the end of the test period. Even though users stated that they found Time Teletext attractive as a service, the expenses in connection with the technology were too high for it to be established as a viable business model [2].

The telephone companies also experimented with videotext-services. Knight Ridders offered the service Viewtron, and Times Mirror launched the service Gateway. Both were based on a box which could connect the telephone line to the television set, enabling the contents to be seen on the television screen. Also in this case, however, the price for the box as well as the subscription was too high to create an adequate customer base, and both services had to be withdrawn. However, surveys did indicate that the consumers were positive minded towards certain types of services and features, among these especially games, several different forms of communication, frequent updates of information, and contents control [2].

In addition to this, telephone companies such as AT&T, also developed services in this period which made it possible for viewers to call a special number and vote or participate in opinion polls during a broadcast. The results from the poll or survey could then subsequently be displayed on the television screen. This technology – where the regular telephone line was used as the return channel for interactivity in connection with television programmes – was used by many broadcasting companies from the late 80s and onwards.

5 Phase 4: Comprehensive Experiments with ITV in the Early 90s

Beginning in the late 80s, but especially pronounced in the 90s, the telephone companies and cable operators began a complex competition and cooperation strategy in an attempt to define the future of ITV. In the period up until 1994, most of these initiatives, however, were based on a limited form of ITV or offered a limited amount of services.

One of the early, relatively broadly focused experiments with programmes and use of ITV was carried out by AT&T in Chicago in the beginning of the 90s. The test was based on a group of 140 employees and spanned a period of two years. Even if the test population was not representative, since it was chosen from employees at AT&T, the conclusions that were derived from the experiment were in many ways interesting and informative; partly because the studies that were carried out in relation to the test were relatively comprehensive and based on scientific methods, and partly because they in many ways accumulate the results from many similar tests.

Among other things, the studies showed that the test persons' reactions were relatively positive and that the main interest was focussed on educational programmes for children, sports programmes, and games, in which the households could compete amongst each other [1]. It was also concluded that in order for a programme to become popular it had to feature "four qualities: entertainment, transaction, information, and communication. In other words, people want to have fun, have something, learn something, and tell somebody about it" [11]. On the other side, it was of less importance "what the service was about (game, storytelling, information and so on) than how it was offered and what it allows people to do" [11]. Here, as in many other places, it was concluded that "the service must be simple to use for the consumer. They want television, not computers, and they do not want to use anything that implies even a hint about complex pc-operations" [11]. Finally, the lesson learned was that: "attractive services are heavily dependant on varied contents". [11]

However, the far majority of the ITV technologies that was introduced and tested up until 1994 were comparatively limited in relation to the services and experiments which were announced and launched in the mid-90s. This period in the mid-90s can in many ways be regarded as if not the golden age of interactive television then, at any rate, the gold rush of interactive television.

The without a doubt most renowned of these test sites for interactive television in the mid-90s was Time Warner's so-called Full Service Network (FSN) in Orlando. The name Full Service Network was meant to indicate that the system covered the full spectrum of interactive services, including: program guides; video-on-demand; music-on-demand; news; shopping; classified ads; games; t-learning; t-banking; health services; ticketing; transactions with public institutions; municipal authorities and library; terrestrial and wireless telephone services; high speed, two-way communication facilities for exchange of high resolution video and graphics within the business sector, hospitals, schools and so on.

FSN was begun in 1993 and was planned to commence in 1994. However, due to a series of technical difficulties in connection with the digital video servers, software, set-top boxes etc., the launch was postponed to 1995. More than 4.000 homes had access to the service via fibre network [12]. However, FSN did not fulfil the expectations, neither for the users nor for the suppliers. And after a test period of only two years, FSN was closed down in 1997 after – what the press described as – "a cash drain and a technological nightmare" for Time Warner [13]. Swedlow [12] mentions figures of up to 100 million dollars.

Swedlow is, however, relatively positive in his assessment of the experiment as he quotes Levy, an executive staff member on the project: ""It was far too expensive, but we knew that when we began", he [Levy] says. Contrary to the public opinion of the project, "we knew that FSN would only be taken into use much later. It was not a waste of energy: we learned a lot". Levy also points out that a few things gathered from the experience were invaluable: These were: 1) the service itself must be available to the consumers free of charge; 2) different gradual price models do not work; 3) VOD is a very popular use of the technology; 4) people want really simple interactive possibilities" [12].

Just like FSN, many of the mid-90s experiments were either cancelled or reduced in relation to the original plans. Thus, by the last half of the 90s, much had changed.

To build advanced ITV systems had turned out to be very difficult, just as it had turned out to be difficult to develop the services and the contents that made sense for the consumer in everyday life.

6 Phase 5: Convergence of Television and Internet in the Late-90s

Seen from a mid- and late-90s perspective, there seems to be three (information) highways that interactive television could travel down. The first was – what could be called – the B-ISDN/Full Service Network, as mentioned above. The second was digital broadcasting. Finally, the third was of course the internet.

ITV and the story of the internet are in many ways diametrically opposed. As the internet came into existence from the end of the 60s to the mid-90s, it more or less happened by cheer chance. From the beginning, no one had planned the internet as it appears now, no media concerns operated with massive investments and development projects in order to further it, and no one believed in it as a medium and commercial possibility. Especially after the launch of the World Wide Web and the introduction of the graphic browser in 1993, the internet grew at an incredible rate. By the latter half of the 90s, the internet and WWW reached growth rates that surpass all earlier known media technologies.

If ITV can be seen as the greatest technological failure of the last fifty years, then the internet is the foremost success of the 90s – maybe even of the entire twentieth century. And in many ways, these developments are more or less connected, since the inert development of ITV and repeated setbacks during the 90s to some extent can be seen as an expression of the fact that the technology, as regards contents development and contents supply was overtaken by the internet media.

An obvious strategy for ITV was therefore to follow in the footsteps of the internet and profit from some of the headway produced by the internet's incredible drive. During this phase, a prominent concept for interactive television was therefore the convergence of broadcast and internet. M. Krantz's article from 1997 constitutes a significant sign of the times. Under the heading: "Marriage of convenience", he wrote: "interactive television, once a mighty idea, which has resulted in a long line of failures, is back … a mixture of computers, television, and World Wide Web is the recipe for an actual success" [14].

Internet on television or television based internet access should make it possible for the users to carry out many activities via the television set, which are normally carried out on a personal computer connected to the internet, including reading and writing e-mails, participating in chat and discussions groups, searching the internet by key-words or category etc.

In the late 90s, to grant access to popular web pages via the television set was seen by several television service providers as a service that would be very attractive for the consumers. The idea was as follows: If we combine the universal market spread of television and the anarchistic multimedia contents of the web, we have the ultimate killer application. Or, put differently: It is the internet itself that is the killer application. And all that is required is to bring the Web to the 'non-connected' majority of the public who are situated where they already sit and wait – on the couch.

7 Phase 6: Enhanced TV, Personalised Television, and SMS-TV at the Turn of the Millennium and Beyond

A part of the outlined tendencies within ITV during the earlier phases lives on in the new millennium. Particularly, this is the case in connection with DVB. Other tendencies within ITV seem to lose their momentum and succumb to inertia. This is the case, among other things, in connection with the combination of internet and television or internet access via the television set.

Moreover, at the turn of the millennium, several new tendencies surfaced, which seems to revive the interest in ITV. Primarily, it is centred on the following three trends: Enhanced TV, personalised television, and SMS-TV. In connection with all three, the case is one of relatively low-tech solutions and downscaled models for interactivity in connection with television, which focuses on a specific and delimited aspect as opposed to opting for the advanced or full interactivity and the entire spectrum of services.

Enhanced TV refers to any type of contents – mostly, however, text and graphics – which is superimposed, that is, placed on top of the actual video contents, and which may be accessed interactively by the user. Mostly, the extra enriched content is sent continuously via the spare capacity in the actual broadcast signal, where the interactivity consists of the viewer selecting between the information that the operator has embedded in the signal. Seen in this perspective, Enhanced TV can be perceived as an improved teletext service, a sort of super teletext or, alternatively, as a technology that can add web contents to television broadcast. Typically, the user browses the desired information – mostly displayed on the screen together with the actual video stream of the programme. Generally, enhanced TV is thus characterised by taking its point of departure in existing television formats and hence also the traditional qualities of broadcast television in an attempt to further develop and improve these qualities [15].

Another new group of interactive applications, devices or media that have surfaced is personalised television. Personalised TV – in some cases also called individualised TV or customised TV – takes on several different forms. In its most widespread version, personalised television is linked with a piece of hardware in the shape of a Personal Video Recorder (PVR) or systems with equivalent PVR functions. In principle, PVR functionality means that the user achieves the same control over the broadcast stream which earlier was tied to the video recorder. That is, the possibility to pause, rewind, fast-forward, slow motion, frame advance and so on. In this manner, the user can perform time shifts in relation to the broadcast stream. PVR functionality also includes programming facilities for automatic recording of programmes according to title, time slot, actor, theme, keyword, rating, and so on, where the PVR may even adapt itself to possible changes in the broadcasting stations' programming. Hence, a Personal Video Recorder generally makes it possible for the viewer to see what they want, when they want it.

The third new, dominant form of interactive television is cross media interaction. Since most localities still do no have advanced interactive two-way systems, and since most users still do not own advanced, digital set-top boxes let alone Personal Video Recorders, by the end of the 90s and especially at the other side of the millennium, it has become increasingly normal to establish different forms of 'two-channel' interaction in order to produce interactive television programmes or interactive moments in television programmes. This means that another media steps in as a 'return channel' from the television viewer to the programme broadcaster, for instance, the telephone, e-mail, web chat, fax, SMS, and MMS. Compared to fully integrated ITV systems, the obvious advantage of this form of cross media interaction is self-evidently that the so-called 'terminal barrier' has already been pulled down. Hence, cross media interaction does not presuppose any major investments in hardware and software neither as regard the provider of the service, the distributor or the viewer. Among the different cross media formats, it was in particular the SMS services via the mobile telephone as a return channel to live television programmes that reached a great degree of proliferation and popularity during the period. This form of television has already been termed: SMS TV. In a research report that characteristically bore the title: *SMS TV: Interactive Television Reinvented* [17] van Dusseldorp indicates that one of the most unexpected development within the area of ITV at this time was exactly that the television broadcasters threw themselves on SMS text messages as a new return channel by which to enrich their programmes with interactive features. The most popular SMS-TV formats have been polls, games, and chat in the form of SMS messages sent in by viewers.

A common feature for these tendencies in recent years is that we are dealing with relatively written down and low-tech experiments that are based on existing media and technologies. They are – as Tan Ee Sze has formulated it in another context – "not quite the holy grail of two-way interactive television" [18]. Rather, they express what, in some contexts, has been called an 'evolutionary' approach to interactive television, as opposed to the 'revolutionary' approach, which perhaps primarily characterised the mid-90s. A less ambitious strategy that develops simple services based on existing, tested and thus relatively cheap technologies.

8 A Happy Ending to the Story?

As is appears from the above, interactive television has a long and rich history behind it with numerous changing phases characterised by different dominant strategies, technological forms, and types of content and services. The majority of these versions of interactive television, however, failed or only had a limited amount of success, and the far majority were taken off the market again fairly quickly. Thus, interactive television has been arriving for the past 50 years.

Overall, the history of interactive television as a novel technology, as a novel media or a novel service is, hence, the history of a failure. A history of a long list of 'false starts' of a media technology that yet could not manage to succeed in the market because: it was not technologically mature; it was not supported by viable business models; the context of other concurrent, competing technologies and supply were unfavourable; content and services did not have enough added value in relation to competing content and services to drive the demand forward, etc. Hence, after five decades with experiments with interactive television and commercial roll-outs, there still is no sure model for establishing interactivity in connection with television viewing, let alone any sure knowledge on what the user demand is.

In spite of this all but glorious past history, there is still much that now indicates that interactive television as well as digital television will enter into a growth phase in the years to come. According to projections from, among others, Forrester Research, the proliferation of digital television in Europe will increase by 50 percent in 2009.

So, to the question of whether interactive television has the status of either a ’holy grail’ or ‘vapourware’, the straight answer probably has to be: neither one nor the other. On the one hand, there is nothing that indicates that suddenly it will be managed to find the irresistible, magical killer application that can power a swift mass proliferation of the technology. Interactive television is, on the other hand, already more than just an ephemeral figment of the imagination – it is already out there in multiple variants and early formats such as SMS-TV, programmes with web communities, and actual digital and interactive television platforms and services.

References

1. Carey, J.: The Interactive Television Puzzle (no longer available on the internet) (1994)
2. Carey, J.: Winky Dink to Stargazer: Five Decades of Interactive Television. In: I-TV 1996, Edinburgh (1996)
3. Schwartz, E.: People Are Supposed to Pay for This Stuff? Wired, 3.07 (1995)
4. Rise, A.: Why Interactive Television has no Future, adage (2004)
5. Jensen, J.: ’Interactivity’ – tracking a new concept. In: Mayer, P. (ed.) Computer Media and Communication. Oxford University Press, Oxford (1999)
6. Jensen, J.: The Concept of ‘Interactivity’ in ‘Interactive Television’ and ‘Interactive Media’. In: Jensen, Toscan, G. (eds.) Interactive Television, TV of the Future or the Future of TV? Aalborg University Press, Aalborg (1999)
7. Jensen, J.: Interactive Content, Services and Applications. In: Brown, A., Picard, R.G. (eds.) Digital Terrestrial Television in Europe. Erlbaum, Mahwah, NJ (2004)
8. Jensen, J.F.: Interactive Television: New Genres, New Formats, New Content. In: Australasian Conference on Interactive Entertainment. Sydney (2005)
9. Jensen, J., Toscan, C. (eds.): Interactive Television. TV of the Future or the Future of TV? Aalborg University Press, Aalborg (1999)
10. Moyers, A.: Air Force Communication Agency Office of History (1997), `http://www.bellsystemmemorial.com/telephones-picturephone.html`
11. Van Tassel, J.M.: Advanced Television Systems. Focal Press, Boston (1996)
12. Swedlow, T.: Interactive Enhanced Television: A Historical and Critical Perspective (2000), `http://www.itvt.com/etvwhitepaper.html`
13. Latta, J.: The birth of interactive TV. In: Multimedia monitor (1995)
14. Krantz, M.: Marriage of Convenience, in Time, November 10 (1997)
15. Weapon7: iTV - A View from the Trenches (2002), `http://www.broadbandbananas.com/download.html`
16. Macklin, B.: What Every Marketer Needs to Know about iTV, eMarketer (2002)
17. Van Dusseldorp: SMS TV: Interactive Television Reinvented, van Dusseldorp & Partners (2002)
18. Sze, T.E.: Looking for Winky Dink. In: Computerworld (2002)

Interactive Narratives: Exploring the Links between Empathy, Interactivity and Structure

Stacey Hand and Duane Varan

Interactive Television Research Institute (ITRI), Murdoch University, Western Australia
s.hand@murdoch.edu.au

Abstract. Interactive narratives have long been advocated as having the potential to create more immersive and transformative experiences for audiences by adding the pleasure of agency. In practice, however, finding the balance between sufficient interactivity for agency and narrative structure has been difficult. This paper proposed a model of interactive narrative, which encased interactivity within a strong narrative structure, as the best model to increase audience entertainment. In order to test the hypothesis that such a model would out perform its linear counterpart in terms of audience entertainment- two separate audience studies were conducted. The first study tested an interactive drama for television and the second study tested two interactive narrative based advertisements. When anaylsing the two studies there was a surprising result- in each case audience empathy was consistently and significantly increased by the addition of interactivity. This discovery suggests links between empathy and interactivity, and has exciting implications about the role of empathic stress and structure in the creation of future models of interactive narratives.

1 Literature Review

Interactive narratives have long been advocated by those excited by their potential to deliver more personalized, transformative and engaging experiences for audiences. "Just as storytellers of old made what, when and how much they told, conditional upon the audience's 'good listening', in order to heighten the tension, entertainment level and mental participation; so now interactive stories solicit audience participation in the story and thus help them to more intensely internalize the material." [23: P.177] The ongoing process of television digitization, and its resultant new interactive capabilities, has sparked a review of interactive narratives. This renewed interest is further piqued by changes in audience preference, particularly among youth audiences, for more interactive and personalized forms of communication. [19]

1.1 Interactive Narratives: Three Schools of Thought

The concept of interactive narratives is a contentious one and thus there are varying opinions on its form and on its effectiveness. This section outlines three main schools of thought on interactive narratives.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 11–19, 2008.

1.1.1 Interactive Narratives Don't Work

While there has been much hype surrounding the possibilities of interactive narratives, in reality success with this form has been rare. While there has been a litany of one-off interactive drama projects, the genre has yet to fulfill its promise. This lack of practical success has led many to argue that interactive narratives are simply an unworkable concept. "As a theory, this narrativistic colonialism might seem aesthetically problematic (Aarseth 1997, chapter six), as well as technologically unachievable (Bringsjord 2001)…" [1: P.49] In this way interactive narratives have become a much maligned and oft dismissed concept.

1.1.2 Interactive Narratives Only Work for Certain People

– Vorderer, Knoblach and Schramm [24] suggest that interactive narratives are not unworkable but are only beneficial, in terms of entertainment, to audiences with higher cognitive capacities. Vorderer, Knoblach and Schramm [24] conducted an audience research experiment which tested three versions of an interactive movie-one where there were multiple opportunities of interaction, one with one opportunity for interaction and a final linear version. Voderer, Knoblach & Schramm [24] tested the audience to determine interactivity's effect on suspense, empathy and entertainment. The conclusions of this study suggest that interactivity only heightened empathy, suspense and entertainment for those participants with higher cognitive capacities and that the opposite was in fact the case for those with lower cognitive capacities.

1.1.3 Interactive Narratives Must Encase Interactivity within Narrative Structure

Research at the Interactive Television Research Institute (ITRI) of Murdoch University, Western Australia has also found interactive narratives to be a workable concept-but only under the condition that the interactivity is encased within a narrative structure. [8]

Much of the literature in the field suggests that the main problem behind creating successful interactive narratives lies in achieving a balance between interactivity and narrative structure. "In fact, achieving the ideal relationship between a compelling story and meaningful interaction is the single most critical make-or-break feature of all interactive fiction applications." [11: P.38]

Hand & Varan [8] hypothesized that a model of interactive narrative which encased interactivity within a strong narrative structure would best heighten audience enjoyment of a narrative.

Specifically Hand & Varan [8] advocated a 'yo-yo' model of interactive narrative which provides a branching form of interactive narrative allowing the audience to interact but which always uses bridging storylines so as to bring the audience back to the central storyline. Similar structures of interactive narrative have also been outlined by Meadows [12] as the 'nodal plot structure', by Garrand [7] as the 'branching from scenes' structure and by Lugmayr, Niiranen & Kalli [10] as the 'conditional branching: bottlenecking" structure. The main difference between these structures and the 'yo-yo' structure is the identification, in the 'yo-yo' structure, of the bridging scene.

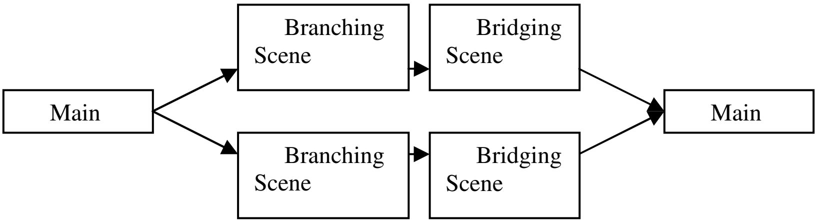


Fig. 1. 'Yo-yo' model of interactive narrative

1.2 Exploring the Interactive Audience: Why does Interactivity Work?

There is very little published audience research in the area of interactive narratives and thus the effect of interactivity in terms of audience entertainment is largely unknown. As Hand & Varan [8] demonstrate, interactive narratives can outperform linear narratives. But perhaps the more pertinent question is *why* interactivity affects the audience. If we can establish the key ways in which interactivity can enhance audience entertainment then it will be easier to develop programmes which best foster these interactive advantages.

1.2.1 Exploring the Possibilities of Interactivity's Effect on Audience Entertainment

In Murray's seminal work on interactive narratives entitled "Hamlet on the Holodeck" [15], Murray hypothesised that interactivity could add to audience entertainment in three main ways: through a deepening of audience immersion, transformation, and through adding a sense of agency. These three 'aesthetics of the medium' have received the most attention in terms of interactive narratives. And each shall be explored briefly below

1.2.2 Transformation

Narrative fiction most often involves the transformation of the protagonist through a series of story events. Arguably the most intriguing aspect of narrative fiction is its ability to transform not only the characters within the story world, but it ability to in turn transform an audience's real world beliefs [17; 5]

Murray [15] argues that interactivity deepens the transformative experience of audiences in terms of fiction narratives. By interacting, the audience enacts rather than witnesses the story and in this way the audience more deeply internalises and personalizes the story events in as it is they who choose to enact certain story events and thus the consequences of those events are felt more deeply. "Enacted events have a transformative power that exceeds both narrated and conventionally dramatized events because we assimilate them as personal experiences" [15: P.170]. Thus interactivity has the potential to create more transformative experiences for audiences.

1.2.3 Immersion

Immersion is an experience which has been described by several researchers under different labels, thus under the heading of immersion this section will also explore literature on telepresence, transportation and absorption.

Immersion describes the audience's feeling of being submerged in the story world of the narrative, the narrative reality thus surrounds the audience member and they are wholly focused upon it. [15]

Immersion, or transportation, has been particularly associated with interactive media. "Interactive media may be particularly transporting, and thus particularly enjoyable, forms of entertainment because they allow users to easily leave their physical and psychological realities behind and become fully immersed as an active participant in the narrative of an alternative, 'virtual' reality" [4].

1.2.4 Agency

Agency is highlighted by Murray as perhaps the most vital of her three 'aesthetics of the medium'. Murray defines agency as "…the satisfying power to take meaningful action and see the results of our decisions and choices" [15: P.126]. Thus in order to give audiences a sense of agency, an interactive narrative must allow for interactivity which is meaningful in terms of the narrative and, in turn, the narrative must acknowledge the audience's interaction by showing change.

The provision for audience agency in interactive narratives has led to some debate. Some authors argue that unless audiences can affect the plot by interacting, agency cannot be achieved [9: P.21]. The notion that interactivity must affect story plot is a point of contention however as Murray herself suggests that agency can arise from interactivity that does not affect the narrative at plot level. [16: P.10]

1.2.5 Empathy

Empathy is not one of the 'aesthetics of the medium' outlined by Murray [15]. Empathy has, however, been outlined as a key process in audience entertainment and thus it is important to examine it here and the implications that interactivity has on it.

Empathy, in terms of narrative, is defined by Neil as: "In empathizing with another, whether she be real or fictional, one imagines the situation she is in from her point of view; one imaginatively represents to oneself her beliefs, desires, hopes and fears, and so on as though they were one's own" [18: P.191].

Empathy has been clearly identified as vital to audience enjoyment of linear forms. "The crux is that storytelling of any kind is a principal forum of empathic reactivity. It would appear, in fact, that empathic engagement is what fuels interest in tales… the empathy concept can thus be considered pivotal to any interest in, and likely gratification from, storytelling via the media of communication." [27: P.152]

Thus while empathy has been identified as being intrinsic to audience entertainment within linear narratives, it has rarely been explored in terms of interactive narratives. Vorderer, Knoblach & Schramm [24] suggest that empathy could in fact be negatively affected by allowing audiences to interact with a narrative- hypothesizing that making interactive decisions may serve as a distraction from simply witnessing events and thus distract the audience from their ability to develop empathic feelings towards the characters. "Based on the background of affective-disposition theory, however, it may be argued that media users are only distracted and overwhelmed by the possibility of interacting with a narrative. Instead of purely witnessing the events in order to develop sympathy, empathy and suspense toward the hero, they have to struggle with new technology and making decisions about the ongoing narrative of a movie." [24: P.346]

In Vorderer, Knoblach & Schramm's [24] study on interactive narrative, empathy was a variable which was tested and the results suggest that empathy was only increased by interactivity for those audience members with high cognitive capacity. This finding is difficult to assess, however, as empathy was tested using a single item measure.

A counter argument to Vorderer, Knoblach & Schramm's [24] hypothesis suggests that once the structural difficulties of new technology are overcome, interactivity can in fact increase empathic feeling towards the characters of a narrative. By asking the audience to make decisions for the characters through interaction and thus to experience the consequences of those events the audience empathy with those characters will, in turn, increase.

A study by Rockwell & Bryant (1999) explored children's responses to interactive media and found "Surprisingly, children who had more chances to interact with the program also reported liking the characters more compared to the low-interactivity condition" [14: P.589]. While character 'liking' is not the same as empathy, these findings are a positive sign that interactivity could increase empathic feeling towards the characters.

2 Interactivity's Effect on Audience Entertainment: Examining the Links between Three Audience Research Experiments

Three audience interactive stories were designed in order to discover how structure and interactivity could best be balanced to heighten an audience's experience of a narrative. Three narratives were tested in two separate audience studies. Each interactive narrative allowed for interactivity encased in a strong narrative structure and was tested in direct comparison to its linear counterpart.

2.1 Outlining the Audience Research Experiments

Three separate interactive narratives were used in two audience studies with the aim of testing whether interactivity could enhance audience entertainment. The first study used a single narrative with two versions, one linear and the other interactive, in order to evaluate whether the interactive model could enhance audience entertainment. The second study used two different narratives, each with a linear and interactive version, in order to ascertain the effect of interactivity on audience entertainment.

There were 180 participants in each of the audience research studies, for a total of 360 participants. Participants were drawn from a database of members of the public who regularly participate in television studies. The participants were divided into three main age groups: 18 to 34, 35 to 54, and 55+. Measurements were collected using a post-test questionnaire.

The audience research experiments used, in total, three forms of content. Each form of content varied in production style- the first experiment used animation, the second used live action and the final used animatics. The content also varied in terms of form: the first study used a drama while the second study used two narrative-based advertisements.

What remained the same in each case was that an interactive version of the content was tested against a linear version, and that all the forms of content followed the 'yo-yo' structure outlined in the literature review. In this way the effect of the interactivity cannot be said to be singularly related to a specific form of content or production style.

2.1.1 Interactive Drama Study: Design and Findings

A controlled audience experiment was used to investigate the effects of an animated interactive drama on audience entertainment, empathy, immersion, appetite and narrative difficulty. The experiment had a 2 X 1 design. There was one between subjects condition: Interactivity with two levels- the interactive programme vs. the control [linear] programme. This audience study was outlined in full in Hand, S and Varan, D [8].

The results from this study were very interesting in terms of interactive narratives. As a whole the interactive version proved to have significantly higher entertainment (.024 p value), immersion (.005 p value), and appetite (.017 p value) scores. The empathy score was highly significant in favour of the interactive narrative (.008 p value). There was no significant difference in terms of narrative difficulty (.483 p value).

2.1.2 Interactive Narrative-Based Advertisements Study: Design and Findings

A controlled experiment was used to investigate the effects of interactivity on two narrative-based advertisements. The experiment had a 2 x 2 design. There was one between-subjects condition: Interactivity (2 levels: Interactive version vs. Control [linear] version), and Ads (two levels).

The study tested for empathy, sympathy, product category involvement, immersion, difficulty, attitude toward the ad, attitude toward the brand, and purchase/donation intention.

With the exception of purchase intention, all the variables were measured using a 7-point Likert scale. Sympathy was measured using the mean of a five item developed by Escalas, J and Stern, B [2]. Empathy was measured using the mean of a further five item scale developed by Escalas, J and Stern, B [2]. Product category involvement was measured using the mean of five semantic differential items [26; 13). Immersion was measured using the mean of a six item scale developed by Taeyong, K and Biocca, F [22]. Difficulty was measured by the mean of four semantic differential scales developed by Keller and Block [6]. Attitude toward the ad was measured by the mean of four semantic differential scales, developed by Perrien, Dussart and Paul [19]. Attitude toward the brand was measured by means of four semantic differential items [3]. Purchase intention was measured using a one-item 11-point Juster scale. [20]

The results from this study found significant interaction effects between the two advertisements in terms of the control (linear) and treatment (interactive) groups on two variables. These two variables are immersion (p value= .022) and empathy (p value =.052). There was also a nearly significant variable in terms of product category involvement (p value = .066).

2.1.3 Examining the Links between the Audience Research Projects

The audience research projects set about to discover how structure and interactivity could best be balanced in order to heighten an audience's experience of a narrative. While the results for both these audience studies varied across several variables- there was one variable which was consistently heightened by the addition of interactivity to a narrative, the variable of empathy.

The above research findings suggest a unique link between interactivity and empathy. In each of the studies, regardless of the production style or content form used, the addition of interactivity to the narrative significantly enhanced the audience's empathy with the main character. This is an interesting and unexpected finding.

3 Discussion: Exploring the Links between Empathy, Interactivity and Structure

Much of the literature within the field of interactive narratives has focused on interactivity's ability to heighten audience immersion, agency and/or transformation. Little attention has thus far been paid to audience empathy.

The results of the two audience research projects outlined in this paper, however, suggest that the links between empathy and interactivity need to receive much more attention. What this paper suggests is that the question 'why does interactivity work' may be closely tied to audience empathy.

3.1 Introducing the Narrative Structure Link

The literature review outlined the ongoing debate between those interested in the field of interactive narrative on how to balance interactivity and narrative structure. This paper suggests that the discovery of interactivity's ability to heighten audience empathy may be important to this debate.

Vorderer links narrative plot and audience empathy through the 'affective-disposition' theory. Here Vorderer suggests that the building of plot leads to an audience's sense of empathic stress [23]. Vorderer, Knoblach & Schramm explain empathic stress as follows "Whenever TV viewers care about a fictional character in such a drama, they not only hope for the protagonist's positive outcome and fear a negative ending, they also consequently share all emotions along with the character. In a typical drama, when the character's failure becomes likely they may even feel empathic stress, a rather negative emotional experience, better known as suspense (Vorderer & Knoblach, 2000). Unpleasant as it may be during exposure, when the conflicts are finally resolved and the onlookers are presented with a positive outcome for their beloved protagonist, former distress turns into relief. It is exactly this relief that, in retrospect, is often described as the core of the viewers' entertainment experience (Vorderer, Wulff, & Friedrichsen, 1996)." [24: P.344]

Vorderer (2003) thus tentatively suggests that if audiences are allowed to interact with the plot of a narrative then the plot will no longer be able to create in the audience a sense of empathic stress and ultimately will be less entertaining. "Media users… want entertainment to be more and more interactive. They want to be included through involvement and immersion; but at the same time, they are not really attracted to making decisions about the plot. Why not? Maybe because the plot, irrelevant of what the audience wants or expects, or can decide, is the major prerequisite for feeling empathic stress as described by affective disposition theory." [23; P.146]

While Vorderer's views may seem to spell trouble for interactive narratives, this thesis argues that they simply imply that due to the link between plot and empathy, interactive narratives cannot allow audiences to change the plot or narrative structure [24].

In this way empathy becomes the link between interactivity and structure in that interactivity's ability to heighten audience empathy can only be furthered by maintaining narrative structure, which in turn guarantees empathic stress.

This view thus supports this thesis' hypothesis that the best form of interactive narrative is one which encases interactivity within a narrative structure.

4 Future Research

It is important to note that the 'yo-yo' structure of interactivity was not the only link between the content used in the two audience studies. In all three forms of content used within these audience studies the form of interactivity was the same- audiences were asked to interact by making choices for the main characters.

This form of interactivity, where audiences are asked to make decisions for the main characters, could also be argued to increase audience empathy perhaps over other forms of interactivity. By asking an audience to make a decision for the main character we are in essence asking the audience to directly empathise with the character. In this way the audience are placing themselves in the character's predicament in order to make a decision whose consequences the audience may feel more deeply.

Thus it becomes evident that the form of interactivity used in an interactive narrative may be a vital area of future study, especially in terns of its effect on audience empathy.

References

1. Aarseth, E.: Genre Trouble: Narrativism and the Art of Simulation. In: Wardrip-Fruin, N., Harrigan, P. (eds.) First Person. MIT Press, Massachusetts (2004)
2. Escalas, J., Stern, B.: Sympathy and Empathy: Emotional Responses to Advertising Dramas. Journal of Communication Research 29 (2003)
3. Gardner, M.P.: Does Attitude toward the Ad Affect Brand Attitude under a Brand Evaluation Set? Journal of Marketing Research 22(2), 192–198 (1985)
4. Green, M., Brock, T., Kaufman, G.: Understanding Media Enjoyment: The Role of Transportation into Narrative Worlds. Communication Theory 14(4) (2004)
5. Green, M., Garst, J., Brock, T.: The Power of Fiction: Determinants and Boundaries. In: Shrum, d.L.J. (ed.) The Psychology of Entertainment Media: Blurring the Lines Between Entertainment and Persuasion. Lawrence Erlbraum Associates, New Jersey (2004)
6. Keller, P., Block, L.: Vividness Effects: A Resource-Matching Perspective. JCR 24(December), 295–304 (1997)
7. Garrand, T.: Scripting narrative for interactive multimedia. Journal of Film and Video 49(1/2) (1997)
8. Hand, S., Varan, D.: Exploring the Effects of Interactivity in Television Drama. In: Cesar, et al. (eds.) Interactive TV: A Shared Experience. Springer, Berlin (2007)
9. Laurel, B.: Response by Brenda Laurel. In: Wardrip-Fruin, N., Harrigan, P. (eds.) First Person- New Media as Story, Performance, and Game, MIT Press, USA (2004)

10. Lugmayr, A., Niiranen, S., Kalli, S.: Digital Interactive TV and Metadata. Springer, Berlin (2004)
11. Mahoney, D.: Interactive Fiction. Computer Graphics World. PennWell Publishing Co. 24(2) (2001)
12. Meadows, S.: Pause & Effect: the art of interactive narrative. New Riders, Indiana, USA (2003)
13. Mittal, B.: A Comparative Analysis of Four Scales of Consumer Involvement. Psychology & Marketing 12(7), 663–682 (1995)
14. Mudorf, N., Laird, K.: Social and Psychological Effects of Information Technologies and Other Interactive Media. In: Bryant, J., Zillman, D. (eds.) Media Effects: Advances in Theory and Research. Lawrence Erlbaum Associates, Mahwah (2002)
15. Murray, J.: Hamlet on the Holodeck- The Future of Narrative in Cyberspace. MIT Press, USA (1997)
16. Murray: From Game-Story to Cyberdrama. In: Wardrip-Fruin, N., Harrigan, P. (eds.) First Person. MIT Press, Massachusetts (2004)
17. Oatley, K.: Emotions and the Story Worlds of Fiction. In: Green, M.C., Strange, J.J., Brock, T.C. (eds.) Narrative Impact: Social and Cognitive Foundations. Lawrence Erlbraum Associates, New Jersey (2002)
18. Onega, S., Landa, J. (eds.): Narratology: An Introduction. Langman, London (1996)
19. Perrien, J., Dussart, C., Paul, F.: Advertisers and the Factual Content of Advertising. Journal of Advertising 14(1), 30–35 (1985)
20. Rossiter, J.R., Percy, L.: Advertising Communications and Promotion Management, 2nd edn. McGraw-Hill, New York (1997)
21. Spero, I., Stone, M.: Agents of Change: How young consumers are changing the world of marketing. Qualitative Market Research 7(2) (2004)
22. Taeyong, K., Biocca, F.: Telepresence via Television: Two Dimensions of Telepresence may have different connections to memory and persuasion. Journal of Computer-Mediated Communication 3(2) (1997) (accessed 25/09/2006), `http://jcmc.indiana.edu/vol3/issue2/kim.html`
23. Vorderer, P.: Entertainment Theory. In: Bryant, J., Roskos-Ewoldsen, D., Cantor, J. (eds.) Communication and Emotion: Essays in Honor of Dolf Zimmerman, Lawrence Erlbraum Associates, New Jersey (2003)
24. Vorderer, P., Knoblach, S., Schramm, H.: Does Entertainment Suffer From Interactivity? The Impact of Watching an Interactive TV Movie on Viewers' Experience of Entertainment. Media Psychology 3, 343–363 (2001)
25. Wand, E.: Interactive Storytelling: The Renaissance of Narration. In: Rieser, M., Zapp, A. (eds.) New Screen Media- Cinema/Art/Narrative, British Film Institute, London (2002)
26. Zaichowsky, J.L.: Measuring the Involvement Construct. Journal of Consumer Research 12(3), 341–352 (1985)
27. Zillman, D.: Empathy: Affective Reactivity to Others' Emotional Experiences. In: Bryant, J., Vorderer, P. (eds.) Psychology of Entertainment. Lawrence Erlbaum Associates, Mahwah (2006)

The Rise and Fall of Juan Perez: Key Pressing Behaviour and Children's Responses to Interactive Television

Nisha Weeramanthri

Interactive Television Research Institute (ITRI), Murdoch University
South Street, Murdoch, Western Australia 6150
itvnisha@itri.tv

Abstract. This paper outlines usability evaluation procedures in an industry-linked research study of preschool children and interactive television. Preschool children's responses to interactive television prototypes were assessed in a controlled laboratory setting situated within a familiar environment. It reports on the analysis of key pressing behaviour as a quantitative measure of preschool children's responses to interactive technologies. The relationship between key pressing behaviour and levels of motivation, mastery and breakdown is explored.

1 Introduction

Interactive television is a medium capable of combining both 'push' and 'pull' elements [1]. This analogy might be extended to describe preschool children's experience of interactive television, which can pull them into a more participatory viewing world and (gently) push them to create a more personalized experience. Interactive television has the power to build what are known as 'intrinsically interesting learning environments' [2]. However, MIT MediaLab's Mitchel Resnick cautions: "If we give children technology that is too complicated and they don't feel they can express themselves, that is a mistake" (quoted in [3]). It is likely that the Lego Mindstorms' 'High Ceiling, Low Threshold' design mantra is best applied whereby children can easily enter the interactive experience, within which multiple layers of complexity have been embedded [4]. There is little research, however, that empirically examines children's responses to a variety of interactive interface designs, particularly in relation to key pressing behaviour. The significance of even one press (*Juan Perez*) should not be underestimated.

The goal of this research was to assess preschool children's responses to Interactive Television (iTV) Prototypes viewed in a controlled setting (a mobile laboratory), situated within a familiar environment (their preschool). The use of a mobile laboratory which traveled between preschools facilitated access to a large numbers of participants, thereby allowing the collection of sufficient data for quantitative analysis of preschool children's responses.

The author was part of a 36-month Australian Research Council funded project entitled "Enhancing the Content and Experience of Children's Interactive Television", referred to as the Children's Interactive Television (CITV) Project. The CITV Project was conducted at the Interactive Television Research Institute (ITRI) at Murdoch

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 20–29, 2008.

University, Western Australia. The CITV Project linked a multidisciplinary team of university researchers (CITV Team) with industry partners: the Australian Broadcasting Corporation, Nickelodeon, Channel Nine Australia / Kids Like Us, the West Australian Department of Education and Training as well as TV New Zealand. The first three industry partners each authorised the CITV Project to create interactive versions of their current, highly successful children's television programs: *Playschool* (live action), *Dora the Explorer* (animation) and *Hi-5* (live action) respectively.

The CITV Project was fortunate to receive the expert advice of Professor Ellen Wartella, a pre-eminent researcher in the field of children and interactive technologies. Professor Wartella posed the question of whether *choice* is the correct modality for preschool-aged children. She reasoned that preschool children love *repetition* and familiarity, whereas 'choosing' is usually the domain of older children with power over their environment [5].

The CITV Team first clarified the project goals of the academic and industry partners. The author and three other postgraduate researchers then engaged in a six-month review of the literature on preschool children and multimedia. Reportage of their combined findings is beyond the scope of this paper, although interested readers may access summaries of many studies relevant to the CITV Project in comprehensive reports by the Markle Foundation (http://markle.org). Based upon the literature review, the CITV Team embarked upon a stringent refinement process of Key Researchable Issues (KRIs). It was eventually decided to examine 'meta-issues' and 'sub-issues' which best encapsulated the KRIs most relevant to the CITV Project.

Of nine iTV Prototypes created, six were assigned the meta-issues of Choice, Repetition or Customisation. Two forms of each meta-issue (sub-issues) were examined: *Incidental Choice* and *Central Choice*, *Exact Repetition* and *Variation Repetition*, *Subject-Centred Customisation* and *Character-Centred Customisation*. The remaining three iTV Prototypes were designed to examine the supplementary issues of *Haptic Interaction, Active Participation* and *Reward*.

In accordance with the overall division of responsibilities on this collaborative project, researchers conducted tasks according to their specific disciplines. As the Usability postgraduate researcher, the author initially carried out heuristic evaluations of low-fidelity and high-fidelity versions of the prototypes, Usability Focus Groups with preschool children and User Eye-Tracking Trials to trial a creative usability evaluation protocol, as described in Weeramanthri & Turk [6]. Production and modification of the iTV Prototypes was the responsibility of the Production postgraduate researcher. The mobile laboratory phase was organised and conducted by the Psychology postgraduate researcher in conjunction with a Research Assistant.

2 Research Methodology

2.1 Production of the Prototypes

A non-interactive Control Prototype was created for each partner program from file footage. This segment was then modified in three different ways, according to the CITV research issue being examined, to create three iTV Prototypes. The allocation of CITV research issues to the partner programs was partially based on the style and themes of the partner programs, but largely governed by the CITV Project's access to archival material and additional production resources for each partner program.

Following data analysis of the Usability Focus Groups, the Usability researchers strongly recommended that the iTV Prototypes incorporate training instructions. Training instructions were considered essential to facilitate preschool children's ease of interaction with the remote control device and understanding of the interactive task/s. In the absence of further access to partner program resources, the author proposed the creation of a *generic puppet character*, unrelated to the themes or content of any existing children's program, to host training segments and training prompts for all the iTV Prototypes. The use of a generic puppet character would: avoid the need to hire 'pretend presenters' or create new animation sequences; allow greater consistency in the training segments and prompts for all prototypes; allow the CITV Project the freedom to add extra training segments at a later date if required; be consistent with the mindset of a preschool audience; and fit within the model of traditional children's television. In fact, the author believes **the use of a generic puppet for training purposes is an excellent paradigm for future interactive television interfaces for preschool children**.

The nine iTV Prototypes and three Control Prototypes utilised a combined MacroMedia Director / DVD delivery platform. All prototypes were run through the researchers' notebook computers to a television monitor. The remote control device used by participants was a modified Macintosh multimedia remote control device - the size of a preschool child's hand and with very few keys, including four coloured keys. All iTV and Control prototypes were approximately 15 minutes in duration. At the start of each iTV Prototype, a training segment with the CITV Puppet demonstrated the layout of the remote control device, which of the coloured keys were necessary to interact and invited participants to engage in a practice task. Just prior to any interactive task/s, a training prompt reminded participants about the chance to interact and how to do so. Brief segments with the CITV Puppet – which invited participants to "join in and play along" – also appeared at the start of the Control Prototypes.

Eight of the iTV Prototypes incorporated between two and six opportunities for interaction. The ninth iTV Prototype (*Haptic Interaction*) incorporated multiple opportunities for interaction. *Incidental Choice* invited participants to press one of two coloured keys (for example, yellow or blue) to ***choose the colour of wagon*** the characters traveled in, whereas *Central Choice* invited participants to press one of two coloured keys to ***choose the characters' next destination***. *Exact Repetition* invited participants to press the red key if they would like the presenter to sing the ***same song again***, whereas *Variation Repetition* invited participants to press the red key if they would like to hear a ***different song about the same subject matter***. *Subject-Centred Customisation* invited participants to press one of two coloured keys to ***choose the subject matter*** of the next story, whereas *Character-Centred Customisation* invited participants to press one of two coloured keys to ***choose the presenter*** of the next story. *Active Participation* invited participants to press the red key ***whenever they observed an on-screen match*** between caricatures of people with different jobs and their work bags, with four possible matches. *Reward* invited participants to press the red key in order to ***'collect' an on-screen playing card*** as it appeared, with one card available for each of the five presenters. *Haptic Interaction* invited participants to press the red key ***whenever the story heroine's nemesis appeared***, for an interaction period lasting 30 seconds.

2.2 Mobile Laboratory Procedure

Participants for the mobile laboratory phase ranged in age from 4 years and 1 month to 5 years and 9 months. According to the results of a pre-evaluation questionnaire, the Psychology postgraduate researcher allocated participants to the twelve viewing groups. Her sample was stratified according to age, gender, socio-economic status, hours of television viewed weekly and prior use of remote control devices. Participants were only assigned to a television program with which they were already familiar. Raw demographic data from the pre-evaluation questionnaire was passed on to the author.

The mobile laboratory was a custom-fitted caravan consisting of two identical viewing rooms with an observation room situated between them. Each viewing room contained 'bean-bag' seating for the participant, a television, a coffee table, books and toys, a laptop computer and chair for the researcher and the remote control device. Each evaluation session was filmed via a black and white, wireless, infrared camera and recorded onto DVD.

Each participant watched a single prototype (viewing session) during their 30-minute evaluation session. Regardless of whether the prototype was interactive or not, participants were told that they would be watching some television, they could use the remote control if they wanted to and they could play with the puzzle and books if they wanted to. At the completion of the viewing session, the participant was asked questions relating to comprehension and enjoyment.

2.3 Data Log of Key Pressing Behaviour

The Usability researchers requested that the iTV Prototypes be linked to computerized data logging files, in order to automatically record the nature and timing of key presses on the remote control device for all participants. Unfortunately, ITRI personnel were unable to achieve this outcome without adversely affecting basic playback of the iTV Prototypes. Instead, generic software (commonly referred to as 'spy ware') was installed on the notebook computers used by the mobile laboratory researchers. Camp Software's 'Keystroke Recorder' logged all key presses on the remote control device and the computer keyboards. Although the software did not generate a specific time-code for the start or finish of each key press, it generated a general time-stamp and text output file for each viewing session.

In order to run the iTV Prototypes, the mobile laboratory researchers pressed a regular series of keys on their keyboards, including commands to start and quit the iTV Prototypes. In the Keystroke Recorder text output file generated, the letters 'r', 'g', 'y' or 'b' corresponded with participants' use of the red, green, yellow or blue keys on their remote control device. Using the researchers' start or quit commands as markers, it was possible to determine whether participants pressed the coloured keys before, during or after the viewing session.

3 Research Hypotheses

Following the literature review and CITV Team's determination of the CITV research issues, the author derived a number of usability-related hypotheses for various phases of her research (laboratory-lounge phases / mobile laboratory phase / home phase). The following hypotheses are those which relate to results reported in this paper.

H1: **The medium of interactive television affords participation by preschool children.**

H2: **The nature of the interactive task affects preschool children's responses to interactive television.**

H3: **The level of difficulty of the interactive task affects preschool children's responses to interactive television.**

H4: **Prior experience of interactive media affects preschool children's responses to interactive television.**

H5: **Gender affects preschool children's responses to interactive television.**

4 Results and Discussion

4.1 Data Analysis – iTV Prototypes

Keystroke Recorder data was obtained for 218 participants who viewed the iTV Prototypes (of whom 182 interacted). Raw demographic data was received for 205 of these participants (of whom 177 interacted). The key pressing behaviour of each participant was compared against the expected patterns of key pressing behaviour (based on invitations to interact) for the iTV Prototype they viewed. The following five dependent variables were examined:

1) **Interact:** Whether or not the participant interacted with the iTV Prototype;
2) **Correct Number of Presses:** Whether or not the participant utilised the correct number of key presses (with non-interaction assessed as an incorrect response);
3) **Correct Colour Presses:** Whether or not the participant utilised the correct coloured keys (with non-interaction assessed as an incorrect response);
4) **Number of Press Deviations:** For those who interacted, the number of press deviations (plus or minus) from the correct number of presses;
5) **Number of Colour Deviations:** For those who interacted, the number of colour deviations (plus or minus) from the correct number of colour presses.

The author would posit that these five dependent variables are measures of participants' levels of Motivation (Dependent Variable 1), Mastery (Dependent Variables 2 and 3) and Breakdown (Dependent Variables 4 and 5). Table 1 outlines the dependent variables (in italics) in relation to the usability issues and sub-issues.

Motivation might be influenced by factors such as participants' interest in less familiar technology, desire to participate in a new experience, novelty of the environment or prior influence of significant others (such as caregivers or friends). Mastery and (conversely) Breakdown might be influenced by factors such as participants' understanding of training instructions, satisfaction with the interactive task/s, level of engagement with the iTV Prototypes or level of distraction.

Table 1. Dependent Variables and Usability Issues and Sub-Issues

Usability Issue & Dependent Variable	Usability Sub-Issue
Motivation	
Interact	Engagement with Interactive Task
Mastery	
Correct Number of Presses	Compliance with Training Instructions for Interactive Opportunities
Correct Colour Presses	Compliance with Training Instructions for Interactive Choices
Breakdown	
Number of Press Deviations	Satisfaction with Number of Interactive Opportunities
Number of Colour Deviations	Satisfaction with Number of Interactive Choices

The independent variables examined were divided into two groups: characteristics related to the *experimental condition* (such as iTV Prototype viewed, form of choice or level of difficulty of the iTV Prototype) and characteristics related to the *participant* (such as gender, experience with a remote control device or experience with interactive media). A priori, it was decided to use an alpha level of 0.05 for all statistical analysis. The (two-tailed) tests of significance employed were One-Way Analysis of Variance (***A***), One-Sample T-Test (***T***), Chi Square analysis (***C***), Pearson Correlation Coefficient and Tukey HSD Post-Hoc analysis. Discussion of the author's findings relate to the usability issues and sub-issues identified in Table 1.

4.1.1 Motivation

For all participants who viewed the iTV Prototypes (N=218), 83.5% of participants interacted with the iTV Prototypes. That is, the majority of participants were engaged with the interactive task. Significant differences were observed for the dependent variable *Interact* (***C***: p value = .000). Hypothesis 1 is therefore strongly supported.

The effect of Form of Choice (*Incidental* vs *Central*) was not significant for any of the dependent variables, possibly caused by the low number of participants (N=45). It is also possible that the interactive tasks did not differ sufficiently (see Section 2.1) and the latter concept was no more 'central' to the story than the former. Engagement with the interactive task, however, was relatively high for both *Incidental Choice* and *Central Choice* (61.5% and 78.9% respectively). Similarly, when comparing responses to *Character-Centred Customisation* and *Subject-Centred Customisation* (which also involved overt, content-related choices) engagement with the interactive task was 87.5% and 68.2% respectively. These findings suggest that **choice is an appropriate modality for preschool-aged children viewing interactive television**.

The author assigned each of the interactive prototypes to one of four levels of difficulty, ranging from Most Easy to Most Difficult, based upon her professional understanding of the likely cognitive and social skills of this age-group. Surprisingly, participants were most engaged with the interactive task in the Moderately Easy and

Most Difficult categories. These findings suggest that **motivation is affected by the nature of the interactive task more than difficulty of the interactive task** and that salient challenge is an important part of the equation. The effect of Difficulty of Prototype (N=218) was significant for *Interact* (***C:*** p value = .039). Hypothesis 3 is therefore supported.

According to Use of Television Remote Control Device, participants were almost equally engaged with the interactive task if they were usually unsupervised in their use of a television remote control device or if they did not use a television remote control device at all. Interestingly, participants were least engaged with the interactive task if they were usually supervised in their use of a television remote control device. It thus appears that preschool children can and do internalise ‘rules’ and transfer them from one setting to another, even when a regular caregiver is not present. It may be, however, that **a modified or dedicated remote control device for children’s interactive television, which cannot change channels or expose children to unsuitable content, might appeal to caregivers**. Alternately, instructions could be incorporated into an interactive program (or distributed to caregivers) to specify that it is ‘OK’ for children to use the television remote control device on special occasions. The findings suggest that some form of **training will facilitate preschool children’s response to interactive television**. The effect of Use of Television Remote Control Device (N=205) was significant for *Interact* (***C:*** p value = .048). Hypothesis 4 is therefore supported.

4.1.2 Mastery

For all participants who viewed the iTV Prototypes (N=218), 62.4% of participants complied with training instructions for interactive choices. Only 28.4% of participants, however, complied with training instructions for interactive opportunities. Significant differences were observed for *Correct Colour Presses* (***C:*** p value = .000) and *Correct Number of Presses* (***C:*** p value = .000).

According to iTV Prototype viewed, participants who viewed *Haptic Interaction* demonstrated the greatest compliance with training instructions for interactive opportunities. As the only iTV Prototype in which there was no maximum number of key presses, however, this undoubtedly contributed to high levels of mastery. Participants who viewed *Active Participation* demonstrated the least compliance with training instructions for interactive opportunities. Interestingly, this iTV prototype allowed six opportunities for interaction (the most after *Haptic Interaction*). These findings suggest that simply **adding more interactive opportunities will not automatically assist mastery** if other variables, such as difficulty of the interactive task, are in play. The effect of iTV Prototype viewed (N=218) was significant for *Correct Number of Presses* (***C:*** p value = .000). Hypothesis 2 is therefore supported.

According to Difficulty of Prototype, participants in the Most Easy category demonstrated the greatest compliance with training instructions for interactive opportunities whereas participants in the Most Difficult category demonstrated the least compliance with training instructions for interactive opportunities. The effect of Difficulty of Prototype (N=218), was significant for *Correct Number of Presses* (***C:*** p value = .003). Hypothesis 3 is therefore supported.

4.1.3 Breakdown

For all participants who interacted with the iTV Prototypes (N=182), there was a strong tendency towards press deviations and colour deviations greater than zero. Most participants seemed to want more interactive opportunities and more interactive choices. Significant differences were observed for *Number of Press Deviations* (***T:*** p value = .000, Mean = 4.98, SD = 12.18) and *Number of Colour Deviations* (***T:*** p value = .000, Mean = 0.48, SD = 0.944).

According to iTV Prototype viewed, participants who viewed *Haptic Interaction* were ostensibly the most satisfied with the number of interactive opportunities. As, however, there was no maximum number of key presses and only 1/18 participants chose not to interact with this prototype, the mean and standard deviation scores were virtually zero. Participants who viewed *Exact Repetition* were the least satisfied with the number of interactive opportunities (Mean = 14.86, SD = 24.76). Interestingly, *Variation Repetition* did not produce a similar level of dissatisfaction with interactive opportunities (Mean = 5.29, SD = 10.27). It is likely that *Variation Repetition* was more satisfying due to the provision of unseen content, whereas frustration with repetitive content led to greater levels of breakdown in *Exact Repetition*. The combined findings suggest that **interactive television for preschool children should incorporate multiple interactive opportunities and interactive choices, as well as opportunities which have varied outcomes**. The effect of iTV Prototype viewed (N=182) was significant for *Number of Press Deviations* (***A:*** p value = .012). Hypothesis 2 is therefore supported.

Satisfaction with the number of interactive opportunities was significantly higher for females (Mean = 2.27, SD = 6.55) than for males (Mean = 7.03, SD = 14.94). Satisfaction with the number of interactive choices was significantly higher for females (Mean = 0.30, SD = 0.75) than for males (Mean 0.60, SD = 1.02). The effect of Gender (N=177) was significant for *Number of Press Deviations* (***A:*** p value = .008) and *Number of Colour Deviations* (***A:*** p value = .032). Hypothesis 5 is therefore supported.

Satisfaction with the number of interactive choices was significantly higher for participants who used the Video Cassette Recorder (Mean = 0.36, SD = 0 .81) than for those who did not (Mean = 0.72, SD = 1.11). These findings again suggest that some form of training will facilitate preschool children's responses to interactive television. The effect of Use of the Video Cassette Recorder (N=177) was significant for *Number of Colour Deviations* (***A:*** p value = .019). Hypothesis 4 is therefore supported.

4.2 Data Analysis – Control Prototypes

Keystroke Recorder data was obtained for 76 participants who viewed the Control Prototypes (of whom 12 interacted with the remote control device). Unlike the iTV Prototypes, there was no 'predictable' key press behaviour for Control Prototypes. The following five dependent variables were examined:

1) **Interact**: Whether or not the participant interacted with the remote control device;
2) **Number of Presses**;
3) **Number of Colours Pressed**;
4) **When Pressed**: If keys were pressed before, during or after the program;
5) **Colours Pressed**: What combination of coloured keys was pressed.

The author would posit that Dependent Variable 1 measures Counter-Intuitive Motivation and Dependent Variables 2 to 5 measure Spontaneous Mastery. Motivation is termed ‘counter-intuitive’ because participants were not expected to use an interaction device during a non-interactive program. Counter-Intuitive Motivation might be influenced by factors such as distraction, the expectation that a remote control device must be in the room for a reason, an eagerness to please the researcher, the expectation that television screened in this unique environment is somehow different to ‘normal’ television, the belief that changes to content might occur via use of the remote control device or the desire to change channels. Mastery is termed ‘spontaneous’ because participants did not receive training instructions about how to use the remote control device. Spontaneous Mastery might be influenced by factors such as the chance to use a remote control device without co-viewers, the absence of supervision from a regular caregiver or experience with interactive media.

Examination of a number of independent variables did not yield conclusive results largely due to the low number of participants in experimental cells.

4.2.1 Counter-Intuitive Motivation and Spontaneous Mastery

For all participants who viewed the Control Prototypes (N=76), 84.2% of participants were satisfied with non-technological interaction. Significant differences were observed for *Interact* (***C:*** p value = .000). Approximately 1 in 6 participants, however, were apparently dissatisfied with non-technological interaction and utilised the remote control device. Hence, for some participants, it may be that **the television viewing experience is intrinsically altered by the mere presence of an extra technological device**. Designers could incorporate this understanding into building a holistic interactive experience for preschool children which extends beyond the traditional interactive television interface. This might include re-designing the interaction devices or incorporating other objects (new or pre-existing) into the interactive task/s. For example, the infra-red flash of the remote control device could be used to trigger responses from an interactive toy, the set-top box or an alternate interaction device.

For all participants who interacted with the Control Prototypes (N=12), significant differences were observed for *Number of Presses* (***T:*** p value = .001, Mean = 7.08, SD = 5.07) and *Number of Colours Pressed* (***T:*** p value = .000, Mean = 1.83, SD = 1.03). These findings suggest that some **preschool children will experiment even in the most unexpected situations**. Interactive television designers should harness this adventurous spirit by providing children with challenging interactive tasks in a creative viewing environment.

When examining the timing of technological interaction (*When Pressed*), 66.7% of key presses were observed during the viewing session, which indicates that most key presses were program-related. Of all possible combinations of coloured key presses (*Colours Pressed*), 41.7% of key presses were for the red key alone. Given that the red key is traditionally the first of the coloured keys on an interactive remote control device, this is an important finding for designers of interactive television interfaces for preschool children.

5 Conclusions

It is hoped that these empirical findings (bolded in previous sections) may be seen as a vital step in the process of creating a comprehensive database of findings relating to children's interactive television. The author's qualitative analysis of preschool children's responses to the iTV Prototypes in an uncontrolled setting, their homes, will be discussed in a later paper. It is envisaged the interplay between these complementary research phases will yield even more conclusive outcomes. It is undoubtedly as valuable to learn less with greater certainty as to learn more with less certainty. Using the synergistic relationship between quantitative and qualitative analysis, perhaps the academic researcher can have the best of both worlds.

Acknowledgements. The author gratefully acknowledges assistance from the ARC, the ABC, Nickelodeon, Channel Nine / Kids Like Us, DET, TVNZ, the CITV Team, Adjunct Associate Professor Andrew Turk and Timothy P. Ryan.

References

1. Kelly, K., Wolf, G.: PUSH! Kiss Your Browser Goodbye: The Radical Future of Media Beyond the Web. Wired Magazine (5.03) (1997)
2. Calvert, S.L.: Production Features as Scaffolds for Children's Informal Learning from Digital Technologies: Lessons for Instructional Design. In: Neuman, S. (ed.) Literacy Achievement for Young Children from Poverty, Brookes (2008)
3. Tellzen, R.: Designs on Playthings for Thinking Kids. In: IT Alive, The Australian, October 7, 2003, p. 10 (2003)
4. Ackley, J., Dooley, M.: Interactive Toys: The Child As Programmer or My Robot Can Beat Up Your Action Figure! In: Games Developers Conference 2000, San Francisco (2000), Proceedings archive: http://www.gamasutra.com/features/gdcarchive/2000/index.htm
5. Wartella, E.: Response to Key Issues. Address, Children and New Media International Forum, Interactive Television Research Institute, Western Australia (2003)
6. Weeramanthri, N., Turk, A.: A Creative Adaptation of Eye Tracking Analysis for Children's New Media Research. In: Lugmayr, A., Golebiowski, P. (eds.) Interactive TV: A Shared Experience, TICSP Adjunct Proceedings of EuroITV 2007, Tampere International Centre for Signal Processing, pp. 291–295 (2007)

From Time-Shift to Shape-Shift: Towards Nonlinear Production and Consumption of News

Henrik Larsson, Inger Lindstedt, Jonas Löwgren, Bo Reimer, and Richard Topgaard

School of Arts and Communication, Malmö University, 20506 Malmö, Sweden
{henrik.larsson, inger.lindstedt, jonas.lowgren, bo.reimer, richard.topgaard}@mah.se

Abstract. People depend on news to make sense of happenings in the world, but current digital news products do not live up to their potential in this regard. Interactivity in relation to news is often seen as a way to give the consumer control over when to consume something and on which platform. Less attention has been placed on *what* should be consumed and *how*. Within the project MyNews-MyWay, a news service has been constructed that makes possible a more in-depth and varied media consumption than what traditional news services offer today. In addition to time-shifting, the service makes it possible to shift also the shape of the material consumed.

1 Introduction

The media are crucial for the upholding of a meaningful public sphere in contemporary societies. But this is not something that functions automatically. Due to commercial and political pressures, there is an obvious risk that the media do not offer the kind of material that is necessary for people to make sense of their everyday lives. A crucial genre in this context is news. In times of increasing globalization – a "shrinking world" – people depend upon the news to inform them about things happening far away. However, news is a difficult genre. The criticism that news producers are not fulfilling their responsibility to society is heard with increasing frequency. Instead of providing news consumers with material that helps them make sense of difficult happenings all over the world, the news output focuses on sensational events, presenting them in simplified ways. Not enough time is spent on contextualization, on putting events within meaningful perspectives, and there is an emphasis on consequences rather than on backgrounds. As Ignatieff [5] puts it, "television news bears some responsibility for […] the feeling that the world has become too crazy to deserve serious reflection".

To address the current situation, hopes have been high for the interactive media [3, 7]. Through *interactivity,* people should be able to use available material in more meaningful ways, question top-down models of news production and take part in the news production process on a more equal basis [4]. But does it work? A recent review of the history of online journalism paints a bleak picture [11]:

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 30–39, 2008.

> The trend towards speed, constant repackaging, and superficial topics facilitated by the web make a mockery of 'contextualized journalism' promised by prominent new media analysts. Multimedia, hyperlinks, dynamic updates, and interconnection with archives of related contextual content have great potential in the abstract but amount to very little if the logic behind business operations runs against the grains of such innovations.

As of now, interactivity is not a major feature of most online news services, at least not if interactivity is to be seen as something seriously transforming the way we look upon news production and news consumption [1, 2, 9]. Put in other words, digital media hold a potential that is not being fulfilled. What can be done about this?

The objective of the project MyNewsMyWay (MNMW) has been to take part in this discussion in a constructive way. Based on both theoretical and practical work, the objective of MNMW has been to develop new ways of working with the genre of news and new ways of making it available for the engager (which is our term for what traditional TV calls the viewer). This has meant questioning the traditional views of what the genre "is". It has also required significant work on how different engagers would want to access the material, and what kinds of experiences they would like to get out of the encounter. In many discussions of interactivity in relation to news, the main selling point seems to be to give the consumer control over when to consume something, and to give the consumer the control over which platform to consume the material on. This seems to us trivial. Less attention has been placed on what we regard as far more important points: *what* should be consumed, and *how* it should be consumed.

MNMW explores the situation where the audience become critical engagers; when the content constitutes new dimensions of flexible nonlinearity and enables new potentials for comprehension. MNMW therefore addresses questions dealing with the kinds of experience that arise from the process of *consumption*. How can one make the news output available in such a way that it leads to more meaningful experiences than the ones consumers are able to gain today? However, we are also interested in changing the way news is *produced*. Is it possible to create production tools that will facilitate a more in-depth and varied media output? Is this possible to achieve without giving rise to unreasonable amounts of extra work for the journalists and editors involved?

In this paper, we present the approach and design of MyNewsMyWay. We discuss its potential and limitations, and address how it might be generalized for the emerging many-to-many communication patterns of the interactive media.

2 The MyNewsMyWay Approach

The project MyNewsMyWay (MNMW) was conducted within the framework of the EU-funded program New Millennium, New Media (NM2). The objective of the program as a whole was to construct new software production tools and a new delivery system for non-linear media, thereby enabling the creation of media genres where the story told is adapted to the preferences of the engager. This meant going beyond the notion of "time-shifting", thinking instead in terms of "shape-shifting", making it possible to, for instance, seamlessly and in real-time modify a program, adding new material or changing the length of a program [12, 13, 14, 15]. Importantly, the project entailed collaboration between technical partners and artistic/cultural partners, who provided the requirements for the production tools to be developed in the project. Eight experimental productions were produced simultaneously with the construction of the tools.

Within the framework of the NM2 program, the objective of MNMW was *to develop new means of news production and consumption*. The work was carried out in collaboration with SVT (public service Swedish television), who provided the project with professional archive material and contributed their news-production expertise.

Every genre has its specific characteristics, and an initial step for the project was to identify the characteristics of news that were crucial to take into account when developing a new news service. The first thing noted was that the genre is unusually *rigid*. Even though fictional elements are sometimes used in relation to news, for example, it is still the case that the Western world shares a fairly strong consensus on what a news production should look like. This means that consumers know very well what they are to expect when watching a news program.

Another point of departure was the fact that, for many people, the consumption of news is part of the *daily routine* of everyday life. People turn to news at certain times, just as much to uphold daily routines as to learn something new about what is happening in the world. Thus, the challenge for the project was to find a way to make news consumption more meaningful, while respecting the conventionalized and ritualistic nature of the genre.

In other words, we had to create a service which would make it possible for viewers to take part of news stories much like they do today, yet offering a more in-depth and involving experience when desired. Given this, we aimed for a *TV experience* rather than a PC experience. News consumption is linked to the routines of everyday life. We did not want to break that link but rather to support it. News is, of course, already today consumed through the use of a computer, and this will become even more common in the future. However, aiming for a TV (living-room) experience was a way of building on the social experience of watching news. It was a way of trying to keep the consumption communal, making it possible for consumers to discuss what they were watching at the time of watching.

Secondly, we aimed for keeping news watching as a *program experience* rather than as an experience consisting of watching isolated stories. The service was therefore constructed so as to provide the feeling of watching a whole program, albeit a program that could vary in length and content depending on the engagers' choices.

3 Designing MyNewsMyWay

Given the approach described above, the design of MyNewsMyWay is based on the notions of *directness* (visual representation of information, user recognition), *consistency* (predictable, similar throughout the system, drawing on common television and Internet skills), and *simplicity* (simple but not simplistic, based on engagers' existing experiences).

One of the major challenges was to create flexible and interesting interaction functions that were easy to use. We needed to accommodate a full spectrum of engagers, ranging from interactive-media novices to highly advanced users of interactive and participatory media. Our goal was to combine a low adoption threshold for interactive-media use with interesting and relevant functionality.

Since the production was situated on the border between television and Internet, certain decisions had to be taken early on. Firstly, a *remote control* was used for executing the interactions rather than a keyboard or a mouse; secondly, the *television screen* was the medium to design for rather than the computer monitor; and thirdly, the design would not demand audience activity but rather *encourage* the audience to interact within the general framework of lean-back consumption.

The remote control assured that the interactions were available on a device with which everyone was familiar, and which most people found easy to use. It also encouraged/forced the designers to develop simple interactions that could be executed through a simple button click. The television screen is a device in front of which audiences have gathered for decades. Consumption patterns may be changing but the desire to enjoy moving images in a large screen format is still strong. An important characteristic of news is discussions and debates in the micro-environment of the living room, which was another reason for designing for the television screen.

Research has shown that the news genre is losing its younger audience [10]. By allowing the audience to engage with the news material, having the power to interact and shape their own news experience, MNMW wished to explore whether younger audiences can find their way back to the news genre. At the same time, there was no wish to discourage audience groups who prefer a professionally produced news program. Therefore the need to design an interface that allowed both active and passive news consumption was given great attention.

4 News Consumption with MyNewsMyWay

The starting point of MyNewsMyWay is the Entry screen (Figure 1), a graphical screen showing main headlines for each of the different sections as 20-second loops of thumbnails. Headlines and dates are applied on top of the thumbnails. The Entry Screen gives the engager an overview on what is currently making the headlines.

From the Entry screen the engager can go to the MyNews section (Figure 2), the General News section, the Sports section, the Culture & Entertainment section, the Economy Section, the Home & Leisure section or do a search and edit the profiles.

The MyNews program is compiled automatically based upon the engager profile and the metadata tags added to each news story. It gives the engager the latest news updates on the topics that he or she finds interesting. An engager can ideally have as

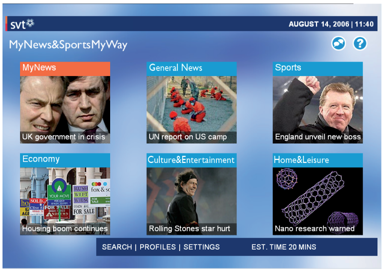


Fig. 1. The Entry screen

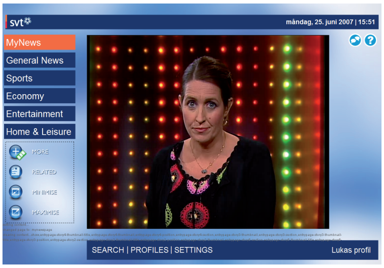


Fig. 2. MyNews

Fig. 3. General News

many predefined profiles as desired, but only one at a time is active. The profiles can be constructed to cover different topics, program durations and breadths of scope. The flexibility allows the engager to create the rules for his or her own news program. Here, as well as in the other sections, it is possible to watch the clips in full screen. MyNews has a default playing time of 20 minutes but the engager can change the duration even during playback.

The other sections — the General News section (Figure 3), the Sports section, the Culture & Entertainment section, the Economy Section and the Home & Leisure section — contain one Latest Edition and four Single Clips. The Latest Edition is equal to the broadcasted prime-time news show about the specific section topic. General News, however, contains two Latest Editions – one is national and the other is a local news program. The section screen shows the four latest (or prioritized) Single Clips and one Latest Edition as looping thumbnails. The engager can select and show the Latest Edition, or a single video clip. If no interaction has been recorded while watching any video clip, the program will automatically return to the Section screen.

There are three major interaction modes within the system: More, Related and Increase/ Decrease Time. The *More* function allows the engager to add clips to the currently playing Single Clip, Latest Edition program or MyNews program. The function is always available during playback, and allows the engager to get an in-depth view of news items and create an understanding for the context of an event in real time. If More is pressed during a Latest Edition or a MyNews program, a clip is added after the currently-playing-clip has ended; it starts playing automatically. After the added clip has ended, you return to the same point as before within the program. If More is pressed during playback of the added clip, another clip will be added in the same manner as above. If the function is pressed during a Single Clip, a clip will be added after the current clip has ended and it starts playing automatically.

The *Related* function allows the engager to view a list of clips, each of which is broadly related to what he or she is currently watching. This gives the engager the possibility to engage with different aspects of the news item context. The function is always available during playback. If Related is pressed during a Latest Edition or a MyNews-program, after the currently-playing-clip has finished, then the engager is presented with a list of related stories to choose from. Having watched one of these related stories, the engager is then returned to the point in the original MyNews or LatestEdition program from which they departed and the original program continues.

The *Increase/Decrease Time* function allows the engager to change program duration. The function is available both in advance of watching and while watching a program, and the engager sets the desired duration by pressing the (+) or (-) buttons on the remote. The function is available for Latest Edition-programs and MyNews-programs, but not when playing Single Clips.

5 How the System Works

First, it must be noted that the current prototype of the system runs on top of a selection of news clips from the Meta archive of Swedish Television. We performed the selection manually according to the principles of professional journalistic work: What events are involved, how are they structured internally, what is the chain of events leading up to the current event? In an ideal full-scale system, the content generation mechanisms should run on top of the full archive. The rest of this section introduces the mechanisms implemented in MyNewsMyWay for automatic content generation and selection.

The post-production of MNMW is exclusively done in the NM2 Tools, i.e. no additional software is used to manage, trim or describe the video content.

In order for the system to be able to find the most relevant clips for a specific engager profile and viewing context, an *ontology* has to be constructed. In this particular case, the ontology was based on the SVT archive keyword directory, but modified and added to throughout the description process. The SVT keyword directory offers ten top-level classes: Economy, Leisure, History, Culture, Philosophy, Medicine, Science, Politics and Technology. There are hundreds of subclasses and the total number of categories is approximately 1500. We also added the top-level category Location which holds all the world's nations and, for Sweden, all major regions and cities, as well as places mentioned in news clips in the MNMW database. Moreover, we increased the granularity of the ontology by adding names of persons and groups, companies and institutions.

Selection rules determine which video clips are selected when the engager selects More or Related. The first version of the rules was based only on finding the clip with the highest count of Topic matches. This led to a bias towards clips with many annotated instances (i.e., many meta tags), and clips with fewer annotated instances were never selected. In the second iteration, the rules look at the level of congruence between media items rather than the specific number of instances that match. This resulted in increased relevance of the selected clips, even though the results were still highly dependent on how many clips were on that topic in the database. (Refer to [12] for details on the narrative structuring and other aspects of content generation in the NM2 Tools.)

6 Evaluating MyNewsMyWay

MNMW has been demonstrated and evaluated a number of times. A formal evaluation was carried out within the framework of the New Millennium, New Media program. MNMW was presented to a number of respondents who experienced the production themselves after a brief introduction, and then were interviewed. Moreover, a session was held at Swedish Television where professional journalists, editors and other staff used and evaluated MNMW. We also presented it at the Moving Images conference in Malmö and collected observations from the attendees' use of the system.

To summarize, the outcomes were quite positive. The respondents in the formal evaluation appreciated the possibilities of going in-depth into news stories that interested them, and they found the remote-control interface simple and easy to learn. The professional journalists were enthusiastic about the production tools and saw new possibilities to reach a younger audience. However, the evaluation also identified two main issues to be addressed in the future development of MNMW and similar systems: the question of a news anchor and the disadvantages related to increased interactivity.

News Anchor Presentation. The evaluation of MNMW use showed that MNMW programs lack the familiarity of a traditional news show. The news genre on television depends on its presentation, not only that news anchors give you a warm welcome but also that they introduce and explain every news clip and connect one news story to another. A personalized news program like MNMW does not provide these relationships, which represents a major drawback in terms of engager experience.

To improve the MNMW service, all clips in the database should be conjoined with their corresponding news anchor presentations; this would make it clear which news show the clip comes from and also make the MNMW program more similar to a traditional news show. There should also be a customized news anchor presentation, telling the engager that he or she is about to watch the personalized news program. The disadvantage of joining a news clip with its anchor presentation is that it will be outdated much sooner; studios change stage set design and news anchors come and go. A standalone clip is more durable than a story presented by a news anchor.

In producing and maintaining an interactive news service, an editor must at all times have the choice to enable or disable the presentation of the news clip; i.e. story and presentation must be separate entities. For example, when a particular news program refurbishes the newsroom, the interactive editor can choose to disable the anchor presentations dating from before the refurbishment.

Engager Competence and Skills in Relation to Increased Interactivity. There is a tension between passive and active modes of news consumption. The disadvantage of handing the interaction choices over to the engager is mainly that it requires much more activity (reduces the lean-back experience) and demands more from the engager in terms of searching, selecting and browsing competence.

Fig. 4. Tribe member lists and communication functions available while watching a clip.

Fig. 5. Zoom-in on a search result, showing a couple of tribe members and some of their significant clips.

An interactive news service must at all times provide passive modes of interacting with the material, i.e. engagers should not need to be active every time they watch news. Weighted annotations can be used to provide More and Related interactions that are adequate for most of the engagers on most occasions.

7 OurNewsOurWays: A Wider Perspective on News Consumption

The work reported so far has concentrated on production and consumption of nonlinear TV news originating from one established producer. However, it is easy to see how the scenario of nonlinear TV generalizes to a situation where the archives of audiovisual material contains clips from many sources, some professional and some user-generated, where comprehensive annotations and ontologies cannot be assumed and where the sheer volume of the material precludes systematic editorial work.

This scenario is analogous in many ways to the situation in other domains of digital information disseminated through the Internet (such as text, images and general video clips). In those domains, it seems that the most successful navigation and access mechanisms emerging in the online communities all draw heavily on *social* mechanisms and participation. Consider, for example, the use of tags in Flickr or the editing schemes emerging organically around the Wikipedia. We argue that nonlinear TV news can be approached in similar ways, and to this end we explore the possible design of a *social navigation layer* which we envision as a sensible addition to systems like MNMW once they move into the many-to-many communication contexts of the digital media.

The social navigation layer, which we call OurNewsOurWays, is designed around the core idea of tribes as social units [6, 8]. Broadly speaking, a tribe consists of a relatively small group of people who know each other and are willing to do things for each other, secure in the knowledge that their altruism will be reciprocated. This sets tribal systems apart from large-scale anonymous social navigation mechanisms such as the oft-cited "People who bought X also bought Y" Amazon feature.

The design of OurNewsOurWays is expressed in a concept demo (available from http://webzone.k3.mah.se/k3jolo/ONS), illustrating a typical situation in which a user receives current and relevant news about an upcoming skateboarding competition, as well as relevant related material, by browsing, searching and communication with his tribal peers. The systems supports his actions through autonomous relevance assessments (refer to figures 4–5).

The key concept underlying the OurNewsOurWays design, enabling the navigation and access functions illustrated in the scenario, is the notion of *social metadata*. As stated above, we cannot assume correct and complete metadata in the clip archive which would seem to make the task of locating relevant material virtually impossible. However, given the assumptions on the nature of tribes roving the archives, and specifically the assumption that tribe members will be prepared to share with fellow members, we can mine metadata from the use of each clip, including *explicit links* that have been made to clips (indicating that the clip was deemed interesting by the link-maker), text and audio *comments* to the clips, and the amount of *activity* that the clip has received. By keeping track of the tribal structure (in terms of "friends" and "pathfinders" lists), we can ensure that reasonable integrity requirements are met. Some clips will certainly have conventional metadata such as *content annotations* and information about *date*, *time* and *location*. In those cases, the conventional metadata is used together with the social metadata to increase the chances of providing relevant and appropriate material.

The view of future nonlinear TV engagement provided in OurNewsOurWays is characterized by many small tribes with only partially overlapping interests, exploring and shaping the collections of digital TV material in a multitude of ways ranging from searching, commenting, producing, debating and other highly involved practices to lean-back viewing by following the trails of fellow tribe members. Whether this is a credible view of future TV remains to be seen, but we argue that it represents a logical continuation of systems like MNMW as the communication structures shift from traditional TV broadcast to the many-to-many patterns of the interactive media.

8 Conclusion

Interactivity is a frequently used concept both within academia and in the media. However, the possibilities offered by "interactive" media services are often limited. The MNMW project has tried to improve on this situation in relation to TV news. We have demonstrated the possibility to give engagers more valuable, in-depth experiences within the context of existing production and distribution structures. Without much extra effort, it is possible to focus upon a news story and follow it historically as well as contextualize it. And this can be done within the framework of everyday life. MyNewsMyWay does not break with people's everyday life routines but rather aims at adding to them and giving them more substance. Finally, the work represented by OurNewsOurWays illustrates a way to generalize the concept in the direction of nonlinear news within the many-to-many communication patterns of emerging interactive media.

Acknowledgments

The authors gratefully acknowledge the contributions by Amanda Bergknut and Tobias Nilsson to the design and development of MyNewsMyWay and OurNewsOurWays. The work was financially supported by the European Commission under the Information Society Technology research program.

References

1. Chung, D.S.: Profits and Perils: Online News Producers' Perceptions of Interactivity and Uses of Interactivity. Convergence (1) (2007)
2. Deuze, M.: The Web and Its Journalisms: Considering the Consequences of Different Newsmedia Online. New Media and Society (2) (2003)
3. Downes, E.J., McMillan, S.J.: Defining Interactivity: A Qualitative Identification of Key Dimensions. New Media and Society (6) (2000)
4. Gunter, B.: News and the Net. Lawrence Erlbaum, Mahwah (2003)
5. Ignatieff, M.: Is Nothing Sacred? The Ethics of Television. In: Ignatieff, M. (ed.) The Warrior's Honor, Henry Holt and Company, New York (1997)
6. Johnson, G., Ambrose, P.: Nemtribes: The Power and Potential of Online Communication in Health Care. Communications of the ACM 49(1), 107-113 (2006)
7. Kiousis, S.: Interactivity: A Concept Explication. New Media and Society (3) (2002)
8. Maffesoli, M.: The Time of the Tribes: The Decline of Individualization in Mass Society. Sage Publications, London (1996) (Originally published in French in 1990)
9. Oblak, T.: The Lack of Interactivity and Hypertextuality in Online Media. Gazette (2) (2005)
10. Patterson, T.E.: Young People and News. A Report from the Joan Shorenstein Center on the Press, Politics and Public Policy. John F. Kennedy School of Government, Harvard University (2007)
11. Scott, B.: A Contemporary History of Digital Journalism. Television & New Me dia (1) (2005)
12. Shapeshift.tv, `http://shapeshift.tv`
13. Ursu, M., Kegel, I., Williams, D., Thomas, M., Mayer, H., Zsombori, V., Tuomola, M., Larsson, H., Wyver, J., Zimmer, R.: Shapeshifting TV - Interactive Programmes. Multimedia Systems Journal, special issue on Interactive Television (forthcoming) (to be published)
14. Williams, D., Kegel, I., Ursu, M., et al.: NM2, New Media for a New Millenium. In: Hobson, P., Izquierdo, E., et al. (eds.) Knowledge-Based Media Analysis for Self-Adaptive and Agile Multi-Media, Proceedings of the European Workshop for the Integration of Knowledge, Semantics and Digital Media Technology, EWIMT 2004, London, UK, November 25-26 (2004)
15. Williams, D., Cook, J.J., Engler, M., Kegel, I., Lohse, L., Stevens, T.S., Ursu, M., Wyver, J., Zsombori, V.: Shapeshifted TV: A Real Opportunity for Broadband. International Broadcast Convention (IBC). In: IBC 2006 Conference Publication, pp. 401409 (2006)

ShapeShifting Documentary: A Golden Age

Vilmos Zsombori[1], Marian F. Ursu[1], John Wyver[2],
Ian Kegel[3], and Doug Williams[3]

[1] Department of Computing, Goldsmiths, University of London, SE14 6NW, London, UK
[2] Illuminations, London, UK
[3] Future Content Group, BT, Adastral Park, Ipswich, IP5 3RE, UK
{m.ursu, v.zsombori}@gold.ac.uk

Abstract. ShapeShifting TV denotes an approach to interactive television programmes that can adapt during delivery to the preferences of the active viewers. ShapeShifting TV is based on a declarative language called Narrative Structure Language (NSL) and is accompanied by a set of genre and production independent authoring and delivery tools. ShapeShifting TV was validated via a number of productions. *A Golden Age* – a configurable interactive documentary – is one of them. This paper presents the production of *A Golden Age* within the ShapeShifting paradigm.

Keywords: itv, interactive storytelling, interactive narrative, documentary.

1 Introduction

ShapeShifting TV [1] denotes an approach to interactive television programmes that can adapt during delivery to the preferences of active viewers. They are automatically edited at viewing time and can result in different narrations (narrative threads) for different viewings. ShapeShifting TV is based on a declarative language called Narrative Structure Language (NSL) [2] and is accompanied by a set of authoring and delivery tools, genre and production independent, which have already been successfully used in the making of a number of interactive productions [3]. *A Golden Age* is one such production. This paper presents the main ideas behind *A Golden Age*, describes its production using the ShapeShifting TV tools, and provides a short evaluation.

A Golden Age is an educational documentary about the arts of the Renaissance in England, configurable at viewing time through interaction. Focussing on the last two decades of the sixteenth century, it explores the culture of Elizabeth I's world.

A Golden Age aims to allow viewers to explore video and audio materials in a television-like form that can be iteratively customised. It is designed to work on a Windows Media Centre PC with the remote control as the primary interface. As an active viewer experiences *A Golden Age*, he or she can indicate the aspects of the subject that are of particular interest to them, and also the degree of their interest in the choices made, thus influencing the form of the programme as it develops. This is implemented simply with themes appearing, at times, on the screen and the viewer selecting their interest in them by pressing the OK button of the remote control a number of times – more for more interest – (refer to Fig. 1). The experience combines coherently structured information with aesthetic value. It is a sit back experience, with the programme adapting *in good*

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 40 – 50, 2008.

time to the choices of the viewer, and not jumping to new content as soon as a new choice is made, thus preserving the continuity of the narration. Any individual experience is that of a well edited linear documentary. Nevertheless, as a configurable documentary, *A Golden Age* offers the active viewer a potentially more useful and more compelling experience than a conventional linear programme for television.

Fig. 1. Snapshot of the interaction interface (Architecture is the theme)

ShapeShifting TV – i.e. the approach, the computational model, the tools and the productions – has already been analysed in the context of related research in [1] and [4]. However, two representative examples to interactive documentaries, *Terminal Time* and *Vox Populi*, are worth being mentioned here for a brief contextualisation.

Terminal Time [5] is a narrative system that produces historical documentaries which respond to the audience's appreciation whilst they are being delivered. It focuses on the automatic generation of the narrative text, through a combination of knowledge based reasoning, planning, and natural language generation. Video clips are subsequently attached to the narration through term matching based on TF/IDF. *Terminal Time* implements a "show me more on this topic" interaction mode and, similarly to A Golden Age, responds to interaction *during* the delivery of each documentary.

Vox Populi [6] generates video documentaries based on interviews about controversial topics. Via a web interface the user selects one of the possible topics and a point of view, and the engine assembles video material from the repository to satisfy the user request. Rhetoric annotations describe the verbal information contained in the audio channel of the video segments. A specialised model encodes how these annotations are to be used to automatically generate video sequences for user specified arguments. *Vox Populi* implements a "show me this topic from this point of view" interaction mode, but it only responds to selections made right at the beginning of the delivery; once compiled, a video argument cannot be interacted with. *A Golden Age* implements the "show me this topic" interaction mode, but allows for a continuous interaction with the programme.

Both *Terminal Time* and *Vox Populi* substantiate the interactive documentary format, in which programmes are automatically compiled from a media pool or database of pre-recorded material on the basis of the viewers' interaction, but achieve this by using dedicated software. Thus the concepts they implement remain hard-coded. *A*

Golden Age is a demonstrator for the *genre and production independent* ShapeShifting paradigm. It is a validating example for the representation language NSL, the authoring tools, and delivery system.

The remainder of this paper presents *A Golden Age*. The focus of the presentation is on authoring. Concepts related to the ShapeShifting paradigm, to the authoring and delivery software, and to NSL will be used without explanations. For clarifications and details the reader is kindly referred to [1], [2] and [4].

2 Authoring

Overall Narrative Structure

The content of *A Golden Age* comprises filmed interviews, works of art and architecture, and performance, together with additional audio recordings of music and of a scripted narrator's voice. There are a number of predefined *themes* according to which the material is structured and therefore could be explored, such as *Architecture*, *Church*, *Theatre*, *Shakespeare*, etc. At appropriate times, such keywords appear on the screen, one at a time, and the active viewer has the ability to express a certain level of interest in them. The cumulated set of interests up to that point determines the way in which the narration unfolds.

The basic unit of structure is the *narrative arc*, which is a (short) standalone unit of narration, communicating one aspect of the documented historical period. Configured documentaries are compiled automatically by the system choosing and sequencing different narrative arcs according to the viewers' expressed interests through the selected themes. Each individual viewing experience – narrative thread – is a sequence of narrative arcs (refer to Fig. 2). The narrative arcs are indivisible elements of narration – once chosen for play, they have to be played to the end.

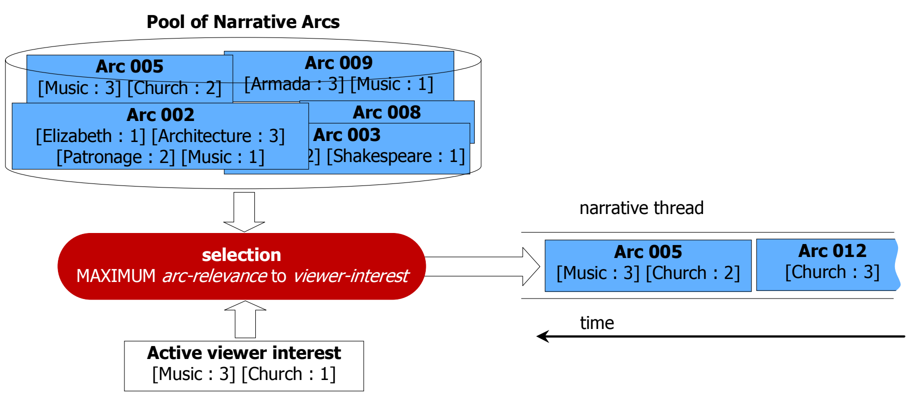


Fig. 2. Iterative generation of narrative threads

The themes to which each narrative arc is relevant are expressed as keyword annotations (of the entire arc). Levels of relevance (of themes to arcs) are also captured as numerical values (integers). The same set of keywords used in arc annotations is also used in eliciting viewers' interest.

Narrative threads are generated iteratively. There is a pool of narrative arcs and a selection mechanism. After a selected arc has been completely played, a new arc is selected from the pool, namely the one that is found most relevant to the viewer's choices up to that point. The relevance of each arc to a given choice is calculated on the basis of its annotations. An example of this is given in Fig. 2: the current viewer's interest is represented by 3 for Music and 1 for Church; Arc005 is the most relevant to this choice (from the ones shown) with relevance values of 3 for Music and 2 for Church.

Narrative arcs have simple internal structures. They are made of *clips*. Clips are the atomic elements of content, the smallest possible but meaningful building blocks for narrative arcs. *A Golden Age* was built mainly in a bottom up approach. The recorded material was first organised into clips. Clips then were used to create the narrative arcs, where one clip could be used in more than one narrative arc. Finally, the main narrative structure was devised using the arcs as building blocks. More about clips, arcs and the main narrative structure will be said in the subsequent subsections.

Creating the content

The majority of the footage was created specifically for the production, although some use was made of existing archive footage from Illuminations, the production company. Filming for *A Golden Age* began in November 2005 and continued to May 2007. During this time 13 interviews were recorded with academics and historians to provide the core content that added up to approximately 50 hours of rushes.

Throughout the filming the production team aimed to produce footage that could be used in a range of different contexts and juxtapositions within the configurable form of *A Golden Age*. Concerns to ensure this during the filming at times mirrored those used during the filming of linear documentary material, such as the request to interviewees to answer questions with complete sentences and without assuming the inclusion of the question, or the need for camera movements to begin and end with frames held for a number of seconds so as to facilitate editing.

The key difference at this stage from linear production was that a "linear" interview would usually be more focused as the production team would have a clear idea of the eventual narrative of the film. Here, the production team was shooting to create material across a "map" of the subject territory and not just for one path through it.

Editing began with transcriptions of each of the 13 interviews. The production team undertook a "paper edit" of the transcriptions, striking out superfluous material and sections in which the interviewee failed to put across a point successfully. Significantly less material was discarded at this stage than in a linear production, with approximately 2/3 of the recorded material being regarded as useful for inclusion. In a linear production probably only 1/8 of an interview might be retained at this stage. The transcript was further divided up into paragraph sections, each of which ideally related a particular event or made a single important point.

The interviews were then edited using a conventional Avid DV ExpressPro editing package. First, a "sync cut" was made of just the interviewee's words, and this was refined to remove "ums", "ahs" and hesitations. Appropriate images were then added, both to cover the jump-cuts within the sync and visually to enhance and extend the impact of the words. For some elements, music or additional location sound (bird noises on garden exteriors, for example) was also added. The final edited element was then exported as a Windows Media Video clip. The format and the bitrate of the clips were influenced by the Delivery System. More than 500 media clips were created for the production, typically between 20 and 90 seconds in length. Each clip features only one contributor, who is seen in vision at least once, and their words may be accompanied by anything from 3 to 20 further shots. The completed clips were then ingested as Media Items (potential Atomic Narrative Objects) in the Authoring Tools.

Creating the Narrative Arcs

The media items (ingested clips) were organised into narrative arcs such that the arcs themselves could respond to the interest of the active viewers. The development of the arcs was driven by the set of selected themes, represented as *keywords*. More than 100 narrative arcs were created on the basis of the keywords. Each arc normally contains between 3 and 6 media items, and each develops a topic, an approach or some sphere of interest, which are more general categories than the themes. Consequently, each arc is annotated with a number of themes/keywords. Topics that overlap or embrace several subjects can be realised more coherently in this way.

The internal structure of the arcs was explicitly represented as NSL link structures employing conditional links and decisions points. Using only explicit structures in narrative arcs allowed the editors to completely check each of them via the Preview Tool, and assess their effectiveness, both in terms of content development and aesthetic pleasure. However, some arcs were also designed to include random elements of content, by using selection groups. In *A Golden Age*, the structures of the narrative arcs were rather simple, but nothing in principle prevents such documentaries to employ more complex structures.

Each arc is annotated with a *[keyword : relevance-factor]* pair. The relevance factor is an integer: 1 for some, 2 for average, and 3 for high relevance. This proved sufficient for A Golden Age, but more sophisticated representations could have been used. The determination of the relevance was editorially determined by the production team and the annotations were added manually to each arc. Fig. 2 shows a number of narrative arcs (in the pool) together with their annotations: for example, Arc002 is annotated to have *some* relevance to Elizabeth and Music, *high* relevance to Architecture, and *average* relevance to Patronage. These annotations are expressed as NSL editing rules associated with the arcs, as illustrated in Fig. 3.

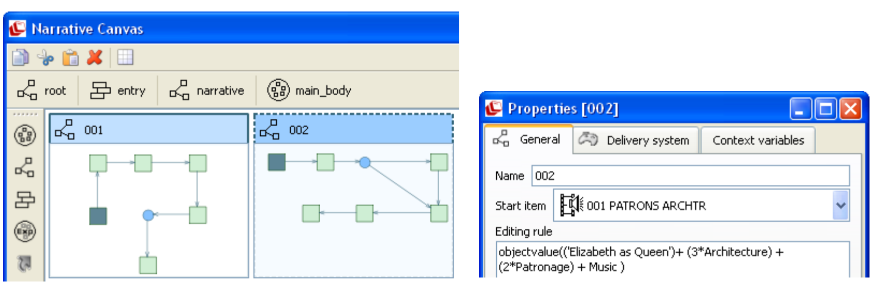


Fig. 3. Narrative arcs on the canvas, and relevance to topics as editing rules

Note that each rule is in fact an integer expression, which represents the overall relevance of the arc to any given choice made by the active viewer. For example, if the choice is 3 for Architecture and 1 for Patronage, then the relevance of Arc002 to this choice is 11 (1*0+ 3*3+2*1+1*0).

Developing the Main Narrative Structure

There are two aspects of the narration that are essential to the implementation (authoring) in NSL: the elicitation of the active viewers' interest and the subsequent selection of the relevant narrative arcs. Recall that both the cues used in the elicitation mechanism and the terms used in annotating the narrative arcs are based on the same set of keywords.

The cues given to the active viewers are expressed as *temporal annotations*, more specifically as *triggering regions*, associated with the media items. Each media item can potentially have parts which, if reached in the narration, are appropriate for asking the question "would you be interested in this theme/topic, and if you are, how interested are you?". For example, whilst a media item which describes religion in the golden age is played, when the images overlaid on the narrator's voice are architectural details of cathedrals, the keyword *Architecture* can appear on the screen and invite the active viewer to express their interest in this theme. This part of the media item would be annotated as a triggering region with the keyword Architecture. An example of a media item annotation is given in Fig. 4. The "Start Time" represents the moment when the cue, expressed as a keyword, appears on the screen and the interaction mechanism is activated, and the "End Time" represents the moment when the cue is removed and the interaction disabled. The annotations are made at the level of the media items themselves, but are validated by viewing the narrative arcs. Not all the media items can be used as cues for engager input and certain media items can be used to make more than one invitation for interaction.

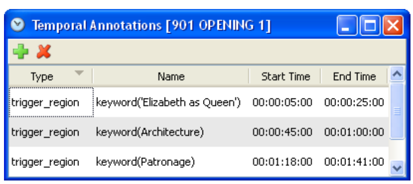


Fig. 4. Media bins, media clips and temporal annotations

Though made with keywords from the same set, it is important to note the difference between *temporal annotations of media items* and *annotations of the overall narrative arcs*. The former describe parts of the material (media items) that are appropriate for engaging the viewer in expressing their interest in themes, whereas the latter describe the relevance of the material (narrative arcs) to the given themes.

The interaction mode is implemented/expressed in NSL using triggered and looping selection groups, layers, and triggering region annotations. The pool of narrative arcs, called main-body, is represented as a looping selection group, and is placed on the leading layer of a layer structure (refer to Fig. 5). The other layer on the layer structure, called interaction, is a triggered and looping selection group with one element only, namely an interactive atomic narrative object, iANO[1].

Narrative arcs are iteratively selected from the main body and played. If the currently played media item contains a triggering region annotation, say

trigger_region(start($time_1$), end($time_2$), keyword(Architecture))

then at $time_1$ after the item started to play, the interaction selection group is triggered, resulting in the iANO being activated[2]. The iANO receives its duration, ($time_2$ – $time_1$), and the keyword Architecture, and sends them to the user interface. If the viewer expresses a level of interest in Architecture (1, 2 or 3), then this choice is retuned to the Realisation Engine (RE) and the iANO is terminated. In turn, this results in the interaction selection group looping back to its start and waiting for another trigger. If the active viewer does not interact, the iANO's default value, zero, is sent back to the RE, as an indication of lack of input, and, as above, the interaction group loops back to its beginning.

Each keyword used as cue for interaction has a corresponding context variables (with the same name) initialised with zero at the beginning of the narrative. When the active viewer makes a level choice for a certain theme/keyword, the corresponding context variable is set by the RE to the respective level choice value. All the non zero context variables are subsequently used by the RE in the selection of a next narrative arc. As it was described above (e.g. refer to Fig. 3), each arc has an expression on whose basis its relevance is calculated. The general formula is

$$Relevance = \sum K_i\_relevance * K_i\_level$$

where K_i_relevance is the value representing the relevance of the arc to the keyword K_i (for example this is 3 for Literature and Poetry and 1 for Sidney for arc001 in Fig. 3) and K_i_level is the value of the context variable corresponding to the keyword K_i. At each iteration, the arc that has the larger Relevance (to the choices made until then) is selected for delivery/play. For both relevance and level *A Golden Age* uses values between 0 and 3. For example, high interest in a theme (3) and high relevance of an arc to the respective theme (3) make a sizeable contribution to the overall relevance factor (9), whereas low interest (1) and relevance (1) make only a minor contribution. If relevance is zero, then the respective term is not present in the expression.

Choices made in the past are less and less influential on the currently chosen content: the more recent the choice, the higher its emphasis on the selected content. This was expressed via a decay function (available in NSL). Each new level of interest choice determines all the other non-zero keyword context variables to be decremented by 1.

There is an ambiguity when a keyword cue appears immediately or very near after it has already been presented to the viewer and the viewer does not select a level. It is

---

[1] This structure is a simplified version of what has actually been used in *A Golden Age*.

[2] It is important to note that the NSL structures are declarative representations and are interpreted. Obviously, the interpretation of an object is not the same with its play. This explanation uses procedural terms and the interpretation appears to happen in sync with the delivery/play. Whereas this improves readability, it also introduces some inaccuracies. However, the principles are correctly described.

Fig. 5. Concept of main narrative structure

not clear whether the user has lost interest in the topic or whether they are happy with the previous choice. In the former case, the corresponding context variable should be set to zero. In the latter, it should be left to its initial value. The best compromise we found was to decrement only the context variable corresponding to the keyword, leaving all the other unchanged. However, by slightly changing the interaction mode, this ambiguity could have been avoided.

Experimentally we found that the relevance formula combined with the decay of the past choices express simply and well the link between interest and the content of the arcs, and, furthermore, that it also preserves the coherence and continuity of the narration.

The main body of the narrative is augmented at the beginning with an opening that explains what the viewer is about to see, followed by an introduction, expressed as a selection group containing a number of arcs, one being chosen randomly (refer to Fig. 6). The narration is terminated when the active viewer chooses “End Soon”. This cue is implemented as all the other keywords – namely as temporal annotations of media items – but the RE moves the narration to its conclusion, after completing the arc, when “End Soon” is selected, rather than set a corresponding context variable.

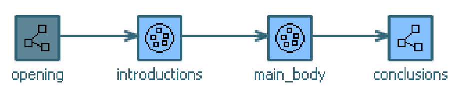


Fig. 6. The top-level story organization

3 Delivery

The functional architecture for delivering *A Golden Age* over broadband is shown in Fig. 7. The Delivery System is production independent, and together with the Narrative Space and the Media Pool are server-side components. The Client Application is production specific. It resides on the server, but it is downloaded via an Application Server (Apache Web Server) to and runs locally on the end-user’s device. The Client Application runs within Internet Explorer on a Windows Media Center, which enables interaction via a relatively simple remote control unit. Video is displayed in full screen at a fixed resolution with widescreen aspect ratio (16:9).

The Realization Engine compiles a dynamic stream of SMIL playlists according to the incoming interactions. The Realisation Engine, an NSL interpreter, is byte-code compiled Prolog and it is wrapped in a multi-threaded C++/Qt console application that handles the timing and communication with the Interaction Manager. The Interaction Manager is in charge of mediating the viewers' interaction between the Realization Engine and the Client Application. It provides functionality for managing the viewers' sessions using XML-RPC and it also maintains user profile data in a relational database. The Content Synthesiser is a media composition engine that interprets dynamically-changing SMIL playlists generated by the Realization Engine and performs rendering operations in order to deliver a continuous audiovisual stream to the Streaming Server.

The Client Application is initiated from within a web-browser. It loads an XHTML document in which both a Windows Media Player component, for rendering streamed video, and a Flash ActionScript component are embedded. The Flash component communicates via XML-RPC with the Interaction Manager, the latter directing the process. The Media Player is superimposed over the Flash component.

Difficulties were experienced with managing the synchronisation of the keyword prompts with particular events in the video. This was due to the variability of buffering and network delays when streaming video over the Internet which caused the de-synchronization of the UI manager and the media player in the Client Application. An incomplete solution to this was derived to allow the production to be partially tested. We are currently working on a complete solution to this.

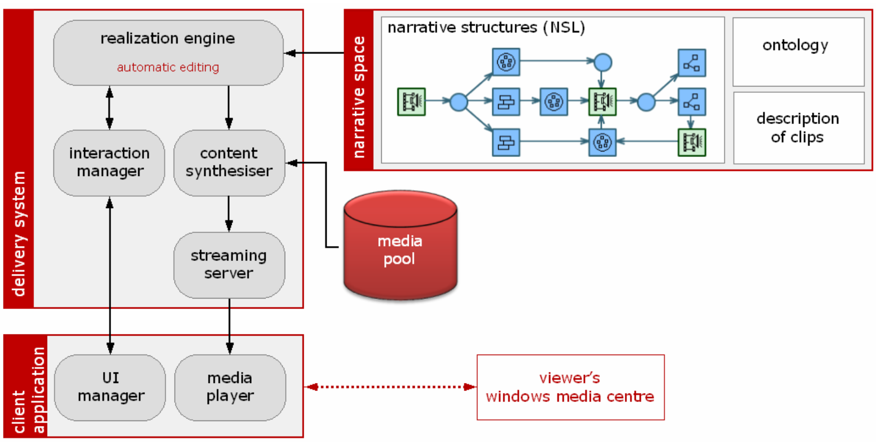


Fig. 7. Delivering *A Golden Age*: functional architecture

4 Discussion and Conclusions

A Golden Age demonstrates the possibility of developing a documentary form that allows configuration in real time, whilst remaining coherent and aesthetically pleasing, using the *generic* ShapeShifting paradigm, and not dedicated software. Due to the problems with the delivery system, extensive viewer evaluation was not yet possible;

we carried out experiments only with a limited number of selected viewers. However, extensive evaluations were carried out by the production team, using the Authoring System's Preview Tool. The programme was found to generate coherent narratives which responded correctly and without discontinuities in the narration to different topic choices. Whereas a partial evaluation of the aesthetic qualities could be carried out via the Preview Tool, and was successful, a full evaluation on the intended delivery platform was not possible, due to reasons just mentioned. However, overall, in terms of the final output, the ShapeShifting paradigm could be declared as having been successfully applied to the production of a good quality interactive configurable documentary.

The production team reported that the Narrative Structure Language (NSL) and the ShapeShifting TV software represent entirely new ways of approaching narrative construction, which don't fit tidily into established roles in the current media production chain. However, they also acknowledged that the narrative structures of *A Golden Age* could not have been developed with any media tools available on the commercial market today, at least not with a comparable development effort.

The overall narrative structure of *A Golden Age* is such that further material could be added, without requiring any amendments to the overall narrative structure. The new material would have to be authored as narrative arcs. Community functionality whereby an active viewer could share an instance of the documentary that they had experienced – a narrative thread – with other people, who could subsequently personalise that experience through their own interactions, is a straightforward development for *A Golden Age*, readily implementable with the current tools.

Te narrative structure of *A Golden Age* is organised on two levels. There is the micro-narrative level – the space of the narrative arcs – and the macro-narrative level – represented by the main narrative body. Refinements to these structures could be experimented with, in a search for achieving more compelling and engaging experiences, on one hand, and to devising more effective methods of authoring, on the other. Narrative arcs could, for example, have more complex structures and employ non-explicit NSL structures, which could make use of the metadata annotations of the media items themselves. The structure of the main body of the narrative could potentially be refined to express, for example, elements of rhetoric, which could generate more compelling narrative threads. The selection mechanism, represented mainly by the relevance formula, could be another avenue for further investigation. For example, the current selection mechanism deals well with expressed choices: it covers the situations when the preferences of the viewer change in time (by responding to the most recent expressed interest, but at the same time taking into the account previous interests); when the viewer has equally high preference in the listed themes (in which case the narration will present a lot of variety, but keeping in focus the most recent choices); and when the viewer has low or no preference for the listed topics (in which case the narration will still present a lot of variety, but through random movement from topic to topic). However, together with the interaction mode, the formula could be further refined for more sophisticated responses. The ShapeShifting technology more or less readily supports this. However, viewer evaluation would be required to assess the effectiveness of more refined selection mechanisms upon increasing the enjoyment value of the viewing experience.

In conclusion, *A Golden Age* is more than a successful proof of concept of an interactive configurable documentary built using the generic ShapeShifting software, and a good basis for further research.

References

1. Ursu, M.F., Kegel, I., Williams, D., Thomas, M., Mayer, H., Zsombori, V., Tuomola, M.L., Larsson, H., Wyver, J.: ShapeShifting TV – Interactive Programmes. Multimedia Systems Journal, SI on Interactive Television (to be published, 2008)
2. Ursu, M.F., Cook, J.J., Zsombori, V., Kegel, I.: A Genre-Independent Approach to Authoring Interactive Screen Media Narratives. In: Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies, Westin Arlington Gateway, Virginia, USA, November 9-11, 2007, pp. 173–180 (2007)
3. Williams, D., Kegel, I., Ursu, M.F., Pals, N., Leurdijk, A.: Experiments with the Production of ShapeShifting Media: Summary Findings from the Project NM2 (New Millennium, New Media). In: Cavazza, M., Donikian, S. (eds.) ICVS-VirtStory 2007. LNCS, vol. 4871, pp. 153–166. Springer, Heidelberg (2007)
4. Ursu, M.F., Cook, J.J., Zsombori, V., Zimmer, R., Kegel, I., Williams, D., Thomas, M., Wyver, J., Mayer, H.: Conceiving ShapeShifting TV: A Computational Language for Truly-Interactive TV. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 96–106. Springer, Heidelberg (2007)
5. Mateas, M., Vanouse, P., Domike, S.: Generation of Ideologically-Based Historical Documentaries. In: Proceedings of AAAI 2000, Austin, TX, pp. 36–42 (2000)
6. Bocconi, S., Nack, F., Hardman, L.: Vox Populi: A Tool for Automatically Generating Video Documentaries. In: Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia Salzburg, Austria, September 6-9, 2005, pp. 292–294 (2005)

HyLive: Hypervideo-Authoring for Live Television

Peter Hoffmann[1], Tobias Kochems[2], and Michael Herczeg[1]

[1] Institute for Multimedia and Interactive Systems, University of Luebeck
Ratzeburger Allee 160, D-23538 Luebeck, Germany
{hoffmann,herczeg}@imis.uni-luebeck.de
[2] Dankwartsgrube 43
23552 Lübeck
tobiaskochems@web.de

Abstract. In this paper we discuss the commonalities and differences of hypervideo and interactive television. We show that, even if the interaction and presentation of both are similar, there are strong differences in handling the content and the link structure of those media. The production and editorial process for live interactive television with hyperlinks and added value including production and editorial workflows is examined. The results led to a concept and a prototypical implementation for an editing tool and a web-based client player for interactive live television with hypervideo structures, called HyLive.

Keywords: hypervideo, live television, interactive television, synchronized additional information.

1 Introduction

Hypervideo follows the basic idea of hypertext replacing the text as a static lead medium by video content as a dynamic one. Currently there are two basic definitions of hypervideo. One emphasises the aspect of augmenting information to video as for example Hesse does [3]. Hyperlinks in the video are activated by the user, who so gets access to additional information synchronised to the actual scene in form of texts, images or even other videos or sequences in the same video. The other one emphasises the influence of the audience on the storyline [4]. This kind of definition has its origins in the field of storytelling. Activating a hyperlink by the user from this point of view means a change from one scene, perspective or aspect to another in regards to the content. So the user has access to the content not only in a linear but also in a non-linear way. The audience takes the step from passive to active viewing building an own individual story-plot with each interaction. None of both definitions is entirely wrong, but both seem to be incomplete. Hypervideo could include both: the usual interaction like starting and stopping the video and the other basis functions deriving from the medium "video" combined with the possibilities of hypermedia and storytelling [5].

Independent from the definition the media content and its structure can be distinguished between hypervideo with a closed structure and hypervideo with an open structure. In the closed structure there is only one video to which all additional

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 51 – 60, 2008.

information is related: so a link to another scene means only that some frames in the video are leapt. The open structure however also includes links to additional information sources outside of the video itself: links can other hypervideos. Changing a perspective in the storyline is easily feasible by leaving one hypervideo and switching to another one.

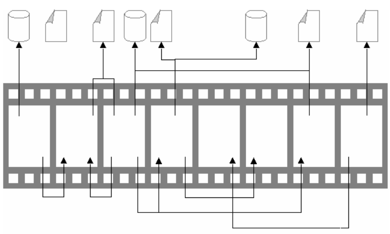


Fig. 1. Closed hypervideo structure: only internal links

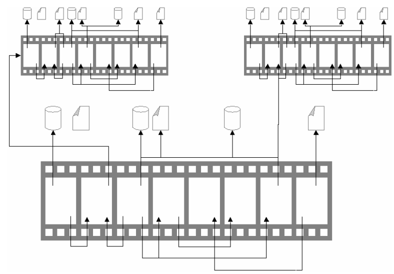


Fig. 2. Open hypervideo structure: internal and external links

In spite of all potential capabilities and benefits hypervideo still is rarely used. At the moment there is no mass media application for hypervideo to be found [7]. The main problem of interactivity in the television context remains the back channel from the audience to the broadcaster. Most concepts are aiming to broadcast more data than only the video information to the audience. The most common technology for this use case is teletext, where the additional information is textual information like national and international news, weather forecasts and TV schedules. The information is encoded into the blanking interval of the TV signal. This approach does not allow any synchronisation of the additional information and the actual content of the video. Few teletext pages seem to be in a synchronous relation to the video content like subtitles or live comments. The "synchronisation" is achieved by human editors who watch the programme and publish the texts just in time [9].

While the analogue technical systems going to be replaced with digital systems some new approaches of adding informational and interactive values to the programmes are upcoming. This technologies offer the chance to extend synchronised information as well as more interactivity to the individual audience. Most popular examples for those applications are the Electronic Programme Guide and the Multimedia Home Platform. While EPG offers mainly information about the programme schedule, MHP concentrates much more on interactivity.

At present there is no standardized combination between the different concepts of hypervideo and added value systems. Instead of single different media like video and teletext a new media approach is conceivable, "interactive live added value systems". This means interactive telemedia, where the actual content of the video stream is synchronized to related additional media, which can easily be reached by the use of links as known as in any internet browser.

2 State of the Art

Since specified by the BBC [1] in 1974 teletext [6], as the most common added value TV system, is almost unmodified. Only minor changes like HiText for a higher resolution have been developed. But even such small development steps were not considered by all broadcast service providers offering teletext.

With the digital technologies some new services appeared. Especially commercial broadcasters providing programmes with specific content were interested in creating new interactive services. As a result, many mainly private stations offer set-top-boxes. This has been used for sporting events mainly, like soccer or car racing events.

Fig. 3. Internet programme n-TV plus (www.n-tv.de/n-tvplus)

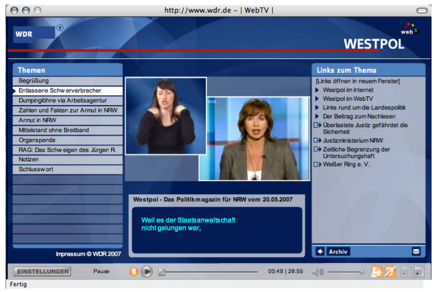


Fig. 4. Internet media archive WDR WebTV (www.wdr.de/themen/homepages/webtv.jhtml)

Since the internet is becoming a more often used medium for broadcasting information some television stations are offering their programmes in the internet, too. While most of those services are not more than the pure television programme, some providers are interested in offering more features. The most often used method shows some similarities to the hypervideo approach: the actual programme, as seen on any TV set, is streamed through the internet, accompanied by links which are grouped into several categories like “current news”, “sports news”, etc. (see Fig. 3). Selecting a link opens a new frame with additional information, mostly a web page. The links themselves are not bound to the visible information in the video image. Any constraints, like the installation of a special software application, are critical for the acceptance of the service by the audience.

Instead of enriching the regular programme, videos can be archived for programme-on-demand purposes, so that the audience is able to watch the programme at any time. In some cases the archived programme is enriched by links to additional information. The links are synchronised to the video content and change over time. The additional information is edited to the programme after its transmission but before archiving. The internet application WDR WebTV of the German WDR (Westdeutscher Rundfunk) as an example (Fig. 4) allows watching archived programmes on demand, navigate the programmes, include subtitles and get additional information to the programme in relation to the actually shown content.

The research and development of hypervideo applications, which do not connect to a live transmitted television programme, can be found in several publications [8][2]. However, nearly all of these applications are based on archived and post-produced video material. The main focus of our current research lies on the question how hypervideo can become dynamic in a live video stream.

3 Live Interactive Hypervideo-Based TV

In this chapter a strategy for combining the ideas of hypervideo and interactive television will be introduced. The differences of classical television, classical added value systems, online and multimedia editing and hypervideo will be examined and a feasible production process and work flow will be described. The goal is to find practical strategies, how additional informational and interactive elements can be added to live television. We have done some earlier research and development in this respect in a project called HyPE, a hypervideo environment with an open API [5]. HyPE should as well be a stand-alone hypervideo application, but it should also be possible to adaptively control the hypervideo player and the behaviour of the hypervideo links by an external application as for example a narrative presentational logic. The external application should be able to control the presentation of hypervideo links in a hypermovie in relation to the profile and the behaviour of a user.

Due to some hypervideo productions done with HyPE there appeared questions about the practical use of HyPE in a live television context. The main question is whether hypervideo links could be defined, enabled and included in a running hypervideo just-in-time.

3.1 Differences between Hypervideo and Live Interactive TV with Added Value

As already discussed there are two different structures of hypervideo, an open structure and a closed one. However, while being different in the handling of associations

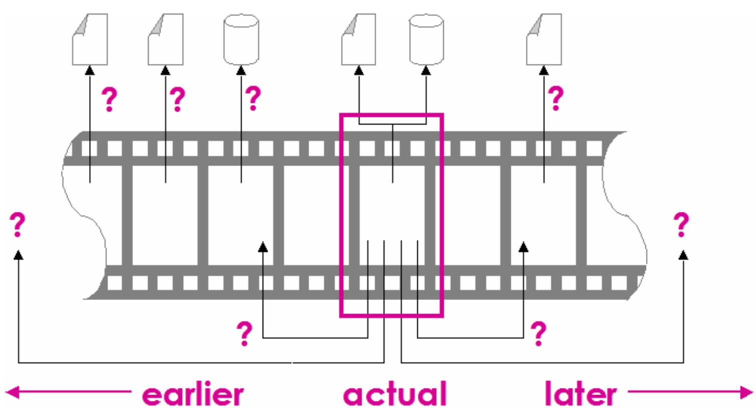


Fig. 5. Structural Question for live hypervideo links

between video sequences and additional information, both types deal with videos, which have a well-defined beginning and end. No hypervideo link can refer to a video frame before the beginning or after the end. In the context of live interactive television this is different.

A live television stream does not allow intercontextual links, which refer to other frames in the video stream. Even if a live programme is well scheduled and planned it is impossible to link to future parts of the programme. Intercontextual links can be possible if earlier parts of the programme were stored and access is given in an appropriate authoring process.

The other possibility for intercontextual links is similar to the open hypervideo structure (Fig. 2) and is independent of storage and archiving. It means that there are hypervideo links which do not refer to the same video stream but to another one, which is independent from the first one.

Besides intercontextual links there are intermedia links, which refer to related additional information which are not included in the video, for example related web-pages or the like. The difference of such links in live interactive television to hypervideo is the definition of the links. Intercontextual links as well as intermedia links can be shown in the transmitted frame. So the problem is how to prepare, store and present those intermedia links.

3.2 The Users and Their Tasks

The structural differences between hypervideo and live interactive television cause differences for their users as well. In general there are two classes of users, the audience at one side and the editors on the broadcasting service side. The goal of the audience is mainly to be entertained or edutained. For live interactive television with an information layer added this means to follow the programme and access the additional information from time to time.

A user class we focus especially are the experts in the editorial department of the broadcast service. They enrich the video sequences with interactive elements related somehow to their visible content. Interactive elements in this context are both, elements for interactions like voting and hypervideo links which refer to additional information about the content. This general goal is to be divided in several subtasks:

- plan and schedule the programme,
- assemble additional information and arrange references,
- schedule, sequence and arrange the interactive elements,
- design the appearance of the interactive elements and
- test the arrangement and its functions.

So far the production of live interactive television with associated information does not differ from hypervideo. The differences will be visible with the start of the live transmission:

- watch the live stream and apply the predefined interactive elements,
- activate and publish the links,
- deactivate the links in dependence of the streamed content and
- interprete the audiences' feedback.

These tasks can not be performed by a single person. Several groups are involved in this process as there are the TV editorial department, which takes care about the television part of the programme and the online editorial department which takes care about the interactive elements. Three particular persons are involved in this editorial department:

- the online editor for the online investigation and the preparation of the additional information,
- the screen designer for the design of the interactive elements and their appearance and
- the live editor who has to publish and deactivate the interactive elements just-in-time and has to react to unexpected changes in the programme schedule.

This task structure is the result of discussions with persons in charge for multimedia and online services at the german NDR (Norddeutscher Rundfunk) as well as of some other research on how online journalists work.

3.3 The Editorial Workflow

From the subtasks for live interactive television described above follows an information flow and a workflow where both editorial departments are highly interconnected with each other.

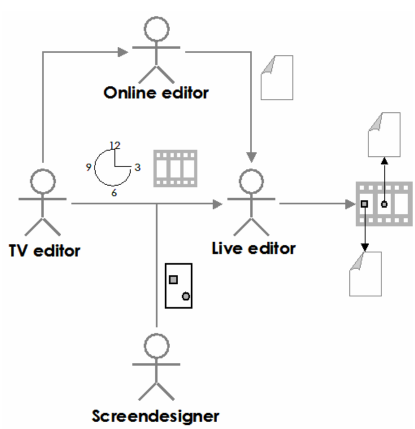


Fig. 6. Information flow in the production of live interactive television

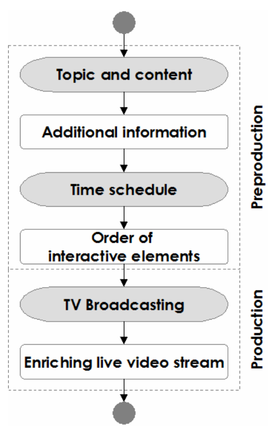


Fig. 7. Production workflow for live interactive television

The editorial workflow starts with the TV editor, who is in charge to prepare a live event. This event shall not only be a television programme but also include options for interactive add-ons. The TV editor schedules and plans the programme. All plans and the additional information go in parallel to the screen designer and the online editor. The screen designer uses certain tools for designing the interactive elements. The online editor does an investigation on related information which can be integrated into the interactive programme and produces specially needed media and/or web-pages. The information from the TV editor, the online editor and from the screen designer is merged by the live editor. With the detailed information about the programme schedule he or she is able to prepare and position the interactive elements. When the transmission of the live programme starts the live editor follows the programme and places, publishes and deactivates links and other interactive elements just-in-time (Fig. 6).

The information flow influences the production workflow. TV editorial department and online department work together in the same production. Figure 7 depicts the production process, divided in preproduction and production phases.

4 The HyLive-Application

The analysis together with production professionals lead to the development of the HyLive-application for live interactive television.

The system architecture of the HyLive system follows a modular approach as to be seen in Figure 8. The part of the system which is used by the live editors consists of a computer with a TV board and an optional TV display. This computer encodes the television signal and transmits the encoded signal to the Media Server. Encoder and Media Server both are standard applications. The live editor follows the television programme and selects with the help of the HyLive Editorial Tool the hypervideo links which are to be published. The hypervideo links are pre-produced in form of a RSS-Feed. Each single hypervideo link is defined by one item in the RSS-Feed. This allows to generate a new feed with prepared entries as well as to use an existing RSS-Feed. Every item in the feed can be used as a hypervideo link. The advantage is a very flexible use of prepared and existing information elements. The HyLive Editorial Tool is connected to the same Media Server as the TV Encoder. The Media Server

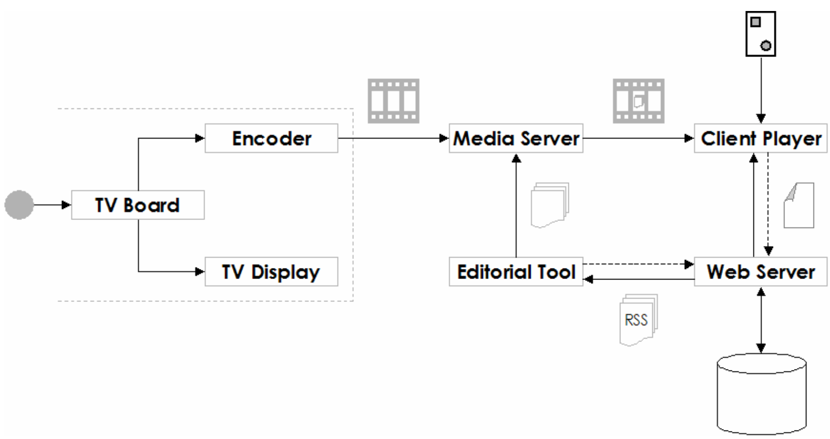


Fig. 8. System architecture of HyLive

combines and synchronises the television signal with the editorial information. The additional interactive information is encoded into the television stream. This signal is finally broadcasted from the Media Server to the Internet.

The audience receives the live interactive television programme via the Internet on their computer in a Player Client. Because this player is implemented as a Flash player it is not necessary to install any new application on the computer to receive and interact with the programme.

The HyLive Editorial Tool can be run within a standard web browser. The user interface is designed as a simple and transparent system that can be used efficiently (see Fig. 9). A toolbar toolbar with buttons for the basic functions at the top and three columns for the editorial workflow are the parts of the user interface. In the left column the connected RSS-Feeds are listed. These are the resources from which the live editor can pick the feed items for the live programme.

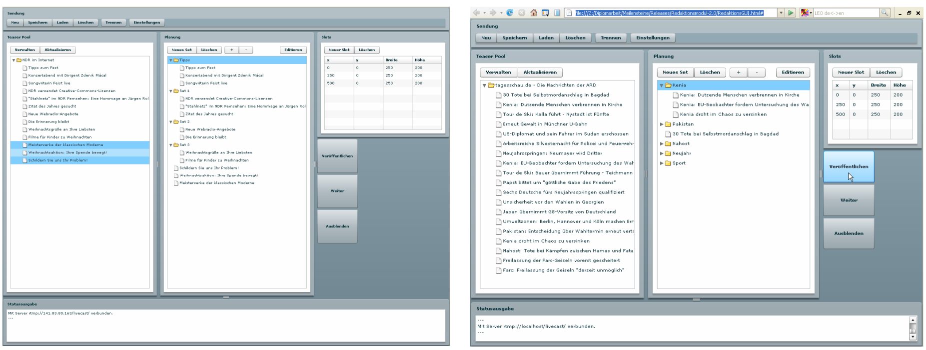


Fig. 9. The HyLive-Editorial-Tool (left: Stand-alone Application, right: Web-Application)

The chosen items are dragged into the middle column where the editor can arrange a first order of the links and group items in sets, which shall be published at the same time. The right column is for the publishing functions. If any corrections in the positions of the hypervideo links have to be done, this takes place in a coordinate field. With one click on one of the central buttons the editor is able to publish a selected hypervideo links or a whole set of them. The published hypervideo link or set is highlighted in the middle column. Using the “Next”-Button allows to deactivate the currently published hypervideo links and to publish the next ones from the ordered list.

The HyLive Player Client on the side of the TV audience will be usually integrated into a web-page of the broadcast service. But it can also be run as a stand-alone application if Flash has been installed on their computer. The HyLive Player Client receives the combined signal of the live interactive television programme from the Media Server of the broadcast service and displays it. The appearance of the interactive elements is independent from the player client. The design of the interactive elements is transmitted through the signal stream. So it is possible to change the design of the hypervideo links and interactive elements at run time of the programme by the editor. This includes also the position of each element within the video frame

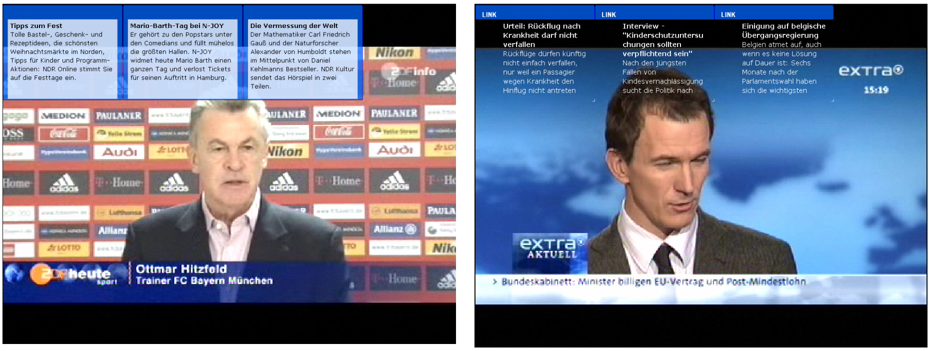


Fig. 10. HyLive Player Client with different designs of integrated hypervideo links

5 Further Work

While the basic functions of the HyLive system are working well and the development of the Editorial Tool has already been finished, for the player client some more functions will be integrated in the next release. Besides the active support of subtitles most important interactive functions like voting etc. will be realised. An evaluation with real audience is planned for the first broad and public use in a Eurovision Programme in cooperation with the german NDR.

Acknowledgements

We would like to thank the department “Programmbegleitende Dienste/ Multimedia” of the NDR (Norddeutscher Rundfunk) for their support and cooperation.

References

[1] Specification of standards for information transmission by digitally coded signals in the field blanking interval of 625-line television systems. Joint BBC, IBA, BREMA publication (October 1974)

[2] Doherty, J., Girgensohn, A., Helfman, J., Shipman, F., Wilcox, L.: Detail-on-demand hypervideo. In: MULTIMEDIA 2003: Proceedings of the eleventh ACM international conference on Multimedia, November 2003, pp. 600–601. ACM Press, New York (2003)

[3] Hesse, W.: Didaktisches Design, Mediengestaltung, Visualisierung (2005) (20.03.2006), `http://www.e-teaching.org/didaktik/gestaltung/visualisierung/hypervideo`

[4] Heintze, J.v., Bode, C.: Emergent Stories - Eine neue Herangehensweise zum Erleben und zum Authoring interaktiver Geschichten. In: Spierling, Ulrike (Hrsg.) Digital Storytelling - Tagungsband, pp. 13–28. Fraunhofer IRB Verlag, Stuttgart (2000) (Computer Graphik Edition 02)

[5] Hoffmann, P., Herczeg, M.: Hypervideo vs. Storytelling Integrating Narrative Intelligence into Hypervideo. In: Göbel, S., Malkewitz, R., Iurgel, I. (eds.) TIDSE 2006. LNCS, vol. 4326, pp. 37–48. Springer, Heidelberg (2006)

[6] Mothersole, P.L., White, N.W.: Broadcast Data Systems: Teletext and RDS. Butterworth & Co., London (1990)
[7] Richter, K., Finke, M., Hofmann, C., Balfanz, D.: Hypervideo, Hershey, USA (2007)
[8] Zahn, C.: Wissenskommunikation mit Hypervideos. Untersuchungen zum Design nicht-linearer Informationsstrukturen für audiovisuelle Medien, 1st edn., Waxmann, Münster (2003) ISBN 3830912706
[9] ZDF: Live-Untertitel. So funktioniert es. ZDF Mediathek (visited, December 20, 2007), http://www.zdf.de/ZDFmediathek/content/Live-Untertitel_-_So_funktioniert_es/7236?inPopup=true

Composer: Authoring Tool for iTV Programs

Rodrigo Laiola Guimarães[1], Romualdo Monteiro de Resende Costa[2], and Luiz Fernando Gomes Soares[2]

[1] CWI: Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
rlaiola@cwi.nl

[2] PUC-Rio: Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente 225
22453-900 Rio de Janeiro, Brazil
{romualdo, lfgs}@inf.puc-rio.br

Abstract. This paper presents Composer, an authoring tool to help creating interactive TV programs for the Brazilian Terrestrial Digital TV System. In Composer, several abstractions are defined creating different document views (structural, temporal, layout and textual). One of these views, the temporal view, preserves as much as possible the timeline paradigm, so popular in TV program editing. Using this view, authoring can be done by placing media objects on a time axis, however, preserving the relative relationships among them. Moreover, non-deterministic time events, like viewer interactions and content adaptations, can also be represented in the temporal view. In addition, the occurrence of these unpredictable events can be simulated, and the resulting TV program played, from any starting point. Besides other facilities provided by its four views, Composer also supports third-party views created as add-ons, and live program editing.

Keywords: Authoring tool, Digital TV, Interactivity, Temporal and spatial synchronization, Composer, NCL, SBTVD.

1 Introduction

Interactive digital TV programs are particular cases of hypermedia applications and must deal with temporal and spatial synchronization of different media assets, besides the main video and audio. These programs may be specified using a textual programming language, but authoring tools are useful to abstract the complexity of such a textual language.

Different abstraction types can be adopted in a hypermedia authoring tool based on the different views offered. As each abstraction view has a set of advantages and disadvantages with regards to others, the joint use of complementary views is often necessary to reach a desired functionality. This approach is used in the authoring tool presented in this paper, called *Composer*, which besides its textual view for explicit specification of programs using NCL (*Nested Context Language*) [1], the standard language for specifying declarative applications in the Brazilian Digital TV System (SBTVD-T), offers a graphic structural view, a graphic layout view, and a graphic temporal view.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 61 – 71, 2008.

Using a graphical representation, the structural view abstracts the main entities [1] defined in NCL. In this view, the author can create, edit and delete media objects, compositions (set of objects and their relationships), and links (relationships among objects). The layout view abstracts the initial spatial representation of media objects on different exhibition devices. In this view, an author can create, edit and remove spatial regions associated with the initial exhibition of objects. Finally, the temporal view abstracts the temporal relationships among media objects providing a graphic solution for these relationships, allowing the representation and simulation of unpredictable events, like viewer interactions and content adaptations, and enabling the resulting NCL document playback.

Composer's views work synchronized, in order to provide an integrated authoring tool. Besides that, *Composer* also allows the development of new views and supports live program editing [2]. In a digital TV environment, *Composer* can be installed in the server-side, allowing broadcasters to produce live editing commands, as well as in TV receivers, allowing live editing from viewers. On the server-side, *Composer* still has a formatter (an NCL player) [1], which makes possible presenting a document in authoring time.

This paper presents *Composer* authoring tool and is organized as follows. Section 2 outlines authoring tools related to this work. Section 3 describes the authoring views available in *Composer*, and how *Composer* supports third-party add-ons. Section 4 presents the features offered to support live edition. Finally, Section 5 presents the final remarks.

2 Related Work

Usually digital TV open standards provide Java-based[1] and XHTML-based (*eXtensible HyperText Markup Language*) languages to specify the synchronism and the interactivity present in their programs. On the one hand, authors need to know standardized libraries and object-oriented programming to create Java applications. On the other hand, XHTML-based languages are more intuitive and easier to use, however, for more elaborated tasks, including media synchronization, script languages are necessary. Given the shortcomings of using Java-based languages and XHTML + script, specially for non-programming users, authoring tools can be very helpful.

JAME Author[2], *Cardinal Studio*[3] and *AltiComposer*[4] are authoring tools to create iTV (*interactive TV*) applications based on DVB-J [3]. Although these tools allow the use of an imperative language, they follow the declarative archetype, showing up the importance of this programming paradigm in developing digital TV applications.

JAME Author intends to facilitate the construction of programs providing an abstraction composed of pages, similar to Web pages (*page-based services*), in which the viewer can navigate. Although this tool is useful for creating programs, it does not allow the specification of spatial and temporal synchronism among media, and the interactivity is restricted to the definition of page references.

---

[1] http://java.sun.com
[2] http://www.jame.tv/
[3] http://www.cardinal.fi
[4] http://www.alticast.com

Cardinal Studio is an intuitive authoring tool that has *acts* as its top level abstraction. *Acts* are used to structure an application and they can be considered as program scenes. *AltiComposer* uses a model consistent with the film and TV industry, in which a *scene* has *planes*, and a *plane* contains *shots* and *actors*. In *AltiComposer* and in *Cardinal Studio* synchronism and interactivity specifications demand writing a script code.

Declarative languages focused on the specification of media synchronism have also been considered as alternatives for authoring digital TV applications. In these languages, spatial and temporal relationships can be defined without detailing the algorithmic implementation as opposed to Java and XHTML + script. Among such languages we can highlight SMIL (*Synchronized Multimedia Integration Language*) [4], a W3C (*World Wide Web Consortium*) standard for authoring interactive multimedia documents for Web and mobile devices, and NCL, adopted in the Brazilian Terrestrial Digital TV System.

GRiNS [5] and *LimSee2* [6], authoring tools for SMIL, have integrated views (temporal, spatial, textual etc.). In *GRiNS* the temporal view is the most important one and it can be used to compose and manipulate the presentation of a document on a time axis. In *LimSee2*, the temporal and the spatial view together are powerful to create applications. Although *GRiNS* and *LimSee2* prioritize temporal synchronization, their temporal views do not represent interactive and adaptative documents.

HyperProp Editor [7] is an authoring tool for creating applications in NCL language profile. In this editor, which was used as the base to develop *Composer*, the abstractions are also defined in different views. Nevertheless, in its temporal view it is only possible to place media objects on a time axis and the specification of deterministic temporal synchronism (user interactions and content adapting are not supported, similar to SMIL-based editors). Besides that, *HyperProp* demands a high level of NCL expertise from authors.

Adobe Flash[5] is a commercial authoring tool widely used for creating interactive content, e.g. Web pages, games and movies. In *Flash* the synchronism and the interactivity specification is possible through writing ActionScript [8] codes.

In summary, for all tools cited in this section, when they provide a way to specify synchronism and interactivity, they do not represent unpredictable events on the timeline. Moreover, these authoring tools do not support live editing (*GRiNS*, *LimSee2* and *Adobe Flash* were not designed for the TV context), an important requirement in interactive digital TV systems.

3 Digital TV Authoring with Composer

The efficient creation of interactive digital TV programs depends on many factors. Among these, it is desirable that the specification language adopted has a high expression power. Besides that, it is important to minimize the specification complexity, allowing authors to focus on the creative part of the process. Declarative languages make advantage of their high level of abstraction. When their abstractions match the author's intentions, they are ideal as specification tools. As temporal and spatial

[5] http://www.adobe.com/products/flash/

relationship specifications among media assets, content and presentation adaptations, and support for multiple exhibition devices are usually the focus of iTV applications, a declarative language should provide support for these facilities. This is the case of NCL that, in addition, provides support for imperative scripts (Lua[6] scripts in the case of SBTVD-T) in order to enhance its computational power. Textual specification of NCL documents (iTV programs) is supported in *Composer* through its textual view.

However, as before mentioned, *Composer* enhances the advantages of NCL offering high level abstraction graphic views, making easier the authoring process.

In *Composer*, there is an associated DOM (*Document Object Model*) tree for each NCL document. Whenever a view modifies this tree, it notifies an observer that calls interpreters of other views for updating. Interpreters are in charge of creating the abstractions used on *Composer*'s views, as presented in sections that follow.

Composer allows many NCL documents to be edited simultaneously within a project. Figure 1 presents *Composer*'s main window, as well the textual view, the structural view and the layout view. In this figure, number 1 highlights the region where projects and their documents are managed. Number 2 shows the working area in which views are enabled. Number 3 presents an auxiliary area used for informing authors about the authoring process (e.g. error messages). *Composer*'s user interface is still compounded by a menu bar and by a tool bar which offer several common and view-specific functionalities.

The following subsections present the views available on *Composer*, paying more attention on its temporal view.

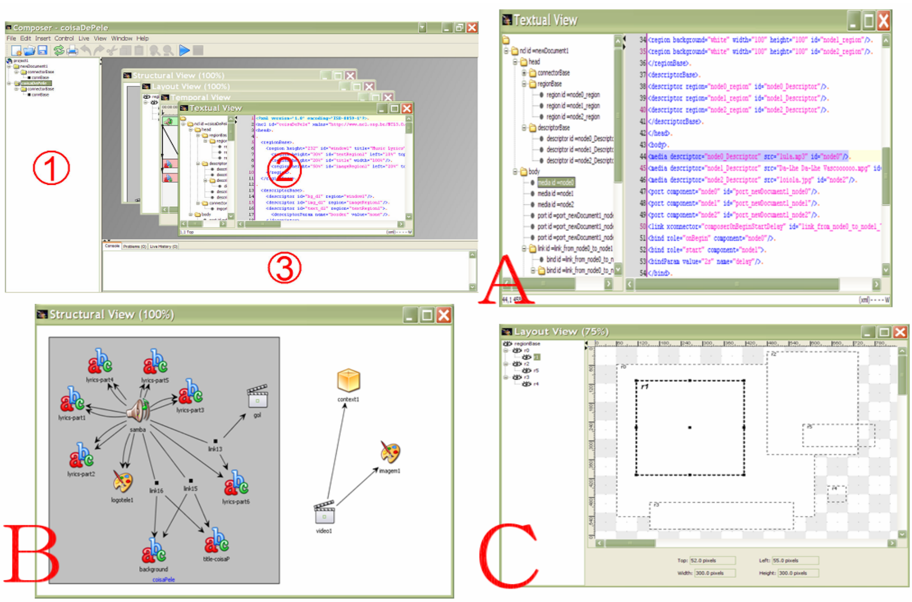


Fig. 1. *Composer* authoring tool and some views

[6] http://www.lua.org/

3.1 Textual View

Although only offering abstractions embedded in the language itself, the textual manipulation can be extremely useful, mainly whether the author wants to explore in depth all language capabilities. *Composer*'s textual view allows authors to create NCL programs by writing textual code explicitly.

The textual view interface is divided into two parts, as shown in Figure 1-A. The left side displays the NCL program as an XML (*eXtensible Markup Language*) tree, in which elements can be expanded or collapsed. The right side presents the NCL code. If an author selects an element on the XML tree, the corresponding NCL code is automatically selected and focused, as illustrated above.

During the textual authoring, an NCL specification is verified based on XML rules, and validated against the XML schemas of the language. The textual view offers several helpful features as auto-complete, printing and source code customization (authors can change text features like font size, font color, space between lines etc.).

3.2 Structural View

The structural view enables authors to create the logical structure of NCL programs. These documents (programs) are logically composed by nodes, which can be compositions or media objects, and relationships among these nodes. Media nodes (NCL objects) represent any specific media content (video, audio, text, imperative code etc.), and compositions group other nodes, which can be media or composite nodes, recursively, and also relationships among these nodes.

In NCL, compositions do not have embedded semantic relationships. Spatial and temporal relationships are defined using links that associate events defined over NCL objects. Events are time occurrences (e.g. the exhibition of a node content, the selection of a node content and the assignment of a value to a node property). In the structural view, links are represented by edges between nodes (composition or media). A media node is illustrated by an icon representing its content. A composition is drawn as a box (cube), and, when it is opened (expanded), as a rectangle, with children nodes and links inside. *Composer*'s structural view is shown in Figure 1-B.

In this view an author can graphically create, edit and delete nodes and links. They can also expand and collapse compositions. When a document is opened, an initial node arrangement is calculated. Nodes can then be dragged to obtain a node placement that improves the program visualization. Any change in the node distribution is saved and associated to the NCL program in order to keep the last node placement when the document is reopened.

3.3 Layout View

The NCL Layout module defines elements and attributes used to specify how media objects will initially be presented in regions of display devices. *Composer*'s layout view (illustrated in Figure 1-C) enables authors to create, edit and delete regions graphically.

The interface of the layout view is divided into two parts. Considering that a region can be specified as child of another region, the left side of the layout view (letter C of Figure 1) shows the hierarchical organization of regions. It must be remarked that

NCL allows more than one display device. The right side presents their spatial representation on a selected display device. It is important to highlight that when a parent region is dragged, all its children are also moved, in such a way to preserve their relative positions.

3.4 Temporal View

The three previous views were inherited, with some changes and simplifications, from *HyperProp* Editor [7]. The fourth view, the temporal view, was only introduced by that editor and it was completely rebuilt and expanded in *Composer*.

The temporal view uses a model called Hypermedia Temporal Graphs (HTG) to represent the temporal behavior of a program. The HTG data structure is formed by directed graphs, which specify time moments for transitions on event state machines. As aforementioned, events are defined by part of a node content (anchor) exhibition, the selection of a node anchor or a value assignment to a node property. Event state machines represent the life cycle of each of these event types, as shown in Figure 2.

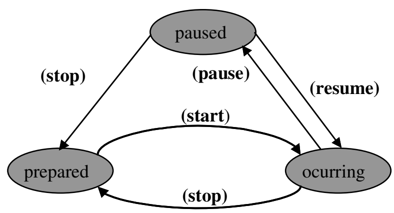


Fig. 2. Event state machine

HTG preserves relationships among events and allows event placements on a time axis, without losing the ability for expressing important structures present in iTV programs, like objects of indefinite duration, interactive events and adaptative contents.

In HTG, graphs are defined by triples (V, A, C), in which:

• $V = (v_0, v_1,.., v_{n-1})$ is a finite set of vertices, where each vertex represents a transition on an event state machine. More precisely, each vertex is represented by a triple: the transition name; the corresponding event type; and the corresponding anchor or property;

• $A = (a_0, a_1,..., a_{m-1})$ is a finite set of edges that individually represent a relationship among transitions. An edge "a" is a pair (v, w) $\in V \times V$, where v and w are called, respectively, source and target vertices;

• $C = \{c_{ij}\}$ is a finite set of conditions associated with edges. A condition c_{ij} is associated with the edge $(v_i,v_j) \in A$ and must be satisfied in order to trigger the target vertex (transition) of the edge.

A simple condition can be defined by a temporal interval that must be spent in order to fire the edge transition; by a variable that must be evaluated in relation to a desired value; or by external actions, such as user interactions. A compound condition can also be defined through logical operators binding two or more conditions.

Target transitions whose conditions are defined by unpredictable actions, such as user interactions, are represented by vertices without output edges. In this case, each transition must also begin another graph, modeling the program behavior from the transition fire on. Thus, when a program contains unpredictable events, a set of graphs is necessary to model the entire flow.

As an example, Figure 3 illustrates the graphs resulting from a program where a video object, called *node0*, is presented during 17.5s (vertex 1 to vertex 3). The video object has a temporal anchor (*anchor0*) whose associated presentation event starts when the video presentation reaches 1.5s (vertex 4). This anchor lasts 5s. The end of its presentation (vertex 5) fires, after 3s, the beginning of an image (*node1*) presentation (vertex 6). During the *node0* presentation, users can interact. This unpredictable event is represented by vertex 2, which also initiates another graph (Graph B). If this interaction happens (vertex 2 of Graph B), the transition representing the end of *node1* presentation (vertex 7) is fired.

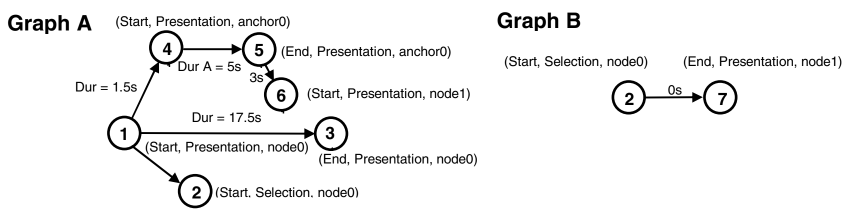


Fig. 3. Hypermedia Temporal Graphs model

Time intervals in each graph can be calculated taking into account the time required to satisfy conditions from the graph entry point to the desired vertex. Therefore, a temporal chain can be defined traversing the graph from its entry point on, in ascending time order. In the temporal view, when a NCL program is opened, the initial chain (called main chain) is graphically illustrated, as in Figure 4, for the example of Figure 3.

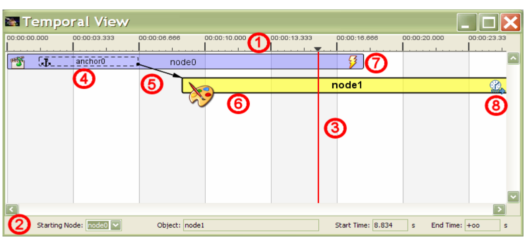


Fig. 4. *Composer*'s temporal view

In Figure 4, the relative time axis (scale) is shown at the top (number 1). In this scale, values are automatically updated depending on the zoom factor chosen by an author. At the bottom (number 2) there is a status bar containing fields that show the name of a selected object and its initial and final presentation times. Since an NCL program may have different entry points, the status bar allows an author to change the graph entry point. A time bar (number 3) is provided in order to facilitate the visualization of what is happening in a specific time.

Still in Figure 4, the temporal anchor *anchor0* (number 4) and its relationship (number 5) with the beginning of the presentation of an image object (number 6) are also shown. The video object (*node0*) also contains an icon (a "lightning") representing an unpredictable event (number 7): the start of the selection event represented by vertex 2 in Figure 3. Finally, the image object *node1* contains an icon representing a "clock" (number 8), which indicates that this object has an undefined duration.

Composer allows authors to simulate the occurrence of unpredictable events in its temporal view. In the HTG model, when an unpredictable event occurs, the transition vertices that model the occurrence are merged. As a consequence, the secondary time chain starting from this vertex on is added to the main chain. Figure 5 shows the graphical representation of the resulting main chain of the example of Figure 3 after the selection event of *node0* at 15s.

Using the temporal view, any time chain (either resulting from the simulation of unpredictable event or not) can be played from any starting point, through the NCL Formatter [1] integrated to the *Composer* tool.

In the temporal view, relationships among events can be edited dragging the corresponding edge ending points. For example, in Figure 5 the author could drag the image *node1* in order to modify its presentation start time related with the end of the anchor *anchor0*. Of course, when events change, HTGs are updated, as well as the textual specification in NCL.

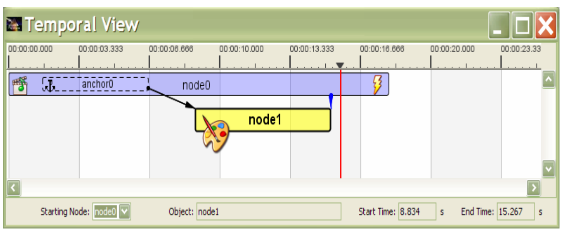


Fig. 5. Temporal view after interactivity simulation

Relationships in the temporal view can also be removed or inserted graphically. To remove a relationship, an author must simply select its graphical representation and delete it. To insert predictable relationships among events two dialog boxes are offered. One allows authors to create relationships to trigger pause and resume transitions, and another one allows authors to create relationships to trigger start and end transitions. To define relationships triggered by unpredictable events, the dialog box shown in Figure 6 is offered.

Using the dialog box shown below, an author can specify the remote control button (Key) that will trigger the action or if the action will be triggered by a pointing device. The author must also specify the NCL object (Relative to) and its point of interaction, that is, its anchor (Anchor). The anchor can be the whole content of the NCL object,

Fig. 6. Specification of interactive relationships

as in the figure. Finally, the author must specify the NCL object (On) to which the specified transition (Action) will be applied. The transition could be start, end, pause or resume.

3.5 Extending Composer's Views

Although the existing *Composer*'s views offer fine abstractions to easy the authoring process, authors could prefer other specific ones. For this sake, *Composer* provides an API (*Application Programming Interface*) that allows third-party developers to create new views "on demand", which can be coupled to *Composer* without modifying the existing views.

In the current *Composer,* implemented in Java, the API is offered as an abstract class that may be inherited by new views. When the *Composer* authoring tool is launched, it checks, on a specific directory set in *Composer*'s configuration menu, for new view code files. After loading these files, new features (e.g. menus and toolbar) are automatically incorporated to *Composer*'s interface. In particular, the view name is added as a new item to the *Composer*'s view menu.

Composer encapsulates how its views interact among themselves, as introduced in Section 3, using a mediator module. When a specific view modifies a document, and as a consequence its DOM tree, the mediator process is notified, each view interpreter is then called, and all views are updated.

4 Live Editing in Composer

In *Composer*, a program can be edited on-the-fly. Live editing can be performed on the broadcaster-side as well as on the viewer-side (end-user). If the live editing mode is activated, edit commands are monitored and generated accordingly to the interface defined in [2].

On the broadcaster-side, as live editing commands are created, they are presented in the message area (region 3 in Figure 1). In a conventional digital TV system, the content provider is in charge of broadcasting the audio and the video streams multi-plexed with other data. Reference [2] discusses how DSM-CC (*Digital Storage Media*

Command and Control) object carousels and event descriptors are used to transport live editing commands in MPEG-2 transport system. Note that live editing commands issued by the server-side are sent by broadcast to all receivers tuned in the corresponding channel.

On the viewer-side, when integrated to a TV middleware, *Composer* tries to extend the traditional television-watching experience allowing end-users to enrich the broadcast content. New contents can be added, new links etc., without changing the original content received. Such an enriched content may be shared with other viewers within a distributed peer group, using for example a return channel as the means to transport editing commands. Figure 7 presents a general view of living editing with *Composer*.

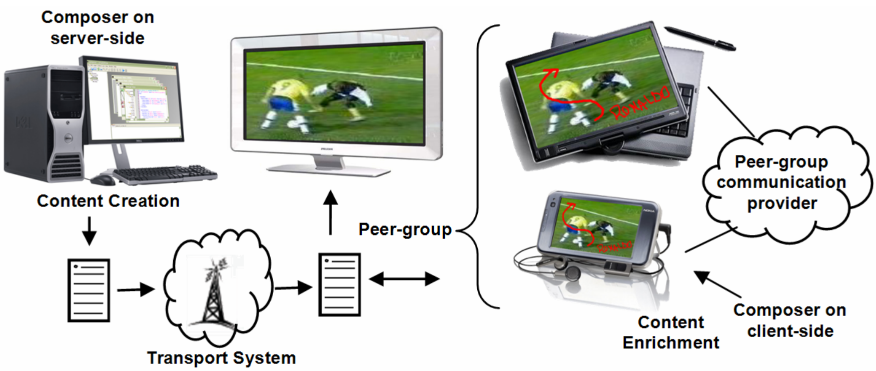


Fig. 7. Live editing with *Composer*

Note that the *Composer*'s module for live editing is almost identical, either implemented in the server-side or in the client-side. The unique difference is the network used to transport editing commands to remote receivers.

5 Conclusions

This paper presents *Composer*, an authoring tool conceived to make easier the design of iTV programs in SBTVD-T. In *Composer,* several graphical abstractions are defined creating different document views. Among these views, the temporal one offers a graphical solution for specifying temporal relationships on a time axis, without the restrictions imposed by the timeline paradigm. The proposed solution, which uses temporal graphs as its data structure, allows the design of interactive and adaptive programs preserving as much as possible the timeline paradigm.

Composer also supports live editing operations. On the server-side, *Composer* can dynamically construct live editing commands, in agreement with the SBTVD-T standard. These commands can then be transmitted in order to modify on-the-fly an interactive TV program. On a viewer-side, editing commands can be used to enhance program content. These enhancements can then be shared as community content.

When not in live production, *Composer* allows the simulation of unpredictable events during the authoring phase. In addition, a simulated execution of an entire program (with or without simulated events) can be commanded. Moreover, if necessary, an author can choose any specific temporal point to start the presentation.

Composer is available on several platforms and is freely distributed[7] under an open source license. The *Composer*'s source code has received some important contributions from the open source community. This paper's authors hope that several contributions could still come from this open source community regarding the design of new views, and that very soon a repository of shared third-party views can be available.

References

1. Soares, L.F.G., Rodrigues, R.F.: Nested Context Language 3.0: Part 8 – NCL Digital TV Profiles. Technical Report – Department of Informatics – PUC-Rio. n. 35/06 (2006) ISSN 0103-9741
2. Soares, L.F.G., Rodrigues, R.F., Costa, R.R., Moreno, M.F.: Nested Context Language 3.0: Part 9 – NCL Live Editing Commands. Technical Report – Department of Informatics – PUC-Rio. n. 36/06 (2006) ISSN 0103-9741
3. ETSI European Telecommunications Standard Institute, TS 102 819 v.1.3.1: Globally Executable MHP (GEM) Specification 1.1.1 (2007)
4. W3C World-Wide Web Consortium: Synchronized Multimedia Integration Language - SMIL 2.1 Specification (2005)
5. Bulterman, D.C.A., Hardman, L., Jansen, J., Mullender, K.S., Rutledge, L.: GRiNS: A GRaphical INterface for creating and playing SMIL documents. In: WWW7 Conference, Computer Networks and ISDN Systems, Brisbane, Australia, vol. 30, pp. 519–529 (1998)
6. Deltour, R., Layaida, N., Weck, D.: LimSee2: A Cross-Platform SMIL2.0 Authoring Tool. In: The European Research Consortium for Informatics and Mathematics – ERCIM News. n. 62 (2005)
7. Soares, L.F.G., Rodrigues, R., Muchaluat, D.: Modeling, Authoring and Formatting Hypermedia Documents in the HyperProp System. ACM Multimedia Systems Journal 8(2) (2000)
8. Williams, M.: ActionScript Coding Standards. In: Macromedia White Paper (2002)

[7] http://www.softwarepublico.gov.br/dotlrn/clubs/ginga/

Ubiquitous Interactive Video Editing Via Multimodal Annotations

Maria da Graça C. Pimentel[1], Rudinei Goularte[1], Renan G. Cattelan[1], Felipe S. Santos[1], and Cesar Teixeira[2]

[1] Universidade de São Paulo, Brazil
[2] Universidade Federal de São Carlos, Brazil
{mgp,rudinei,renan,fss}@icmc.usp.br, cesar@dc.ufscar.br

Abstract. Considering that, when users watch a video with someone else, they are used to make comments regarding its contents – such as a comment with respect to someone appearing in the video – in previous work we exploited ubiquitous computing concepts to propose the *watching-and-commenting* authoring paradigm in which a user's comments are automatically captured so as to automatically generate a corresponding annotated interactive video. In this paper we revisit and extend our previous work and detail our prototype that supports the *watching-and-editing* paradigm, discussing how a ubiquitous computing platform may explore digital ink and associated gestures to support the authoring of multimedia content while enhancing the social aspects of video watching.

1 Introduction

Typical video[1] authoring tools may be considered not worth learning by typical users because the tools have not been designed for non-expert users: they assume trained users with complex authoring tasks at hand. By specialized authoring tools we mean, for non-interactive video, simple tools such as *Movie Maker*[2] and *iMovie*[3], as well more sophisticated ones such as *Premiere Pro*[4]. Properly trained users are also the target of interactive video authoring tools such as the one proposed by Ursu et al. [25], and the one available to author documents for the Brazilian Interactive TV System: the NCL Composer [9].[5]

There are many opportunities for non-specialists to produce interactive video if we can take advantage of the following scenario:

[1] In this text we use *video* to refer to digital (non-interactive) video, and *interactive video* to refer to digital interactive video.
[2] http://www.microsoft.com/windowsxp/using/moviemaker
[3] http://www.apple.com/imovie
[4] http://www.adobe.com/products/premiere
[5] http://www.ncl.org.br/index_.html

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 72–81, 2008.

Watching a video is a task many users are comfortable with. Moreover, when users watch a video, they have many opportunities to (pause the video and) make comments with respect to the video – for instance when the video is one they have captured themselves.

The area of ubiquitous computing investigates alternatives for providing services to users in a transparent way – that is, without taking their attention from their main task [26]. Three recurring and complementary themes in ubiquitous computing research are natural interfaces, context awareness, and capture and access applications. The term *capture and access* refers to *the task of preserving a record of some live experience that is then reviewed at some point in the future* [1].

Considering the scenario outlined above, we previously proposed that *watching-and-commenting* tasks be automatically captured to generate a corresponding annotated interactive video [21]. In this paper, we generalize that approach by proposing and demonstrating the *watch-and-edit* (WaE) paradigm, and generating interactive video in NCL, the Brazilian Digital TV Standard for declarative interactive programs.

In this paper, we first present related work (Section 2), followed by discussions on our previous experience with multimodal annotations and with ink playback (Section 3). We then detail a prototype that creates an interactive video as a result of a user *watching-and-editing* a video file (Section 4). We finish by presenting our final remarks and perspectives of future work (Section 5).

2 Related Work

Kirk et al. studied the use of video production by a varied user population, observing that their main aim is usually to share the video in the capture device itself (in the case of portable phones, for instance) and that they usually do not use the editing options embedded in the devices themselves. They also observed teenage users who opted by explicit capture (with digital cameras) and some level of editing (with computer software) [16]. It is worth noting Kirk et al.'s observation that the point is not whether users are or not able to use a particular software, but whether they perceive it as useful tool to learn and deploy.

The literature reports several efforts targeted at non-expert users, as in architecture proposed by César et al. [8]. While we exploit an ubiquitous capture and access approach, several authors report using explicit authoring techniques (e.g., [14,11,15]). Targeting at the home user, Hua and Li's LazyMedia [14] aims at facilitating the editing and sharing of home videos using techniques such as content analysis in association with authoring composition and presentation templates. In an earlier work, Girgensohn et al. [11] analyse video contents, for instance in terms of the type and amount of camera motion, to select frames which are presented to the user, who can then use a drag-and-drop approach to position the desired video in a storyboard. Video and music content analysis are also used by Hua et al. to automatically collect and align video clips, photographs, music and lyrics to compose a Karaoke video [15].

The use of metadata to support authoring video has been also reported (e.g., [18]), in particular in the context of the TV-Anytime [6]. User and context information has also been discussed (e.g. [10]), including efforts to improve TV Based Communication [13].

While our proposal aims at achieving authoring via transparent interaction, other authors opt to the explicit manipulation of tangible objects representing video artifacts. In the multi-user Tangible Video Editor reported by Zigelbaum et al. [27], users edit video clips by manipulating active handheld computers embedded in plastic cases; in the mediaBlocks system children capture, edit and display media by manipulating passive wooden blocks [24].

Processing the digital video to support authoring has been extensively investigated. In fact, Truong and Venkatesh provide a survey on techniques used to process video information to generate video summaries [23].

Digital inking systems as we propose in this work, are part of many ubiquitous computing platforms. The user's natural interaction via pen, gestures, audio, video or sensors, for instance, can be captured so as to transparently produce associated multimedia documents that can later be reviewed in an integrated and synchronous manner. Focusing on the desktop, systems such as ScreenCrayons [19] and OneNote[6] provide a universal note taking facility.

Several capture and access applications automate video capture and produce some related artifact (e.g. [2]). In particular for home scenarios, the Video Family Archive [3] exploits the semi-automatic annotation of video files. In our work we exploit explicit annotation, which means the manipulation of the video is explicit but we believe it is one the user is comfortable with.

Liu and Chen [17] address explicit and implicit correlations among various media streams in a composite document. As in our work, their system considers temporal and spatial relationships to allow the replay of a video (corresponding to a captured lecture) as a synchronized multimedia document. The main difference is that their techniques, such as the adaptable handwriting that groups handwriting and text for example, are applied during the capture phase while we focus on both capture and reviewing phases.

A few projects focus on helping users to develop an understanding of complex information by exploring the relationships underlying digital ink annotations, its representation and related medias. Bulterman elucidates the requirements for an environment supporting user-centered analysis of annotations [4], supported by work on a SMIL[7] player, the Ambulant Player, and its companion Ambulant Annotator, aimed at allowing interactive multimedia SMIL documents. In work targeting interactive digital TV, Bulterman and his colleagues address the viewer-side enrichment of multimedia content [5,7]. We move toward this goal by giving the user increased control for customizing the review of annotations.

In the context of the interactive video, authoring tools include VideoClix[8], which allows users to segment and track objects in a video stream, as well as add

[6] http://office.microsoft.com/onenote
[7] http://www.w3.org/AudioVideo
[8] http://www.videoclix.com/technology/software/overview.html

annotations and metadata. VideoClix is an authoring tool based on annotation of video objects. The NCL Composer [9] allows the authoring of declarative documents via alternative views (structural, temporal, layout and textual). In our work we exploit the capture and access ubiquitous computing approach to provide the user with a transparent authoring while she watches and comments the video.

3 Capture and Access: Previous Work

In previous work we deployed an extensible software infrastructure to investigate many problems in ubiquitous computing applications [20]. For example, *iClass* is a system able to record a variety of information produced during a traditional lecture or seminar, including ink strokes and slides presented in an electronic whiteboard, as well as audio & video, so as to produce, at the end of the lecture, an XML document integrating the different media streams. As in many pen-based applications, *iClass* allows users to review the ink annotations as static HTML pages, or to play it back so that the ink is presented in a speed equivalent to that in which it was written and, when audio or video was captured, they are played synchronously.

We have also explored the ubiquitous computing paradigm to associate user annotations to video streams. We designed the *Multimedia Multimodal Annotation tool*, which supports annotations in two complementary methods: context-based metadata association and content enrichment [12]. The tool comprises a multimodal interface which allows both live video capture and annotations. Annotations can be made by means of pen-based digital ink or via voice recognition and are, in both cases, converted to text. At the end of the capture process, XML documents are generated as a composition of references to all captured media: video, audio, images, slides, ink strokes and text.

Our experience has given us the opportunity to observe situations where intermediary stages and derived representations of a document annotated with digital ink might be relevant to users. We then formalized *Inkteractors*: operators that can be applied to ink strokes so as to allow the generation of documents containing alternative views of the original interaction process [22]. Inkteractors collapse or expand annotated pages into one or more derived versions we call *snapshots*. We implemented the *InkPlayer* tool to allow a user to interact, by selecting the desired operators, with digital ink-annotated documents while reviewing them. The aim is to produce alternative views of the inking activity according to the user task at hand, which is achieved by processing the underlying representation of the ink.

4 Watch-and-Edit (WaE) Tool Support

Leveraging up our previous experience, we have proposed elsewhere that the *watch-and-comment* (WaC) paradigm be supported by a capture and access platform [21]. In this paper, we generalize that approach by proposing the *watch-and-edit* (WaE) paradigm and by demonstrating our current prototype tool that

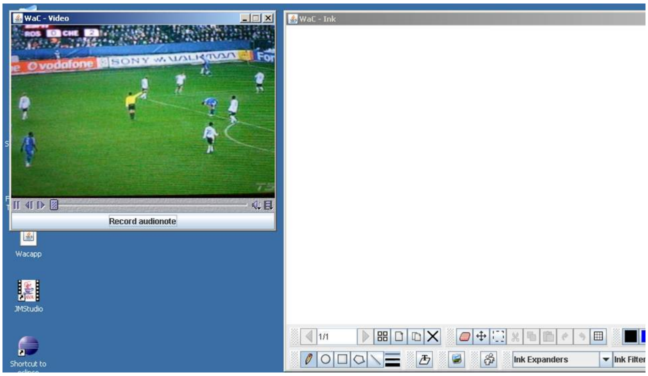


Fig. 1. Left: Video playback window; Right: Ink window for pen-based annotation

allows a user, as a result of a *watch-and-edit* session, to generate a piece of interactive video in NCL. An important feature, from the digital rights perspective, is that editings and annotations are kept separate from the original media, i.e. they can be distributed independently from the video.

4.1 Watch-and-Edit: Input

The WaE tool has options to open an existing video file in several formats – we use Java Media Framework and its supported formats. We are working on a version with options for opening a folder containing a set of pictures and to capture images directly from a camera. We are also working on a version that allows the video to be open associated with previous annotations – so one can edit and extend existing annotations.

When the user starts the tool and selects a video file: the WaE tool presents two panels as illustrated in Figure 1: the *playback window* for video (left), and the *ink window* for pen-based annotation (right). The usual buttons for *play/pause/stop*, as well as a *Record audionote* button, are located under the playback window.

Once the video is loaded, the user starts a *watch-and-edit session* by pressing the *play button*, which causes the video to start in the playback window (Figure 1). At any moment during the playback of the video, the user may capture audio and/or ink notes, or add *skip* indicators.

To create an *ink comment*, the user taps on the playback window which causes a copy of the corresponding video frame to be copied into the ink window, as shown in the example in Figure 2 (left).

Whenever a frame has been grabbed and is presented in the ink window, the user can make pen-based annotations such as free-form drawings as illustrated in Figure 2 (right), or a handwritten note such as someone's name in the scene; it is also possible to include typed text or other images (such as a photo).

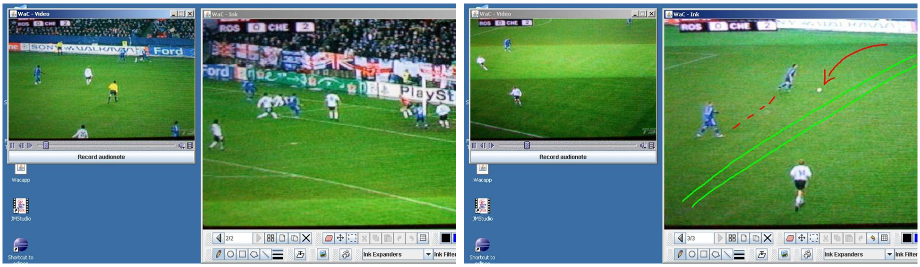


Fig. 2. Left: A user tapping on the *playback window* causes a copy of the current frame to be presented on the *ink window*; in this example, the video continues its playback (which explains why the frame in the *ink window* is different from the frame in the playback window). Right: Once a frame is available in the ink window, the user can make many types of annotations.

The *skip operation* allows a user to indicate a portion of video that can be skipped when the interactive video is later watched. Using our tool, when the user identifies the start of such a portion, she taps on the left bottom corner of the playback window; when the end of the video portion to be skipped is reached, the user taps on the right bottom – both tapping operations are indicated in Figure 3 (left).

In the current version, we shift by 3 seconds the skip time indicated by the user: the idea is to acknowledge that, when the user taps on the playback window to indicate the skipping, the intended start and end times have been, in fact, presented a few seconds earlier.

To record an *audio comment*, the user selects the *Record audionote* button at the bottom of the playback window (as shown, for instance, in Figure 3 (left)): the button label changes to *Stop recording* (as shown in Figure 3 (right)) to indicate that the audio is being recorded; a tap on the *Stop recording* button stops the recording and the button label changes back to *Record audionote*.

When the user decides to stop, an *annotation file* is saved containing the corresponding XML-based markup for the annotations (e.g., ink coordinates and time information), as well as references to the audio file containing all the recorded comments.

4.2 Watch-and-Edit: Output

The tool automatically generates an interactive digital TV specification in the form of a NCL document. Presentation of the resulting interactive video in an interactive digital TV platform, as emulated by the Ginga-NCL Player, is achieved as follows:

– at each occurrence of an *ink comment*, a *thumbnail* of the corresponding annotated frame is presented on the bottom right corner or the video window (as illustrated in Figure 4 (left)) and associated with one of the buttons on the remote control: if the user presses that button, the video is paused and

Fig. 3. Left: The user taps on the left and right bottom corners of the playback window to indicate, respectively, the starting and ending points of a portion of video to be skipped when later reviewed as an interactive video. Right: Clicking on the *Record audionote* button activates the recording of audio comments and changes the button label to *Stop recording*; a second click on the button stops the recording and the button label changes back to *Record audionote*.

the annotated frame is presented (as illustrated in Figure 4 (right)) until another control remote button is pressed to resume video playback;

– at each occurrence of a *audio note*, an *audio icon* is presented on the upper right of the video (as illustrated in Figure 5 (left)) and associated with one of the buttons on the remote control: if the user presses that button, the volume of the original audio is lowered and the audio commentary is played back – the original audio is resumed at the end;
– at each occurrence of a *skip note*, a *skip icon* is presented on the bottom left of the video (as illustrated in Figure 5 (right)) and associated with one of the buttons on the remote control: if the user presses that button, the portion of video indicated by the user is skipped.

Although much processing can be done to extract data from the video (and audio) streams to provide useful information for a viewer of the interactive digital TV, the watch-and-edit paradigm, as illustrated, can be also exploited to enrich the user interaction. The inkoperators, for instance, have already been exploited in the skip example above. For the sake of illustration, other operators that are under construction in the current prototype include:

– FilterByAuthor( ) and FilterByAuthorRole( ): used to show icons indicating annotations of a particular user or class (role) of users, which applies when more than one person has contributed with annotations to the same original video;
– FilterByArea( ): used to indicate annotations associated with specific portions of the screen.

Fig. 4. Left: A thumbnail of the annotated frame is presented to indicate that an ink note is available. Right: The ink note is presented when the user selects the proper button on the digital TV remote control.

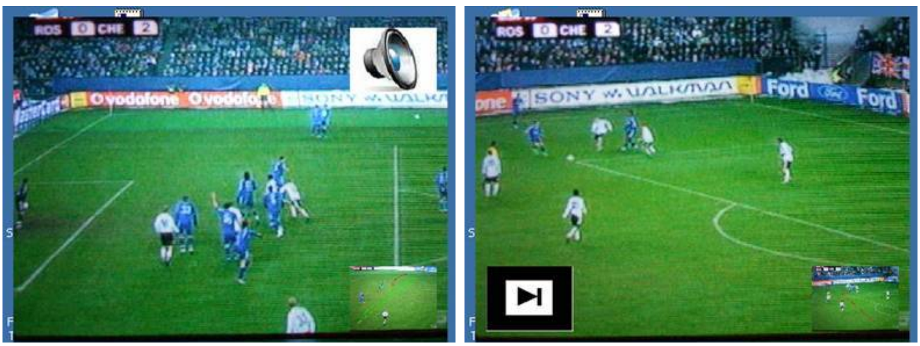


Fig. 5. Left: An audio icon indicates that a audio comment is available and is played back when the user selects the proper button on the remote control. Right: an skip icon indicates that a portion of the video can be skipped if the user selects the proper button on the remote control.

5 Conclusions

Authoring new media from existing content may be considered a difficult task for the average user because it demands specialized tools that are themselves not trivial to use. By mixing concepts of ubiquitous computing with video annotation, we propose the *watching-and-editing paradigm* for the transparent authoring of interactive multimedia content. We illustrate our proposal by describing our current prototype and discussing opportunities to exploit digital ink and associated gestures for authoring multimedia content.

The ubiquitous aspect of our system refers to the integration of automated capture tools and natural pen-based input with a (soon expected to be widely available) digital TV platform, presenting novel ways a user can seamlessly author interactive video streams. Similarly, context information relative to metadata extracted directly from the capture session can be used as interaction alternatives in the interactive digital TV. For example, *who* made the annotation (e.g., a menu with the authors can be always available and, if selected, causes the video to skip until the annotation from the given author is found), *when* the annotation was made (e.g., menus may link to the point in the video containing annotations in the listed day/month), *where*, if some sort of location information is available, it can be used for providing options for navigation on the video.

Moreover, our approach enhances the social experience of watching a video with others. When watching a video in groups, making comments with respect to its content is a practice many people enjoy and feel comfortable with. Our current prototype allows collocated users to collaborate in a *watching-and-editing* session via voice and ink comments. We are working on the implementation of a version that supports remote collaboration and content sharing via peer-to-peer connections.

Another short term plans include having a version of our prototype that also generates SMIL documents and is robust enough to perform user evaluations. As future work, we are currently defining a layered model for content enrichment and developing a device-to-device communication scheme for off-line multimedia distribution.

Acknowledgments

We thank the following organizations: FINEP, FAPESP, CAPES and CNPq. We thank Dick Bulterman and Luis F.G. Soares for great discussions on this topic. Renan G. Cattelan is a PhD candidate supported by FAPESP (03/13930-4). Felipe S. Santos is a MSc candidate supported by FAPESP (06/53783-9).

References

1. Abowd, G., Mynatt, E., Rodden, T.: The human experience. IEEE Pervasive Computing 1(1), 48–57 (2002)
2. Abowd, G.: Classroom 2000: An experiment with the instrumentation of a living educational environment. IBM Systems Journal 38(4), 508–530 (1999)
3. Abowd, G., Gauger, M., Lachenmann, A.: The family video archive: an annotation and browsing environment for home movies. In: Proc. ACM SIGMM Intl. Work. Multimedia Information Retrieval, pp. 1–8 (2003)
4. Bulterman, D.: Using SMIL to encode interactive, peer-level multimedia annotations. In: ACM DOCENG 2003, pp. 32–41 (2003)
5. Bulterman, D., César, P., Jansen, A.: An architecture for viewer-side enrichment of tv content. In: ACM MULTIMEDIA 2006, pp. 651–654 (2006)
6. Butkus, A., Petersen, M.: Semantic modelling using tv-anytime genre metadata. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 226–234. Springer, Heidelberg (2007)
7. César, P., Bulterman, D., Jansen, A.: The ambulant annotator: empowering viewer-side enrichment of multimedia content. In: ACM DOCENG 2006, pp. 186–187 (2006)
8. César, P., Bulterman, D., Obrenovic, Z., Ducret, J., Cruz-Lara, S.: An architecture for non-intrusive user interfaces for interactive digital television. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 11–20. Springer, Heidelberg (2007)

9. Costa, R., Moreno, M., Rodrigues, R., Soares, L.: Live editing of hypermedia documents. In: ACM DOCENG 2006, pp. 165–172 (2006)
10. Van den Ende, N., de Hesselle, H., Meesters, L.: Towards content-aware coding: User study. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 185–194. Springer, Heidelberg (2007)
11. Girgensohn, A., Boreczky, J., Chiu, P., Doherty, J., Foote, J., Golovchinsky, G., Uchihashi, S., Wilcox, L.: A semi-automatic approach to home video editing. In: ACM UIST 2000, pp. 81–89 (2000)
12. Goularte, R., Cattelan, R., Camacho-Guerrero, J., Inácio, V., Pimentel, M.: Interactive multimedia annotations: enriching and extending content. In: ACM DOCENG 2004, pp. 84–86 (2004)
13. Hemmeryckx-Deleersnijder, B., Thorne, J.: Awareness and conversational context sharing to enrich tv based communication. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 1–10. Springer, Heidelberg (2007)
14. Hua, X.-S., Li, S.: Interactive video authoring and sharing based on two-layer templates. In: HCM 2006: Proc. ACM Intl. Work. Human-centered Multimedia, pp. 65–74 (2006)
15. Hua, X.-S., Lu, L., Zhang, H.-J.: P-karaoke: personalized karaoke system. In: ACM MULTIMEDIA 2004, pp. 172–173 (2004)
16. Kirk, D., Sellen, A., Harper, R., Wood, K.: Understanding videowork. In: ACM CHI 2007, pp. 61–70 (2007)
17. Liu, K.-Y., Chen, H.-Y.: Exploring media correlation and synchronization for navigated hypermedia documents. In: ACM MULTIMEDIA 2005, pp. 61–70 (2005)
18. Madhwacharyula, C., Davis, M., Mulhem, P., Kankanhalli, M.: Metadata handling: A video perspective. ACM TOMCCAP 2(4), 358–388 (2006)
19. Olsen, D., Taufer, T., Fails, J.: Screencrayons: annotating anything. In: ACM UIST 2004, pp. 165–174 (2004)
20. Pimentel, M., Baldochi, L., Cattelan, R.: Prototyping applications to document human experiences. IEEE Pervasive Computing 6(2), 93–100 (2007)
21. Pimentel, M., Goularte, R., Cattelan, R., Santos, F., Teixeira, C.: Enhancing multimodal annotations with pen-based information. In: Workshop on New Techniques for Consuming, Managing, and Manipulating Interactive Digital Media at Home, pp. 207–213 (2007)
22. Pimentel, M., Prazeres, C., Ribas, H., Lobato, D., Teixeira, C.: Documenting the pen-based interaction. In: Brazilian WebMedia 2005, pp. 1–8 (2005)
23. Truong, B., Venkatesh, S.: Video abstraction: A systematic review and classification. ACM TOMCCAP 3(1), Article 3, 37 pages (2007)
24. Ullmer, B., Ishii, H., Glas, D.: mediablocks: physical containers, transports, and controls for online media. In: ACM SIGGRAPH 1998, pp. 379–386 (1998)
25. Ursu, M., Cook, J., Zsombori, V., Zimmer, R., Kegel, I., Williams, D., Thomas, M., Wyver, J., Mayer, H.: Conceiving shapeshifting tv: A computational language for truly-interactive tv. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 96–106. Springer, Heidelberg (2007)
26. Weiser, M.: The computer for the 21st century. Scientific American 265(3), 94–104 (1991)
27. Zigelbaum, J., Horn, M., Shaer, O., Jacob, R.: The tangible video editor: collaborative video editing with active tokens. In: TEI 2007: Proc. Intl. Conference on Tangible and Embedded Interaction, pp. 43–46 (2007)

Unobtrusive Dynamic Modelling of TV Program Preferences in a Household

Elena Vildjiounaite, Vesa Kyllönen, Tero Hannula, and Petteri Alahuhta

VTT Technical Research Centre of Finland, Kaytoväylä 1, 90580, Oulu, Finland
{FirstName.LastName}@vtt.fi

Abstract. Majority of recommender systems require explicit user interaction (ranking of movies and TV programs and/or their metadata, such as genres, actors etc), which requires user time and effort. Furthermore, often such ranking is done separately by each person, while merging these manually acquired preferences in multi-user environments remains largely unsolved problem. This work presents a method to learn a model of multi-user environment in intelligent home from implicit interactions: the choices which family members make together and separately. In tests on TV viewing histories of twenty families, acquired during two months, the method has achieved prediction accuracy comparable with the accuracy of systems which require explicit user ratings: a set of TV programs, actually viewed during each test session (average set size was 2.2 programs per viewing session), was recommended among five top choices in 60% of cases on average, despite training on small data sets.

Keywords: User Modelling, TV recommender system, Smart Home.

1 Introduction

Development of context-aware intelligent homes can affect recommender systems. Since such homes are capable of recognising people and logging their actions, these logs can be used by recommender systems for reducing user effort and for adaptation to each single user as well as to multi- user situations. On the other hand, the more data is collected by a home, the more privacy concerns arise, and families become less willing to share viewing data. Thus, many researchers (including this work) aim at developing systems which could work based only on data of each single household.

For example, Personalized Electronic Program Guide [1] recommends programs by combining outputs of implicit dynamic model (Bayesian Network, trained on 10 months viewing history), explicit model (constructed from users' ratings), and stereotypical model (constructed from demographic data, also entered explicitly). The system was evaluated on Italian TV viewing data of 62 subjects, and prediction accuracy of the combined user model was very good: MAE (Mean Average Error) between predicted and real user ratings was 0.3. Unfortunately we did not find in the paper the accuracy of implicit dynamic user model compare to the models requiring explicit user input. Furthermore, the system predicted ratings of each test subject separately; prediction of user choices in a multi-user environment was not studied.

TV personalisation system in the work [2] also combines explicit and implicit (Bayesian Classifier and Decision Tree) recommendation methods. The system was

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 82–91, 2008.

trained and tested on the viewing data of eight participants, collected during three years. The participants were asked to rate "would watch", "may watch" and "wouldn't watch" a set of approximately 300 TV shows, which resulted in a set of 1348 rated shows. Average Mean Squared Error was 0.2-0.3, but again, performance of purely implicit user model was not presented; and viewing behaviours when users were alone and when in a company of other family members were not compared.

Current research on modelling multi-user environments is mainly concerned with merging profiles of users whose individual preferences were collected manually beforehand. The work [3] studies how humans try to solve this problem, and presents several fairly different strategies used by the test subjects. Company of a user was taken into account in collaboration filtering– based movie recommender system in work [4]; and creating a group model from subgroups of similar users has shown to improve satisfaction of museum guides usage [5]. The work [6] also proposes a method to merge individually acquired user profiles. The method was tested on 25 users and 200 TV programs, movies and advertisements, and one of the conclusions was that the method worked well for fairly homogeneous groups of users, but members of heterogeneous groups were not satisfied with the recommendations.

Since family members often do not constitute a homogeneous group of users, we propose to build a model of a household as a whole by observing which choices family members make together and separately. With this approach we avoid the need to decide how to deal with contradicting user preferences: we simply follow the family practices, which can differ in different families. The drawback of this unobtrusive approach is that satisfaction of family members remains unknown.

Method of fairly unobtrusive modelling of TV viewing patterns of a whole family was proposed in the work [7]: the system called FIT uses stereotypes-based models and tracks the viewing history. Based on time, FIT infers which subset of family members is watching TV, and recommends programs for the inferred subset. FIT was evaluated on data of three family groups; each one consisted of three similar families. For one group, predictions were created at random; for another group – by FIT, and for the third group the knowledge of who is actually watching TV was used. FIT prediction accuracy in the test was 14% lower than that in the third group. However, the number of programs and program genres, used in this study, was not presented.

The contribution of this paper is the following: first, we present a method to learn a joint model of a multi-user environment by observations of choices made by different subsets of the group. Second, we present and analyse the results of evaluating our approach on TV viewing data of 20 families (62 persons), collected during two months. To the best of our knowledge, this is the first work which presents results of evaluation of a dynamic model of multi-user environment on real life data.

2 Dynamic Unobtrusive Modelling of Household Preferences

Since not all families are willing to share their viewing data, and since ways of resolving conflicting interests of family members can differ between families, we propose to learn household model based on this household data only. We propose to learn dependency of TV programs choice on context (if some preferences are valid in all contexts, this will be learned too), and we express context as a set of context descriptors. An example of context set is: "TimeOfDay: *Morning*; DayOfWeek: *Sunday*; SocialContext: *Adult1, Adult2, Child2*". Components of context-aware TV recommender system are presented in Figure 1. Such recommender system can work purely unobtrusively, and it can also utilise user feedback whenever users give it.

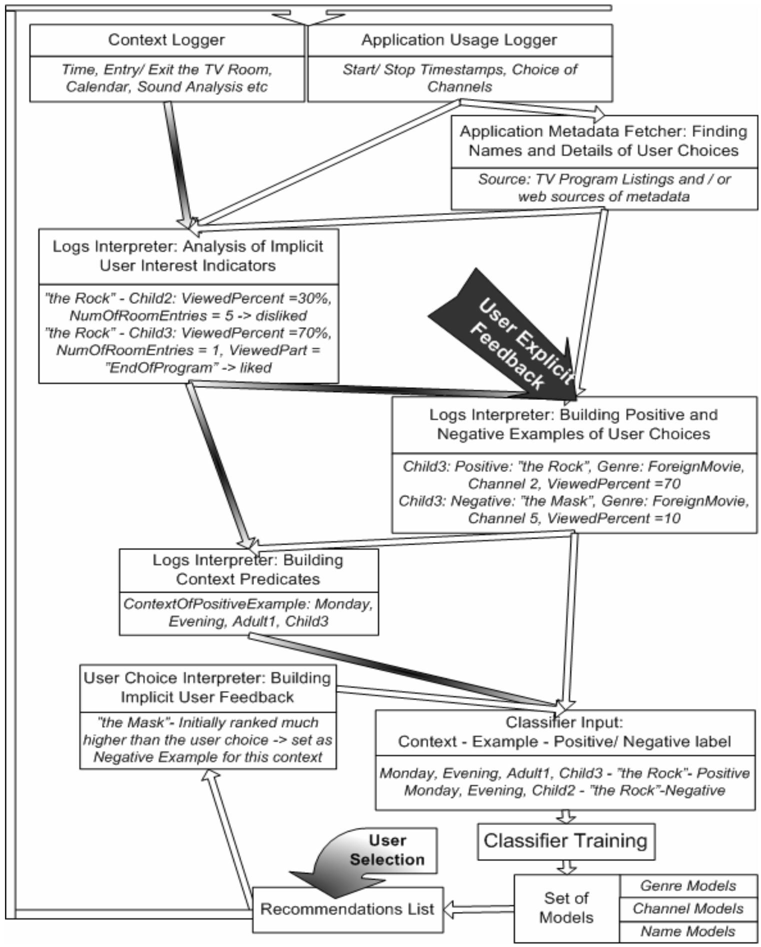


Fig. 1. An overview of unobtrusive context-aware recommender system. The black arrow "user explicit feedback" denotes optional component which requires explicit user interaction.

Context Logger: logs time and changes in physical context (e.g., user's entry/ exit).

Application Usage Logger: each time when users switch to another channel, the channel number is added to the log along with the timestamp.

Application Metadata Fetcher: based on timestamps and channel numbers, finds a program name from TV program listing and retrieves available metadata (genre, actors etc) from a web service. We found, however, that in our non-English speaking TV often only a broad genre category of a program is available, but no other details.

Logs Interpreter – Analysis of Implicit Indicators of User Interest: degree of user interest can be deduced, first, from application context by such indicators as: percent of time a program was viewed and if it was only a part, whether the users started to watch the program and switched to another program, or ended up with watching this program. (Programs which were not viewed at all can be used as negative examples.) Second, indicators can be derived from physical context: number of entries to TV room, approaching user appointments (stop of watching because of the appointment does not mean program dislike), sound analysis etc. In our tests we only had user presence data, and we have not yet evaluated affect of each indicator on prediction accuracy. We decided that viewing less than 40% of a program denotes lack of user interest and over 80% denotes user interest, and we applied a set of common sense-based rules to a combination of other indicators for the range of percents in between.

Logs Interpreter – Building Positive and Negative Examples of User Choices: after estimation of a degree of user interest in an item, this degree needs to be assigned to all metadata descriptors of the item. We used a straightforward approach and assigned the same degree of user interest to all metadata descriptors, although dislike of some program does not necessarily mean that its genre is unsuitable for this context. Additional negative examples were selected from programs, not viewed at all.

Logs Interpreter – Building Context Predicates: we build Context Predicate for each program as a set of semantic context values. Since most systems recognise users' locations at a room granularity, social context is derived from timestamps of entering and leaving a room. We collect into a group of "being together" all users, whose feedback for this item is positive. If feedback of other family members is negative, we treat this case as a separate (negative) example of these persons' preferences.

Building Histories – Classifier Input: history is a list of consecutive viewing sessions; each session consists of context information, users' choices and their labels (positive/ negative). Since for many programs genre is the only available metadata, for each household three histories were built: "name history", "genre history", and "channels history".

Classifier Training: minimising prediction errors on the training data. For each family SVM (Support Vector Machines) and CBR (Case-Based Reasoning) were trained separately on "name history", on "genre history" and on "channel history". We used these classifiers because they have good generalization capabilities (implicit indicators of user interest create noise, e.g., users can watch uninteresting programs or miss interesting ones) and allow different weighting of examples (e.g., if users started to watch a program and switched to another program soon after that, this first program is more certainly negative example than a program not selected for viewing at all).

Set of Models: we trained classifiers to provide recommendations on each metadata item (TV program name, genre and channel) separately in order to allow fast re-training. Thus, classifiers built distinct models for all items: own model for predicting whether the user is interested in "foreign movies" genre or not in different contexts; own model for predictions regarding "channel 1"; own model for predictions regarding "program name 1" and so on. Models were produced only for items viewed more than two times during training period. Altogether we had for each family 20 genre models, from 10 to 15 channel models and from 20 to 250 name models.

Additionally, we compared learning of models in a "global" and in a "local" way. Global way is based on the assumption that all possible choices compete with each other, and reflects the attitude of users who decided to spent time in front of TV and tried to find the best program for this time. Local modeling reflects a goal-driven way to choose activities, e.g., "now I want to see something funny and I am not in a proper mood to see even a famous drama." For high-level metadata descriptors (genres and channels) only global modeling is feasible, but for differentiating between choices below these generic descriptors (program names in our case) we tested both ways.

Recommendations list: outputs of all models are combined in order to rank all available choices. The ranked outputs of all name models are combined into one list, and similarly another ranked list of genre – channel combinations is created and used for ranking of programs. After that outputs of both lists are normalised and combined again. In these experiments we created separately recommendation lists produced by CBR and recommendation lists produced by SVM because we were interested in pros and cons of each method, and we combined the outputs of SVM and CBR only after that. SVM recommendations are created by feeding current context to all SVM models, so that each model outputs "to recommend" or "not recommend" decision for the current context, and also ranks this decision. Recommendation with CBR is done by comparing current context with all cases in the viewing history. Most similar cases are retrieved along with their ranks; ranking is based on the similarity between current and stored contexts and on frequency of occurrences of a metadata item in the history. Similarity between contexts is based on Cosine measure (1), commonly used in Information Retrieval (C1 denotes current context, and C2 denotes stored context).

$$\boldsymbol{sim}\ (\mathrm{C1,C2}) = \frac{\sum_{i=0}^{N} W_i \ * Sim(C1_i, C2_i)}{\sqrt{\sum_{i=0}^{N1} C1_i^2} * \sqrt{\sum_{i=0}^{N2} C2_i^2}} \tag{1}$$

User Choice Interpreter – Building Implicit User Feedback: after the user selects for viewing some items from the recommendations list, comparison between the ranks of the selected items and the ranks of the top recommendations shows how well classifiers predicted the user opinion. After the user makes choices, the whole process of logging physical and application contexts starts again, and we add user choices to the history as soon as logs analysis labels them as positive or negative. We employ the following strategy for implicit user feedback estimation: we add the top non-viewed recommendations to the history as negative examples for this context if initial ranks of these programs were significantly higher than the highest rank among positive user choices (users can view several programs in one viewing session). Correct ranking is not specially rewarded: we just add new positive examples.

In the process of dynamic modelling, described above, the number of parameters to set when viewing history is created from log files is too large to learn all of them. Often it is needed to rely on common sense of domain experts. Acquiring an explicit user feedback would be of great help, but unfortunately users do not give feedback always. Thus learning of user preferences would anyway employ some set of common sense-based rules of the kind "if the users watched the program and did not complain, perhaps they liked it". In multi-user settings acquisition of explicit user feedback is even more tricky problem because the users often adjust own feedback to opinions of others if they want to be polite or are afraid to look in disagreement with others. Thus, unobtrusive modelling makes sense not only because it significantly reduces user effort, but also because it is not easy to make the user effort to be really useful.

3 Experiments

The proposed unobtrusive dynamic modelling approach was tested offline on real life in-house data of 20 families (62 test subjects altogether), collected during two months in a normal course of family lives by Finnpanel OY [8]. Social context (presence of family members near TV) was acquired via remote control, with each member of the family having his or her own push button on the remote control. Family members were asked to push the button each time they entered or left TV room. Application context consisted of channel numbers and timestamps of switching to these channels.

We used TORCH machine learning library [9] (with a Gaussian kernel for SVM), and we trained and tested classifiers on the data of each family separately.

3.1 Data and Experimental Protocol

Data pre-processing was done as described in Section 2 and resulted in over 5000 viewing sessions, contained 12500 viewed programs (average number of programs, viewed for longer than 40% of the program duration during one session, was equal to 2.2). The term "viewing session" in this study denotes a statement in a form: timeslot - time context - social context - list of viewed programs along with their channels, watched percents and metadata. (In this study the only available metadata was quite broad genre categorisation: 20 genres such as "foreign movies", "domestic movies", "educational and science programs".) "Social context" is a set of people (family members and guests) who watched TV. "Timeslot" is a time period when TV is on and no changes in social context occur: we attempted to generate a new recommendation list each time when social context changed, even though in real life people often join other viewers without checking what's going on in other channels. "Time context" includes both day of week and time of day (semantic) information.

Over 35% of sessions were viewed in a multi-user environment. Significantly uneven distribution of viewing sessions during data collection, caused by weather changes, presented additional challenge: average number of viewing sessions during first six weeks of histories was 120, while during last two weeks it was equal to 70.

Experiments were conducted for each family in a following way: classifiers were trained on first two weeks data, and the remaining 1.5 month data was used for testing and incremental learning as follows: from the first viewing statement in the data which was not yet used in training we took semantic time and social context, and for this context each classifier created a ranked list of programs for the next two hours from the starting timestamp (we feel it to be long enough for sitting in front of TV). Recommendations were compared with the list of actually viewed programs, prediction accuracy was estimated, and new examples were added to the training data.

Recommendations ranked significantly higher than the new positive examples were added as additional negative examples. After that classifiers were re-trained, and the process continued.

We tested Sum and Voting fusion methods to combine recommendation lists created by different models and classifiers. Sum fusion sums up normalised ranks from each list. Voting method sums up the places of recommended items.

3.2 Experimental Results

We present here results in a form of "recall-in-top-N": percent of relevant recommendations among N top items. In our data average number of positive examples per viewing session was equal to 2.2, and in all families this number was close to average. Thus we were most interested in "recall-in-top-five" (so that average precision would be approximately 45%). "Recall-in-top-five" of different classifiers, calculated for each family and averaged over 6.5 weeks of incremental learning, is presented in Figure 2. It shows that the difference between families is fairly large, and that success of predictions largely depends on success of name-based models (with a few exceptions). Thus unobtrusive dynamic modelling approach, proposed in this paper, learns regular activities quite well if the metadata allows to differentiate well between the items. Fairly poor prediction accuracy of genre-channel models (and consequently little improvement by fusion of outputs of name-based and genre-channel based models) is due to lack of metadata: programs were categorised in 20 genres, and there were no other program details in our data. Nevertheless prediction accuracy for all families was significantly better than that of random predictions.

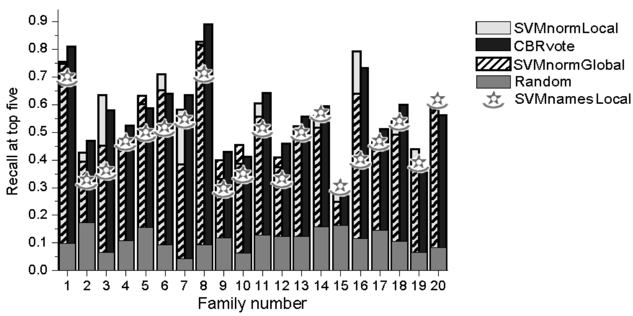


Fig. 2. Averaged over 6.5 weeks recall at top five for each family. *SVMnormLocal* stands for fusion of local name-based models with global genre-channel models with Normalized Sum method; *CBRvote* stands for fusion of global name-based and genre-channel based models with Voting method; *SVMnamesLocal* stands for local name-based models.

An interesting exception is family number 16: for this family recall of name –based models was almost twice lower than that of genre-channel models, and consequently the recall of the merged recommendations. Close look at this family data revealed significant differences in genre preferences of family members' subsets. Thus we can conclude that the proposed method works well for heterogeneous groups of users.

Figure 2 shows two overlapping bars for SVM, for comparison of global and local learning approaches in name-based models of SVM. SVM model is a decision surface between positive and negative examples. In local approach negative examples are taken from the programs of the same group (same genre for name models), while in global approach – from any group. In all CBR models and in SVM genre and channel models we used global approach. Figure 2 shows that for families with good accuracy of predictions local approach was better than or similar to a global approach, while for three families with bad prediction accuracy the global approach outperformed the local one significantly. Probably viewing habits are less consistent in these families, and thus leaning of their models is more difficult. (Other four families with bad prediction accuracy had 1.5–3 times shorter histories than the other families).

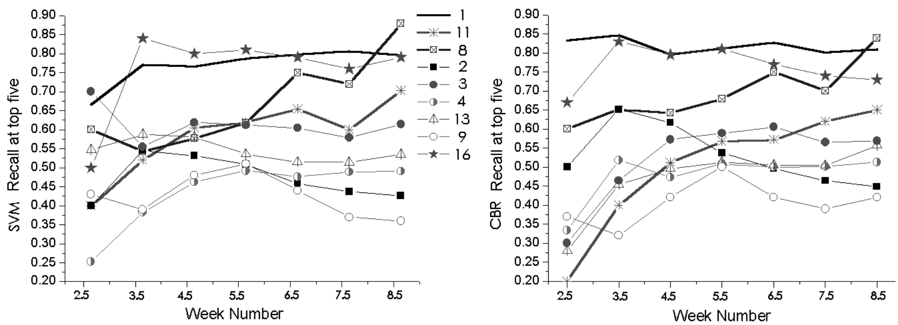


Fig. 3. Time dependency of prediction accuracy in incremental learning by SVM and CBR. Numbers in the legend correspond to families' numbers in Figure 2.

Figure 3 presents learning curves for some families, which show that already after one month of learning prediction accuracy is not too bad. It is seen that for many families prediction accuracy drops at week number 7.5, which was caused by weather changes (when the weather is bad, people watch TV longer and more randomly).

Figure 4 presents an average recall over all test cases in the viewing histories. It shows very similar accuracy of SVM and CBR. Figure 4 also shows that name-based models can predict almost 60% of viewed programs: surprising finding because

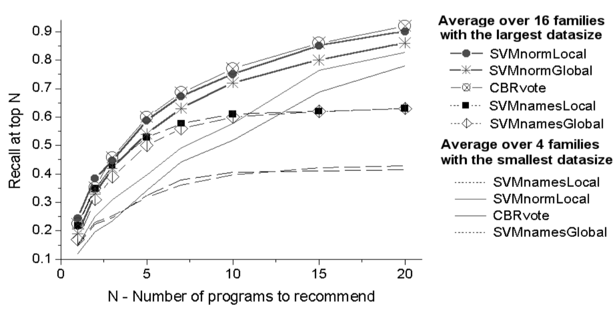


Fig. 4. Recall at top N for different N: the thin lines present an average over 4 families with the smallest size of data; the lines with symbols present an average over the other 16 families

dynamic modelling in TV recommender systems is usually based on genre and channel information [6]. Figure 4 also shows that local approach outperformed global approach, which is a pleasant discovery because local approach is faster to re-train.

4 Conclusions and Future Work

Despite an active research on recommender systems, providing recommendations for multi-user environments remains a problem [6]. This work presents a method to learn models of multi-user environments unobtrusively by observations (the method is capable of using explicit user feedback as well). This method has the following advantages: first, it adapts to family practises (which depend on a family and a culture) instead of enforcing a practice chosen by an application designer; and second, it does not force the users to invest time and effort into ranking TV programs or to specify preference values for programs genres, actors and other metadata descriptors. This method has disadvantages as well: first, learning requires time; second, as in any purely unobtrusive approach, satisfaction of family members remains unknown. On the other hand, acquisition of explicit user rankings in heterogeneous multi-user environments is a problem in itself because often people adapt their ranks to ranks of others and thus true user satisfaction can remain unknown anyway.

Regarding the problem of learning time, the experiments on real life TV viewing data of 20 Finnish families have shown that for 50% of families two months time period of incremental learning is sufficient for achieving decent recommendation accuracy: 60 - 90% recall among top five recommendations. Lower prediction accuracy for four of the other families was caused by significantly smaller size of data (1.5 – 3 times shorter viewing histories) than data of the other families. Reasons for lower prediction accuracy for the rest of families can be less regular viewing habits and higher importance of metadata details. Additionally, in all families an average number of viewing programs per day was significantly higher in the end than in the beginning of the data collection period, and this change in viewing habits (caused by weather changes) presented additional challenges for machine learning methods.

Unfortunately we are unaware of any other works presenting results of unobtrusive modelling of multi-user environments and testing them on real life data. We were able to find only four works to compare our numbers; in each of these works explicit user effort was required. First, PTV [10] is a web system which creates personalised program listings for individual users. PTV combines content-based and collaborative filtering and was evaluated as “good” by 61% of its users. This result suggests that 60% of successful recommendations is an acceptable number for current state-of-the-art. Second, the work [11] presented collaborative filtering – based system with precision in top five recommendations by different methods in a range 65% - 75%, while precision of random recommendations was 35.6%. Our method has shown similar accuracy for families with sufficiently long viewing histories, but outperformed random recommendations by factor of 4.5 on average (which means that our dataset was more challenging). Third, FIT [7] is the system which builds a model of household preferences by combining stereotypes-based and dynamic modelling. Additionally, FIT requires users to specify their preferred watching times. In tests with 6 families (data size is not described in the paper) FIT outperformed random algorithm by factor of 4.2 on average, which is similar to our unobtrusive modelling results. Fourth, in the work [6] accuracy of explicit profile merging algorithm was presented as 75% recall at 45% precision (we attempted at predicting 2.2 viewed programs among top five recommendations, thus our precision was equal to 44%). However, members of heterogeneous groups were not satisfied with the recommendations in the work [6], while our method suits well for such groups.

Usability of recommender systems with not-so-high prediction accuracy remains to be studied. We think that such accuracy could be acceptable, provided that the users are given means to quickly correct the system's outcomes, in applications which automatically record (and delete non-viewed programs after some time period) possibly interesting TV programs to a home computer or to a user's mobile device. Recording to a mobile device is feasible before a user's trip because on the one hand storage capacity of mobile devices constantly grows, and on the other hand wireless connections are still unavailable in many places, e.g., in airplanes or in the wild.

Nevertheless we consider our first experimental results encouraging. In our tests average recommendation accuracy for multi-user environments (about 25% of sessions were viewed by two persons together, and about 10% of sessions were viewed by three and more persons together) was close to that of single-user viewing sessions, despite significantly smaller amount of multi-user training data, which confirms the feasibility of learning a joint model of multiple users.

References

1. Ardissono, L., Gena, C., Torasso, P.: User Modeling and Recommendation Techniques for Personalized Electronic Program Guides. In: Ardissono, L., Kobsa, A., Maybury, M. (eds.) Personalized Digital Television, Targeting programs to individual users, pp. 3–27. Kluwer, Dordrecht (2004)
2. Zimmerman, J., Kuparati, K., Buczak, A., Schaffer, D., Gutta, S., Martino, J.: TV Personalization System. In: Ardissono, L., Kobsa, A., Maybury, M. (eds.) Personalized Digital Television, Targeting programs to individual users, pp. 27–52. Kluwer, Dordrecht (2004)
3. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. User Modeling and User-Adapted Interaction 14, 37–85 (2004)
4. Adomavicius, G., Sankaranarayanan, R., Sen, S., Tughilin, A.: Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. ACM Trans. Inf. Syst. 23(1), 103–145 (2005)
5. Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: Personalized Recommendation of Tourist Attractions for Desktop and Handset Devices. Applied Artificial Intelligence 17(8-9), 687–714 (2003)
6. Yu, Z., Zhou, X., Hao, Y., Gu, J.: TV program recommendation for multiple viewers based on user profile merging. User Modelling and User Adapted Interaction 16, 63–82 (2006)
7. Goren-Bar, D., Glinansky, O.: FIT-recommending TV programs to family members. Computers & Graphics 28, 149–156 (2004)
8. Finnpanel: People meter data, `http://www.finnpanel.fi/english`.html
9. TORCH, http://www.torch.ch/
10. Smyth, B., Cotter, P.: A personalized television listings service. Communications of the ACM 43(8), 107–111 (2000)
11. Adomavicius, G., Kwon, Y.: New Recommendation Techniques for Multicriteria Rating Systems. IEEE Intelligent Systems 22(3), 48–55 (2007)

An Ethnographic Study on Recommendations in the Living Room: Implications for the Design of iTV Recommender Systems

Regina Bernhaupt, David Wilfinger, Astrid Weiss, and Manfred Tscheligi

ICT&S Center, Universität Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg Austria
{Regina.Bernhaupt, David.Wilfinger, Astrid.Weiss, Manfred.Tscheligi}@sbg.ac.at

Abstract. Today recommendations are used to improve the quality and the number of interactive TV services offered by providers all over the world. Their main goal is to recommend TV shows and TV events, sometimes combined with an automatic recording function. With the growing number of IPTV offers, the usage of recommendation systems will increase and help to personalize and individualize the TV viewing experience. This article studies recommendations for the home context investigating daily living habits and routines in 40 households in depth using playful and creative cultural probing. Study results are present design recommendations for the development of new forms of recommendation systems. Main results find that users prefer individualized recommendations in-time, either automated or user-oriented, but mostly prefer individualized recommendations for each member of a household rather than personalized for the whole household.

Keywords: interactive TV, IPTV, recommendation system, ethnography, playful probing, design.

1 Introduction

Digital video recorders (also known as DVRs or PVRs) are becoming commonly available in households worldwide. Today they include most often the functionality of a recommendation system – also called 'Tivo-ing' in the US, coming from the known Tivo recommendation system [13]. This work investigates what aspects should be taken into account, when introducing recommendations in the German speaking market. One special focus is the possible integration of recommendation systems together with individualized TV experiences. An example of this individualized TV experience use is finger-print recognition to identify the user of the system [18].

1.1 Recommendation Systems for Interactive TV

Recommendation systems are representations of user preferences used for suggesting items which should be purchased or examined by the user. Originally defined as systems to which people provide recommendations as input that are then distributed to other appropriate people by the system [17], recommendation systems today have a

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 92 – 101, 2008.

broader connotation. They describe “any system that produces individualized recommendations as output or guiding the user in a personalized way to interesting and useful objects in a large space of possible options” [5]. Recommendation systems are used in various (Internet) services, supporting and increasing sales of various products (e.g. Amazon), assisting in all aspects of entertainment selections. An example of such use is the music recommendation system of iTunes. What separates recommendation systems from other search engines is the user-oriented notion of an “individualized” service and the perception of being useful and interesting to the user of the system [5].

In general two recommendation strategies are dominant today: content-based recommendation systems relying on rich content descriptions of items and collaborative filtering recommendation strategies (also called social filtering) relying on human ratings of the content [16]. From the user perspective, the level of influence on the recommendation system also has to be taken into consideration. Fully automated recommendation systems learn the users profile based on the TV viewing habits, whereas other systems allow the user to actively specify their profiles [10].

With the introduction of new technical ways of TV distribution like IPTV and new forms of (input) devices and services (remote controls including a fingerprint reader to solve personalization issues [18] or PVRs), a (repeated) success of a variety of recommendation systems seems realistic in the area of interactive TV and IPTV. Various examples of TV recommendation systems have been developed to support users in their daily selection of TV content enabling the reception of recommendations via E-mail, SMS, WAP or MMS. The number of TV recommendation services as an integral part of interactive TV is also steadily growing. One of the most commercially successful recommendation systems is Tivo. Anecdotes from Tivo usage [21] continue to show that recommendation systems still need improvements to really address user needs.

(TV) Recommendation Usage in Daily Life

Recommendations can be found almost everywhere in everyday life. To inform the design and development of a new recommendation system and to investigate possible (new) areas for recommendations in iTV/IPTV services, we studied daily recommendation habits and recommendation usage and related concepts like automation, personalization and security. We wanted to see to what extent people today benefit from processes that are computed by the system based on their preferences, what aspects of peoples leisure time should be supported (including various forms of entertainment) and to what extend recommendations play an integral role in daily life.

We describe in sections two and three the ethnographic study including its methodology. Section four shows the findings related to concepts like automation and recommendation and section five presents the general design implications and development recommendations based on the results.

2 Method

For deducting design implications for TV recommendation systems, we investigated the usage of recommendations in every day life and in the special context of the living room, to find out how services and devices can support the existing structures in the area of iTV/IPTV.

The study of technology usage in the home has increased considerably during the last 10 years. Ethnographic studies [1, 7] have become an important method in human-computer interaction to investigate technology usage and acceptance in the home context [4, 2]. An overview can be found by Haddon [9]. To gather additional in-depth data during these ethnographic studies, various methods have been developed and improved to address appropriately the context home. Gaver et al. [8] invented cultural probing, which were followed by various other probing methods [12] like technology probes [11] or playful probing [3].

We used a methodological variation of cultural probing for the study improving creative and playful aspects of the probing method [2, 3]. The selection of the method was first motivated by the fact that two different research areas should be investigated namely recommendations and the area of personalized and individualized services and security. Second, the goal was to gather user designs for various aspects of recommendation systems and third, we wanted to focus on family lives including children. The probing method consists of a introduction interview, the ethnographic period of two weeks (including the later described packages) and an in-depth final interview with household members after the study.

The probing method uses a package of (creative and playful) material to support the self-observation of participating households. The package includes so called creative cards for each research topic (see Figure 1). Typically, one research topic is addressed during one week of the study. An example for one of the questions on these creative cards is related to the household's preferences on "recommendations for me or for the whole household". The study participants then can either comment on the question, draw some aspects they feel are important or use the additional material (like cameras) to document their behavior and opinions. To support the playful experience and the integration of children in the study, at least one of the questions addresses a task where modeling clay should be used. Additionally, a card game was developed including research related questions. A detailed description of the playful probing methodology is given in [2], a similar study set-up can be found in [15, 3].

Fig. 1. CCP cards used for the study

3 Study Setup

The ethnographic study was conducted during August 2007 with 40 participating households within a larger project on the development and evaluation of a new (automatic) recommender system together with Siemens Nokia Networks. We only recruited households with at least two household members to enable the study of critical multi-person usage of recommendation systems and related topics like personalization and security. In total 126 participants (55 female, 71 male), with 23 children up to the age of 18 years, 15 boys and 8 girls participated. Households represented a typical demographic sample of Austria in terms of living area, income and education. The study lasted for two weeks with various activities for the participating households. Besides the given probing material described above, 20 households received an additional multifunctional device with photo, video and audio recording functionality, while the other 20 households received an additional analog single usage camera.

In a first household visit we explained to the participants the content of the probe package, then a first short narrative interview was conducted to investigate the prior knowledge of the participants regarding the research topics. During the two weeks, participants filled in the creative probing cards and documented their daily media usage with photo and video material. During an in depth final interview, the probing materials were discussed with the participating households.

Participants were asked to record important aspects of their daily life concerning recommendations and other topics given by the creativity cards. Additionally all households received modeling clay and were asked to model devices and artifacts they invented for certain functionality. 20 randomly chosen households received a card game that is well known in the study area (called "Uno") which was adapted by the researchers with questions concerning the usage of recommendations and recommendation systems. We also included post-it stickers and pens in the probing material to enable the households to communicate other ideas that may not fit onto photos or creative cards.

Another part of the study included the documentation of the TV viewing behavior and the usage of recommendations based on a system under development at Nokia Siemens Networks. We asked all households to log their TV viewing habits, providing a website with a form where household members typed in the TV program watched. This data was used to find out about the viewing habits of the participating households and also for generating recommendations based on TV viewing. These (automatically generated) recommendations were then sent to the selected households in an email to see what impact they would have in the households.

4 Study Results

Before describing the study results, we want to emphasise that the ethnographic study results are presented in terms of qualitative data using grounded theory [19] to cluster responses and using a reporting technique inspired by ethnographic field notes [6]. The high number of participating households is related to the methodological development of the creative and playful probing aspects. Neither the methodological analysis is part of this publication, nor are the investigated aspects on personalization, individualization and security.

Learning from Recommendations in everyday life

When asking people in what aspects of their daily life do recommendations play an important role, we can see that almost all areas of living are affected by recommendations. Most often, book recommendations were named (10) followed by health care related recommendations (9) holiday (8) or travel (6) recommendations. Six (6) households stated that recommendations concerning the TV program play an important role in their everyday life. Participants mentioned recommendations concerning profession and work such as newspaper articles and schedules, concerning leisure time and hobbies (e.g. hotel or fashion tips) and recommendations for cultural issues like operas and plays. Another important aspect of everyday life, family and friends, were also involved in recommendations as important source for them.

When asked for the sources of these recommendations, social relations (friends (14), acquaintances (9)) and newspapers (14) were mostly named. Important in this context to note is that participants used mainly sources that had a social connotation (friends) or that were coming from independent institutions. Sources like advertising (6) are named, but were commented as less “trustworthy”.

Recommendations for interactive TV thus should reflect the social aspect or indicate clearly the independence of the recommending authority from other (less trusted) sources(Guideline 1).

Participants were also asked what recommendations they would like to receive from their TV systems. The most frequently named items were strongly TV related as 49 statements (multiple answers were possible) were related to broadcasts and 13 to movie ratings. Other features were also strongly related to other leisure time activities such as events (20) or travel and holiday tips (6) indicating that participants wish recommendations that fit to their present situation, mainly leisure and relaxing when spending time in the living room in front of the TV. Only one participant mentioned job offers as a desirable function.

Interactive TV recommendations should, in general, stay within the entertainment oriented services, and should not be extended to work-related topics (Guideline 2).

Insights on TV Recommendations

A central question was also how novel or random recommendations should be in relation to the data that was gathered to create the recommendation. Our study confirmed Herlocker et al. [10] in terms of participants expressing their wish to receive new recommendations for TV viewing, not TV shows they are already familiar with. This leads to the fact that recommendations have to be given with care knowing that users expect novelties but also information that fits their taste. There is a thin line between useless recommendations because they are already known and the ones that are too far away from the personal needs and taste. Recommendations have to be something one would not find for oneself easily. An already known advantage of recommendations is that they give confidence towards the system as they are a proof for its reliability. This is important but should not outnumber the unfamiliar recommendations as this would make the recommendation system useless.

HH21[1]: "*I partly liked the recommendations although recommendations that I would have seen anyway are no recommendations for me. I also did not like recommendations for 1 o'clock am and for shows I already knew.*"

HH22: "*Sometimes recommendations proposed a good alternative program, very interesting and informative, one could then very well identify with the recommendations.*"

Recommendation system should carefully balance novelties with information that fits the user taste and usual TV usage (Guideline 3).

Another aspect discovered in the ethnographic study concerning TV recommendation was the necessary self-reflection on the TV viewing behaviour. As humans tend to suppress bad conscious for example because of too much or too late TV viewing, the recommendations may bring these feelings back to peoples minds. These unpleasant feelings will be negatively reflected towards the system that provides recommendations and should therefore be avoided. The design of the user interface should allow to mask such negative emotions, allowing the user to delete recommendations or change settings (Guideline 4).

Findings on Design Implications for Recommendation Systems

To further inform the design of recommendation systems beyond existing work [20], we asked participants to what extent automatically performed tasks play a role in their daily life. We wanted to learn how people perceive these with respect to TV, and how to best support their mental model of these automatically performed tasks within the design.

When asking the participants what in their households should work automatically answers were mostly time motivated. People want to save time in every kind of way. Especially regarding things they don't enjoy doing. Over 60 percent of the participating households would like to have some automatically performed tasks for cleaning, washing and cooking. TV is not at all a key technology that is seen as something to be automated.

For design, we then summarize that the user's mental models of automatic tasks are not related to TV. Design must take into account this fact, and thus inform the user about what the (automatic) recommendation process is about, how to influence this process, and how to access the results from this process. Design of recommendation systems must help the user to learn about any form of automatic recommendation, he should be guided in how to access this kind of information, and be reminded about this functionality whenever the system is taking an automatic action. (Guideline 5).

Recommendation systems try to give personalized recommendation. In front of the TV typically several people might want to receive their recommendation. There is existing work on recommendations for modeling groups [14]. To investigate this aspect from the user side, we prompted participants to comment on recommendations to be completely individualized (per person) or personalized (per household). Interestingly enough, households preferred individualized recommendations arguing that in different interests of the household members (19), formats that might not be reasonable for special users in a group (children) or the higher probability to recognize the recommendation.

---

[1] As the whole study was conducted in German, the quotations used in this paper are translated to English.

HH14: *"Everybody has different interests, also the generation gap is important and the different times of the day one is watching TV. Therefore it is an advantage as recommendations are more suitable and personalized."*

Fifteen households preferred group recommendations arguing that with a better overview on the recommended programs the TV set is a joint possession and interest.

HH19: *"One knows, what could be interesting for the others. This helps to watch TV together, there is also more conversation about the recommendations."*

The aspect of individual vs. group recommendation was also controversially discussed within the final interviews. To help support contextual influenced usage, the design should clearly indicate if only group recommendations are given (so the users can adopt to the possibly as "wrong" perceived recommendations). Based on the users comments we recommend to use an individualized recommendation system (for each identified user) and to additionally allow the system to be personalized for user groups (e.g. by using a family profile that can be selected whenever the whole family is watching TV). Overall design must clearly communicate the intended user or user group of the recommendation (Guideline 6).

One of the cards used as probing material included the request to draw the TV set and the way recommendations should be given to the viewers. Different answers were given, but most households saw recommendations on the TV as something only visually appearing on the screen (25), six (6) mentioned acoustic signals in combination with visual feedback and one (1) household preferred only acoustic information. Four (4) households wanted to receive their recommendations on other media such as the mobile phone.

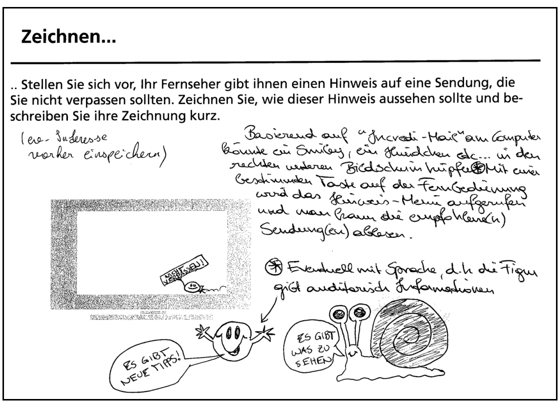


Fig. 2. Example of Creativity Card with Drawn Recommendations (in German)

Cards like the one depicted in Figure 2 indicate that as TV viewing is still an activity that is strongly connected with fun and amusement. Therefore the recommendations received should also be friendly, funny and likeable. In this example the participants would like to get the notification whenever the system has any form of new recommendation and then decide to press a button to get more information or not. For recommendation systems working with an automatically learned user profile that can not be influenced by the user, design should then be informative about new recommendations as soon as they are started to be recorded by the system (Guideline 7).

Following Recommendations
20 households received recommendations via email, which began one week after the beginning of the study when the households had logged enough TV shows to create recommendations for them. For this purpose we used both a commercial recommendation system and recommendations created by researchers. The recommendations contained ten TV shows for a maximum of three days in advance and were sent via email. Every day there was an update sent for the next day. Overall 811 recommendations were given and after comparing recommended and viewed broadcasts, it appeared that only 17 recommended programs had been really watched. This discrepancy can be explained with various facts: the media break – the recommendations were not easy available (stored in the e-mail inbox), when the TV program selection was made (in front of the TV), missing awareness of participants to note what shows where really seen on TV (as logging was a cumbersome task, including reflection on what has been seen on TV).

We learned that usage of recommendation systems should be studied with a real prototype in the field, and that the low-fidelity prototype (including a media break) was not helpful to investigate real usage of recommendations.

5 Summary and Interpretation

Studying recommendations in daily life, we could find interesting habits, opinions and views on TV recommendations. In general recommendations do accompany us in our everyday life in all aspects from health to education, from entertainment to buying decisions. Recommendation systems can be either automatic or based on social recommendation. What is important is the clear information of the user about the abilities and limitations of the system. Based on the insights we have formulated the following guidelines that can help improve (automatic) recommendation systems.

- Recommendations for interactive TV should reflect the social aspect or indicate clearly the independence of the recommending authority from other (less trusted) sources.
- Interactive TV recommendations should in general stay within the entertainment oriented services and should not be extended to work-related topics.
- Recommendation system should carefully balance novelties with information that fits the user taste and usual TV usage.
- The design of the user interface should allow to mask negative emotions evoked from the “mirror effect”, allowing the user to delete recommendations or change settings.
- Help the user to learn about any form of automatic recommendation, guide the user on how to access the recommendations and help the user remember about the functionality whenever the system is taking an automatic action.
- Clearly communicate to the intended user or user group of the (automatic) recommendation system.
- Inform about a new recommendation as soon as it is started to be recorded by the system.

To summarize, this study presents qualitative findings on current recommendation usage in middle Europe and presents guidelines for the design of recommendation systems. The results are currently used to implement a recommendation system used within a larger iTV pay-service for middle Europe. We will continue our research on recommendations with a larger field trial to investigate possible contextual factors of the recommendations. We are currently working on the integration of individualized recommendations in a system using fingerprint recognition to identify TV users to enable integration of contextual influences in a final system.

Acknowledgements

We would like to thank Martin Altmaninger, Sarah Kepplinger, Philipp Losbichler and Michael Pirker for their engagement and commitment during this study, doing the difficult part of data collection and transcription. Special thanks go to Nokia Siemens Networks for enabling this study due to their financial support and making their automatically learning recommender system available.

References

1. Atkinson, P., et al. (eds.): Handbook of Ethnography. Sage, London (2001)
2. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tscheligi, M.: Trends in the living room and beyond. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 146–155. Springer, Heidelberg (2007)
3. Bernhaupt, R., Weiss, A., Obrist, M., Tscheligi, M.: Playful Probing: Making Probing more Fun. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 606–619. Springer, Heidelberg (2007)
4. Blythe, M., Monk, A.: Notes Towards an Ethnography of Domestic Technology. In: Proc. DIS 2002, pp. 277–281. ACM Press, New York (2002)
5. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. J. of User Modeling and User-Adapted Interaction 12, 331–370 (2002)
6. Emerson, R.M., Fretz, R.I., Shaw, L.: Writing Ethnographic Fieldnotes. University of Chicago Press, Chicago (1995)
7. Fetterman, D.M.: Ethnography: step by step, 2nd edn. Sage Publications, Thousand Oaks (1998)
8. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural Probes. Interactions 6.1, 21–29 (1999)
9. Haddon, L.: The Contribution of Domestication Research to In-Home Computing and Media Consumption. The Information Society Journal 22, 195–203 (2006)
10. Herlocker, J., Konstand, J., Terveen, L., Riedls, J.: Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems 22(1), 5–53 (2004)
11. Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B.B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., Eiderbäck, B., Lindquist, S., Sundblad, Y.: Technology Probes: Inspiring Design for and with Families. In: Proceedings of CHI 2003, pp. 17–24. ACM Press, New York (2003)
12. Jääskö, V., Mattelmäki, T.: Observing and Probing. In: Proceedings of the International Conference on Designing Pleasurable Products and Interfaces DPPI 2003, pp. 126–131. ACM Press, New York (2003)

13. Kamal, A., Wijnand, v.S.: TiVo: making show recommendations using a distributed collaborative filtering architecture, pp. 394–401. ACM Press, New York
14. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. J of User Modeling and User-Adapted Interaction 14, 37–85 (2004)
15. Obrist, M., Bernhaupt, R., Tscheligi, M.: Interactive Television for the Home: An ethnographic study on users requirements and experiences. In: Proceedings of EuroiTV 2006, Athens, Greece, May 25-26, 2006, pp. 349–358 (2006)
16. O'Donnovan, J., Smyth, B.: Trust in Recommender Systems. In: Proc. of IUI, San Diego, January 9-12, 2005, pp. 167–174. ACM Press, New York (2005)
17. Resnick, P., Varian, H.R.: 'Recommender Systems'. Communications of the ACM 40(3), 56–58 (1997)
18. Ruwido (last visited 11-01-2008), `http://www.ruwido.com/products/voco-media/1/`
19. Strauss, A., Corbin, J.: Basics of qualitative research: grounded theory procedures and techniques. Sage Publications, Newbury Park (1990)
20. Swearingen, K., Sinha, R.: Interaction Design for Recommender Systems. In: Proceedings of DIS 2002, ACM Press, London (2002)
21. Zaslow, J.: If TiVo Thinks You Are Gay, Here's How To Set It Straight – Amazon.com Knows You, Too, Based on What You Buy; Why All the Cartoons? The Wall Street Journal sect. A, 1 (November 26, 2002)

Recommender System for the Multi-channel TV Production

Janez Zaletelj

University of Ljubljana, Trzaska 25, 1000 Ljubljana, Slovenia
janez.zaletelj@ldos.fe.uni-lj.si

Abstract. This paper presents the concept of content recommendations for the production of multi-channel TV shows. Within the IST project "LIVE – Live Staging of Media Events" [1] we are developing a production support system which will have a functionality of content recommendations and will support production of multi-channels programs. The paper outlines a concept of a recommender system for the multi-channel TV production and presents basic architecture and workflows within the system. The recommendation of the archive content for a given channel is personalized by taking into account the profile of the target audience.

Keywords: recommender system, personalization, iTV production.

1 Introduction

Today's TV production workflows are highly standardized and adapted to produce linear, broadcast media format. The production of TV programs typically contains the pre-production, live production and postproduction phases. In the preproduction, the planning of the live event is done and editor selects and prepares background material on the event from the archives. During the live production the director can use the background material to add additional audio-visual clips to the program.

The first problem of the standard workflow is that it does not provide the necessary flexibility which is needed for the coverage of the live event such as sport competition. If something unexpected happens, for example an unknown athlete wins a race, the director does not have the necessary tools to react to this event. Director is not able to search and retrieve additional background information which is needed during live production. The second drawback is that the standard workflow produces a single stream for all viewers, and thus the variety of audience preferences is ignored. The digital TV broadcasting allows for additional flexibility of transmission of multiple channels within one program, and each channel can be targeted to different class of viewers. However such scenario imposes further requirements on the production workflow and equipment.

Within the 6th Framework project "LIVE : Live Staging of Media Events" [1] we are developing new concepts of TV production [4] based on iTV technologies. The key challenges that need to be addressed are how to efficiently support the production

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 102 – 106, 2008.

team to be able to produce several interlinked thematic channels simultaneously, and how to enable reuse of AV material from vast TV archives on the fly, during live production.

The concept of content recommendation system within the TV production is being developed to address the issue of the on-the-fly content selection from TV archives and its reuse during live production. In this paper we start by defining the roles of production recommender system, we outline the different workflows and system architecture of the recommender system, and we present how the recommender system will work in the production workflow. Finally, the concept of personalization of TV programme to the target audience is presented.

2 Roles of the Production Recommender System

One of the main goals of the LIVE project is to develop new tools and concepts to support the TV Director in the authoring of new type of live TV content, which is generated live and substantially influenced by the viewers' preferences [2]. The Director should be able to react in real-time to the live events by including related information and AV material into the program. At the same time, he should be able to review Consumers' feedback on the program, which would influence his future decisions on the program content. These functionalities will be supported by the Production Recommender System.

The main role of the Recommender System will be to support new production workflows of the LIVE system by providing recommendations of content suitable to be included in the program. The Recommender System implements automatic methods to find and recommend archive content, which will be used to help the Director in finding the related audio-visual content from the archives, and will enable him to react to live events. Because of the huge amounts of AV material available in the TV archives, the content selection and finding functionalities of the Recommender System can be also successfully employed during program preparation phase, where the editor would get automatic recommendations of archive content for the given subject.

On the other hand, the Recommender System is the system component through which information links to the Consumer systems are established. The Consumer System is represented by the application running on the consumer's set-top box or other end device, which is connected to Recommender system through suitable communication channel (also called the Feedback Channel). The availability of consumer feedback information is one of the central requirements to the LIVE system. The feedback from each Consumer will be collected, stored and analyzed by the Recommender System services. Analysis of feedback will provide two kinds of information. First, the analysis of the feedback from the single consumer will be performed, which will result in a detailed user model of the Consumer. This information will enable personalized services for the consumer. The second goal of the analysis is to calculate statistical information on the whole TV audience, which will be available to the Video conductor. Feedback from many (or all) of the TV consumers will be taken into account to derive statistical information on the audience.

3 Workflows and Usage of the Recommender System

Recommender System is part of the general LIVE Production Support System. Its main goal is to provide support for the production of live, interactive, multi-channel TV programs. The Recommender system thus provides professional users with information and services necessary to build interactive, personalized live TV formats.

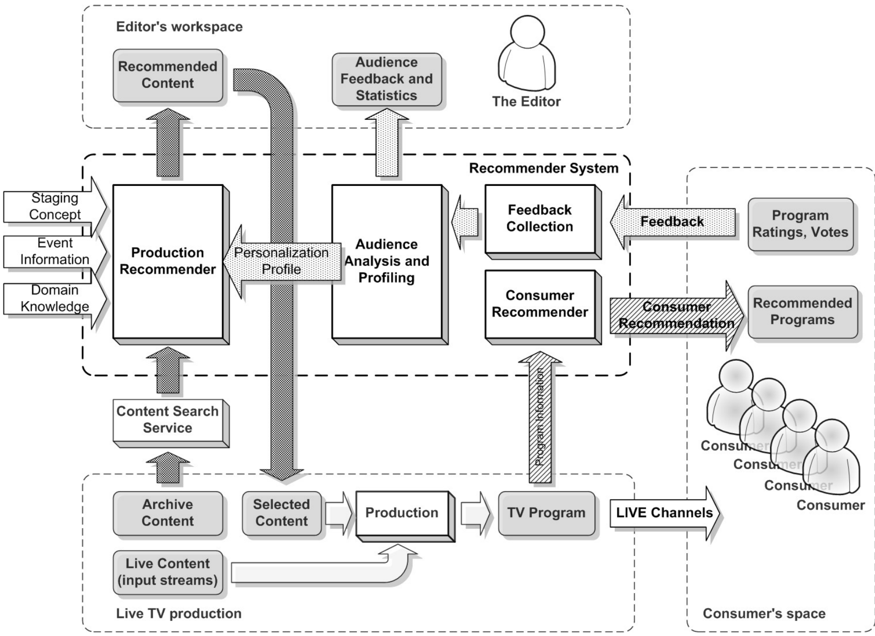


Fig. 1. Architecture and main workflows within the TV Production Recommender system

Main workflows within Recommender System are shown on the Fig.1:

- **Selection and organization of content from the Archive** during live production. Inputs to this workflow are a staging concept, event information, domain knowledge, search specification and content from the archive. An output are **recommended content items (clips, segments) from TV Archives**, targeted to Editor and Director who can select from these recommendations the content for their show.
- **Collection of user feedback.** Feedback is collected from consumers and stored to the feedback database.
- **Analysis and presentation of feedback:** this workflow is a logical sequel of previous workflow. Stored feedback is analysed and presented to the professional user. Within the same workflow user profiles (consumer profile and audience profile) are also built or updated from the collected feedback and stored to the user profile database. **Audience profile and channel statistics** are computed and

displayed to the Video conductor in real-time which allows him to evaluate the reaction of viewers to his show.

- **Generation of consumer recommendations:** this workflow is coloured in orange. Consumer recommendations are created on the basis of the individual consumer profiles and broadcasted program information. Recommendations are delivered to the consumer in form of personalized EPG or Programme Alerts.

3.1 The Production Workflow Using Content Recommendations

The workflow involves two actors. The Editor is using the Recommender system interface to search for content, and the Video Conductor is reviewing the materials and deciding on the inclusion of content to the live program.

The role of the production recommendations within the future production workflows is shown in Figure 2. The workflow starts in the pre-production phase by content planning, where the topics of the future show are defined. During this phase, the Recommender system assists the Editor in performing searches on the available archive material in the long-term AV archive. The result of this phase is that the production archive for live production is prepared, including for example background information and interviews with athletes appearing in the event.

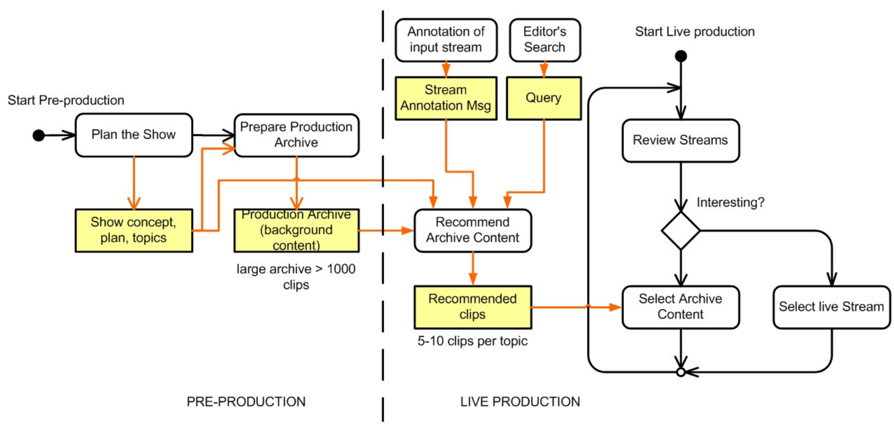


Fig. 2. The production workflow using archive content recommendations provided by TV Production Recommender system

During live production the input camera streams are annotated by human or automatic annotations. The Recommender system analyses these messages and automatically produce recommendations of the illustrating material. For example, if an athlete takes a lead, a director immediately receives recommendations of background archive material on this athlete. In case of an unpredictable or unexpected event, the Editor can use the search interface to specify query and search in the archive for the AV material. The final archive recommendations are evaluated by the Director and selected for the program. The Recommender System User Interface was developed for the Editor which allows him to perform queries and review the recommended content.

3.2 Personalization Scenario

In the personalized TV scenario the TV content is produced according to the profile of the target user group (individual TV viewer or group of people). On the other hand, personalization might also include personalized presentation of content to the user, for example through personalized user interfaces or personalized selection of content.

The concept of program personalization in LIVE project is based on group or massive personalization, where the recommendation takes into account the preferences of a target group of consumers, specified by the **target audience profile**. The concept of **audience personalization** means the selection of the content according to the audience profile, so that the acceptance of content is maximized within the audience. For example, the audience profile will include statistical preferences of the general TV audience towards specific TV program types, different sports, different commentators etc. It will be generated and updated by collecting information on consumer actions during watching. The details on the computation of recommendations are given in [3].

4 Conclusions

The proposed concept of recommendation system integrated into the TV production will enable a significant step towards enabling live production of personalized multi-channel TV shows. The integrated user tracking and profiling enables personalization of the program to different target audiences, and gives a Director a new capability of observing the audience feedback in real-time. The Recommender system will be evaluated in the field trial during Olympic Games 2008 in real production environment. We expect that this field trial with 1000 households receiving multi-stream ORF Olympic channel will prove the proposed concept.

Acknowledgment. This work was partially funded by the European Commission within the 6th framework of the IST under grant number FP6-27312. All statements in this work reflect the personal ideas and opinions of the authors and not necessarily the opinions of the European Commission.

References

1. LIVE – Live staging of media events (last visited 1.03.2008), project web-site: `http://www.ist-live.org`
2. Wages, R., Williams, C.M., Grünvogel, S.M., Trogemann, G.: Video Composer and Live Video Conductor: Future Professions for the Interactive Digital Broadcasting Industry. In: Proceedings of the EuroITV 2006: Beyond Usability, Broadcast, and TV - Fourth European Conference on Interactive Television, Athens, Greece, May 25-26, 2006, pp. 32–38 (2006)
3. Zaletelj, J., Wages, R., Bürger, T., Grünvogel, S.M.: Content Recommendation System in the Production of Multi-Channel TV. In: Proceedings of the AXMEDIS 2007, 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution, Barcelona, Spain, November 28-30 (2007)

A New Approach for a Lightweight Multidimensional TV Content Taxonomy: TV Content Fingerprinting

Javier Recuenco[2], Noelia Rojo[2], and David Bueno[1]

[1] Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga
[2] AbyPersonalize
javier.recuenco@abypersonalize.com,
noelia.rojo@abypersonalize.com,
bueno@lcc.uma.es

Abstract. In order to support the demands of fast and precise TV categorization for a personalized Electronic Programming Guide (EPG), the authors propose a new multidimensional taxonomy for the TV content. In addition to being much more lightweight in nature than the approach proposed by TV Anytime, with this method, a much more streamlined format is generated, which attempts to balance sensitive and detailed categorization criteria with computer efficiency in order to fulfill the demands of the recommender system. Furthermore, the authors propose a mechanism to obtain a quick and efficient real-time comparison of specific TV content with an alternative system called TV content fingerprinting

1 Introduction

The number of TV channels and programs available is constantly increasing. This often means that viewers receive much more information than they can actually manage, which may lead them believe that they are missing programs that could possibly interest them.

In order to organize this TV programming information, in many cases, traditional American categories analyzed in [4] are still in use. For example, the term "foreign film" is still used as a category, in spite of its ambiguity. In [1] there is an analysis of the program content on BBC channels One, Two and Three using the standard classification of TV-Anytime [2]. Some metadata has been defined related to the content: *title*, *free text synopsis*, *keyword* lifting and structured *Genre* information combining terms from the *Intention*, *Format*, *Content*, *Intended Audience* and *Atmosphere*.

Recommender systems help the user to choose when there is a very large number of possibilities on offer. Basically, they can be content based and/or collaborative, depending on whether the recommendations are done on the basis of the content of the document or on the basis of documents which were of interest to users with similar tastes. The best solution benefits from both content and collaborative information. This is called the hybrid model and a good review can be found in [5][6]. Related to The Electronic Programming Guide we can find myTV project [7] which contains information on TV programming and services, AVATAR [8], which is an EPG with a content based recommender, or TPTV [9], which combines several recommender algorithms to obtain a more accurate recommendation of TV programming.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 107–111, 2008.

If the programs are well categorized and taxonomized this information could be used as another recommender algorithm. In section 3 we will propose a new multidimensional content taxonomy that completes the recommendations of TV-Anytime. This proposal will be used to create a fingerprint of each program characterizing it and it will ease in the future for any TV recommender system the process to determine the degree of similarity between programs.

2 Taxonomy Proposal

The main question we tried to answer in order to define the categorization model of the TV contents has been: "How organized is the heterogeneous world of contents offered by the current Television Broadcasters?". We have analyzed the contents the EPG programs provided by Spanish channels and have obtained the following initial conclusions:

- Although broadcast on different channels, program names are normally preceded or followed by some indicators of genre, for example: news, cartoons, cinema, sports (football, tennis, etc.), children's programs, chat shows, debates, quiz shows, etc.
- On other occasions, we may find only the name of the program. This normally happens in the case of those programs that are so popular than the name is enough to inform of the content, for example: Big Brother.

If these are the tags used by the TV companies, it is understandable that the general public use the same classifications. For example, if viewers are asked about their favorite type of TV program they will generally say something like: the news, films, TV series, documentaries, etc.

We have also based our study on the standard classification of TV-Anytime [2], which establishes 18 different taxonomy criteria to fully identify the nature of the content. Among these we can find: "*IntentionCS*", "*FormatCS*", "*ContentCS*", "*OriginationCS*", "*IntendedAudienceCS*". From this we obtain that the content of TV-Anytime may be from different sources (*OriginationCS: Cine, Radio, TV …)*, not only TV. Also the metadata *ContentCS classify the program according to the content or the topic, and sometimes different classification criteria are put together. For example some of the categories refer to the function of the information programs, others to the content of the sports programs, others to the kind of spectator (e.g. children), and others related to their structure "magazine", etc.*

The TV-Anytime taxonomy has been developed conscientiously and covers most of the topics that can be used to categorize a program. Nevertheless, being so general also makes it difficult to use. Imagine how many people it would require and how long it would take to categorize all the TV programs and for each one to select 17 dimensions. In the case where we wished to categorize all this data to use it in a recommender system, the sheer amount of information would make it difficult to manage. In order to overcome these problems, we propose a simplified categorization model. This will allow us to identify univocally for all the TV programs available, any content included, without mixing different category criteria. We have identified eight main dimensions detailed in the following table [Table 1].

Table 1. Dimensions proposed to classify TV programs

Dimension	*Objective*	*Values*
Public	Identify segments of objective public (~IntendedAudienceCS)	Children, young, adult, any public
Functional	Identify the purpose or main function of the program[11] (~IntentionCS)	Information, entertainment, and combinations (info. & enter.), (+info,-enter.), (-info,+enter.)
Content	Identify the topic/subtopic of the program (~ContentCS)	Sport, religious, reality shows, culture, comedy,…
Temporal	Identify the temporal nature of the content	Auto content, serial (daily, weekly) or exceptional
Format	Identify the format of the content (~FormatCS)	Film, documentary, series, broadcast, quiz show, debate, information (news, report, interview, weather,…), music festival, music concert,…
Treatment	Place the content in a context reality-fiction	Fiction (situation comedies), reality, tv conditioned reality.
Synthesis	Identify the generation of the content	Real, synthetic, Hybrid (real+synthetic)
Origin	Identify the origin of the content	Local, Regional, National, International (clarifying the country/region)

3 The Concept of Audiovisual Content Fingerprinting

Fingerprinting is a well know science (at least in its basic form) that provides us with a powerful metaphor of what are we looking for. Generally, a fingerprint is a combination of variables which characterize one individual member or an individual class of a group. These are in some cases also called signatures.

We propose a clear fingerprint for every unit of content. Not with the purpose of making them unique, but like in early forensic fingerprint, to provide a quick way to identify individuals. Fingerprinting used in forensic science provides us both with a powerful image of what we are looking for, and before the introduction of computer imaging and scanning; fingerprint pattern acquisition using the Vucetich System [10] used a very similar mechanism to the one we are proposing here. Before AFIS and all the CSI panoply entered our lives there was already a system that used 256 possible values for a total of 10 fingers, which would rule out a concrete person out of more than a million chances, more than enough to build a solid case of identification in the pre-computer era. The whole idea of a specific fingerprint associated to our audiovisual content is in order to reach a particular objective: quick content identification beyond merely a rough approach. We would replace the ten fingers used on the original Vucetich System for our current dimensions, producing a vector of numbers associated according to the values of every dimension. The resulting vector would be pretty much a good characterization fingerprint of the content involved. The objective is to quickly provide an idea of the more relevant characteristics of any specific audiovisual content in order to produce relevant associations with other content. The classic scenario in most recommendation systems is based on Collaborative Filtering scenarios. We wish to make recommendations which are also based on program similarities according to an established fingerprinting criterion. Our taxonomy allows us to reduce the huge task of comparing countless individual criteria to one of the fastest operations a computer can perform: checking hexadecimal digits. The format we propose lets to compare at different levels of deep in the tree. For example, for the dimension *Functional* (IntentionCS in AnytimeTV) we have four levels of deep. We can represent it with 4 bytes (one for each level in the tree). We could compare four programs with the values of AnytimeTV 1.1.1, 1.1.2, 1.1.1, 1.2. Any similarity measure would say that the programs with the classification 1.1.1 are the most similar, but that the 1.1.2 is more similar to 1.1.1 than 1.2. We can apply the same method to all the Dimensions. Only for the *Content* dimension we will use 12 bytes instead of 4, with the intention of giving the people the possibility of classifying the content up to three different content categories (i.e. a program can be Theater (3.1.4.3) and Romance (3.4.3)) We believe that the current version of the signature can be generated using 4*7+12*1=40 bytes. A signature associated with a specific program is a vector like this (public, functional, content1, content2, content3, temporal, format, treatment, synthesis, origin): **04020205 01030200 03010403 03040300 00000000 …**≡ age 45-54, Lifelong/Further education, theater, romance…

Therefore we assume that one of our programs has a specific fingerprint and another program shares a relevant percentage of this fingerprint, then we can assume that the similarities between the content would be noticeable. Not all dimensions produce a similar degree of contribution to the overall similarity (content contributes much more than other dimensions to the overall degree of similarity, for example). The whole idea behind the Audiovisual digital fingerprint is to provide a quick and effective “cloud” of close and related contents, in a quick and computer manageable way. The results of a triple check between collaborative filtering, profile matching and digital fingerprinting would produce much more precise recommendations than the existing one-dimensional recommendation mechanisms.

4 Conclusions and Future Work

The approach adopted by TV-Anywhere is by its nature so holistic and complete that it is rendered useless in the context of generating viable fingerprinting. We have selected a subset of relevant dimensions in terms of true content, defining criteria regarding a concrete fingerprint that could provide us with a powerful and quick mechanism to locate similar content, with a clear possibility to pivot from either of the available dimensions to search affinity in terms of several criteria. Our subset not only makes the concept of digital content fingerprinting valid and quick, but it provides a fast way of sorting content data in a relevant, easy to understand way. At the same time provides enough depth of content description and is robust enough to help end users to categorize TV contents. Our current approach does not exclude the progressive inclusion of new dimensions, new values and/or the reorganization of the available dimension values. Nevertheless it is based on the main objective of obtaining a quick, well balanced and complete content match. Where TV Anywhere has created a clear standard, we will follow suit. Future work will be carried out in several fields: 1) Improving the algorithms involved in quick determination of the degree of correlation degree between contents using a digital fingerprint. 2) Work on "loose" and "precise" correlation algorithms in order to produce close associations and loose relationships for soft search. 3) The integration of the whole subsystem into a TV recommender system and the creation of a basic interface for users in order to help categorize contents. 4) The provision of a sensible contribution model for the overall degree of similarity of each dimension involved (Some dimensions like content contribute much more to the degree of similarity than other ones more generic).

The overall idea of the taxonomy system is to be a mean to an end (Easy, robust and quick content matching for our recommender system) and not a purpose, but we are pretty sure that our streamlining efforts would be fairly useful in order to provide functional taxonomy criteria that the general public can abide by without the need for a complicated, unmanageable, slow and non functional approach.

References

[1] Butkus, A., Petersen, M.: Semantic Modelling Using TV-Anytime Genre Metadata, pp. 226–234 (2007)

[2] ETSI TS 102 822-3-1: TV-Anytime; Part 3: Metadata; 1 Sub-part 1: Part 1 - Metadata schemas (2006)

[3] Pogacnik, M., Tasic, J., Meza, M., Kosir, A.: Personal Content Recommender Based on a Hierarchical User Model for the Selection of TV Programmes. User Modeling and User-Adapted Interaction 15(5), 425–457 (2005)

[4] Castañares, W.: La televisión y sus géneros,?'una teoría imposible? CIC, Servicio de publicaciones UCM (3), 167–181 (1997)

[5] Burke, R.: Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction 12(4), 331–370 (2002)

[6] Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)

[7] Pogacnik, M., Tasic, J., Meza, M., Kosir, A.: Personal Content Recommender Based on a Hierarchical User Model for the Selection of TV Programmes. User Modeling and User-Adapted Interaction 15(5), 425–457 (2005)

[8] Fernandez, Y.B., Arias, J.J.P., Nores, M.L., Solla, A.G., Cabrer, M.R.: AVATAR: An improved solution for personalized TV based on semantic inference. IEEE Transactions on Consumer Electronics 52(1), 223–231 (2006)

[9] Bueno, D., Conejo, R., Recuenco, J.G.: TV Recommender System Architecture. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 117–122. Springer, Heidelberg (2007)

[10] The Science of Fingerprints, available at Project Gutenberg (2006), http://www.gutenberg.org/etext/19022

[11] Baggaley, J.P., Duck, S.W.: Dynamics of Television. Westmead, Farnboroug, Hants: Saxon House, Tekfield Limited (1976)

Modeling Moods in BBC Programs Based on Emotional Context

Michael Kai Petersen and Andrius Butkus

Technical University of Denmark, DTU Informatics,
Building 321, DK.2800, Kgs.Lyngby, Denmark
{mkp,ab}@imm.dtu.dk
http://www.imm.dtu.dk

Abstract. The increasing amounts of streaming and downloadable media becoming available in converged digital broadcast and next generation mobile broadband networks will require intelligent interfaces capable of personalizing the selection of content according to user preferences and moods. We propose an approach to automatically generate atmosphere-like metadata from BBC synopsis descriptions, by applying LSA latent semantic analysis to define the degree of similarity between textual program descriptions and emotional tags in a semantic space.

Keywords: personalization, emotions, LSA latent semantic analysis.

1 Introduction

Since 2005 the broadcaster BBC has made their program listings available as XML formatted TVA TV-Anytime [1] metadata, which allows for describing media using complementary genre aspects, atmosphere as well as synopsis. We have in a related paper [2] analyzed how these metadata features may complement each other when applying more genre dimensions in parallel, and thus increase the number of relevant recommendations, by capturing similarities across the traditional divide of categories. In particular the TVA genre dimension atmosphere seemed able to identify programs that might be perceived as similar even though they belong to different genre categories. Extending this approach we propose in the present paper a method to automatically generate atmosphere-like metadata using the synopsis of TV programs. We outline in the following sections a framework for modeling emotional context using *last.fm* tags as markers in a semantic space, the methodology for extracting latent semantics, the retrieved results followed by a discussion of our early results based on BBC synopsis descriptions.

2 Affective Terms

When investigating how unstructured metadata can be used to describe media, the social music network *last.fm* provides an interesting case. Despite the idiosyncratic character of tags defined by hundred thousands of users, recent studies

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 112–116, 2008.

within music information retrieval have revealed that *last.fm* users often tend to agree on the emotional terms they apply to music. This correlation between social network tags and the specific music tracks they are describing, makes it possible to define high-level categories, which provide a simplified mood ground-truth reflecting the perceived emotional context of the music [3][4].

With point of departure in these findings we hypothesize that it might be possible to extract the emotional context of a TV program by projecting its synopsis into a semantic space, and use *last.fm* tags as affective buoys to define the textual description within emotional context. Drawing on psychological studies [5], establishing that emotional assessment can be reduced to a semantic differential spanned by the two primary dimensions of *valence* and *arousal*, we use these two axes to outline an emotional plane for a *last.fm* semantic tag space. The first of the these two dimensions describes how pleasant something is along an axis going from happy to sad, whereas the latter dimension captures the amount of involvement ranging from passive states like dark or soft to active aspects of excitation as reflected in tags like angry or sexy.

3 Latent Semantics

As a machine learning technique which resembles cognitive comprehension of text, LSA *latent semantic analysis* [6][7][8] extracts meaning from texts by modeling the usage patterns of words in multiple documents and represent the terms and their contexts as vectors in a high-dimensional space. To retain only the most essential features the dimensionality of the original sparse matrix is reduced to around 300 dimensions. This reduced LSA space makes it possible to compute the semantic relatedness of synopsis and affective terms as the cosine of their vectors, with values towards 1 signifying degrees of similarity between the items and low values close to zero or negative signifying a random lack of correlation. In this semantic space a synopsis text and words which express the same meaning will thus be represented as vectors that are closely aligned, even if the terms are not literally co-occurring within the same context.

4 Results

Taking a selection of short BBC program descriptions as input, we compute the cosine similarities between a synopsis text vector and each of the selected *last.fm* emotional words. An analysis of the program "News night", based on the short description: *News in depth investigation and analysis of the stories behind the day(s) headline*, triggers the tags funny and sexy which might not immediately seem a fitting description, probably caused by these emotional terms being directly correlated with the occurrence of the words stories and news within the synopsis. The atmosphere of the lifestyle program "Ready Steady Cook!" might be somewhat better reflected in the synopsis: *Peter Davidson and Bill Ward challenge celebrity chefs to create mouth watering meals in minutes*, which triggers the tag *romantic* as associated with meals. Another singular emotion can be retrieved from the documentary “I am boy anorexic”, which based on the synopsis: *Documentary following three youngsters struggling to overcome their obsessive relationship with food as they recover inside a London clinic and then return to the outside world*, triggers the affective term dark. We find a broader emotional spectrum reflected in the lifestyle program “The flying gardener” described by the text: *The flying gardener Chris travels around by helicopter on a mission to find Britain(s) most inspirational gardens. He helps a Devon couple create a beautiful spring woodland garden. Chris visits impressive local gardens for ideas and reveals breathtaking views of Cornwall from the air*. The synopsis triggers a concentration of passive pleasant *valence* elements related to the words soft, mellow combined with happy. In this context also the tag cool comes out as it has a strong association to the word air contained in the synopsis, while the activation of the tag aggressive appears less explainable.

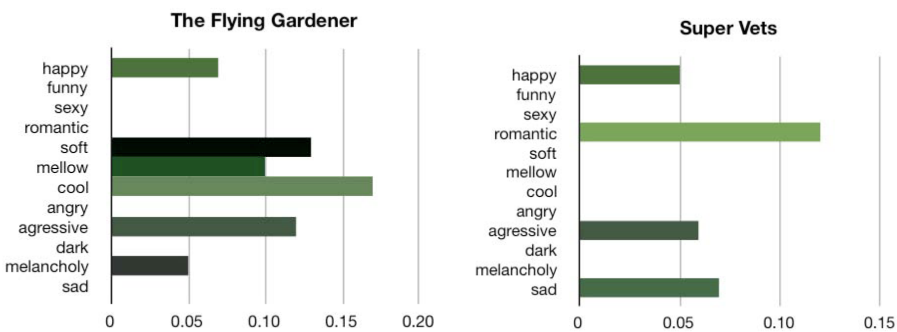


Fig. 1. LSA cosine similarity between the synopsis descriptions of “The flying gardener” and “Super Vets” against 12 frequently used *last.fm* affective terms

This cluster of pleasant elements is lacking in the LSA analysis of the program “Super Vets” which instead evokes a strong emotional contrast based on the text: *At the Royal Vet College Louis the dog needs emergency surgery after a life threatening bleed in his chest and the vets need to find out what is causing the cat (..)fits*, where both pleasant and unpleasant active terms like happy and sad stand out in combination with strong emotions reflected by the tag romantic. And as can be seen from programs like “The flying gardener” and “Super Vets” (Fig.1) the correlation between the synopsis and the chosen tags might often trigger both combinations of complementary elements as well as contrasting emotional components rather than a single monochrome feeling.

We proceeded to explore whether we could sum up a distinct pattern reflecting an emotional profile pertaining to a TV series, by accumulating the LSA values of correlation between synopsis texts and emotional tags over several episodes. For this purpose we chose the soap “East Enders” and the comedy “Two pints of lager” and analyzed descriptions of six consecutive episodes from each series.

Even when only comparing the synopsis and emotional tags over six episodes (Fig.2), it appears that the accumulated LSA correlation values in the soap “East

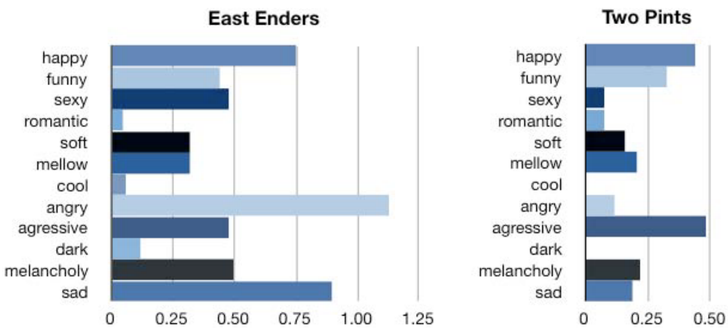


Fig. 2. LSA cosine similarity of the soap "East Enders" and the comedy "Two Pints" against 12 frequently used *last.fm* affective terms accumulated over six episodes

Enders" are roughly twice as high as in the comedy "Two pints of lager". The contributions of affective components in both histograms are unbalanced, but whereas the former series has a bottom-heavy emphasis on angry and sad emotions, the balance is reversed in the latter with a shift towards predominantly happy and funny elements complemented with soft and mellow aspects. These patterns can similarly be made out when considering the emotional components plotted over time for the soap and comedy respectively (Fig.3). The distribution in "East Enders" is much more dense and emotionally saturated reflecting aspects of *arousal*, while the character of "Two pints of lager" seems mirrored in a pronounced clustering of lighter elements of positive *valence* and an overall sparsity of excitation within the matrix.

Fig. 3. LSA cosine similarity of the soap "East Enders" and the comedy "Two Pints" against 12 frequently used *last.fm* affective terms accumulated over six episodes

5 Discussion

Projecting BBC synopsis descriptions into an LSA space using *last.fm* tags as emotional buoys, we have demonstrated an ability to extract patterns reflecting combinations of emotional components. Analyzing the emotional components reflected in the synopsis descriptions over a sequence of episodes, we have been able to separate these aspects into patterns defined by the sparsity and character of the distribution. While each synopsis triggers an individual emotional response related to a specic episode, general patterns still emerge when accumulating the LSA correlation between synopsis and emotional tags over consecutive episodes, which enables us to differentiate between a comedy and a soap based on a textual description alone. We therefore propose that emotional components describing the content of media might be retrieved as latent semantics by using affective terms as sensors in a semantic space, and we suggest that LSA might be applied to extract structural patterns from synopsis descriptions as a basis for automatically generating mood-based recommendations. Though the synopsis descriptions trigger both combinations of complementary elements as well as contrasting emotional components rather than a monochrome affective response, they nevertheless pertain to distinct patterns which we speculate might be used as a basis to build emotional patterns capturing user preferences.

References

1. ETSI: TV-Anytime. Part 3. Metadata 1. Sub-part 1. Part 1 - Metadata schemas TS 102822-3-1 (2006)
2. Petersen, M., Butkus, A.: Semantic modelling using TV-Anytime genre metadata. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 226–234. Springer, Heidelberg (2007)
3. Levy, M., Sandler, M.: A semantic space for music derived from social tags. In: Proceedings of the 8th International Conference on Music Information Retrieval, pp. 411–416. Austrian Computer Society (2007)
4. Hu, X., Bay, M., Downie, S.: Creating a simplied music mood classification groundtruth set. In: Proceedings of the 8th International Conference on Music Information Retrieval, pp. 309–310. Austrian Computer Society (2007)
5. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The measurement of Meaning. University of Illinois Press (1957)
6. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science 41, 391–407 (1990)
7. Landauer, T.K., Dumais, S.T.: A solution to Platos problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review 104, 211–240 (1997)
8. Dumais, S.T.: LSA and information retrieval: Back to basics, Handbook of latent semantic analysis, pp. 293–321. Lawrence Erlbaum, Mahwah (2007)

Predicting Future User Behaviour in Interactive Live TV

Martin Gude[1], Stefan M. Grünvogel[1], and Andreas Pütz[2]

[1] Cologne University of Applied Sciences, Betzdorfer Str. 2, 50679 Köln, Germany
martin.gude@alumni.fh-koeln.de, stefan.gruenvogel@fh-koeln.de
[2] Pixelpark AG, Friesenplatz 25, 50672 Köln, Germany
andreas.puetz@pixelpark.com

Abstract. Recommender systems are a means of personalisation providing their users with personalised recommendations of items that would possibly suit the users needs. They are used in a broad area of contexts where items are somehow linked to users. The creation of recommendations of interactive live TV suffers from several inherent problems, e.g. the impossibility to foresee the contents of the next items or the reactions of the user to the changing programme.

This paper proposes an algorithm for building personalised streams within interactive live TV. The development of the algorithm comprises a basic model for users and media items. A first preliminary evaluation of the alogithm is executed and the results discussed.

Keywords: recommender system, interactive live TV, multistream.

1 Introduction and Related Work

Recommender systems are a means of personalisation providing their users with personalised recommendations of items that would possibly suit the users needs. They are used in a broad area of contexts where items are somehow linked to users. Recommender systems are used in a great variety of contexts (e.g. [1], [2]). last.fm [3] and Lifetrak [4] bring recommender system research to continuous media consumption.

This paper proposes an algorithm for building personalised streams within interactive live TV. The development of the algorithm comprises a basic model for users and media items. Additionally, means are introduced to overcome the problems inherent in the character of live TV.

Recommender systems have been strongly researched in the last time (cf. [5], [6]). In [7] and [8] it has been shown that hybrid recommender systems perform better than either content-based or collaborative filtering algorithms. Tagging, a new flexible means to classify items, allows categorisation of items with loosely associated keywords. This can be used in order to provide descriptions for content-based filtering [9], [10]. It can also help to overcome the vocabulary problem (cf. [11]).

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 117–121, 2008.

2 Requirements for Recommender Systems in the Context of Live-Transmissions

We are considering here interactive TV in the context of live transmissions, which are aimed at mass events like the Olympic Games. There several things happen at the same time, and generally broadcasters are provided with a multitude of parallel audio visual feeds from a central broadcasting company. Instead of creating only one interactive TV stream, these streams will be used to broadcast several interactive TV streams live in parallel. This is already planned in the research project LIVE [12] and will be tested during the Olympic Games in Beijing 2008. The iTV show is to customise itself to the viewer. The aim is to prevent a mere "zapping" behaviour of the consumer at home. This will be accomplished by interlinking the parallel live streams to guide the consumer and enhance the streams with archive material to ensure a dramaturgical integrated user experience.

The requirements of a recommender system in live iTV differ significantly from those of a non-live environment. Implementing a personalised EPG does not suffice. In contrast, rating the programme and recommending a new programme item must be an ongoing process. A recommendation system for live iTV can provide different user experiences to the viewer: In a rather passive lean-back approach it would liberate the user from having to decide consciously which programme to watch, a lean-forward approach would require the user to decide whether to follow a recommendation or not.

Recommender systems supporting live broadcasts have to face some key problems due to the nature of live events. Content-based filtering suffers from the impossibility of providing appropriate content descriptions. As you cannot precisely predict what will happen in the audio visual media streams on the short or long term, you can only provide raw descriptions afore. Furthermore, these descriptions must be associated in real time. Collaborative filtering relies on a critical mass of viewers that must be available. It is impossible to take all registered viewers as a basis, but only consumers currently watching the programme. Hence, you have to induce an action from imperfect data.

3 Development of the Recommender System

As a base for its predictions, our recommender system uses the content-boosted collaborative filtering algorithm developed in [7].

A crucial aspect for calculating recommendations is the accurate description of the user and his relations to the media items. The user profiles are built using two means of user feedback: watching statistics (watching and zapping) and asset rating (favour and dismiss). Both means are bijectively mapped to a benchmark variable as shown in Figure 1. Each media item is tagged with a finite number of tags, where the number of tags may vary between the media items. The rating of a media item is mapped on the asset's tags, i.e. the tags are rated with the same value as the whole asset. The user profile $U_{u,t}$ for user u and tag t is calculated

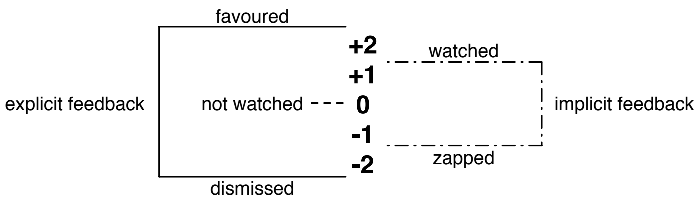


Fig. 1. Values of the benchmark variable used to describe watching and rating of live media assets

as arithmetic mean of the ratings user u gave to items associated with tag t. If u did not rate any item tagged with t, $U_{u,t}$ would have the value of 0.

A content-based prediction for one viewer and one item is calculated as arithmetic mean of the vector containing the mean ratings the viewer gave to the tags associated with that media item. The content-based predictions are calculated according to the profile that has been built up at this time. They serve as a basis for the pseudo user ratings. Figure 2 illustrates the construction of the pseudo user ratings for one example viewer. Each row denotes a live media stream which is divided into a continuous sequence of media assets with some fixed duration.

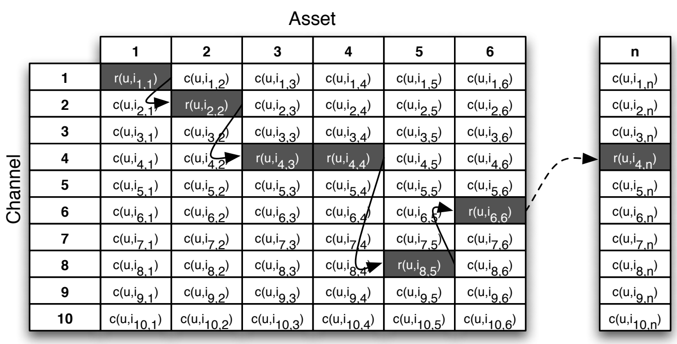


Fig. 2. Construction of the pseudo user ratings for user u. The black background indicates the assets watched by u.

Collaborative filtering for a hybrid recommender is implemented as in [7], where we have omitted any reliability components of the predictors. However, they are planned to be implemented after a field trial in Mid 2008.

In a live broadcasting context, it is impossible to predict the contents of the items that are to be broadcast and the behaviour of the neighbours, i.e. how the neighbours react to that items, in the future. As a replacement, it is assumed that the contents of two sequential items do not differ significantly. Hence for practical reasons, the next item is described in our approach with the description of its preceding item.

4 Evaluation of the Proposed Algorithm

The algorithm is evaluated with the help of a small test application and data derived from watching the interactive DVD "Vision Europe" produced in the EU IST Project MECiTV [13].

Vision Europe provides ten different parallel channels with several predefined switching points. At each switching point, another channel is recommended to the user. In the test case, a media item is defined as 10 second part of the live stream. The test case is rather far away from the real usage scenario because of these editorial switching points.

The algorithm is evaluated through two different metrics: the statistical accuracy metric mean absolute error (MAE) and the decision support measure Receiver Operating Characteristics (ROC). Both metrics compare the actual rating a user gave to an item with the predictions made by an algorithm [7].

The tests have been conducted in October 2007 with 13 employees of Pixelpark Agentur Köln (10 masculine, 3 feminine). As a consequence of the rather small amount of test persons and the not optimal interactive video, the data gathered within the test cannot be classified as good.

Although this is not a representative study, first results can be stated. Both metrics induce that the Content-Based predictor performs better than the hybrid predictor. This might be caused in the bad base data, but the Content-Based predictor already performs quite well with this bad data. The worse performance of the hybrid predictor could be caused by the small number of test persons.

5 Conclusion

The simple but powerful model of the complex system for live recommendations used in the algorithm seemed to suit the requirements of interactive live television. It leads to high quality recommendations for the consumer based on low quality data. It can serve as a basis for further research in live recommender systems. The algorithm will be re-evaluated during a field trial scheduled within the LIVE project [12] during the Beijing 2008 Olympic Games.

Acknowledgements

This work was partially funded by the European Commission within the 6th Framework of the IST under grant number FP6-27312. All statements in this work reflect the personal ideas and opinions of the authors and not necessarily the opinions of the European Commission.

The LIVE project ('Live Staging of Media Events' [12]) is an integrated, multidisciplinary initiative that will contribute to the IST strategic objective 'Semantic-based Knowledge and Content Systems' and 'Exploring and bringing to maturity the intelligent content.

References

1. Amazon.com: Site features:recommendations (2007)
2. Schachter, J.: del.icio.us: people who like recommendations also likes. (2005)
3. last.fm (2007), http://last.fm
4. Reddy, S., Mascia, J.: Lifetrak: music in tune with your life. In: Proceedings of the 1st ACM international Workshop on Human-Centered Multimedia, HCM 2006, Santa Barbara, California, USA, October 27-27, 2006, pp. 25–34. ACM Press, New York (2006)
5. Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., Riedl, J.: Is seeing believing?: how recommender system interfaces affect users' opinions. In: CHI 2003 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, April 05-10, 2003, pp. 585–592. ACM Press, New York (2003)
6. O'Donovan, J., Smyth, B.: Trust in recommender systems. In: IUI 2005: Proceedings of the 10th international conference on Intelligent user interfaces, San Diego, California, USA, pp. 167–174. ACM Press, New York (2005)
7. Melville, P., Mooney, R.J., Nagarajan, R.: Content-boosted collaborative filtering (2001)
8. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)
9. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proceedings of the 16th international Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 08-12, 2007, pp. 211–220. ACM Press, New York (2007)
10. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In: Proceedings of the Seventeenth Conference on Hypertext and Hypermedia, HYPERTEXT 2006, Odense, Denmark, August 22-25, 2006, pp. 31–40. ACM Press, New York (2006)
11. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Communications of the ACM 30(11), 964–971 (1987)
12. LIVE: Live staging of media events (2007), http://www.ist-live.org
13. MECiTV: Media collaboration for interactive tv (2004), http://www.meci.tv

Does Mobile Television Challenge the Dimension of Viewing Television? An Explorative Research on Time, Place and Social Context of the Use of Mobile Television Content

Marinka Vangenck, An Jacobs, Bram Lievens, Eva Vanhengel, and Jo Pierson

Researcher - IBBT-SMIT, Vrije Universiteit Brussel, Pleinlaan 2 – 1050 Brussel, Belgium
{marinka.vangenck,an.jacobs,bram.lievens}@vub.ac.be
{eva.vanhengel, jo.pierson}@vub.ac.be

Abstract. Television is one of the last media technologies to become disconnected from a fixed place like home. Media like newspaper, radio, audio recordings and computing have already been introduced in a mobile setting since some time. However with the wide spread and strong domestication of mobile devices, and the dominant character of television in households, it seems that the transition to mobile television is inevitable. Viewing patterns and behavior are strongly determined by dimensions of time, place and social context and it is exactly on those dimensions that the mobility aspect has a great impact. The question we therefore address in this paper is to what extent that relationship between mobility and these dimensions will influence the uptake and usage of mobile television.

Keywords: Mobile television, DVB-H, television experience, user practices, proxy technology assessment, living lab, Maduf.

1 Introduction

Just by looking around it seems that the mobile phone is omnipresent. Wherever you are people call, write text messages, listen to music etc. In most parts of the Western and Asian world the mobile phone is domesticated in such a way that for most people it is perceived natural to use their mobile phone at anytime and anyplace regardless their context and setting. Weber describes this phenomenon as *public-zation*. This refers to the fact that a practice, in this case 'making phone calls', which previously exclusively belonged to the private domestic sphere, is entering new consumption spaces. 'Listening to music' has gone through a similar process with first the Walkman and then the MP3 player (e.g. i-Pod) [1].

But this process of *public-zation* is not limited to listening to music or making phone calls, other practices are also subject of *public-zation*. The mobile phone became a facilitator to perform different types of activities in a hybridized environment [2]. One of the next practices to go public is '*watching television*'. This process is enhanced by the digitalization of the television signal as well as the increasing convergence. Technology

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 122–127, 2008.

is available to make broadcast television on mobile phones possible without the need of a fast (and often expensive) internet connection. DVB-H (Digital Video Broadcasting for Handhelds) - recently promoted by the European commission as the standard for mobile broadcasting technology – is one of this technological solutions enabling the broadcasting of television signals to all kinds of handheld devices, to some extent including portable computers.

There is a growing number of literatures on the subject of mobile television that covers different aspects of the technology: legal and regulatory issues, usability aspects, adoption, technological aspects etc … [3, 4, 5, 6, 7]

In this paper we want to contribute to this literature by focusing on the interaction between everyday practices and the use of mobile television. More specifically, to what extent the dimensions of time, place and social context – all three playing an important role in traditional television usage – are challenging the uptake and usage of mobile television. How do the users integrate the mobile television application in their daily activities?

2 Method

The paper is founded on the research project 'Maximize DVB Usage in Flanders' (MADUF) between 2006 and 2008. The objective was to generate an optimal model for providing mobile television services in Flanders (Belgium) via DVB-H and developing a proof of concept to be tested extensively by users.

In order to address the research question we have investigated mobile television from two perspectives. On the one hand we looked at existing functionalities and services of mobile television using 3G. This includes aspects as interactivity, video-on-demand and use of the electronic program guide. On the other hand the broadcasting element and high video resolution was investigated via DVB-H. In order to capture the real life experience of users – with the objective to identify the relation of the different dimensions of watching television – a multi-method ethnographic approach has been followed.

We did two kinds of studies. First we did a conceptual study by using the Proxy Technology Assessment (PTA) [8] approach with 12 users giving them an experience of interactive (video-on-demand) mobile television via 3G. 18 months later the same respondents tested the developed interface based on their recommendations and experienced shortly mobile television via DVB-H during a group interview. Secondly, we did a large iterative evaluation study, consisting of different consecutive phases, within a Living Lab setting in a specific locality [9]. In total about 70 users were provided with either a DVB-H enabled mobile phone (including a headset) or a specific DVB-H receiver was installed on their portable computers. The different respondents were selected via purposeful sampling. Using different selection criteria within the domains of mobility, viewing behavior, ICT skills and socio-demographic, a phenomenal variation was created. Next to these we have also taken some technological limitations into account (like area of coverage, the use of specific devices, …) This approach enabled a comparison between different types of user groups, but also to look at the differences on TV-viewing between mobile phones and portable computers.

Within both settings a multi-method approach - combining structured questionnaires, probing cards, diaries, auto collecting pictures, logging data and in-depth interviews - was used to get a user-oriented picture of the potential experience of the usage of (mobile) television.

3 Results

3.1 The Reference Point

The comparison of the results on different moments with the same participants enabled us to see possible changes in the experiences. Most remarkable is the ***change of the reference point***. During their first acquaintance with mobile television the respondents referred to their fixed television, which caused an overall negative evaluation of mobile television. The small display of a mobile phone, the limited audiovisual quality, the need to hold your mobile phone in your hand while you watch, … makes mobile television very different from the “real” lean back television experience. However, during the second experimental phase (1,5 year later) the respondents referred to their previous quite negative experiences with mobile television and no longer to their fixed television in order to evaluate the application. This lowered their expectations and contributed to satisfaction due to the small improvements. Respondents experiencing the applications developed within the project for the first time in the evaluation phase did not take this positive attitude.

Because of the mix in devices we used, we saw a different experience and reference point. First, the mobile phone is so compact it is very fit to use anyplace and anytime. Second, the reference to other practices with the portable computer results as well in a different perception and usage of mobile television service. On the mobile phone people easily watch television in a very short period of time. The usage of the portable computer itself is already very much limited to the place, time and social context. In addition to that, actions necessary to watch mobile television on the portable computer is more demanding compared to the mobile phone. Mobile television on the portable computer is, in the minds of people, more related to the standard television set in contrast to the mobile phone. This is also expressed in the usage. While the use on the mobile phone is about short time slots in order to kill time, the portable computer is used more frequently as a normal television screen to watch a movie or show.

These results illustrate that ***reference point*** of users/viewers can change quite fast. In addition we can conclude that mobile television should not be positioned as a television you can put in your pocket, because the reference point will be fixed television with its own high expectations.

3.2 The Socio-Cultural Habits of the Television Experience

The possibilities of watching television on a mobile handset promise to breach the current home oriented routines, by making other time slots and places within reach.

We state that the possibility of watching television on the move will not be adopted if current cultural norms are not appropriated into these kinds of applications. Our research shows that cultural ideals on time, social context and location/place in general and on watching television in particular guide the attitudes and behavior towards watching television on a mobile handset, at least during the first phases of introduction.

First, in Western societies the cultural ideal of ***time*** efficiency is paramount: making optimal use of our time and finding a good balance between leisure and work. Since our free time is limited, television goes in competition with all the other daily practices. In the daily life setting of users television is often combined with other activities (multitasking), either in terms of time efficiency or just because some people like to combine activities. In that case the medium can serve as '*wallpaper*', when television is secondary as a background while doing something else. Users can also alternate their attention between watching television and other activities. In that case, television is not subordinate to other activities. These practices apply to both regular as mobile television. For example some respondents watched mobile television while working or playing on the computer and alternated their focus constantly. However, not only the total amount of time, but also the attention span of people is limited. Due to this fact, watching television alternating or in the background, is not always and for everybody a possibility. Therefore, although the infrastructure is available, people can ignore the possibility of watching (mobile) television.

Second, next to the ideal of time, the willingness to watch television, fixed or mobile is also determined by the ideals about the appropriate ***social context*** to watch. Television is often perceived as an individual, lean back, leisure medium that leaves little room for social contact. Therefore television often comes on the second place when social contact is available. Our research even indicates that a lot of respondents perceive mobile television as asocial and a possible jamming station at public places, even when they had the opportunity to use a headset. This is one of the reasons why mobile television is mostly used at home (home watching).

However, there is also the ideal of watching television together as a social family activity. Yet when it comes to mobile television, the display of a mobile phone is perceived too small for this. Although when no social contact is available, television can even serve as company. With television on a mobile not only at home, but also in for example a restaurant having lunch on your own or during dead moments waiting for someone or something if no social contact was possible. For example one respondent watched mobile television while waiting all by herself for her dinner at a restaurant.

Because of the asocial connotation, mobile television could function as a 'draught screen', to create a personalized private sphere in a public space. In contrast to listening, watching television asks for a high concentration level, which allows viewers to seclude themselves from their physical and social surroundings.

In general, because of the connotation '*asocial*' and '*a noise factor*' in the social space, people withheld from watching.

Third, strongly intertwined is the factor of the ideals on ***place*** and watching television. Mobile television enables users to watch real-time television at any place and at any time. In spite of its mobile potential, we found that the application is not mainly used on the move, but nomadic at home, as previous studies have also indicated.[1] This is stipulated by several factors. First, a feeling of insecurity among the respondents is playing, though the feeling is very individual and situational. The current high device price and its high degree of gadgetry, contribute to this feeling. The gadget status of mobile television also causes scruples and makes people feel uncomfortable at public

[1] See the following research projects: Helsinki Mobile TV project 2003, Broadcast Mobile Convergence project in Berlin 2004, Paris Mobile TV Customer Trial 2006.

places. Furthermore, noises from the surroundings are often perceived as disturbing and contribute to the dominant usage at home. In addition, as already mentioned the respondents did not want to disturb their environment with the sounds of television. The fact that most of the respondents did not think about using the headset is quite remarkable and could indicate that at this moment watching television is not associated with headsets. Old television habits indeed die hard. Finally, watching television at home does not imply that everyone perceives the home setting as the most perfect place to watch mobile television. This became obvious when the respondents talked about appropriate and inappropriate places for mobile television during a photo sorting session. Situations such as waiting moments or commuting would be more appropriate to watch television on a mobile. This may seem to be a contradiction with the previous results, but for us it actually highlights the difference between the actual current usage of mobile television and the expected future usage. In other words, when the respondents talked about mobile television in the future they often described other contexts where they would like to use mobile television beyond the home setting. So the home was perceived as a safe and quiet place to test the new mobile technology, but not necessarily as the most appropriate place for future usage once the technology has been adopted.

4 Conclusions

During this project we were able to set up different studies on the interaction of users with possible mobile television services, using different devices and technologies. We aimed at a broad variation of users, to get the opportunity to look for maximal variation in user patterns.

Surprisingly, very homogeneous patterns were seen and we were able to show a relation between the some cultural ideals on social context, time and place and the use of mobile television. Currently these factors are constraining the time people want to watch mobile television. In addition to these factors, people do not want to spend their time on watching advertisements or programs they are not interested in. Considering the factors described above and the fact that today people are not in the ability to choose the content (in case of DVB-H), we can conclude that for this moment mobile television does not comply with expectations of the consumers.

Although our research set up giving some of our participants another period of experience with an improved service, as well as the PTA approach comparing experiences with different devices, we are aware that the point of reference should be as far as possible away from a fixed television experience to make a mobile service a success.

Acknowledgements

This paper is the result of research performed within the Maximize DVB Usage in Flanders (Maduf) project, a joint research project of the Interdisciplinary Institute for Broad Band Technology (IBBT) and several industrial partners. The research was conducted by Nokia Siemens Networks, Belgacom, Telenet, Option, Proximus, Scientific Atlanta and VRT in cooperation with the IBBT research groups: Catholic

University of Leuven (CUO and ICRI), the University of Ghent (IBCN, WICA, MMLab and MICT), the Free University of Brussels (ETRO and SMIT), and IMEC (DESICS). Within this project IBBT-SMIT was responsible for the qualitative user research.

References

1. Weber, H.: On “DOMESTICATION” or: Who is domesticating what or whom. In: Design and Consumption: ideas at the interface. Durham University, UK (2005)
2. Lievens, B., Van den Broeck, W., Pierson, J.: The Mobile (R)evolution in Everyday Life: A Cross Border between Public and Private Sphere. In: Conference Proceedings of Mobile Media 2007, Sydney, Australia, July 2-5 (2007)
3. Curwen, P., Whalley, J.: Mobile television: technological and regulatory issues. Q Emerald Group Publishing 10(1), 40–64 (2008)
4. Knoche, H.: A user-centred mobile television consumption paradigm. In: Human Centred Technology Workshop, Brighton, June 28-29 (2005)
5. Nolan, D., Keen, B.: Mobile Digital Television (MDTV) – The coming Handheld Revolution. Screen Digest Ltd., London (2005)
6. O’Hara, K., Slayden Mitchell, A., Vorbau, A.: Consuming Video on Mobile Devices. In: Conference Proceedings of CHI 2007, California, USA, April 28-May 3 (2007)
7. Södergård, C.: Mobile television – technology and user experiences. Report on the mobile TV project. VTT, Finland (2003)
8. Pierson, J., Jacobs, A., Dreessen, K., Lievens, B., Van den Broeck, I., Van den Broeck, W.: Walking the interface: uncovering practices through proxy technology assessment. In: EPIC 2006, Portland, USA, September 24-26 (2006)
9. Pierson, J., Lievens, B.: Configuring living labs for a ’thick’ understanding of innovation. In: Conference proceedings of EPIC 2005 (Ethnographic Praxis in Industry Conference), pp. 114–127. NAPA (National Association for the Practice of Anthropology) - Microsoft - Intel, US (2005)

Relax or Study?: A Qualitative User Study on the Usage of Mobile TV and Video

Koji Miyauchi, Taro Sugahara, and Hiromi Oda

Hewlett-Packard Japan, Ltd., Tokyo, Japan
{koji.miyauchi, taro.sugahara, hiromi.oda}@hp.com

Abstract. The usage of mobile TV and video devices is now spreading in Japan as well as in other countries. We conducted a user study to know when, how and why people were using the devices through qualitative interviews. In this paper, we showed one of the findings from this user study, that is, different attitudes concerning the usage of live mobile TV compared with that of mobile video on commuter buses or trains.

1 Introduction

Many products to enjoy mobile TV and/or video are emerging to the market. Examples of such products include HP iPAQ, Apple Video iPod and Sony PSP. Multiple cell phones have both TV and video functions. “1seg” service (a mobile terrestrial digital audio/video/data broadcasting service in Japan) officially started in April, 2006. The shipment of cell phones with the “1seg” function showed a rapid growth in the market of TV-enabled cell phones[1].

Under these circumstances, we conducted a user study on the usage of mobile TV/video devices in Tokyo, Japan in August, 2007. This was a qualitative user study to know when, how and why people were using these devices, and its aim was not to get statistical results such as “30 percent of people liked comedy”. Rather, we tried to identify the issues from different types of users and devices.

Although we expected similar usages of mobile TV and video devices at first, sharp contrast was found between the usages of live mobile TV and those of mobile video. This paper shows the findings related to this point.

1.1 Previous Works

Everyday practices of traditional TV were reported in [2]. Specifically, they reported highly disengaged viewings and TV channel hoppings in the early evening. Similar behavior patterns were also observed for live mobile TV users on the evening commuter buses or trains.

Various aspects of cell phone usages in Japan, especially by young people, were reported by [3]. Our work can add other findings on the cell phone with mobile TV or video functions, and on the usages by commuters in Japan.

There are already multiple user studies on the usage of mobile TV or video devices in several countries, for example, US and UK [4], Finland [5,6] and Korea [7]. These works deal with either a TV or video device. Our work deals with both devices, and we could compare the usages of them at the same time.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 128–132, 2008.

2 Design of the User Study

We collected volunteers who owned a TV or video enabled portable device and used it at least once a week. They were required to be either an employee of our company or a family member who was living with an employee. We received 27 applications and chose 11 people taking into consideration age, mobile devices they owned, and usage frequency. They were in the range of 15 to 46 or more years old. Seven people were using a cell phone with mobile TV or video functions. About half of the participants used their devices up to 3 times/week, while most of the others used their devices more than 6 times/week. They used commuter vehicles[1] for about 20 to 90 minutes.

We asked the participants to keep a diary to record any mobile TV/video related activities (e.g. "watched a baseball game" or "looked for an interesting TV program") including date, place, circumstance (e.g. "on the train from work") and devices used. After keeping a diary for three weeks, we had a one-hour interview with each participant individually for more details about the circumstances and motivations of the episodes in the diary. Each participant was given a 5000 JPY book coupon as a token of our gratitude after the interview.

3 Usages of Live Mobile TV Devices

Most mobile TV participants watched live TV programs when riding on commuter vehicles, especially in the evening. There were some conditions to determine whether to watch TV with their device.

1. Variety shows were the most popular contents among our participants, and more programs of variety show were broadcasted in the evening.
2. They had more chances to be seated in the evening because the number of commuters in a bus or train became smaller than in the morning. Seating availability was important to enjoy mobile TV by holding a device easily.
3. Some participants also needed a relaxing atmosphere. They could feel relaxed when they sat on a seat, when the vehicle was not too crowded, and when they felt a sense of freedom after work. In the interview, they said:

    > "I don't want to watch TV when I stand on the train. I would like to watch it relaxing on a seat." (M, 36-45)[2]
    >
    > "... I don't watch TV on crowded trains, as I don't feel like watching because I am not comfortable." (F, 26-35)

A variety show consists of multiple short acts such as musical performances and comedy skits, which is suitable for a program on a commuter vehicle. They could start viewing midway through a program and enjoy some acts before getting off. Some of them also watched a drama on commuter vehicles. They didn't

[1] This paper uses 'commuter vehicle' to mean the public transportation such as bus or train to commute.

[2] We use 'M' for male and 'F' for female. The pair of numbers means the range of age.

mind if they couldn't watch a whole drama, because they recorded the drama with the video recorder at home. They didn't watch TV seriously by the mobile TV device, but needed light entertainment.

> "There are multiple acts in one variety show, and it is enough to watch some acts. In that sense, a variety show is a content suited to '1seg'. " (M, 36-45)
>
> "I don't mind if I cannot watch the whole drama, because I record it at home." (M, 14-18)

Most mobile TV participants searched for interesting TV programs by turning on and changing channels without checking TV guides. They seemed to have no specific TV programs they surely had to watch.

> "I watch a TV program which happens to be broadcasted when I turn it on." (M, 36-45)

Several participants enjoyed TV programs such as variety shows as well as news and sports programs with subtitles even without sound. Some participants didn't like earphones. The additional step to take out earphones from somewhere might also make them feel troublesome, because they didn't use earphones for usual usage (talking and e-mail). It seemed that sound was not necessary for the light entertainment by mobile TV on commuter vehicles.

> "I often choose a program which I can enjoy without sound, for example, a variety show, or one which I can enjoy midway through the program." (F, 26-35)
>
> "When I view TV without sound, subtitles appear on the display. I can read the words of a commentator in a baseball game a little after each play. [... omit ...] I have no problem without sound." (M, 26-35)

Although some participants had cell phones with both live TV and video functions, they didn't use the video function except for one user. They knew video preparation including downloading, format-conversion and compression was time consuming. It seemed that mobile TV users felt too much effort was required to prepare video contents for entertainment on commuter vehicles. This doesn't always mean that live mobile TV users are less technology-oriented people. One of them used Sony PSP and LocationFree[3] to watch (recorded) TV programs at a different place from where they were recorded. He said that content transfer to a memory card was troublesome, and that home electronics appliances were more convenient.

> "Honestly, I don't know how to prepare for video contents. I don't want to watch it if I have to do so." (M, 26-35)
>
> "I don't want to watch video contents then [on the train], I like watching live TV. It is troublesome if I have to transfer the contents to a memory card." (M, 36-45)

[3] Sony LocationFree was used to send recorded contents to his PSP via the Internet.

4 Usages of Mobile Video Devices

Half of the mobile video participants watched video contents on commuter vehicles in the morning, and all did in the evening. The number of commuters in a vehicle was one of the conditions to determine the usage. Those who watched videos in the morning could be seated, or rode on a train which was not so crowded.

Although news and entertainment contents were viewed, multiple participants also viewed contents to learn English or business-promotion videos related to their jobs. One participant even created the contents which had recorded his golf swing and swimming style in order to check his condition. Another young participant viewed or listened to musical contents to practice his musical instrument. These contents categories and behaviors suggested that mobile video users had specific purposes and made use of their commuting time for them.

> "English programs are convenient. Their length is suitable. [I learn] English conversation or English words. A program for English words is 2 or 3 minutes long." (M, 46+)
>
> "My hobbies are golf and swimming. [... omit ...] I took golf lessons and reviewed my golf swing and swimming style. I made video contents of my performance." (M, 36-45)
>
> "I listened to musical contents outside, and viewed them at home. In my case, I use it [Video iPod] by connecting it to the speakers, and practice the instrument viewing the contents with sound." (M, 14-18)

Multiple participants reduced the length of the contents by fast-forwarding the contents or skipping unnecessary parts such as commercial breaks. These usages were useful to adjust the content length to their commuting time, or to repeat the necessary parts. These behaviors also suggested active viewings for their purposes.

> "Well, I view the content [a news program] by fast-forwarding. The playing speed is about twice the normal speed." (M, 26-35)
>
> "Because animated cartoons are about 25 minutes long, I move the cursor to skip unappealing parts." (M, 36-45)

Compared with the case of live mobile TV, mobile video viewings have more steps such as searching for contents, collecting the contents, format conversion, copying to mobile devices, etc. Although the participants had complaints about the time-consuming downloading and format conversion, those who made these processes automatically at night had no problems. They just started the process using their PCs, and did other things, such as visiting web sites, reading e-mails, and going to other rooms during the process.

> "Actually, DivX converter requires much time. After I start the conversion [by PC] at home, I don't touch it. I always start at night and the conversion is finished by the morning." (M, 36-45)

5 Discussions

One of the important findings in this user study is that there were different attitudes concerning the usage of mobile TV and video devices. Many mobile TV participants viewed entertainment programs, especially variety shows on the evening commuter vehicles, only when they could be seated and when the vehicles were not too crowded. They searched for TV programs by turning on and changing channels without TV guides. In contrast, several mobile video participants viewed contents for self-development such as English lessons or business-promotion videos related to their jobs as well as entertainment. They didn't mind spending some time on content preparations that require much effort, which many mobile TV participants were reluctant to do.

These facts imply that the mobile TV users wanted to relax with light entertainment after work by as little effort as just turning on their devices. On the other hand, the mobile video users tried to make good use of the commuting time for specific purposes such as improvement of their English skills. The mobile video users might want to use their commuting time in order to get the most of the cost of content preparation. We think there were differences in desires for the mobile devices when comparing the two groups of people behind the differences in the contents they viewed, preferences for content preparations, and the usage of the commuting time.

Although we have reported only a small number of findings, we hope they will be useful to design mobile TV and video products, especially the ones with both functions, and to improve user experiences when using those products. We also hope to report remaining findings in other occasions, and to do further investigations related to this study.

References

1. Japan Electronics and Information Technology industries Association: The shipment of the receivers for terrestrial digital TV broadcasting (in Japanese), http://www.jeita.or.jp/japanese/stat/digital/2007/
2. Taylor, A., Harper, R.: Switching on to switch off: An analysis of routine TV watching habits and their implications for electronic programme guide design. usableiTV 1(3), 7–13 (2002)
3. Matsuda, M., Okabe, D., Ito, M.: Personal, Portable, Pedestrian: Mobile Phones in Japanese Life (Japanese translation), Kitaoji-Shobo (2006) ISBN4-7628-2532-8
4. O'Hara, K., Mitchell, A.S., Vorbau, A.: Consuming Video on Mobile Devices. In: Proceedings of CHI 2007 (2007)
5. Repo, P., Hyvonen, K., Pantzar, M., Timonen, P.: Users inventing ways to enjoy new mobile services – The case of watching mobile videos. In: 37th Hawaii International Conference on System Sciences (2004)
6. Södergård, C.: Mobile television – technology and user experiences (Report on the Mobile-TV project). VTT Publications 506, VTT Information Technology (2003)
7. Cui, Y., Chipchase, J., Jung, Y.: Personal TV: A Qualitative Study of Mobile TV Users. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471. Springer, Heidelberg (2007)

Live @ Dublin – Mobile Phone Live Video Group Communication Experiment

Erika Reponen

Nokia Research Center
Visiokatu 1,
FI-33720 Tampere, Finland
erika.reponen@nokia.com

Abstract. Live video has gone mobile. In this paper we present an experiment on mobile phone live video group communication, conducted in Dublin, Ireland. We observed 24 people who self- organized into groups for sending and watching real time internet videos on mobile phones over two days. A total of 49 first person view live videos were sent during the experiment. This paper reports observations on attitudes, opinions, communication and context, as well as technical issues regarding the experiment. Findings include varying preferences between live vs. delayed video as well as between following vs. sending live videos. We describe some of the positive and negative feelings that the experiment caused. Finally, we also discuss implications of this technology for wider user populations.

Keywords: Mobile phones, live video, mobile video, video communication, interaction.

1 Introduction

It is nowadays possible to share live video content from mobile phone to internet and thus to multiple watchers. This means that there is a new interaction method available in everyday environment, with the benefit that communication between group members can be enhanced by live video streaming. Video communication can happen on move, in mobile context, thanks to built-in video cameras and sharing capabilities such as internet browsers of current mobile phones.

In this paper we present the results of an experiment which was conducted in 2007 in Ireland. We wanted to better understand possibilities of live video in the area of mobile phone communication and study peoples' attitudes towards live video sharing and following as well as being recorded for live video stream in mobile context. We describe real time video material which was recorded by study participants and analyze responses towards the experiment based on field observations, a questionnaire and group discussions.

We observed 24 people who self-organized into groups for sending and watching live internet videos on mobile phones over two days. Each person in each group had the possibility to send and follow these videos from a mobile phone or to take part in sending as a member of a group. Videos were shared to the internet which in theory

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 133–142, 2008.

made them available for anyone, though the intended audience was the experiment group members and some people who were invited to observe. General thoughts about the experiment included quite equally positive, negative and neutral opinions. Most of the problems and negative feedback were caused by technical issues.

We were interested in how participants use live video as a part of their group communication. Participants' activities and feelings as cameramen, audience and targets were observed. We note also related work as well as discuss reflections and influences of mobile live video sharing for future group video communications.

2 Related Work

Video calls over the internet are now becoming within the reach of the masses. Recent versions of telephony and messaging software such as Skype and iChat allow shared video conversations, while mobile networks (for instance, 3G) are enabling mobile video calls. In general, video has arguably become a first-class internet data type (witness the rise of YouTube (http://www.youtube.com)). The literature seems void of similar mobile live video group communication experiments like the one presented here. The experiment setup therefore provided a novel possibility to examine in practice the issue that has been discussed but not widely experimented in practice.

2.1 Related Research

There is not much specific research on mobile live video group communication available, but lot of research has been done on capturing and sharing visual content in mobile contexts. The findings there apply to our study as well.

Koskinen et al. report that people familiar to each other rely on mutual trust for controlling sharing of sensitive material [1]. When sharing something to a public internet page, anyone may see the posted content. We note that, since the audience is anynomous, this spreading to secondary context (extended context which is reached via technical tools) of unknown scale and composition creates privacy concern for the creators of the content as well as the targets for recording [2]. Media-archaeology studies by Huhtamo reveal that camera has always had a de-humanizing effect on person carrying it [3]. We observed that the attitudes of target persons towards continuous video capturing changes during one month from excitement to irritation to ignorance [4]. In a mobile still imaging study, Kindberg et al. found that there is major difference between sending an image at time of capture or sharing it later; choices for sending time depend on timeliness communication requirements [5]. A mobile video telephony study by O'Hara et al. suggests that key drivers for video communication in mobile contexts are sharing special occasions and showing things to talk about. Same study discusses also social and practical barriers such as privacy management in public spaces and problems with ambient noise and lightning toward video communication [6]. Jacucci et al. noticed that in large scale events spectators experience the event together also many other ways than watching [7] and that event information, media sharing and awareness between group members are all important, media sharing being the most central of those [8].

Enabling technology. Various video sharing applications are available for current high end mobile phones with internet capabilities. One directional live video streaming from mobile phone to internet, from one to many, can be done for example using software by ComVu (http://comvu.com) and Qik (http://www.qik.com). Also Kyte (http://www.kyte.tv) enables sharing live image material to internet. Live two-way communication is possible via mobile phone video call. Videos recorded with mobile phone can also be sent delayed directly from mobile phone to YouTube (http://www.youtube.com). Many kinds of one and two-way video communication software are available for PCs.

3 Experiment

The participants of the experiment were 24 colleagues from a research department. Because the participants knew each other in advance, they formed a natural group of people, shaping also natural sub groups. Two of the participants were female, others male. The experiment was arranged on a group trip where the participants traveled together to an unfamiliar city. Author was a natural member of the group, taking part into the experiment activities and acting as an observer and also moderator in group discussion but not answering to questionnaire. The participants extended their natural group communication during the trip with live video sharing capabilities of their mobile phones. All participants were experienced mobile communication technology users and some of them had used live video software before the experiment. 23 participants took part in the questionnaire and group discussions.. The participants shared live video streams to public internet pages, without restricting the amount of viewers. Anyone may have followed the videos, but the links to the web pages and streaming time notifications were shared only to the participant group and some other familiar people who were invited to follow the videos from their mobile phones or PCs. These members of extended audience participated to study only in roles of audience, not taking part into the other activities. Some sms feedback (to live video stream senders) from the extended audience are cited in this paper, though.

3.1 Technical Setup

ComVu Pocket Caster live video software was used mostly in Nokia N95 multimedia computers in the experiment. Each participant had suitable equipment but not all of them were able use the software in their mobile phones during the experiment.

3.2 Experiment Setup

The participants got the instructions for experiment by email in advance. Technical specifications for the needed device as well as instructions for installing the needed software for the device were given. We asked participants to use these devices with installed software as their primary mobile phones during the experiment. Group members were also asked to define a group to their phonebooks, including phone numbers of all participants and extended audience, to enable fast and easy group sms sending. They were also guided to create in advance a text message template for announcing upcoming live videos and share a link to the sender's live video web site. We asked participants to send message before each live video stream sending action.

The experiment started with an information sharing session on site right after arriving in Dublin. In this session we announced general schedule for the experiment and gave last minute technical support. Groups started to build up then.

The experiment lasted for two days, from Friday noon to Sunday noon. Participants spent leisure time in Dublin in dynamic ad-hoc groups. People inside groups as well as between groups communicated during the period by normal everyday communication methods (such as sms and call), they are used to use, but extended with live video sharing. The sub groups were dynamic, some participants moved from group to another during the time period but there was still some general stability in the groups. People were encouraged to save live videos for future viewing. Notification group messages were sent before live video streams. Quite often there was also sms feedback from videos coming from the audience, including messages from experiment participants as well as from the extended audience.

As a parallel task to the live video experiment reported here, participants recorded also mobile phone videos which were not shared live but submitted to YouTube.

Before the end of the experiment we arranged a questionnaire and group discussion concerning the experiment. General feelings toward live video in mobile context were asked as well as willingness and possible reasons for this in the everyday life. More specific questions were such as: preferences for sending and following, subjects, places and times; comparison between sending vs. watching and live vs. delayed video; feelings of being a target for recording; problems during the experiment and thoughts of seeing through someone else's eyes. Most interesting questionnaire topics were further discussed in group discussion. Although delayed video sharing is not the subject of this paper, some comparison between live and delayed video experiences is presented in this study, based on the questionnaire answers and group discussions.

After the trip we collected the recorded material that was successfully saved. Also data about sent notification messages and received text message feedback were collected. We also arranged a group discussion three weeks after the experiment. During that we watched part of the saved live videos again. For example, technical quality and differences of requirements for live vs. delayed videos were discussed.

4 Results

Mobile phone live video group communication was successfully tried out during the experiment. Eleven of the 24 participants succeeded to broadcast live video from their mobile phones. Every participant was taking part into sending as parts of groups. 49 live videos were shared during the trip, and 33 of them were successfully saved. All participants saw live videos during the trip and were targets for capturing.

General thoughts about the mobile phone live video group communication experiment were quite equally positive, negative and neutral. Most of the negative feedback was caused by technical issues. When participants were asked to define feelings about watching live during this experiment in couple of words, the received answers were: *fun, excellent, superb, pretty good, relaxing, ok, hard, challenging, technically challenging, frustrating, crappy and boring*. Sending live video was commented to be:

fun, fun and easy "after a lot of mess", meaningful when something fun to send, very nice, ok, new interesting way, quite easy, cumbersome, impossible, still immature tech, technically difficult, boring and too complicated, consuming battery quickly, not working in practice, a violation of privacy and good idea but technical problems. These simple questions and free text answers give support to the preconception that people have widely variable personal opinions towards video communication. Some people like to send and some prefer following.

When video was meant to be shared live, watching it later was considered quite boring because audio quality was always not good and the events were gone and seen already. Possible interesting fun parts were difficult to find again and even if they were found, the quality was not good enough to create a pleasant viewing experience.

More than half of the participants would like to send live video even more actively in the future if the technological and financial circumstances were optimal: 12 participants would share the material with familiar people and 6 with anyone. 3 participants would not like to share live video at all.

Fig. 1. A live video stream frame. One of the groups is sharing their experiences to the audience. One of the participants (left) acts as an ad-hoc commentator in this situation.

4.1 Communication During the Experiment

General sms, mms and voice call communication was happening during experiment just like it would have happened during a normal weekend with the group of friends traveling together. Information such as location details and schedules were messaged or called. ComVu Pocket Caster software was used only for live video sharing.

Live Video Communication. Usually many members were participating in each video, one being a cameraman, very often someone(s) acting as a commentator(s) or interviewer(s) and others being targets for video recording or audience. Sometimes the cameraman acted as a commentator and an interviewer. Based on the field observations and group discussions, videos with commentators were liked most. Commentary helped to understand the context such as location, event and situation. Commentary was often used in the beginning for introducing the location, possible plans, the people present and the person behind the camera. Usually there was also an end speech, reporting again some context information and possible plans.

Sometimes speeches were pointed towards extended audience, and sometimes for certain participant who might be interested in some special information. Those kinds of private messages in public live videos enabled live video communication to function in multiple levels. One example of the targeted live video communication was a situation where a member of one sub group decided to move to another sub group after dinner. Groups shared videos which included information about locations, group composition, cues telling where is “the place for the best party” and navigation information. Videos for extended audience were mostly group greetings and fun making. There were guesses about different weather in cities and mentioning audience members by name. Because of the revealing characteristics of video format, basically all videos, even the “only for entertainment” ones included some information which the others could utilize. One participant mentioned his favorite video being the one from the lobby bar because that way he saw who were there and he decided to join. The video was not mentioned in that purpose specifically but as greetings for the audience not participating to a trip. This result suggests that single live video functions in many communication levels as well as in context sharing.

Video messages were one-directional but sometimes one of the group members sent and one received video, which made it possible to have two-directional video communication, with 2 devices. If many streams were running at the same time, many devices were used in group to follow them all. About 15 seconds delay in starting live video caused some communicational problems. That and delays in group message deliveries is one reason why it was not always clear if there were audience already or anymore. Most of the videos had a fun, comic mood, with any content and context. This resulted from the short use period, leisure characteristics of the trip and excitement for new thing. Even though knowing the audience is important, the action of recording in group situation has a certain value, in shaping the group dynamics [see also 4] which in some situations lowers the influence of the audience.

The experiment results confirmed our initial hypotheses of popular and unpopular contexts for mobile live video communication. Out of 33 saved live videos, 10 were taken outdoors (park, beach, street), 6 at bar, 5 at airport, 3 at museum, 2 at hotel, 2 at bus and one at office while leaving to trip (Because the experiment was made on a trip, home and work locations were not included). Outdoor videos were often recorded on move, which gives an intense first person view feeling. Indoor videos are more static, moving the camera, not the cameraman. Quality is better in most static ones. On move videos are ad-hoc, lasting usually longer, showing mostly walking and talking group, and showing and commenting the surroundings. Indoor videos are often concentrating on interviews, greetings and showing people, special events such as music gigs and museums or bar conversations and joking.

Questionnaire results showed that the most popular context for following live video was the bar (9). 3 participants considered vehicles as the best environment, 2 wanted to see videos anytime and 2 requested a place where they can hear the audio well. The popularity of the bar environment for video communication arises form the fact that there people usually want to show where they are, know where the others are and what are the most popular places. That helps e.g. in planning the continuation of the evening. Other reason is the leisure time and relaxed feeling. One difficult thing in bar environment was considered to be hearing audio. On the other hand privacy concerns caused by public audio communication are bigger is places with less noise.

Live video was commented to be fun addition to group communication during the trip. With less technical problems it would have been used even more. Many fun moments were experienced during sending and following the videos in group and talking about them afterwards. There were even some "legendary" videos (such as a a girl surprisingly appearing from below the bar table) which many of the group members may remember for a long time.

Group SMS Notifications on Upcoming Live Videos. Participants sent a group SMS before starting to broadcast live video. Messages were sent 20 times. This means that messages were not sent each time live video was sent and also it happened that streaming was not working at first attempt and needed to start again so there was no need to send new info message because timeframe was so small between attempts.

The following notification messages were sent (links to personal ComVu web pages are here replaced with "[link]"): "See my video feed NOW at [link]","At Temple bar: [link] "Comvu(name): [link]", "Yo! Go now to [link] if interested seeing entertaining live video by (name)!", "Guinness fifo queue: [link]", "Live from Guinness stockhouse: [link]", "[link]", "Live stream from above the rooftops of dublin. Guinness gravity bar: [link]", "Video stream coming soon: [link]", "Feel the boogie. Live jazzy blues music: [link]", "Yo! Go now to [link] if interested seeing entertaining live video by (name)!".The notification messages show that participants either used a general message just to share the time and web address of their live video or defined more specifically the intended content or location. The fastest way to send an alarm message and video was to use ready-made default notification message. Default messages were used in sudden situations, while specific ones were for static or planned situations.

Feedback from Live Video Audience. Although live videos were one-directional, interaction happened anyway. It seemed to be natural to answer to the live video with sms. Text messages were received from the participants in Dublin as well as extended audience from around the world. Freely translated SMS feedback messages were such as: "Showing well :)", "Rainy here, I would rather be there in Dublin… Have fun!", "I guess we were late from the broadcast" "Do you have a fan club? I'll join". Text feedback was considered very important, giving a feeling of interaction and proves that someone was really following the videos. Simply getting feedback was more important than actual message content.

4.2 Roles as Cameramen, Audience and Targets

The attitude towards video capturing and the part the user likes to act is personal. This experiment proved the assumptions that some people like to be recorded and flirt to camera while others rather disappear from the scene. There are also persons who ignore recording. In this experiment 7 participants liked and 5 didn't like when they noticed being captured on live video. 5 persons ignored being on video and 2 persons commented that they hoped not to say anything stupid when they are on live video. Some people are willing to watch the videos, while others are not so interested in following videos; same applies to recording videos. Ongoing era of social software gives possibilities for showing up as well as spying. There were sympathizers for both these viewpoints in the group of participants.

In the future 8 of the participants would like to follow videos by familiar people (friends, family, peers or traveling company), 5 of the participants by people with common interests, 5 by anyone, 3 by celebrities and one by professionals. 12 participants would like to share the live videos with familiar people (friends, family or colleagues), 6 with everyone and 2 would not like to share them at all. Participants would like to use live video for example for greetings, events, parties, travel and unexpected or exiting situations. Results suggest that video communication between familiar people is most popular but there are also persons who like to have wider audience as well as see live events from around the world, not depending on the recorder, but the content or time.

4.3 Live vs. Delayed and Sending vs. Following Live Videos

15 of the group members liked more following live videos than send them while 8 liked sending more. When we asked about following live vs. delayed videos, 16 answered to prefer following live videos and 5 watching videos later. Asking same live vs. delayed question regarding sending the videos, 13 answered to prefer sending live videos while 5 prefer sharing videos later. Some participants didn't have a clear opinion. The participants didn't put as much requirements for live video as for delayed. Live video has an important ability to give information about what and where is happening "right now", and future plans, while delayed videos should contain interesting content and/or good technical quality to be useful. Based on the study we notice that live video is mostly context oriented while stored, delayed video is focusing more on content. Delayed videos were automatically stored to public web page, while live videos by default were visible to audience only real time. One characteristic of delayed video is the possibility for the cameraman to see the video. Communicatively interesting thing is that seeing videos during group discussion after the trip, made it easier to recall the event and discuss afterwards. Watching the videos later caused some fun moments to the "owner" sub groups but others had some lack of interest in following them. Some discussion about the locations and activities happened, for example from the places which not all sub groups visited (such as coast) and from the places which many groups visited different time (such as park).

4.4 Technology and Communicational Problems

Technology problems were mostly caused by immaturity of the technology. There were also difficulties in installing and signing in to software. Lack of field and high battery consumption caused problems too. Also audio and video quality was complained, but accepted because of the mobile nature of the experiment. Usability problems were ones connected to communicational issues and multitasking. Combining video communication to sms and call was problematic. That was caused on the fact that one directional video took the device for its own use. Form factor of the device caused problems in multitasking, one proof being black videos, recorded with camera lens protector on. Moving between messaging, phone and live video applications caused some usability and user interface problems. Location and context of the other group members was only received via sms, mms, call and live videos.

5 Discussion

Our study in Dublin was just a small scale experiment, but it gave assumption that mobile live video communication has potential for future social media and it will also create new design challenges. This experiment lasted two days so we can't draw any conclusions how the video communication behavior would have developed during the longer period, but experiment results suggest that live video communication brings new level to everyday group communication. Earlier research tells that during longer use-period, targets in many situations start to ignore being captured to one-directional video [see also 4]. We don't have proves yet, but in two-directional leisure time video communication, camera may be ignored as a technology, but considered as a telepresence bridge between places.

Audience has an effect on cameraman as well as on target and audience. For cameraman and target it is important to know the audience to have a feeling of privacy. Most important thing seems to be to know if there is any watcher or not, but the more defined the audience is, the better feeling of privacy is reached. Because audience has an effect, what does it mean if sender or target does not know who watches? This is an important question which raises from open nature of the internet and social media. From audience's viewpoint, it's important to know whether the video is specifically targeted for someone.

There are usually "passers-by" captured to the video, unintended but as part of context. Restrictions for recording in public places vary in different countries, but typically it is not prohibited to record normal everyday life if intentions are not bad against certain people. There are no clear laws for this but it is clear that in some cases these unintended recordings and consequences of that may be unwanted for the captured passer-by. Tapes of the public surveillance cameras are kept strictly confidential if no special reason to use them, but mobile phone video recordings made by citizens on public spaces have no such protections, anyone can post them live or delayed. Ambivalent mobile video sharing has made this privacy thread very real [See also 2 and 4]. Although there are privacy threads, this is not necessarily bad or good thing; it is just a big change in society, we are facing. This is challenge for designers of the devices and services but it is also a challenge for citizens in this new situation. Study participants wanted to share both unexpected and planned situations, and mobile phone video camera made those both possible.

Many technical issues will probably be cured almost automatically during time but certain problems such as communicational issues (such as multitasking with devices) and battery consumption need some problem solving. Battery consumption of live video will not change easily but the mobile phone video should be available in every situation, without charging possibilities. Multitasking such as simultaneous sending and following video and using other communication features of the device at the same time needs solutions in the areas of hardware, software and user interface.

We found out that although experimented live videos were one-directional, communication was two-directional. To create a wanted interaction state, various communication methods, not only video, were used. There is certain need for two way communication, be it video, audio, text or something else or any combinations of those. Even though the participants were familiar with technology, usability problems were found. If masses of people will in the future use mobile live video for group communication in work and leisure, many usability challenges will arise. Live videos are good for communication and context awareness while delayed videos are used more for sharing content. That's the reason why users set less quality requirements for live than non-real time video content. Because of quality and instant characteristics of live video, re-watching it was considered quite boring. Participants liked commentary in live videos because it helped understanding context. Participants would like to use live video mainly with familiar people for sharing greetings, events and special situations, but possible use cases were not limited to those. There are interested people for all roles in live video communication. Study suggests that live video on mobile context gives new possibilities for group communication and future social media environment; most of the experiment participants want to share and follow live video in the future. There is still lot of work ahead to make it happen in large scale and design good system for various communication needs. Live video is not only beneficial for business conference calls and calls between two persons but also for fun leisure time group communication in mobile context.

References

1. Koskinen, I., Kurvinen, E., Lehtonen, T.-K.: Mobiili kuva, pp. 80–81. Edita/IT Press (2001)
2. Reponen, E., Huuskonen, P., Mihalic, K.: Primary and secondary context in mobile video communication. Personal and Ubiquitous Computing journal (2007); Online. Springer
3. Huhtamo, E.: Pockets of Plenty: An Archaeology of Mobile Media. In: Proc. ISEA (2004), http://www.isea2004.net/proceedings/
4. Reponen, E., Lehikoinen, J., Impiö, J.: Mobile Phone Video Camera in Social Context. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551. Springer, Heidelberg (2007)
5. Kindberg, T., Spasojevic, M., Fleck, R., Sellen, A.: I Saw This and Thought of You: Some Social Uses of Camera Phones. In: CHI 2005, pp. 1545–1548. ACM Press, New York (2005)
6. O'Hara, K., Black, A., Lipson, M.: Everyday Practices with Mobile Video Telephony. In: CHI 2006, pp. 871–880. ACM Press, New York (2006)
7. Jacucci, G., Oulasvirta, A., Salovaara, A.: Active construction of experience through mobile media: a field study with implications for recording and sharing. Personal and Ubiquitous Computing journal (2006); Online. Springer
8. Jacucci, G., Oulasvirta, A., Ilmonen, T., Evans, J., Salovaara, A.: CoMedia: Mobile Group Media for Active Spectatorship. In: CHI 2007, pp. 1273–1282. ACM Press, New York (2007)

A Lightweight Mobile TV Recommender

Towards a One-Click-to-Watch Experience

Arian Bär, Andreas Berger, Sebastian Egger, and Raimund Schatz

ftw., Forschungszentrum Telekommunikation Wien,
Donau-City-Straße 1, 1220 Wien, Austria
{baer,berger,egger,schatz}@ftw.at
http://www.ftw.at

Abstract. In this paper we present a novel program recommender for DVB-H based Mobile TV systems. It addresses the restriction of low processing resources on the client devices and is fully compliant to the OMA BCAST[1] standard. There is no learning phase included and hence users can instantly start using the system. Based on the proposed, highly scalable architecture design we implemented a prototype application that employs a powerful, though minimal, mobile user interface.

Keywords: Mobile TV, EPG, recommender, content-based, prototype.

1 Introduction

Current research on Mobile TV usage [1,2,3] has shown that a majority of Mobile TV usage happens in waiting and commuting settings. According to the long term study in [2], the average tolerated clip duration in such settings is five minutes or less. Additionally, users' concentration resources are in short supply in mobile situations. Hence, they are not willing to waste time on selecting the program of their choice. In contrast to TV program selection in stationary living room settings, where frequent zapping and EPG browsing are common, for Mobile TV even zapping just a few channels already consumes intolerable amounts of time. Reading detailed EPG data for e.g. as less as four channels would require more time than the user is willing to spend on a clip in total. In addition, practical restrictions such as text size and representation on the mobile interface have to be taken into account. Therefore, dedicated recommender systems have great potential to improve the Mobile TV experience. We have identified the following requirements for such systems:

- Scalability
- Low request/response latency
- Compliance to current broadcast standards
- Instant adjustability to varying user moods
- Protection of user privacy

[1] http://www.openmobilealliance.org/Technical/release_program/bcast_v1_0.aspx

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 143–147, 2008.

2 An Overview of Recommender Strategies

Present approaches for recommendation or personalization systems can be subdivided in the following two main categories:

- **Collaborative filtering** systems deduce their decision based on a collection of *ratings* or *usage logs* of like-minded users and their match to the present user profile. An overview of such collaborative strategies can be found in [4]. Requirements of such systems are the ability to gather user data, to transmit it (to servers or peers), and to parse these archives when a recommendation is inquired. Hence, scalability is an issue as these implementations often degrade with an increasing number of users [5].
- **Content-based filtering** tries to extract information *from the content* via audio and video extraction methods as [6] or to allocate information *about the content* from different sources as the EPG [7] or content categories. This information is then compared to the user's preferences and a recommendation is given. This approach requires mainly high computational power for cognitive approaches and for e.g. EPG parsing.

Concerning their application in the domain of Mobile TV both types of recommendation systems have their shortcomings. Commonly, either type matches the program content to viewing preferences with the aid of user profiles. To obtain such profiles often a certain learning phase is required, with the goal to map users to a discrete set of certain categories, e.g. age, gender, or social background. Learning phases have to be repeated when the user's viewing preferences change. A general problem of many recommender systems is their incompatibility to established TV standards and their need for extensions to them. Moreover the users' viewing preferences reveal a lot about their personality and hence such systems must protect their privacy.

3 A Standards-Compliant Mobile TV Recommender

Our recommender system employs a generalized text mining approach and falls in the category of content-based filtering systems. It targets Mobile TV services and is compliant to the established standards OMA BCAST and DVB-H.

3.1 Relevant Components and Standards

The Recommender Engine used by us has been originally optimized for home TV set-top boxes [8]. It accepts natural language text input as well as four percentage values reflecting the preferences for *fun*, *action*, *thrill* and *erotic*. The recommendation process is done in four steps: first topics in words are extracted. Second the text is scanned for the emotions it contains. Then distances between the topics are computed and finally a *score* is calculated for each entry. The resulting programs are given back as a list, ordered by this *score*.

When new EPG data is imported into the system every EPG dataset has to pass through the first and second step of the recommendation process. Each EPG dataset is stored in the database along with the computed data as a Semantically enriched EPG Dataset (SED). By the time the user queries a recommendation only the input data has to pass the first and second step. Then the query is compared with the stored SEDs and finally the score is calculated.

The OMA BCAST standardization efforts specify the delivery of service and program guide information. The EPG information is delivered as an XML fragment and the entries are encapsulated by the complex element *Content*. For our system we employ the child element *TargetUserProfile* to carry the additional semantic information for the SEDs.

DVB-H is a recent standard that specifies digital video broadcasting for mobile devices. As it is based on IP it can be used to transport generic data. Due to the use of the FLUTE protocol, file objects can be reliably sent over the broadcast channel. The XML-encoded ESG/EPG information is transported by employing this mechanism.

3.2 Distributed System Architecture

We designed and implemented a recommender system using the components described above for the usage in a Mobile TV setting (see Fig. 1). In the design of the recommender engine we identified a two-stage processing cycle. Firstly, the SEDs need to be generated once for all clients by the Service Guide Generation Server. Secondly, the request for a recommendation can be executed on mobile devices with low computational resources, since the system minimizes the processed EPG data. Consequently, this architecture has no scalability issues and is perfectly suited for Mobile TV clients. Furthermore, the user is able to express her preferences in a fine-grained way: Instead of choosing from a discrete set of categories or moods, the recommender application can precisely reflect the user's current viewing desires by employing a simple user interface (see Fig. 2). The intended process flow consists of the following steps:

1. The recommender server component processes the program descriptions found in the EPG in advance for e.g. a period of one week. The output of this calculation are the SEDs which are then embedded in the channel's ESG.
2. In the targeted DVB-H setup, the SEDs are continuously transmitted within the standard ESG data by the underlying FLUTE-based data carousel. The SEDs approximately double the size of the original EPG and thus account for no significant increase in traffic on the broadband DVB-H channel.
3. As the mobile client is tuned to a particular DVB-H channel, it receives the current ESG, extracts the SEDs and stores them to disk. Outdated SEDs are automatically deleted.
4. Whenever the user wants to watch TV, she makes adjustments to her personal user preferences or accepts the stored settings. According to these preferences, matching recommendations out of the set of currently running programs are computed from the stored SEDs. Finally the user starts watching the most suitable program by issuing a single click.

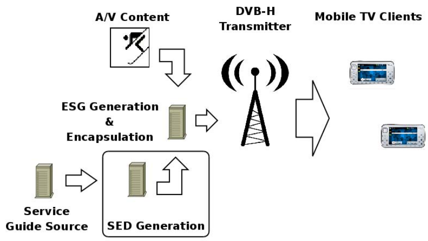


Fig. 1. Integration of the SED Generator in the Mobile TV architecture

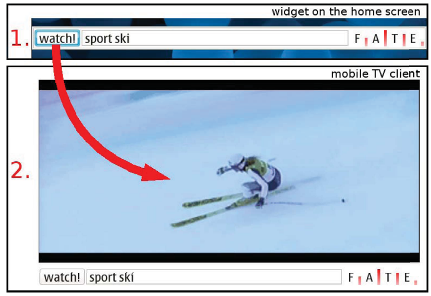


Fig. 2. Clicking "watch!" on the home screen widget starts the mobile TV client

Our approach has several advantages: It only requires a unidirectional broadcast link, thus avoiding the costs and the roundtrip latency of a wireless backchannel. Secondly, the user does not have to perform any registration and profiling steps before utilizing the system. Thirdly, the user's privacy is perfectly protected since no personal information ever leaves the client device. Furthermore, our approach allows the user to instantly adjust her personal viewing preferences anytime, instead of having the recommender system relearn her changed profile. On the operator's side this results in significantly reduced service complexity for user management and accounting.

4 Implementation

Our prototype implementation is based on the AMUSE testbed [9] and Nokia N800 clients. In general, the design of a Mobile User Interface (MUI) has to have a user-centric minimized set of functions [10]. Our application is designed as a widget on the device's home screen and the user has access to all recommender functions with only a single control bar (see Fig. 2). The text field accepts natural language where the user can describe her currently preferred program. Each of the four mood buttons (F for *fun*, A for *action*, T for *thrill* and E for *erotic*) is followed by a bar indicating the intensity of each mood dimension of the ideal program.

After clicking the "watch!" button the entered values are given to the local recommender engine. Within a few seconds a recommendation out of the currently running programs is calculated and the best candidate is directly played in the Mobile TV player. When the user switches to the next program the next best due to the recommender score is shown. Next time the user switches on the device, she gets the recommended program with only *One-Click-to-Watch*, since the application just needs to store the previously entered values.

5 Conclusion and Future Work

In this paper we presented a lightweight recommender system for Mobile TV which is compliant to the OMA BCAST standard. We have identified requirements for such systems and have shown how these requirements can be met by adding low additional complexity to present architectures. Our prototype implementation features a minimal interface enabling a *One-Click-to-Watch* user experience. Our future activities include research on integration of targeted advertisement within this architecture. We currently plan to conduct user tests to evaluate the performance and to verify the design decisions taken for the Mobile User Interface.

Acknowledgements

This work was supported by the Austrian Government and the City of Vienna within the competence center program COMET. We would like to thank the Chair for Artificial Intelligence at the University Erlangen-Nürnberg for providing their recommender engine as well as for extensive support and fruitful discussions.

References

1. Carlsson, C., Walden, P.: Mobile TV - To Live or Die by Content. In: HICSS 2007: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, Washington, DC, USA, p. 51. IEEE Computer Society, Los Alamitos (2007)
2. Soedergard, C.: Mobile Television - Technology and User Experiences. Technical report, VTT (2003)
3. Schatz, R., Egger, S.: Social Interaction Features for Mobile TV Services. In: BTS 2008: IEEE Broadband Multimedia (2008)
4. Adomavicius, G., Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)
5. Aroyo, L., Bellekens, P., Bjorkman, M., Houben, G.J., Akkermans, P., Kaptein, A.: SenSee Framework for Personalized Access to TV Content. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 156–165. Springer, Heidelberg (2007)
6. Wang, J., Duan, L., Xu, L., Lu, H., Jin, J.S.: TV Ad Video Categorization with Probabilistic Latent Concept Learning. In: MIR 2007: Proc. of the int. Workshop on multimedia information retrieval, pp. 217–226. ACM, New York (2007)
7. Yamamoto, N., Saito, M., Miyazaki, M., Koike, H.: Recommendation algorithm focused on individual viewpoints. In: Consumer Communications and Networking Conference, 2005. CCNC, January 3-6, 2005, pp. 65–70. Second IEEE (2005)
8. Ludwig, B., Mandl, S.: Centering Information Retrieval to the User. In: International Workshop on Recommendation and Collaboration at IUI 2008 (January 2008)
9. Schatz, R., Jordan, N., Wagner, S.: Beyond Broadcast – A Hybrid Testbed for Mobile TV 2.0 Services. In: ICN, p. 87. IEEE Computer Society, Los Alamitos (2007)
10. Subramanya, S.R., Yi, B.K.: Enhancing the User Experience in Mobile Phones. Computer 40(12), 114–117 (2007)

Local Communities: Back to Life (Live) Through IPTV

Marianna Obrist[1], Elke Beck[1], Sara Kepplinger[1], Regina Bernhaupt[2], and Manfred Tscheligi[1]

[1] ICT&S Center, University of Salzburg
Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
{marianna.obrist, elke.beck, sara.kepplinger, manfred.tscheligi}@sbg.ac.at
[2] University Paul Sabatier, Toulouse III, IRIT-LIIHS
118, Route de Narbonne, 31062 Toulouse, France
regina.bernhaupt@irit.fr

Abstract. Enhancing local communities is seen as an enrichment of people's life by providing additional possibilities for the building of relationships and identities. The question is how to improve these local social ties and how to promote interaction among people in local communities. Within this paper we will address this question based on a case study, where a local community is enhanced by an IPTV platform to exchange user-generated audio-visual content. The paper outlines major steps to be considered in a community building process based on results from active user involvement throughout the whole design and development process of new applications in an European project.

Keywords: IPTV, user-generated content, local community, usability and user experience evaluation, methods.

1 Introduction

To understand why it would be beneficial for local communities to use IPTV for their purposes, we have to take a closer look on societal changes. As many researchers from various disciplines (e.g. [3]) have stated before that our living environment has changed due to structural changes in several places: societal development not only changed the way we are dwelling, working and travelling, it also affected our experiences of place and our place related relationships. Focusing on the home, research on demographics and social trends shows that household size is declining and a diversification of household types takes place [7].

Although 'traditional' forms of how one is living still remain significant. Nevertheless there is a development towards more spatial mobility and individuation in everyday life. Some examples will make this point more clearly: Working population is commuting from home to workplace, residents migrate to second homes, people decide individually about their path of life, and are (largely) independent

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 148–157, 2008.

from social assurance via networks of family and relatives due to institutionalized social welfare [17,23,15]. The post-modern individual is confronted with a diversification of places, which are all (more or less) relevant to the individual and from where identities are shaped. These developments are related to changes in meaning and occurrence of relationships in local communities.

Different viewpoints exist whether local communities are loosing strength and importance or not. Guest and Wierzbicki [5] have analyzed the social attachment of people towards their neighbours and non-neighbour friends and found out that there is a slight but continuous trend of limited socializing with neighbours and a slight strengthening of relationships with non-neighbour friends. At the same time they state a high quality in social ties among neighbours. These and other findings (see [21]) show that local communities still play an important role in people's life — despite the variety of other communities and places.

This paper is not valuing the changes related to local community ties as good ones or bad ones. Enhancing local communities is seen as an enrichment of people's life by providing additional possibilities for the building of relationships and identities. The main research question is *how to enhance these local social ties* and *how to promote interaction among people in local communities?* In this paper, preliminary qualitative results from first investigations of a rural community and their requirements are presented. Moreover, the main implications for the design and future deployment of the IPTV platform are summarized. First of all the basic cornerstones of our research are described in detail.

2 Related Work

2.1 Characteristics of Local Community

Particularly since the invention of the automobile, sociologists have debated whether communities need to be geographically collocated [1]. George Hillery found 94 different definitions of 'community' used in the literature since 1955 and concludes that most writings have the "basic agreement that community consists of persons in social interaction within a geographic area and having one or more additional common ties" [8, pp.111-123].

In contrast, in 1979 the field of social network analysis was becoming prominent, and lending increasing evidence to support the idea that both weak and strong interpersonal ties can exist over distance — communities don't need to be geographically collocated. Wellman and Leighton make a compelling case that 'neighbourhood' and 'community' are separate concepts, which are both important [22].

In this paper, the term local community is used in the following sense. Local community is not only meant to be the geographically defined meaning of people living close to each other or in a village. We also use the notion of local community to describe a community in its simple sense (social binding, share of interest, social interaction etc.) referring to regional cases and locally based topics.

2.2 IPTV for Local Communities

Nowadays, the TV is the most wide spread technology among European households. Although, access to Internet still exists mainly via the PC and seldom via the TV. However, IPTV will become more important in European households in near future [6] and will provide the technical basis for user-generated content on the TV. The main question will be, why should local communities be interested in user-generated content on TV? Previous research on (online) communities clearly shows, that cohesion in communities is based on sharing the similar interests (bonding) but also on differing attitudes and relationships (bridging) [12] as well as user requirements [13].

In our research we can also build on previous and ongoing projects focusing on community TV and user-generated content trends on TV, such as 'Buntes Fernsehen', 'Lommel TV', or 'SAMBA'. Buntes Fernsehen – literally 'colourful TV' – is a local TV service available on broadband and on the local cable service in a small village in Austria. It offers community members the possibility to produce videos and post them on the local net-based TV station (see http://www.buntesfernsehen.at/). 'Lommel TV' is another project located in a small town called 'Lommel' in Belgium. It's a do-it-yourself community TV channel allowing people to post videos (see http://www1.alcatel-lucent.com/bnd/lommel/). In the SAMBA project – an Interactive Community TV project in Brazil – the main aim is to create a framework for allowing local communities and citizens (including low income population) to access community-oriented content and services by means of iDTV channels (http://www.ist-samba.eu/).

The above examples show that the TV environment is changing, characterized by new technological possibilities (e.g. Broadband). Moreover, it becomes clear that user-generated content concepts based on examples such as YouTube (http://www.youtube.com/) are transferred to the TV. This makes it a challenging research area from a design and HCI (Human-computer Interaction) perspective, as well as from a user perspective, in the sense of changing user behaviours, community-related aspects as well as usability, sociability and user experience factors (see [14]).

2.3 User-Generated Content on IPTV

User-Generated Content (UGC) platforms such as YouTube seem to be the glue that sticks community members together. In particular audio-visual content like photos and videos gain a lot of attention among non-professional users [9]. Producing and distributing media content is nowadays not the exclusive domain of telecommunication and media professionals anymore. End-users are participating in peer-to-peer networks by producing, distributing and consuming content [2]. Communities – no matter if online or local – need social interaction to enable social cohesion. In the case of online communities, social interaction takes place as well through making user-generated content available and consuming this content. Therefore one success factor of online community applications is their ability to provide a platform for content exchange.

Due to the success of online communities where users exchange content, one might ask why there should be user-generated content on TV. To explain the reasons, we will firstly have a short look on online communities as well as different usage types among community members. By using the dichotomy of 'activeness' and 'passiveness', we want to make a clear differentiation between users who mainly consume content (passiveness) and users who also produce and distribute content (activeness). Of course one user can inherent both roles. These two behaviours 'ideal types' are also related to lurking and posting. Considering the extent of both user behaviours, lurking is much more common in most online communities than posting [16]. There is a minority among the community members who contributes content and a majority of community members who consume the content. According to Jacob Nielsen [11] in most online communities there exists a 90-9-1 rule:

– 90% of users are lurkers (i.e., read or observe, but don't contribute).
– 9% of users contribute from time to time, but other priorities dominate their time.
– 1% of users participate a lot and account for most contributions.

"The first step to dealing with participation inequality is to recognize that it will always be with us. It's existed in every online community and multi-user service that has ever been studied" [11].

When now focusing on the TV, we can see similar developments (to the 90-9-1 rule) in interaction behaviour in front of the TV occurred through interactive TV applications, which encourage more user participation. It's also common in social theory to approve the 'passive' television audience some (mental and physical) activity during watching TV [10]. Beside this, interacting with the TV is (still) reduced to rather passively consuming content most of the time. But this traditional consuming behaviour can be used to attract attention of the viewer to audio-visual user-generated content on IPTV applications. Consuming audio-visual content corresponds to the traditional television viewing habits and the TV experience: leaning back and watching content [4].

Moreover, IPTV can attract different user groups (young and old — both present in local communities) by building on existing user experiences related to different media devices: TV and Internet. Television and Internet play different roles and have different meanings and can be seen complementary to each other.

Furthermore UGC offered over IPTV is evolving rapidly and content provider think about the right framework (either graphically or by offering editorial supporting) to make UGC closer related for watching on TV than browsing the Internet. We assume that user-generated content provided on TV will attract users and in particularly reach more user groups with special interests. Social cohesion in local communities can also profit from this relatedness. These aspects are further discussed as part of the case study presented within this paper.

3 The Case Study

3.1 Selected Community

The case study presented here is part of a larger European project, which is investigating new ways of how to exploit the huge amount of UGC in innovative ways to support people in their daily lives and in particular how technology will enable social change. A major part of the project consists in community hosting and how to involve people into this process following a user-centered approach. Three European testbeds are currently established for deploying newly developed applications and providing users a platform for (co-)creating UGC for their community. The one testbed we are focusing at in our research is a rural community in Austria. The community counts 8.000 inhabitants and consists of approximately 30 tiny villages. The lack of one single centre or main-square prevents the social cohesion of the inhabitants.

3.2 Methodological Approach

The project consists of two major phases: (1) A *design phase* including several co-design workshops with members of the local community, exploring interesting design concepts and services and (2) a *deployment and evaluation phase*: This phase includes the installation of the IPTV platform in 100 test households in the local community for further evaluation and collection of user feedback (starting with March 2008). In this paper, we focus on the results from the first phase. We present the results from the co-design sessions with the community, lessons learned and preliminary design implications are reported in the following chapters.

In general within this research process the participatory design approach was used. Co-design means actively integrating end users in the development process of a future product by putting them in the role of designer and stakeholder [20]. Users were involved right from the beginning by giving them a role as designers. Three co-design workshops were conducted to investigate general ideas and assumptions about social interaction and participation in local communities. In the first two workshops (led by Sonovista a project partner, having the first contact with the lead users in the community) we were passively participating. We used the workshops for observing the participation of the different community members, for taking notes of central statements due to the integration of a new technology into their community life and to get in touch with the different people in the community.

The third workshop was actively organized by us with the intention to make the next step in the co-design process. The participants worked out concrete application concepts within small groups of people. The methodological concept used for this co-design workshop was based on previous research in the area of storytelling and participatory design techniques. Storytelling is mainly used to communicate ideas and also to provide a central theme for the workshop.

A story helps participants to follow a common goal, evoke interest in solving a problem and foster activities among participants [19]. The personal knowledge and experiences of each participant served as a starting point for the co-design workshops. In the following section we will present and discuss the main findings from this first project phase in particular towards community-relevant aspects for establishing a community around IPTV.

4 Preliminary Findings from Co-design

4.1 Community Background and Interests

Sharing similar interests is one important element for communities, so that community members have a reason for interacting with each other. Therefore it's first of all important to find out which specific community member characteristics, interests and activities exist. Initial workshops with community members on site showed, that it's important to keep in mind the high diversity of characteristics among the population of the local community, in sense of demographics and personal interests. Moreover, people in the local community are connected via local clubs. There are more than 60 clubs in the village and more than 50% of the whole population, ranging from kids to the older inhabitants, are involved in local clubs.

During the workshops, the expectations of the community members related to the IPTV project were discussed. The community members defined three major goals as most important for their community:

– Get to know each other,
– Strengthen togetherness, and
– Involve the youth in community activities.

These goals are seen as most important requirements for the IPTV platform and can be further described as follows. As this community is a geographically scattered community its members desire to use IPTV for a feeling of togetherness and also to get to know each other. Additionally, they want to bridge a generation gap, which means bringing together young and elderly residents (see Figure 1). A main community requirement, which should be considered in the design and development of services throughout the co-design process, is fostering the bonding of the local community based on the IPTV platform.

Based on this user feedback the IPTV platform was built iteratively to offer several features and services for enhancing community relationships. Similar to common web applications, the IPTV platform enables the user to create for example an own user profile, to organize friend lists and audio-visual content.

4.2 Establishing Contact and Trust with and within the Community

The workshop participants were conscious of the difficulty to bring people together. They suggested using local formal and informal structures of communication for promoting the IPTV project – e.g. by using local media like the newspaper of the local government and other local media, as well as using local

Fig. 1. Young and Elderly in the Co-Design Workshop

events for promoting the new community TV platform – as they called it. They found it equally important to get support from well-known local people (e.g. the mayor) and local lead users, who are very active in the community (e.g. head of clubs) to promote the platform, in particular informing other community members by personal recommendations. The majority of the participants were still critical at the beginning, but also had a lot of sympathy for the offered IPTV platform and content.

Establishing trust takes time and has several aspects, which are interrelated, based on technology, provider, research people and other people and partners involved. Trust in the researchers and telecommunications operators is an important step into the community for setting up the testbed. In this sense, the co-design workshops during the first phase of the project were not only very useful to involve people right from the beginning and creating ideas/concepts together, but also to build up trustful relationships between the community members and the researchers.

4.3 Evoking Motivation for Participating

Another topic discussed in the workshops was related towards the motivation of community members to take part in the IPTV community, generating audio-visual content and sharing it within the community. Participants in the workshop – who were mainly lead users as defined above – were very concerned and worried about how to inform and motivate other people in the community. One workshop participant critically noted that it's not sufficient to announce it in the local newspaper and then hope that a lot of people will take part. One solution was offered for this problem: making Live Events. All workshop participants agreed on the importance of offline events for getting people interested in the project and informing them about it. In the co-design workshop, the participants suggested a combination of offline events and IPTV-related activities.

Another difficulty mentioned in the workshop is how to motivate the groups of participants with different age to build up relationships. To meet the goal of bridging the generation gap, participants said, “It has to be fun”. People should like to do it and it should not be too far away from their interests, but it still has to be a challenging activity that motivates people to participate.

Participants in the workshops also mentioned their fear of work overload due to the participation in another community. As many people in this rural community already take part in local clubs or associations and use their spare time for activities related to their communities, people don’t want another load of time-consuming community ‘work’. Therefore, the IPTV platform must support already existing community activities instead of requiring them to do additional activities. It also needs to create the feeling among community members of being in control and feel free to participate and it has to provide them a clear added-value for participating. In order to keep community members actively contributing to the community, it’s beneficial if activities – done for the community – have a community value as well as a personal value. By sharing pictures over the IPTV platform, users can also use this feature for personal means for managing their photo collections.

The co-design workshops themselves seemed to motivate people to stay interested in the project. The workshops are seen as an opportunity for community members to promote their personal interests and activities.

5 Conclusions and Outlook

Virtual as well as physical communities are complex; there is no one single route to success. Communities have to be carefully designed, implemented, and maintained in order to encourage and support social interaction continually [16].

- To simplify the community building process, it’s useful to build upon existing structures. Therefore, the potential users of the IPTV platform in the local community will be firstly members of local clubs, who could use IPTV for promoting their messages.
- Secondly, retirees, who are hobby and amateur film-makers and have enough time, knowledge and equipment to produce films for IPTV and thirdly, young amateur media-producers or media talents, who would use IPTV to express themselves or to present their audio-visual art work.
- Communities supported by the IPTV technology require similar mediation and conditions like for example online communities. But as an IPTV based platform offers a technology not far away from online technology a lot of requirements are already met. For example, by addressing the user’s interests and offering the right boundaries and hurdles, more users get involved

in the community [18]. As an example for understanding this pertinence, one pattern supporting online community shows how to make it as easy as possible for a user to join a community while still protecting the members of the community of strangers. It is called quick registration [18]. It has to be researched on one of the further evaluation steps, if this is also the case for IPTV based communities.
– In a first usability evaluation of the first version of the IPTV platform (results will be reported in another publication) people were also asked about their motivation to use the IPTV platform, producing and sharing audiovisual content through the TV. The most answers were community related. Namely, the possibility of getting to know other people, look at their photos and videos, and in particular to watch videos and pictures of friends. The platform usability was mentioned to be an important aspect, which still has to be improved, as it affects the overall user experience.

Starting from March 2008, the services developed for the IPTV platform will be deployed into the local community and a long term field study will be conducted focusing mainly on user experience aspects, such as sociability, fun/enjoyment, user engagement and involvement into the creation of audiovisual content as well as on user acceptance aspects.

In future developments, additional attention has to be paid also on the different ways of interaction (instead of remote control) and on this basis, the visualization and the navigation have to be made in a TV adequate way.

Acknowledgments. The described research activities are part of the project "CITIZEN MEDIA", which is funded by the European Union. We would like to thank all our partners within this project for their support and cooperation. Special thanks go to our Austrian partners Sonovista and Telekom Austria and the local Community in Engerwitzdorf.

References

1. Bruckman, A.: A New Perspective on "Community" and its Implications for Computer-Mediated Communication Systems. In: Ext. Abstracts, Proceedings of CHI 2006, Montreal, Quebec, Canada (works in progress, 2006)
2. Bruns, A.: The Future Is User-Led: The Path towards Widespread Produsage. In: PerthDAC: Digital Arts & Culture, Perth, Australia (2007)
3. Castells, M.: The Rise of the Network Society. The Information Age: Economy, Society and Culture, vol. I. Malden, Oxford, UK (1996)
4. Gawlinski, M.: Interactive television production. Focal Press, Oxford (2003)
5. Guest, A.M., Wierzbicki, S.K.: Social Ties at the Neighborhood Level: Two Decades of GSS Evidence. Urban Affairs Review 35, 92–111 (1999)
6. Harris, C.: Interactive Television. In: Grant, A.E., Meadows, J. (eds.) Communication Technology Update, 10th edn. Focal Press, Oxford (2006)
7. Häußermann, H., Siebel, W.: Soziologie des Wohnens: eine Einführung in Wandel und Ausdifferenzierung des Wohnens, Juventa (1996)

8. Hillery, G.A.: Definitions of community: areas of agreement. Rural Sociology 20, 111–123 (1955)
9. Jensen, J.F.: User-generated Content – a Mega-trend in the New Media Landscape. In: Lugmayr, A., Golebiowski, P. (eds.) Interactive TV: Shared Experience, TICSP Adjunct Proceedings of EuroITV2007, pp. 29–30 (2007)
10. Morley, D.: Home Territories. Media, Mobility and Identity. Routledge, London, New York (2000)
11. Nielsen, J.: Participation Inequality: Encouraging More Users to Contribute. Jakob Nielsen's Alertbox, October 9 (2006), http://www.useit.com/alertbox/participation_inequality.html
12. Norris, P.: The Bridging and Bonding Role of Online Communities. In: Howard, P.N., Jones, S. (eds.) Society Online: The Internet in Context, pp. 31–42. SAGE, Thousand Oaks (2004)
13. Obrist, M., Bernhaupt, R., Tscheligi, M.: Interactive Television for the Home: An ethnographic study on users' requirements and experiences. In: Proceedings of EuroiTV 2006, Athens, Greece, pp. 349–358 (2006)
14. Obrist, M., Weiss, A., Tscheligi, M.: Evaluating User-Generated Content Creation across Contexts and Cultures. In: IAMCR Conference (2007)
15. Parent-Thirion, A., Macias, E.F., Hurley, J., Vermeylen, G.: Fourth european working conditions survey, Europ. Foundation for the Improvement of Living and Working Conditions (2007)
16. Preece, J., Nonnecke, B., Andrews, D.: The top five reasons for lurking: improving community experiences for everyone Computers. Human Behavior 20, 201–223 (2004)
17. Scase, R.: Demographic and Social Trends Issue Paper: Mosaic Living. IPTS Futures Project, Series No. 7, IPTS, Sevilla (1999)
18. Schümmer, T., Lukosch, S.: Patterns for computer-mediated interaction. John Wiley & Sons, Chichester (2007)
19. Van Rompaey, V., Hemmeryckx-Deleersnijder, B., Van der Meerssche, B., De Mondt, H., Godon, M.: Beyond Marketing. Applying qualitative user experience research techniques on social media applications. The Journal of The Communications Network 4, 26–30 (2005)
20. Van Rompaey, V., Van Der Meerssche, B., Godon, M., Vanden Abeele, M., Charliers, K.: Connecting the family home: Co-designing new technologies for Community Communication. In: Conference proceedings of the ECCR/ECA conference, Amsterdam, November 25-26 (2005)
21. Völker, B., Flap, H., Lindenberg, S.: When Are Neighborhoods Communities? Community in Dutch Neighborhoods. European Sociological Review 23, 99–114 (2007)
22. Wellman, B., Leighton, B.: Networks, Neighborhoods, and Communities: Approaches to the Study of the Community Question. Urban Affairs Quarterly 14, 363–390 (1979)
23. Williams, D.R., McIntyre, N.: Where heart and home reside: changing constructions of place and identity. In: Luft, K., MacDonald, S. (eds.) Trends 2000: shaping the future: Outdoor recreation and tourism trends symposium, Michigan State University, September 2000, pp. 392–403 (2001)

PresenceRemote: Embracing Ambiguity in the Design of Social TV for Senior Citizens

Tomas Sokoler[1] and Marcus Sanchez Svensson[2]

[1] IT University of Copenhagen
sokoler@itu.dk
[2] Blekinge Institute of Technology
marcus.sanchezsvensson@bth.se

Abstract. This paper reports on our early experiences with the design of Social TV for senior citizens. Even though there seems to be a straight forward match between new interactive TV technologies on the rise and the notion of supporting social interaction amongst elderly TV viewers it is not at all clear how these technologies can accommodate the specific challenges related to the everyday life of elderly people. In particular, using an example concept – the PresenceRemote – we will discuss how the stigma associated with being lonely, an inherent part of senior living, can be addressed by leaving room for unarticulated intentions of use in our design of Social TV. The 'PresenceRemote' is an enhanced TV remote control making it possible for you to notice others and be noticed by peers within your local neighborhood as you watch TV.

Keywords: Senior citizens, social interaction, ambiguity, stigmatization, interaction design, Interactive television, Social TV.

1 Introduction

While much effort has been directed towards the design of assistive technologies that aim to help overcome the physical challenges of growing older there is an increased awareness that more attention needs to be directed towards the design of technology addressing the social and emotional aspects of aging (cf. [1]). These concerns for the social dimension of growing older have led to various attempts to introduce internet-based computer applications such as chat-lines and billboards for network relations among senior citizens ([2], [3]). On this background, the notion of Social TV, combining the familiar and much enjoyed everyday activity of TV watching with the possibility to provide openings for social interaction seems very promising. However, even though the basic TV technology needed to pursue the design of Social TV is (or is about to become) a standard component part of ordinary households it is not clear how this technology can be turned into a resource for social interaction that can accommodate the specific challenges related to the everyday life of elderly people.

As we start looking at Social TV as an assistive technology in the context of elderly living the success and acceptance of such a technology will not only depend on the functionality offered. As pointed out by Hirsch et al [1], the difference between 'can' and 'will' be used is directly linked to the notion of non-stigmatizing versus

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 158–162, 2008.

stigmatizing assistive technology - non-stigmatizing technology meaning technology that can be used without putting the elderly on display as disabled, needy, weak or in any other way pathetic individuals. In a previous study of how older people go about socializing in everyday life [4] we found that despite the fact that a majority of the people in the communities we looked at lives alone - and that loneliness clearly is a theme of great relevance - loneliness was tabooed and something people were reluctant to expose. Hence, we argue, people will not appreciate and consequently not make use of Social TV if the use of this technology singles them out as lonely individuals. Thus, as we move towards design of Social TV for senior citizens it is critical to understand the social stigma often associated and experienced with being lonely. In order to address this stigma we suggest that our design of Social TV should embrace the kind of ambiguity that people already experience and take advantage of in face to face social interactions throughout everyday activities - an ambiguity inherent to the interaction experienced at casual meetings between people throughout everyday life that helps people to interact without directly touching upon the taboo of loneliness. In the following sections of this paper we will use an example concept - The PresenceRemote (PR) - to discuss and demonstrate this approach to the design of Social TV for senior citizens.

2 Example Concept: PresenceRemote

The 'PresenceRemote' (PR) is an enhanced TV remote control making it possible for you to notice others and be noticed within your local neighborhood as you watch TV. Designing for a local neighborhood implies an environment with ample opportunities for social interaction, not least of course, as part of face-to-face joint activities and chance encounters. We think of the PR as a supplement to these already existing ways for the elderly within a community to interact. Furthermore, while other design for Social TV [5] aims to change the activity of TV watching from a 'lean-back' to a 'lean-forward' interactive and engaging online experience we seek to add a dimension of social interaction while keeping the original TV watching activity as intact as possible. We are looking to take advantage of TV watching as an already existing everyday activity and use this activity as a provider of openings for social interactions.

The PR is basically a TV remote control (see figure 1 next page) with the addition of a color display and three extra buttons. Below follows a brief description of the functionality:

- When the PR is OFF it works as your ordinary TV remote control with no extra functionality and no information about TV activities, yours or others, flowing to or from the PR. Switching the PR to its ON state is done by pushing the 'PresenceRemote on/off button' (see figure 1).
- The 'Take me to the most popular channel button' is a simple push button that takes you to the channel currently watched by the majority of people within your community with their PRs turned ON.
- When your PR is ON the PR display and the functionality offered depends on whether your TV is ON or OFF. If your TV is OFF the display shows a pulsating red color indicating how many people in your community has their TV turned ON. Hence, you may notice the community TV activity as you walk by the PR even

when your TV is OFF. This may in turn lead you to turn on your TV on and press the ‘take me to the most popular channel’ button.

- When your PR as well as your TV is ON the PR display will show the names of those of your buddies that currently have their TV and PR turned ON. We will refer to this mode as ‘Buddy Mode’**.** When in Buddy Mode the PR will display the names of your buddies watching TV. It will however not display which channel(s) the buddies are watching. We imagine that the Buddies available to the PR are a subset of the people that you already decided to include in your list of contacts on your cell phone.
- Finally, the transition from noticing that a Buddy of yours is watching TV to actually establishing contact is done by a handover to your cell phone. Handover meaning that you select your Buddy on the list of active buddies displayed on the PR and by pressing the ‘Handover to cell phone button’ the PR (using Bluetooth for example) sends the Buddy ID to your cell phone. The cell phone in turn now enables you to make a regular call or send a SMS whichever you find more appropriate.

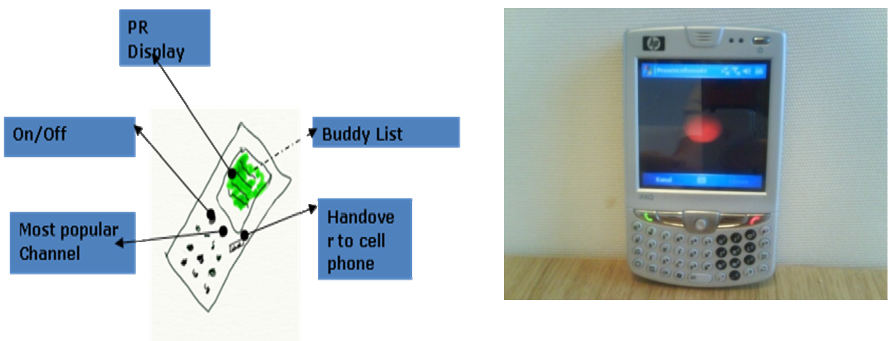


Fig. 1. Sketch of the PR and an early prototype using a standard PDA

In our design of the PresenceRemote we deliberately aim for a kind of technology that does not force people to be explicit about their intentions as to why they use a particular technology at a particular time. Hence, with the PR we allow people to leave it unarticulated whether they in fact are watching TV as an excuse to meet others. First, by design, having and using a PR should not be perceived by others as an invitation but rather as a way of saying: ‘‘I have a PR, and like other people in this community, I don’t mind that you know that my TV is on’. Second, by design, when in Buddy mode it is not possible to distinguish whether a) you are looking for company b) you are simply watching TV or c) both! Also, if you do not show up on your buddies PR display there is no way for them to know whether this is due to a) your TV is off or b) you have chosen to watch TV ‘in secret’.

We choose not to reveal the channel watched by buddies. By not displaying the choice of channel we serve some obvious privacy concerns but just as important by not making the channel explicit we bring forward openings for conversation. We speculate that this mode will stimulate communication along the lines of conversation starters such as “what are you watching?” and “is it any good?” or a simple “hi”. That is, conversation starters that seem to be about the TV activity but in fact may be much more about extending a greeting - about noticing others and making oneself noticed. This would indeed resemble a pattern observed during our field studies. Confirming the presence/existence of others and (re-)establishing the fact that everything seems to be as usual was often expressed as being a very important part of the daily routines amongst our interviewees. However, the existing ways of doing so was not expressed as activities with the explicit purpose of checking in on each other but expressed as part of some other activity like pseudo chance-encounters on the daily tour, the daily swapping of news papers in the afternoon, or other daily routines where it was kept unsaid but quite evident that the routines to a large extend really was about reconfirming that people in your network of friends where doing all right. In the same way we speculate that simply noticing a Buddy on the PR and letting her notice you may be sufficient for you and her to feel that people caring for you are within reach should something happen. Also, the 'offer' made to a person entering a buddy relationship is one of reciprocity in terms of the information that is accessible to you and the information you provide. The PR does not allow lurking on your buddies' TV activities without, at the same time, giving them a chance to notice you. This of course emphasizes that we are dealing with a relationship between peers. Furthermore, we believe the pulsating red color may serve two different purposes. First, we envision that this display of community wide TV activity may provide you with a sense of community belonging even without actually turning on your TV. Second, if combining the activity indicator with the 'Take me to the most popular channel button' we believe that the sense of community can be strengthened further. We speculate that watching the same TV show as the majority of people in your neighborhood can increase the chance to strike up a casual conversation about last night's TV show with people that you run into the next day.

Finally, an important feature of the PR is that it only detects and reports information that is directly related to the single activity of watching TV. Hence, there are clear delimitations to the PR's 'reach' and by simply turning off the PR all detection and exchange of information regarding your TV activities stops. By strictly tying the activity detection to a well-defined activity we aim to keep a strong sense of control and containment.

3 Concluding Remarks

Using an example concept – our PresenceRemote - we have discussed how to circumvent the stigma associated with being lonely in the design of Social TV for senior citizens. Based on our prior studies of elderly living and the design process leading to our PresenceRemote we bring forward the overall suggestion that we, in order to succeed in our design of Social TV for senior citizens, have to leave room for ambiguity and thereby allow people to leave their intentions of use unarticulated. This implies a rather eccentric challenge of designing a technology that even though it is designed to bring people together and support social engagement it cannot reveal that this is why people would use it. By taking on this challenge, we at the same time questions what seem to be the dominant model underlying the design for Social TV - a model implying that there is willingness or maybe even a desire amongst TV viewers to turn the serene 'lean-back' experience of watching TV in private into a 'lean-forward' activity emphasizing the synchronous interaction between the inhabitants of virtually interconnected living rooms. In our studies of the everyday life of senior citizens, it has become evident that we need to think differently about the design of Social TV for this group of people since they do not normally consider themselves as part of technology-mediated networked communities. In consequence, we have explored how Social TV can support a form of social interaction that builds on the subtle and discrete understanding of other people's activities and how Social TV, in this way, can act as a supplementary resource for the circumstantial interactions that plays out during casual encounters throughout everyday life. We realize, as we have not yet been able to fully implement the remote control in actual use that we can only speculate about the actual effects of our design rationale and choices. However, we had the opportunity to work with a group of seniors who participated in workshops to elicit their thoughts about our concept and early prototype. The outcome of these workshops was; firstly, a recognition among the seniors that the less explicit form of social interaction we design for plays an important role in their everyday life as it provides opportunities for informal and spontaneous interaction with other people when for example grocery shopping, taking an elevator ride, standing in a queue, doing laundry in the shared washing room etc; secondly, the group appreciated the idea of a technology that in a subtle way could help 'break the ice' and provide an opening rather than a commitment for social interaction and thereby accommodate rather than control the process of establishing contact during casual encounters.

Acknowledgements

We would like to thank the senior citizen living at Bataljonen in Karlskrona, Sweden for their participation in discussing and evaluating the PresenceRemote concept.

References

1. Hirsch, T., Forlizzi, J., Hyder, E., Goetz, J., Stroback, J., Kurtz, C.: The elder project: Social and emotional factors in the design of eldercare technologies. In: Proceedings of the 2000 Conference on Universal Usability, pp. 72–80. ACM Press, Virginia (2000)
2. Brunette, K., Eisenstadt, M., Pukinskis, E.: Social Well-Being through Place Attachment. In: Proceedings of CHI 2005, pp. 2065–2069. ACM Press, Oregon (2005)
3. Consolvo, S., Roessöer, P., Shelton, B., LaMarca, A., Schilit, B., Bly, S.: Technology for care networks of elders. In: Pervasive Computing, April-June 2004, pp. 22–29 (2004)
4. Sokoler, T., Sanchez Svensson, M.: Embracing ambiguity in the design of non-stigmatising technology for senior citizens. Behaviour & Information Technology 26(4), 297–307 (2007)
5. Vanparijs, F., Coppens, T., Godon, M., Bouwen, J.: Social Television: Enabling rich communication and community support with AmigoTV. In: Proceedings of ICIN 2004, Bordeux, France, Alcatel Technology (2004)

Investigating the Use of Voice and Text Chat in a Social Television System

Joe Tullio, Gunnar Harboe, and Noel Massey

Motorola Labs Schaumburg, IL, USA
{joe.tullio, gunnar.harboe, noel.massey}@motorola.com

Abstract. We tested a social television system with both voice and text communication in the field for one week, in four households. Participants expressed a preference for the freeform communication features over an earlier system with restricted communication options. Use of the communication features was nevertheless sporadic. Text was used more often than voice communication. However, both were usually used in combination. The in-home context had marked effects on the usage patterns, and we observed issues of privacy and identity verification.

Keywords: social television, computer-mediated communication, field study.

1 Introduction

Social television systems promise to enhance the experience of watching television with the simultaneous benefit of socialization with remote parties. These parties could include friends, family, or even strangers with similar tastes in television programming. The communication facilities in current research systems vary, incorporating among other things voice chat capabilities, text chat, simple emoticons, lightweight, predetermined messages, or some combination thereof. Each of these facilities seems to have inherent benefits and drawbacks that may be dependent on factors such as the context of use, user preference, user demographics and relationships, and perhaps even the content being viewed. To this point, there are few studies in the literature that give any guidance as to the use of different communication modalities in a real-world setting for social television systems.

In this paper, we present results from a preliminary study of an experimental social television system called STV3 that incorporates voice and text chat facilities as well as lightweight graphical messages (emoticons). This study builds on earlier work examining a previous version of the system (STV2) that used lightweight messages only [4], and serves as preparation for an upcoming, larger study. We report on how a group of close friends and family used our system in the field for a period of one week. Based on interviews and use logs, we suggest that text chat may work better than voice in the real-world context, but that the combination of both is better still. In addition, conversations seem to be only occasionally related to the TV content, and that identity verification remains a challenge.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 163–167, 2008.

2 Related Work

While there has been plenty of research on different electronic communication modalities, studies looking at social television specifically are scarce. Geerts compared text and voice communication in a within-subjects lab trial of two different systems [2], and found that voice was considered more natural, although young people preferred text. Baillie *et al.* compared voice communication to emoticon-based communication in a similar experimental setup, concluding that enjoyment and social presence was higher with voice [1]. In Harboe *et al.* [3], we reported, based on responses to a system concept in group interviews and also on field trials of an audio-only Social TV system, that participants expressed a preference for voice over text.

Previous work, then, points to voice communication as the most promising modality for social television. However, there is no evidence from the field to support or contradict this finding. There may be important factors of the real-world context that do not arise in the lab or occur to interviewees discussing hypothetical scenarios.

3 STV3

STV3 is the result of several major enhancements to its previous incarnation, STV2. These enhancements address issues voiced by participants in our 2-week field study of STV2. In that study, participants felt that the limited communication options (initially just three emoticons, later also a selection of 20 predetermined text messages) were too restrictive, and wanted the ability to express themselves more freely [4]. STV3 contains both instant messaging-like text chat and multi-party voice call functionality, thus allowing users to communicate in their own words using the modality of their choice.

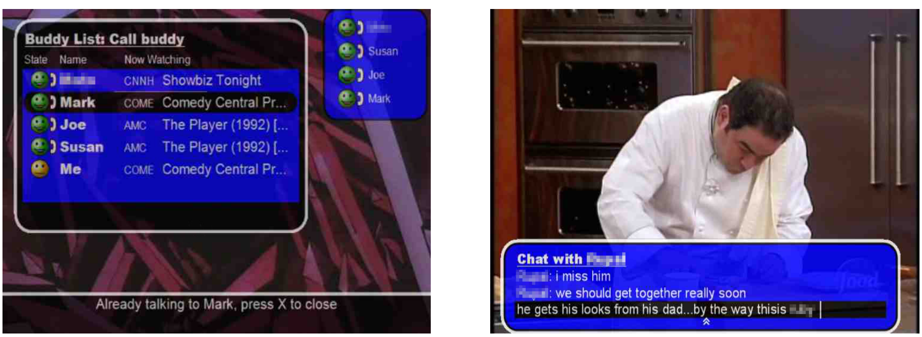


Fig. 1. Screenshots of buddy list and voice call presence (left), and a text chat window (right)

STV3 runs on a laptop PC, is displayed on a regular TV, and is controlled with a TV remote control. An on-screen buddy list allows users to contact friends and family, and also displays their presence, current TV show, and whether they are in a call (Figure 1). The prototype also includes a color-changing light to signal when a call is ongoing, and to indicate others' presence when the TV is off.

Voice calls are supported through the use of an echo-canceling room microphone, while text chats are supported using a wireless keyboard. When a call is attempted, the recipient can choose whether or not to answer. If the call is accepted, the two parties are connected and can begin talking to one another. Any participant in a call can invite others to join. Anyone can leave at any time, and a call ends when there is only one participant left.

For text chats, text messages appear immediately on the recipient's television in a small window near the bottom of the screen. At this point, either participant can type messages using the keyboard and send them. Each text conversation only supports two parties, users can have multiple simultaneous text chats with different buddies.

4 Method

For this preliminary study, we recruited four participants external to Motorola, all of whom lived in different households (A1, A2, A5 and A6). Three of the participants (A1, A2 and A5) took part in the earlier study of STV2. This decision was intended to minimize the novelty effects of using the system, since most participants were already familiar with its basic operation, and focus on the newly added communication features instead. The participants were between the ages of 25 and 35. A2, A5 and A6 were sisters, while A1 was a best friend of A2. All except A5 were married, and all watched TV frequently and regularly discussed certain TV shows.

The system was deployed in their households for a period of one week in mid-December of 2007. An initial interview was used to gather background data, a phone interview was conducted mid-week to investigate participants' experiences with the system during the first half of the week, and a final interview was conducted at the end of the week to examine their experiences in-depth. All interactions with the system, including text conversations, were logged, and voice conversations were recorded. Interview clips were transcribed and grouped along general themes to arrive at common insights regarding the experiences of our participants as a group.

5 Results

During the study period, participants had 4 voice calls and 12 text chats. Three attempts to make calls went unanswered, and 9 text messages (or series of messages) received no response. The first call (A1-A2 on the night of the setup) was a simple 1-minute system test which led into a joking text chat. This was followed by a 40-minute long call later that evening, with all participants joining in. 4 of the text chats took place within or around this conversation. The last two calls, A1-A2 (10 minutes) and A2-A4 (1 minute), took place on the fifth night, and although the calls were separate, all the participants were communicating with each other in 3 simultaneous text chats.

The three participants from the STV2 study all agreed that STV3 was a significant improvement from the previous version, describing it as "a good concept", "more interactive", and "a lot more fun." While there were some complaints that the system, had too many wires and too many peripheral devices, participants understood that the system was still only a research prototype and not yet a commercial product.

5.1 Feature Use and Choice of Modality

The amount of communication through the system was lower than we had hoped. After the first day, voice calls were used on only one occasion, and text chats on four separate occasions. This may be due to busy holiday schedules, lack of interesting programming that week, or it may reflect the natural frequency of use of the system.

While previous research has presented text and voice as contrasting alternatives, we found that participants combined the two. All of the calls in our study took place just before, during or after a text chat between the same participants. Indeed, most of the text chats were associated with voice calls. Participants generally stated that they preferred text chat to voice calls. Their reasons for this choice varied. One participant did it out of concern for the other households in the study, stating “there’s no sound or anything, so it’s not disruptive,” especially at night after children’s bed time. Another participant cited its familiarity: “I guess that’s a technology I’m used to, because I text-message a lot, and that’s all I do at work.”

This is not to say that they had no use for voice calls. One participant said “I would prefer the voice if it was a more lengthy or detailed conversation… it’s a little bit more time consuming to sit and type it.” Another cited her sister’s personal preference: “she would much rather talk and get the personal interaction.”

5.2 Communication and Topics of Conversation

All of our participants communicated daily prior to the study, usually by phone during their afternoon commutes and in some cases via email during the day. Evening, as reported by one participants, was “our time with our significant others that we live with”. The addition of STV3, however, encouraged additional communication outside of their daily routines. One participant said “in the evening it’s a bonus chat…just to say ‘hi, what are you doing?’” Husbands who had not previously been part of the conversation could join in, as one stated jokingly: “Usually I wouldn’t know what they talk about. Here I get more involved, probably more than I wanted to.”

Conversation topics mostly did not relate to the programming being viewed during those conversations. One participant described them as “almost like a side conversation.” Topics were typically about “things in general” and “more casual” than conversations conducted over the phone earlier in the day. However, there were two text chats and one voice call about the soap opera *General Hospital*.

5.3 Transparency and Privacy

Participants recounted a number of experiences concerning privacy, from lighthearted jokes to more disruptive misunderstandings. Twice, the husbands of participants attempted to pose as their wives in text chats, though they were quickly discovered because as one participant put it, “it was the way he said ‘hey dude’, and my sister doesn’t talk to me that way.” Another time, one participant accidentally mentioned a surprise Christmas present in a text chat, not realizing that the intended gift recipient was in the living room watching the TV, causing awkwardness for both participants.

Not knowing exactly who was on the other side of the conversation was a common issue for all participants. One participant likened the chat feature to her IM client at home: “at first I wasn’t conscientious [*sic*] of the fact that…I thought it was just a private conversation. In my mindset I'm comparing it to my Sametime chat I have at work which to me is private." Another participant was wary of using the voice feature, stating "you feel like it's kind of an invasion of your privacy, because they can be listening to you the whole time."

6 Conclusions and Future Work

While our results are only preliminary, they highlight issues that designers of social television systems should be aware of, and that warrant further attention from researchers. Most of these issues have received little if any consideration in previous work, perhaps because observations such as the reluctance to intrude with a voice call, and other household members impersonating the main participant are not likely to be made in a lab. In fact, due in part to such factors, our findings seem to contradict previous studies, by favoring text chat over voice calls.

With our method validated by this small study, we are currently conducting a larger trial where we hope to gather more substantial data. Our prototype has already benefited from a design iteration based on the input collected in this round. We encourage others to test our findings independently to see how well they generalize to other groups and contexts.

The authors thank Frank Bentley, Elaine Huang, Larry Marturano, Crysta Metcalf, Seonyoung Park and Guy Romano for their contributions to this work.

References

1. Baillie, L., Fröhlic, P., Schatz, R.: Exploring Social TV. In: ITI 2007 29th Int. Conf. on Information Technology Interfaces, Cavtat, Croatia, pp. 215–220 (2007)
2. Geerts, D.: Comparing Voice Chat and Text Chat in a Communication Tool for Interactive Television. In: NordiCHI 2006, pp. 461–464. ACM Press, Oslo, Norway (2006)
3. Harboe, G., Massey, N., Metcalf, C.J., Wheatley, D., Romano, G.: The Uses of Social Television. Computers in Entertainment 6(1) (2008)
4. Harboe, G., Metcalf, C., Bentley, F., Tullio, J., Massey, N., Romano, G.: Ambient Social TV: Drawing People into a Shared Experience. In: Proc. CHI 2008, pp. 1–10 (2008)

Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content

Pablo Cesar,[1] Dick C.A. Bulterman[1,2], and A.J. Jansen[1]

[1] CWI: Centrum voor Wiskunde en Informatica Kruislaan 413
1098 SJ Amsterdam, The Netherlands
p.s.cesar@cwi.nl, dick.bulterman@cwi.nl, jack.jansen@cwi.nl
[2] VU University Amsterdam
De Boelelaan 1081a, 1081 HV Amsterdamm, The Netherlands
dca.bulterman@cs.vu.nl

Abstract. This paper investigates a number of techniques and services around a unifying concept: the secondary screen. Far too often television consumption is considered a passive activity. While there are specific genres and programs that immerse the viewer into the media experience, there are other times in which whilst watching television, people talk, scan the program guide, record another program or recommend a program by phone. This paper identifies four major usages of the secondary screen in an interactive digital television environment: control, enrich, share, and transfer television content. By control we refer to the decoupling of the television stream, optional enhanced content, and television controls. Moreover, the user can use the secondary screen to enrich or author media content by, for example, including personalized media overlays such as an audio commentary that can be shared with his peer group. Finally, the secondary screen can be used to bring along the television content. This paper reviews previous work on the secondary screen, identifies the key usages, and based on a working system provides the experiences of developing relevant scenarios as well as an initial evaluation of them.

1 Introduction

The television watching experience can be enriched by using other devices than the traditional television set and remote control. This paper focuses on the usages of the secondary screen. Figure 1 shows the ubiquitous computing sphere of two viewers at different homes. The sphere of each user is composed of a number of devices that can be used for rendering multimedia content or for interacting with such content. Worth mentioning is that the multimedia content can be split onto different devices, each one rendering part(s) of the presentation, while other device can be used for controlling such presentation. As we will see in this article, the sphere of a user is not isolated, but is linked to his social network, thus including a path to other people's spheres.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 168–177, 2008.

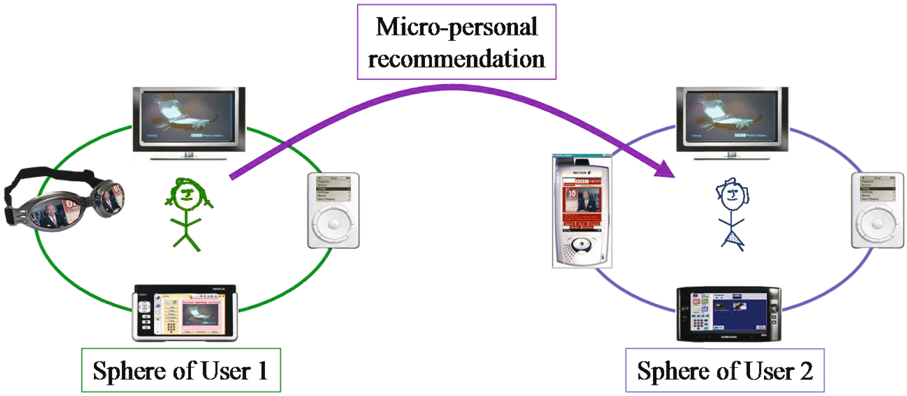


Fig. 1. Ubiquitous Computing Sphere at Two Homes

In this article, we consider the sphere of the user as a television set, connected to a personal video recorder or a set-top box, and his handheld device. The handheld device acts as a secondary screen providing both rendering and interaction capabilities. This article studies four basic scenarios: control, transfer, enrich, and share. We assume that the content reaches the home via various input paths. Once the content arrives we differentiate a number of end user behaviors:

– *Personal content selection/preview*: this is basic viewing functionality that allows a user to navigate through a set of content objects (and, where appropriate, within content objects) to find and activate particular content sequences of interest. The content may be gathered by explicit user scheduling activity (such as tagging a program in an electronic program guide), by indirect user scheduling (via a recommender system) or might be sent as a micro-personal recommendation message from a family member of a social network member.
– *Presentation continuity*: even though television watching normally takes place in the living room, it can happen that a user wants to continue watching the program in her personal device while moving to another room or outside the home. In this case, the current state of the content stream should be stored, the presentation should be streamed to her personal device, and the presentation should be restored there. This process should be as seamless and dynamic as possible.
– *Micro-personal recommendation generation and sharing*: the end user can generate direct recommendation messages. That is, he can explicitly fragment the video stream. This fragmentation is saved separately from the base content and may be used as the basis for a direct recommendation message, together with user-generated media overlays, that one user sends to another within his family or social network.

The rest of the article will elaborate on these representatives scenarios and will report on our experiences implementing them. Moreover, the results of a business analysis and a user evaluation are provided. This paper is structured as follows. Section 2 reviews the related work, indicating previous research on the secondary screen. Section 3 introduces the architecture and the design goals. Section 4 presents the results in terms of implemented services, while Section 5 introduces the initial results as business analysis and user evaluation.

2 Related Work

We cannot claim that the idea of using a secondary screen in the television environment is new. Back in 1996 Roberston et al. [1] presented a system where handheld devices are used for interacting with the television. Nevertheless, as interactive television systems and handheld devices are becoming more popular a body of research is emerging around the usage of the secondary screen.

Two specific areas of research that can take advantage of the secondary screen are interactive learning and content selection. As presented by Fallahkhair et al. [2], non-desktop technologies fit learning activities. The authors use the secondary screen for a number of scenarios such as to provide help for difficult cultural or language items, to provide extra information about specific concepts, and to manage the personal learning sphere. Other active research field for the secondary screen is the electronic program guide (EPG) as exemplified in [3,4,5]. Park et al. [3] provide rough guidelines about the foreseen service, while Cruickshank et al. [4] present a detailed study on system design and implementation, together with a comprehensive user study. They report on a working system in which a PDA is used for displaying a personalized EPG. Moreover, the PDA provides functionality such as volume controls and channel navigation. On the other hand, Karanastasi et al. [5], provide a solution for a ubiquitous personal video recording interface in handheld devices. Their system is capable of recording, deleting, and summarizing recorded television content. Finally, other usages of the secondary screen includes advertisement and commerce [6] and user participation in television quiz games [7].

In terms of social communication, one can highlight the capability of the handheld device for personal, or multi-participant, phone calls. Moreover, the viewer can use his mobile phone for sending messages that will later appear as an overlay of the television content, as a broadcast chat [6], or to allow the end-user to become a participant in the show [8]. For example, the Finnish Broadcast company, YLE, aired a television series, which narrative could be affected by the viewers using their handheld devices [9]. All in all, as Jensen [10] highlighted, the mobile phone is currently the most popular return channel technique and it is used for a number of activities while watching television (e.g., voting).

Apart from enhanced information rendering, content control, and communication, the secondary screen can be used to bring along the content. The viewer is a mobile being, thus when leaving the place in which the television screen is located, he might want the media content to follow him. Previous work in this subject is mainly focused on the enabling technology [11,12,13].

The justification of our work is based on all the mentioned research plus several additional studies. Firstly, Bernhaupt [14] found that remote controls are often considered as unusable, moreover the results indicated that the rendering capabilities of the remote control could be exploited. Cruickshank et al. [4] concluded that 'a more sophisticated form of input and control needs to be introduced for iTV to reach is full potential'. And Seager et al. [15] write that users 'frequently use their laptop to surf the web, use email, or shop online whilst watching television' and 'there was a preference for accessing different services on different display panels rather than overloading one shared display channel'.

3 Architecture

Our research studies new paradigms for multimedia interaction with content that is available to social groups of users in a consumer electronics (that is: non-PC) setting. Our home architecture consists of a home media server which may be implemented in a set-top box, a home networking gateway or a separate server device that stores content that is provided via standard broadcast channels, via peer-to-peer content sharing networks or on high-density optical disks such as DVDs or BluRay HD content. We expect that this content will be fetched on the users behalf using an intelligent recommender system, and may be post-processed in the home to allow differentiated viewing based on the individual interests of family members.

A schematic diagram of the home environment is shown in Figure 2. For a more detailed description of the technological aspects of the architecture, the reader can refer to [16]. One distinguishing characteristic of this architecture is that multiple remote control devices are presented to home users. These personal remote control devices form the basis of a system that supports differentiated content delivery and differentiated personal recommendation delivery and generation. The devices range from a conventional remote control, through low-powered handheld devices like telephones and minimal pen-based devices such as the Nokia N770, up to full-featured (but reduced size) tablets such as the Samsung Q1. In the home environment the devices are connected using a wireless home network, while when transferring a session to a mobile phone, the mobile network is used to retrieve the media content.

The central content storage and management component within our architecture is a home media server. This server can ultimately be implemented in many different forms (as a PC Media Center, as a conventional set-top box, as a network controller hidden in a utility closet). Our concern was not to study the commercial models for home media storage, but to study a model in which multiple control clients could be managed in a home environment. For this reason, we made the pragmatic decision to use a small-size personal computer (in our case, a Mac-Mini) upon which our server infrastructure could be implemented.

Our server architecture has been designed to be aware of DRM issues in the home. All operations on actual media content are abstracted away from the actual media encoding into a higher-layer structure. A portion of this structure uses

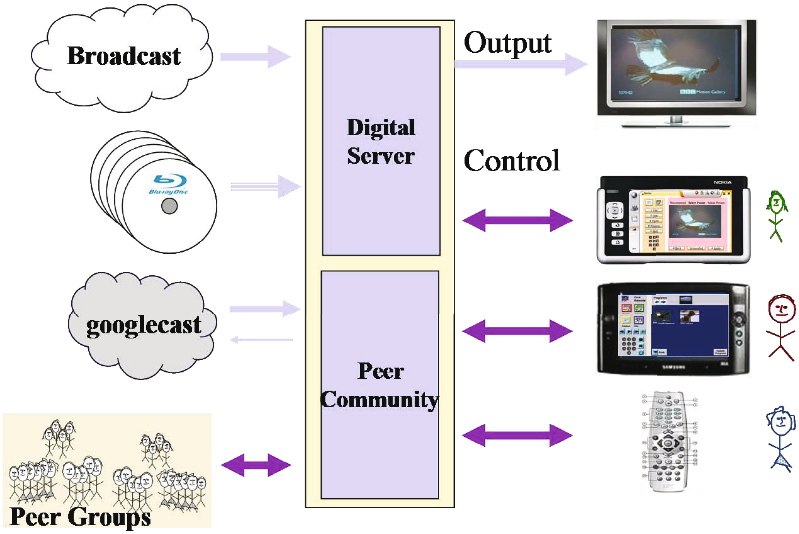


Fig. 2. Schematic Diagram of the Home Environment

the TV-Anytime specification for program and package descriptions. The local user operations are implemented by dynamically generating SMIL presentations that describe the transient structure of content modifications and annotations within a program.

4 Control, Enrich, Share, and Transfer

This section introduces the first set of results in the form of implemented services. These services include control and transfer services, in which the secondary screen is used for rendering enhanced information of the television content and to control the content itself. Moreover, the concept of a 'presentation following the user' is shown. Finally, this section reports on a service for sharing fragments of television content, and possible personal media overlays, with other friends and family members. The contribution of this paper is not so much the implementation of each of the services that can be found in previous articles [16,17], as a study on innovative services for the secondary screen and as an initial evaluation.

4.1 Control and Transfer

The control service is capable of differentiating the shared/personal nature of the media content. For example, the video content is a shared resource that can be shown in a shared screen such as the television set. While the enhanced material is private information that might not be of interest for the rest of the viewers, thus it can be displayed in the private display, that is the secondary screen. In

(a) Content Selection Interface (Nokia 770)

(b) Extra Material Interface (Nokia 770)

Fig. 3. Screenshots of Non-Intrusive Interfaces (Nokia 770)

previous work we refer to it as *Non-monolithic rendering* [16]. Figure 3 shows two examples of such interfaces. The first one (left) is used to navigate, select and preview media within a television program, while the second one (right) is used to view enhanced information about the current active program, about stored content, or about the EPG.

In addition to using the extended remote control for selecting and previewing personal content, for showing enhanced information, and to control the television content, it can be used for presentation continuity. For example, when the user is moving out of the living room, but still wants to bring along the presentation shown in the television set. In this case a dynamic evaluation of the context of the user, in terms of available devices, takes place and the presentation can be transferred to the secondary screen. Figure 4 shows an screenshot of transferring a presentation from a television screen to a mobile device.

(a) Screenshot of the Television Content

(b) Screenshot of the Content Transfer

Fig. 4. Presentation Continuity

4.2 Enrich and Share

The last few years have seen an increasing interest on sharing media material on the web, YouTube and MySpace being two clear examples. Nevertheless, in spite of their success, current systems contain a number of serious restrictions. First, the user is unable to share a bounded fragment of the video, which is what in most of the cases the user wants. Second, the user cannot customize the recommended video by including, for example, a voice commentary or strategically placed line art overlays. Our system allows users to share personalized fragments of television content. These enriched fragments can be sent as personal recommendations to friends within a users social network using a number of messaging technologies such as MMS, e-mail, and blog posting. Figure 5 shows the television media sharing interface. The actual fragmentation of the video and the inclusion of media overlays are done using the secondary screen. Detailed information about this asynchronous social television feature, named micro-personal recommendations, can be found in [17].

Fig. 5. Screenshot of TV Media Sharing Interface

5 Results

This section reports on the results of two independent tests of such services. It is important to notice, though, that the market analysis, technology development and user testing were performed by independent groups, across three countries. In addition to the results presented in this paper, other research in the topic suggests the benefits of the secondary screen in the interactive television environment [4,14,15].

5.1 Business Analysis

In order to analyze the business opportunities we were able to gather a panel of representatives from a European quadruple-offer player, a European equipment provider, an international advertising company, and a major international mobile video service provider. The goal of the focus group was not to jointly design a new interface, but to analyze the commercial prospects of a distributed home control paradigm from a non-technical perspective. The panel was presented with a stylized presentation of the capabilities presented in this paper. The panel was told that the intention of the project was to define a value-added service that could be offered as an enhancement to a home PVR offering.

The participants were globally enthusiastic about the services presented in this article. The majority of them think that this sort of application should be deployed as soon as possible because some of its main features would be covered by major market players within a 24 month period. Nevertheless, the application should be deployed progressively.

It was felt that the micro-personal recommendations should not be restricted to a family or neighborhood circle, but should focus on networked communities. Users should have a choice about being included in a local or global community of recommenders. All participants agreed that a distributed control application should be offered to the end user by a service operator as a part of a larger package. Its functionalities would not only benefit the end user, but also help to expand the base of the operator and to position the operator as a value-added supplier. Finally, one business case highlighted by the panel was the possibility to use the secondary screen for displaying targeted and personalized advertisements.

5.2 User Testing

We modeled a representative user community of up to three people watching television together. The viewing environment consisted of three handheld control devices, a small library of content, a high-definition television set, and the prototype server. Even though it was clear that the system was still in prototype status, the obtained results about its functionality are relevant. Each of the participants, sometimes three of them at the same time, were given a handheld control device, which was a personal device they could carry around as a mobile phone. The goal was to get feedback on the services, so we encourage them to explore the different capabilities, to play around, and to complete a number of predefined tasks (e.g., to share a fragment of a video with some friends).

Due to space restrictions, this paper only includes a summary of the obtained results. Nevertheless, the interested reader can refer to [18], in which detailed information on how the tests were conducted and analyzed, together with comprehensive set of results can be obtained. As a summary, we can highlight that the users were attracted by the possibility of having a personal display that allowed for browsing, personalizing, and enriching content. On the other hand, while sharing content with other people outside home was seen as a value-added service, the end-users did not find appealing to share the content within the home. We can conclude that in order of relevance, the secondary display for previewing and viewing content as well as for accessing enrich information was the most valued usage. Secondly was the possibility of sharing fragments of television content. Finally, most users liked the idea of enriching the micro-personal recommendations with personal overlays.

6 Conclusions

This paper focused on the usages of the secondary screen in an interactive digital television environment. Based on previous studies we have justified that handheld devices will be used in the living room, in conjunction with other consumer electronics appliances, for consuming and manipulating television content. This article proposes four main usages: to control, to enrich, to share, and to transfer television content. One of the differentiating factors of the results proposed in this paper from previous work in the field is that it is not restricted to any specific service (e.g., T-learning or the EPG), but it is a general architecture that can be applied to a variety of end-user situations. Nevertheless, an important topic that should be investigated in the future is which genres or programs types can better be benefited from the secondary screen.

The contributions of the work presented in this article go beyond the description of an architecture or implemented scenarios. We placed this work in a spectrum of activities that included an initial market assessment by professionals in the areas of media creation and distribution, and we subjected our prototype implementation to test by a dozen groups of users in a social setting. So far, the results are encouraging, even though more implementation work is needed and testing work remains to be done.

Acknowledgments

This work was supported by the ICT-FP7 project Ta2, the NWO project BRICKS, and the IST-FP6 project SPICE. The development of Ambulant is supported by NLnet. The business analysis were performed by D. Boullier, S. Kocergin, and A. Visonneau (LUTIN/France) and the user testing by H. Knoche and W. Seager (UCL/UK), to whom the authors would like to thank.

References

1. Robertson, S., Wharton, C., Ashworth, C., Franzke, M.: Dual device user interface design: PDAs and interactive television. In: CHI 1996: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 79–86 (1996)
2. Fallahkhair, S., Pemberton, L., Griffiths, R.: Dual device user interface design for ubiquitous language learning: Mobile phone and interactive television (iTV). In: WMTE 2005: Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education, pp. 85–92 (2005)
3. Park, J., Blythe, M., Monk, A., Grayson, D.: Sharable digital TV: relating ethnography to design through un-useless product suggestions. In: CHI 2006: CHI 2006 extended abstracts on Human factors in computing systems, pp. 1199–1204 (2006)
4. Cruickshank, L., Tsekleves, E., Whitham, R., Hill, A., Kondo, K.: Making interactive TV easier to use: Interface design for a second screen approach. The Design Journal 10(3) (2007)
5. Karanastasi, A., Kazasis, F.G., Christodoulakis, S.: A natural language model for managing TV-anytime information in mobile environments. Personal and Ubiquitous Computing 9, 262–272 (2005)

6. Davis, R., Yung, D.: Understanding the interactivity between television and mobile commerce. Communications of the ACM 48(7), 103–105 (2005)
7. Luyten, K., Thys, K., Huypens, S., Coninx, K.: Telebuddies on the move: social stitching to enhance the networked gaming experience. In: NetGames 2006: Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games, p. 18 (2006)
8. Miller, S.: Taking on the masses with mobile messaging TV. Computers in Entertainment 3(2), 6 (2005)
9. Ursu, M.F., Cook, J.J., Zsombori, V., Zimmer, R., Kegel, I., Williams, D., Thomas, M., Wyver, J., Mayer, H.: Conceiving shapeshifting TV: A computational language for truly-interactive TV. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 96–106. Springer, Heidelberg (2007)
10. Jensen, J.F.: Interactive television: New genres, new format, new content. In: Second Australasian Conference on Interactive Entertainment. ACM International Conference Proceeding Series, vol. 123, pp. 89–96 (2005)
11. Keukelaere, F.D., Sutter, R.D., de Walle, R.V.: MPEG-21 session mobility on mobile devices. In: Proceedings of the International Conference on Internet Computing, pp. 287–293 (2005)
12. Mate, S., Chandra, U., Curcio, I.D.D.: Movable-multimedia: session mobility in ubiquitous computing ecosystem. In: MUM 2006: Proceedings of the 5th international conference on Mobile and ubiquitous multimedia, p. 8 (2006)
13. Shacham, R., Schulzrinne, H., Thakolsri, S., Kellerer, W.: Ubiquitous device personalization and use: The next generation of IP multimedia communications. ACM Transactions on Multimedia Computing, Communications and Applications 3(2), 12 (2007)
14. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tscheligi, M.: Trends in the living room and beyond. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 146–155. Springer, Heidelberg (2007)
15. Seager, W., Knoche, H., Sasse, M.: TV-centricity - requirements gathering for triple play services. In: EuroITV 2007: Adjunct Proceedings of the European Conference on Interactive Television, pp. 274–278 (2007)
16. Cesar, P., Bulterman, D., Obrenovic, Z., Ducret, J., Cruz-Lara, S.: An architecture for non-intrusive user interfaces for interactive digital television. In: Proceedings of the European Conference on Interactive Television, pp. 11–20 (2007)
17. Cesar, P., Bulterman, D., Jansen, A.: Social sharing of television content: An architecture. In: Proceedings of the IEEE Symposium on Multimedia (Workshops), pp. 145–150 (2007)
18. Bulterman, D., Cesar, P., Jansen, A., Knoche, H., Seager, W.: Enabling pro-active user-centered recommender systems: An initial evaluation. In: Proceedings of the IEEE Symposium on Multimedia (Workshops), pp. 195–200 (2007)

An Experimental Platform Based on MCE for Interactive TV

Ping-Yi Liu[1], Hung-Wei Lee[2], Tsai-Yen Li[1], Shwu-Lih Huang[3], and Shu-Wei Hsu[1]

[1] Computer Science Dept., Natl. Chengchi Univ.
{g9527, li, g9204}@cs.nccu.edu.tw
[2] Dept. of Applied Psychology, Hsuan Chuang University
spoon@hcu.edu.tw
[3] Dept. of Psychology, Natl. Chengchi Univ.
slh@nccu.edu.tw

Abstract. In this paper, we propose an experimental platform for the design of interactive TV. On this platform, designers are allowed to emulate broadcasting programs, define control functions, collect user interaction data, and incorporate external application systems. We also use several examples to demonstrate that through this platform one can easily add new functions and set up psychological experiments for evaluating these new functions. Two experiments have been conducted and their preliminary results are reported.

Keywords: Interactive TV, Experimental Platform, Emulated TV, TV Watching Behaviors.

1 Introduction

Many researches about interactive TV focus on program recommendation [5], but we believe that a better design should be based on better understanding of the interactions between the viewer and the television. Therefore, we aim to conduct research on the design of interactive TV at the behavioral level. However, it is found that most commercial interactive TV services, such as TiVo[1], do not allow designers to manipulate the controls and contents of the programs to support interactive experiments. It is not easy to conduct in-depth experiments on TV-watching behaviors without these supports. The lack of a proper experimental platform has hindered the development of psychological experiments for better understanding of the effects of interactive TV. Therefore, in this paper we propose an experimental platform based on the software development kit (SDK) in Windows Media Center[2]. for interactive television that supports the following functions: **(1) Different watching modes.** We provide three TV-watching modes, including a customized video-on-demand (VOD) system, an emulated broadcasting TV service, and a subscribed cable TV service for the demands of different experiments. **(2) TV control.** Customized controls such as

[1] http://www.tivo.com
[2] http://www.microsoft.com/windowsxp/mediacenter/default.mspx

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 178–182, 2008.

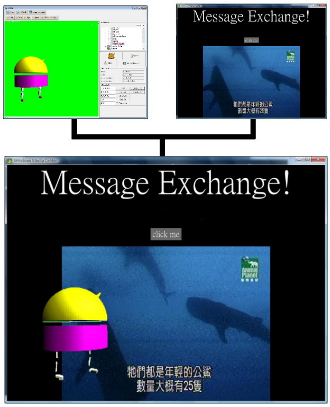


Fig. 1. Integration of external 3D animation system to overlay a virtual character on TV

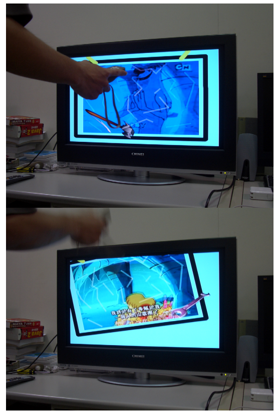


Fig. 2. The viewer uses Wii Remote to express negative emotion

TV interface design and media presentation methods are used to serve as instrumental functions. **(3) Experiment data collection.** The system could automatically record the viewer's operations on a remote controller and signals from other physiological indicators. We will describe the above three functions in more details in the next three sections.

2 Support for Different Watching Modes

In order to control experiments and observe behaviors of viewers in different scenarios, we have designed three watching modes. The first mode is a customized VOD subsystem. In this mode, the programs can be arranged into categories according to psychological states that can be aroused by these programs such that we can observe resulting behaviors and acquire instrumental measures. Second, we have designed a subsystem called SimTV that emulates broadcasting programs with multiple channels. From the user's point of view, the programs displayed are just like in a real broadcasting TV except for that the sources come from pre-recorded videos. A common problem on TV-viewing experiments with a real TV system is that the contents in live broadcasting channels are not the same for experiments done at different times. With the SimTV system, we can precisely control the program schedules for the same experiment at different times. The last watching mode is a subscribed cable TV service, which is a real cable TV environment that can be mixed with the SimTV environment to perform evaluation of a design in a real environment.

3 Customizing TV Control

An important feature of our system is that we can customize the interface, media contents and control functions for the aforementioned watching modes. Designers are allowed to change display functions such as on-screen-display (OSD) position and the overlay of a message dialog. In addition, the functions of keys on the remote controller can be easily redefined to accommodate the needs of experimental designs. Moreover, our platform can be incorporated with external program or devices to create innovative applications that traditional TV can hardly achieve. For example, an external 3D animation system has been integrated with our platform (as shown in Fig. 1) to design an alternative user interface that can provide a more effective way to communicate with the viewer according to his/her affective and/or attentive states. Although the detection of emotional states and the development of suitable treatments are still ongoing studies, the instrumental mechanism has been made ready for this goal on the platform. Another example of application that we have implemented is similar to the work proposed in [2]. We used Nintendo Wii Remote as an input device to create various real-time visual effects on images displayed on TV. For example, the viewer can fire a destroying action to crash the figure in the video by shaking the Remote to express their negative emotions such as anger, as shown in Fig. 2.

4 Experimental Data Collection

During the session of an experiment, the platform can automatically collect responses of the viewer from the remote controller and other devices that can be triggered programmatically. Currently we collect two types of data: (1) triggered events of the remote controller such as switching channels or adjusting the volume, and (2) video contents associated with these events. By exploiting webcams that are commonly available on regular PCs, not only images that are shown on TV screen but also the viewer's body movements and his/her facial expressions are captured in synchronization, as shown in Fig. 3. The recording is turned on and off automatically as the TV is turned on/off and the designer can use the system to examine video clips according to event triggers or markers. On this experimental platform, we plan to incorporate measurements of multiple physiological signals such as electroencephalogram (EEG), skin conductance response (SCR), heart rate, blood pressure and body temperature. We attempt to find physiological signal patterns relevant to specific TV-watching behaviors or the viewer's psychological states [3]. The results will then be exploited to build emotion and attention detectors for the design of further appropriate multimodal interactions with the viewer [1][4].

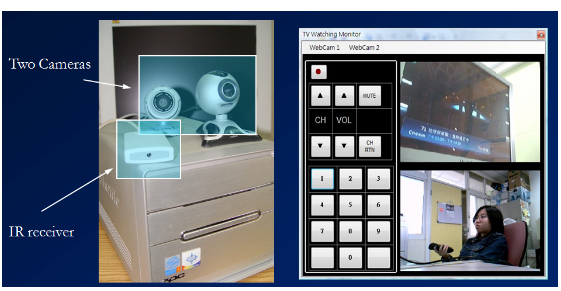


Fig. 3. Setup and Snapshot of event recording tools

5 Examples of Experiments

We have used the functions provided by the experimental platform described in the previous sections to design and evaluate new functions for interactive TV. In this section we will describe two psychological experiments for two functions that are found important by most users in a focus group study.

5.1 Volume Adjustment Experiment

The proper audio level for TV watching depends on programs, channels, and user preference. We aim to design new volume adjustment methods and evaluate their effectiveness through psychological experiments. Five modes for adjusting the volume of TV have been designed. (1) **Traditional mode:** The volume changes step by step constantly according to the number or duration that the adjustment buttons are pressed. (2) **Linear mode:** The adjustment scale is linearly proportional to the difference between current level and a predefined level such that it can adjust the volume more efficiently when the difference is large. (3) **Degression mode:** The adjustment scale for continuous presses decreases gradually for the user to fine tune the volume when the ideal level is approached. (4) **Frequency-dependent mode:** The amount of adjustment depends on the duration between two consecutive presses. The faster the user presses buttons, the more adjustment will be implemented. (5) **Semi-automatic mode:** The user can press a special button to jump directly to a predefined volume level and then fine tune it as usual.

We have taken advantage of the control and scheduling functions to facilitate the design of this experiment. First, we remap the functions of related buttons on the remote controller to corresponding methods in each of the above modes. The experiment consists of five sections for five different modes in addition to the initial instruction and warm-up practice. In each section, we ask the participant to adjust the volume to a most comfortable level for several programs with different initial settings (such as silent, extremely loud, etc.). The arrangement of TV programs in this experiment has been facilitated by using the VOD watching mode of our platform. Control events created by the 20 participants are recorded and the efficiency of each control mode is computed. The preliminary results reveal that the semi-automatic mode is most effective for adjusting volume. It is also found that the linear mode and degression mode are more efficient than the traditional mode.

5.2 Channel Selection Experiment

In this experiment, we aim to understand what kind of presentation can help the user recall the channels that he/she has visited and would like to switch back. We have designed an overlaid popup panel that can be evoked on demand by the user to provide quick links to the previously viewed n channels (n=5 in our experiments). A channel is considered as previously viewed only if the user has stayed at this channel for more than a given period of time. A snapshot of the interface is shown in Fig. 4.

We have conducted experiments to evaluate the effectiveness of three kinds of symbols for denoting the previously viewed channels: *channel numbers*, *channel logos*, and *typical snapshots*. With the help of the SimTV function on our platform, we are able to control the content of each channel while allowing thirty participants to

Fig. 4. Snapshot of the interface showing the Prior-N function on the SimTV platform

watch TV freely at different times. The preliminary results reveal that on average the participants use the Prior-N function to switch to previously viewed channels for 42 percent of time. The feedback collected through a questionnaire after the experiment also indicates that the Prior-N function is accepted as a helpful function. Among the three types of channel symbols, the typical snapshot was expected to provide more relevant link to the program. However, the experiment result shows that the participants take the longest time to switch to their target channels by using this symbol type. We think a possible explanation can be that a snapshot of a just-viewed program could be too complex and probably only exists in short-term memory while the other two types of symbols are simpler and have stronger association with the type of programs displayed on the channels. However, more experiments and analysis are needed to clarify these two possibilities.

6 Conclusions

We have built a flexible experimental platform for the design of interactive TV. The platform serves to provide innovative functions and instrumental measures to facilitate the design and experiments of interactive TV. In the future, in addition to building up more physiological measures, we will continue to conduct various experiments to understand the subjective demands of the viewers during TV watching via various observation channels and analysis methods.

References

1. Jaimes, A., Sebe, N.: Multimodal Human Computer Interaction: A Survey. In: IEEE Int. Workshop on. Human-computer Interaction (2005)
2. Lee, C.H., Chang, C., Chung, H., Dickie, C., Selker, T.: Emotionally Reactive Television. In: Proc. of Intelligent User Interface (2007)
3. Nasoz, F., Lisetti, C.L.: Emotion Recognition from Physiological Signals for User Modeling of Affect. In: 3rd Workshop on Affective and Attitude User Modeling (2003)
4. Pantic, M., Rothkrantz, L.J.M.: Toward an Affect-Sensitive Multimodal Human-Computer Interaction. Proc. of the IEEE 91(9), 1370–1390 (2003)
5. Yu, Z., Zhou, X., Hao, Y., Gu, J.: TV Program Recommendation for Multiple Viewers Based on user Profile Merging. User Modeling and User Adapted Interaction 16(1), 63–82 (2006)

Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience

Radu-Daniel Vatavu and Stefan-Gheorghe Pentiuc

University Stefan cel Mare of Suceava, Department of Computer Science,
str. Universitatii nr. 13, 720229 Suceava, Romania
vatavu@eed.usv.ro
http://www.eed.usv.ro/vatavu

Abstract. Watching television is usually a shared experience allowing family or friends that share the same viewing interests to watch, comment and enjoy programs together. The interaction part however is at the opposite end being reduced to the traditional remote control which by itself proves very limited with respect to the sharing part: although the viewing experience is shared among the group, the control part of the interface only allows one-viewer-at-a-time interaction. We are discussing in this paper a new interaction technique for controlling the TV set using one commonly available shared wide-area interface: the coffee table. By visually designating interaction sensitive areas on the coffee table surface, television control may be achieved via simple hand movements across the surface which may be performed by any of the viewers at any time. The final interface is thus fun, simple, intuitive, and very important, wide-shareable and immediately available for all the participants.

Keywords: interactive surfaces, tv, gestures, hci, computer vision.

1 Introduction

Families and friends usually gather in order to watch, comment and enjoy television shows, movies, news feeds or live sports transmissions for which they share the same interest. The most important fact to note here is that the global experience is a shared one: all the viewers share the same viewing transmission. The opposite happens however when it comes to the TV controlling interface: the traditional remote control only allows serialized one-viewer-at-a-time interactions which drastically limits the interaction with respect to the sharing part of the viewing experience. The interaction is not shared due to the fact that the interface device may be viewed as temporarily-blocked or temporarily-owned by other viewer at one particular time.

We address in this paper the very problem of allowing TV-control interactions via immediately available shared interfaces. We arrive at interactive coffee tables using simple logical implications: interaction should be shared and immediately available for all viewers, simple to achieve (should not add cognitive load or require training), fun if possible (the dimension of fun is introduced by novel

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 183–187, 2008.

Fig. 1. Left: our vision for interactive coffee tables: all the viewers may interact with the TV set using the same wide-area interface via simple hand movements on top of the video-sensitive area of the table. The interactive area is depicted in the center of the table by using a darker color. Right: Several possible designs for the interaction sensitive area on the coffee table accommodating for 2 or multiple viewers.

user-friendly technologies). Also, TV viewing rooms usually have a coffee table. Our vision is illustrated in Figure 1: viewers are watching TV while gathered around the coffee table; a video camera is installed so that it monitors only the surface of the table; also, there is some sort of delimitation between interactive and non-interactive areas on the coffee table surface; TV control is achieved by positioning and moving one or both hands inside the interactive area. Diverse interactions may be achieved using simple hand gestures such as hand-open or hand-closed combined with the relative hands positions with respect to the sensitive area of the table.

2 Related Work

Speech interfaces have been proposed for controlling the TV-set in the context of natural language processing [1] with the main challenges being related to environmental noise and speaker variations. Gesture recognition systems have been tackled before as well [2,3,4]. They make use of computer vision techniques in order to detect and recognize hand gestures performed in sometimes very complex environments. Also, several unusual interfaces have been reported such as tangible [5], affect-input devices [6] or even plush toys [7].

Controlling TV with the bare hands is very intriguing. For example, [2] use hand posture recognition combined with visual feedback of the hand position on the TV screen. Seated on a couch, users manipulate a graphical icon of a hand on the screen. Activating TV menus and performing selections are also reported by [3] in the same working scenario. The ARGUS system [4] was designed to control home appliances (TV included): users point towards a device and control its standard functions such as power on/off, volume up/down, play or stop via hand gestures.

One major disadvantage of all the above gesture-based interfaces is that having a video camera permanently facing and monitoring the viewers while they sit in front of the TV-set may issue problems and freights with regards to intimacy and privacy. Also, as noted in [3], the current setup with subjects seated facing the TV and performing gestures with their arms in mid-air is not suitable from the articulatory and musculatory point of view causing inconvenient interactions due to fatigue settling in. Problems also arise from the complexity of the scene to be analyzed by computer vision algorithms: crowded scenes with multiple objects and colors lead to difficulties in hand tracking with immediate effect on the entire system performance.

By moving the interaction to the coffee table space the mentioned problems are instantly removed. First of all, the interaction process is not fatiguant any more as the hands rest comfortably on the surface of the table and are not held in mid-air. The privacy issue is alleviated as the camera only monitors the coffee table and not the viewers. Also, we pay attention to choosing only two easy-to-perform hand postures (hand open and hand close) which in turn prove to allow for a variety of different interactions. And fourth, very important: the scene to analyze is far less complex as it is reduced to the surface of a table (of usually one single color) instead of processing images of an unknown living-room scenario.

3 Detecting Interactions

Hands may be detected above the surface of the coffee table by exploiting the contrast difference between the light skin color of the hands and the darker color of the table. We assume that a certain amount of contrast is present so that that hands may be segmented successfully. Simple gray-level image thresholding was used in order to segment the hands against the darker background of the table inside the margins of the sensitive area as Figure 2 illustrates.

For the actual interaction process we are interested in recognizing two postures: open-hand (all the fingers are stretched) and hand-closed (all the fingers are maximum flexed). As Figure 2 clearly shows, computing the bounding rectangles for the hands gives information about the two postures: the area, width and height of the hand-open bounding rectangle are clearly greater than the ones

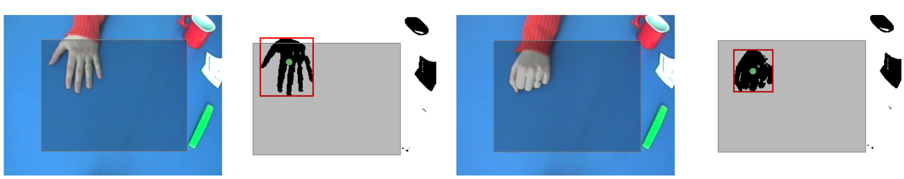


Fig. 2. Hands detection above the coffee table surface is achieved using simple color thresholding inside the margins of the interaction sensitive area

associated to the hand being closed. Hence again, simple thresholding against the geometric dimensions of the bounding boxes allows for easy and fast discrimination between the two postures.

4 Interfacing TV

The open-hand and closing the hand respectively act as a mouse-click event enabling viewers to acknowledge and issue commands. The principle is similar to the TAFFI interface [8]. Using these two hand postures several distinct interactions for interfacing TV-sets may be implemented.

Placing one open-hand in the sensitive area of the coffee table makes the channels selection menu appear as Figure 3 illustrates. Moving the hand across the surface determines the same movement of a hand icon on the TV screen with a permanent visual feedback. Closing the hand actually performs menu selection and the channel is changed. The same principle may act in selecting options from a general TV menu. Placing two hands inside the sensitive area of the coffee table allows controlling slide-like actions, for example modifying the audio volume or changing the brightness level. Similarly, the actual change is only triggered when the hands are closed and the effect is proportional to the relative distance between the two hands. Turning on/off the TV-set may be achieved by simply waving one open-hand inside the sensitive area.

Fig. 3. Selecting TV channels (left) using one hand and modifying the audio volume (right) in accordance with the relative distance between the hands

The system was implemented on a P4 2.66GHz desktop computer with digital mms live streaming TV. Video was processed at 25fps with a resolution of 320x240 pixels. The total CPU load was around 45% out of which the video processor took 35% while the rest was due to the Windows Media Player control. A demonstrative video of the system running is available at *http://www.eed.usv.ro/~vatavu/*.

5 Conclusions and Discussion

We described in this paper an interaction technique for interfacing the TV-set using simple, intuitive and comfortable gestures. Gestures are detected inside a sensitive area on the coffee table. The sensitive pattern we implemented is not unique and other versions may be imagined as Figure 1 illustrates. We have not yet performed actual usability tests as they will be the subject of further work. However, the first results are positive due to the system simplicity and people seem impressed of the technology and its ease-of-use which stimulates continuation of our work.

The main advantages that our system brings with respect to related work are given by: simple video processing (the scene is known in advance and very simple) which leads to better tracking and recognition accuracy; there are no privacy or intimacy issues related to video cameras monitoring viewers as the camera only faces the coffee table; using two easy-to-perform hand postures (open hand and hand closed) a diversity of interactions may be implemented; and, very important, the coffee table allows for a wide-area interface which is evidently shareable and immediately available for all the viewers at any time.

As future work it may be interesting to investigate new options for the coffee table scenario: adding and recognizing objects (such as coffee-cups) inside the sensitive area may lead to new interactions. The coffee table may also act as a personal DVR with different areas of the table 'hidding' the recorded TV shows.

Acknowledgments. This work was supported by the national funding grant Research of Excellence, Ref. No. 131/2006 CEEX, INTEROB.

References

1. Fujita, K., Kuwano, H., Tsuzuki, T., Ono, Y., Ishihara, T.: A new digital TV interface employing speech recognition. In: IEEE Int. Conf. on Consumer Electronics, pp. 356–357 (2003)
2. Freeman, W.T., Weissman, C.D.: Television Control by Hand Gestures. In: IEEE International Workkshop on Automatic Face and Gesture Recognition, Zurich (1995)
3. Lenman, S., Bretzner, L., Thuresson, B.: Using marking menus to develop command sets for computer vision based hand gesture interfaces. In: The 2nd Nordic conference on Human-computer interaction, pp. 239–242. ACM Press, New York (2002)
4. Kohler, M.: Special topics of gesture recognition applied in intelligent home environments. In: Wachsmuth, I., Fröhlich, M. (eds.) GW 1997. LNCS (LNAI), vol. 1371, pp. 285–296. Springer, Heidelberg (1998)
5. Tahir, M., Baily, G., Lecolinet, E.: ARemote: A tangible interface for selecting TV channels. In: ICAT 2007: the 17th Annual Conference on Artificial Reality and Telexistence, pp. 298–299. ACM Press, New York (2007)
6. Lee, C.-H.J., Chang, C., Chung, H., Dickie, C., Selker, T.: Emotionally Reactive Television. In: ACM IUI 2007, pp. 329–332 (2007)
7. Kawasaki, Y., Igarashi, T., Ajioka, T., Honda, I.: Vision-based Gestural Interaction Using Plush Toys. In: ACM UIST 2005 (2005)
8. Wilson, A.: Robust Vision-Based Detection of Pinching for One and Two-Handed Gesture Input. In: ACM UIST 2006 (2006)

YouTube3D: Accessing Web Video Streams through a 3D Interface

Fabio Pittarello and Alberto Narda

Università Ca' Foscari di Venezia
Dipartimento di Informatica
Via Torino 155, 30172 Mestre (VE), Italy
(pitt,anarda)@dsi.unive.it

Abstract. YouTube3D is a 3D prototype for accessing video streams on the web, conceived for stimulating exploration and information refinding, promoting collaboration and enhancing the sense of community. It is based on a 3D metaphorical city built by the collaborative activity of different categories of users, from high-level designers that are responsible for the public areas of the world to ordinary users that populate their private areas according to their specific interests.

1 Introduction

The phenomenon of huge video repositories on the web, built with the primary contribution of a wide number of users (most notably YouTube), has become extremely relevant in the last few years. Scenarios seen so far were different for the number of video available to users (hundred vs. millions of videos), the author (professional film-makers vs. common users) and the content organization (simple lists of streams vs. interfaces for managing a huge number of contributions). Such video repositories are usually available through 2D web interfaces allowing the users to browse videos according to category, popularity and to search the repository using metadata. Users accessing such repositories can add videos, draw comments in relation to what they see or can build networks of friends sharing the same interests. The project YouTube3D takes into account the requirements of this new form of communication, trying to enhance some aspects, such as the perception of a social environment where the users may share their interests and communicate their personality. Besides, YouTube3D tries to enhance some specific information search paradigms, such as exploratory seeking and information re-finding. YouTube3D is based on a 3D multi-level urban space characterized by different zones. Such zones include abstract visualization schemes (e.g., rows of cubes representing the most popular videos) or metaphorical areas, such as thematic pavilions where selections of videos are organized by distinguished designers and private mansions designed by common users. Finally, YouTube3D reflects some aspects of the organization of a real city, where there is a hierarchy of spaces, from public to private locations, populated by different classes of citizens that may manage content or simply explore the city, stimulated by the variety of content produced by other users.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 188–192, 2008.

2 Related Works

Different research areas have been considered for designing the YouTube3D project. Because of the limits of the space, we'll cite only some relevant works for each area. Concerning TV interfaces, most research is focused on providing the users with systems that may help them to select a subset of channels from the huge amount of possibilities offered by the broadcasters. Such recommender systems [1,2] are usually conceived as an enhanced EPG (electronic program guide) where the program information is filtered according to static user profiles or dynamic data coming from users. Such systems offer a valuable help for users that are focused on specific interests, but they don't seem to be appreciated so much by the so-called *TV zappers*, TV users that prefer an explorative search paradigm. While such users are not the target group of the application proposed in this paper, YouTube3D is focused on personal computer users that share the same interest for associative search mechanisms.

Information visualization is another relevant research field for the YouTube3D project. While notable proposals [5] take advantage of abstract visualization schemes, other researchers work on metaphorical schemes that associate information and concepts to objects that belong to the everyday experience (e.g., a room, a chair, etc.). The deployment of such objects in a (virtual) environment stimulates the explorative approach, helps the casual finding of information (i.e., serendipity) and allows the users to build a mental map of the information organization, enabling also faster information re-finding. Such method, known as *method of loci*, has been extensively used in the Western civilization, since the Classical period, for organizing and remembering concepts and things [8].

The *time-pillars world* by Pittarello [4] is built using an hybrid approach: the satellite TV panorama is represented through a number of visual artifacts, called time-pillars, associated to the different channels: each time-pillar uses an abstract visualization scheme (i.e., the spiral) for displaying a selection of recorded video frames related to the specific TV channel. The time-pillars are placed in a metaphorical city, organized in thematic or linguistic zones. YouTube3D shares with the time-pillars world the hybrid approach, but it uses a simpler scheme for representing the video streams (i.e., a 3D cube textured with a single frame taken from the video), because of the higher number of objects to render in real time. Besides, while the time-pillars world is focused on conventional satellite broadcasting, YouTube3D brings the 3D approach to the domain of user-generated content, offering to its citizens higher degrees of freedom for the cooperative design of the environment.

Navigation support is another important issue that has been exploited in YouTube3D, considering both the use of landmarks [7] and the introduction of different navigation modalities, such as the free or the guided navigation, that the users may choose according to their searching needs (e.g., known-item search, information re-finding or exploratory search [6]).

The Web 2.0 identifies a participatory architecture where the role of the site designer is limited to the definition of the general infrastructure and the users are the primary content contributors. YouTube3D is compliant with such

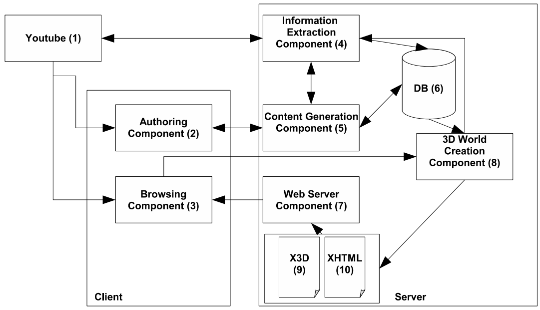


Fig. 1. Architecture of the system

collaborative approach, assigning different privileges for content creation: the urban metaphor is designed by the city major, distinguished designers and common users that manage, respectively, the general city plan, the public areas and the private zones. Such solution combines the advantage of having a general scheme, clearly recognizable by the users, with the freedom of personalizing specific zones according to the different user privileges. The collaborative creation of a 3D environment is a feature that can be found also in other on-line 3D environments, such as Second Life. What distinguishes YouTube3D from Second Life is that, in our approach, the video streams are the primary content objects and both the 3D metaphorical and abstract representation schemes are focused on offering an organic approach to such content.

3 Architecture and Interface

YouTube3D is based on a client-server architecture integrating different technologies and languages: the Apache web server, the MySQL DBMS and the PHP language, on the server side; Ajax3D, that integrates XHTML, Javascript and X3D, on the client side. The application queries the YouTube database for accessing the video streams and related data.

Figure 1 shows the different software components. During the authoring phase, users interact with the authoring component (2) on the client-side for placing on a 2D map the objects that populate the 3D world. Information related to the objects is managed by the content generation component (5) that retrieves from the DBMS (6) information related to the different classes of objects and stores the result of the authoring activity. Information about the video streams (the video id, the preview image, the title and the description) may be already stored in the DBMS. But, if the user wants to place on the 2D map a video that is not available in the DBMS, the information extraction component (4) will query YouTube for retrieving all the needed data.

A part of the 3D world is not directly authored by the users, but is reserved to the visual counterparts of abstract placeholders (e.g., rows made of the most popular videos) that are periodically filled by the system with fresh information.

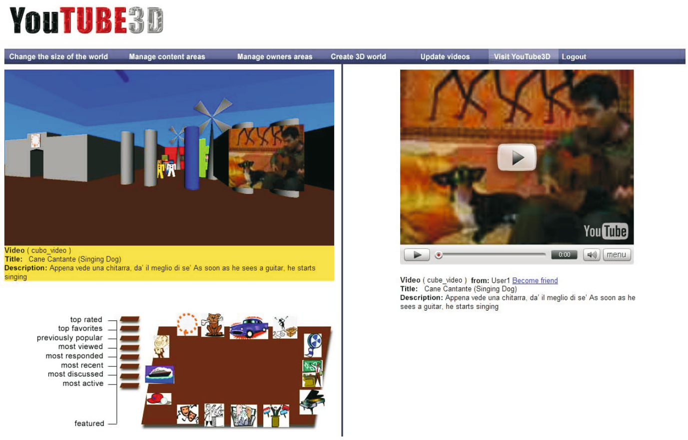


Fig. 2. Browsing interface

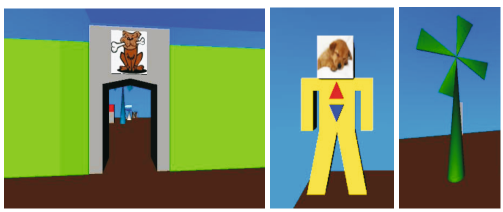


Fig. 3. The arch (a), the avatar (b) and the mill (c)

At start-up the 3D world creation component (8), triggered by the client, generates the 3D world, complying with the X3D standard, and the associated XHTML files, containing information related to the videos and to the other objects of the world. The user will browse the generated environment with an Ajax3D [3] compliant browser, assembling on a unified interface the information coming from the web server and the videos streamed from the YouTube site.

Figure 2 displays the YouTube3D interface. The menu bar on the upper side of the screen gives access to the authoring, world generation and browsing functions. The snapshot displays the browsing interface, that can be accessed also by unregistered users. The window on the left displays the 3D world that can be navigated by the user using a standard mouse for free navigation or clicking the map displayed on the lower side of the interface for guided navigation. The user can walk through the pavilions of the public thematic areas and the private zones situated at the lower floor of the 3D world, but s/he may also move through the upper levels that display alternative types of organization (e.g., the second floor displays, for each theme, the rows of the most discussed videos).

The YouTube3D world contains several classes of objects that have been chosen because they represent landmarks with a well recognized semantics in the real world. Besides, there are other classes of objects that have a metaphorical role and permit to visualize at a glance both static and dynamic properties that may be interesting for the users. The *archs* (Figure 3a) are among the most relevant landmarks. They are placed on the lower floor, at the beginning of each thematic zone, to introduce the public pavilions. Both public and private zones are populated by *avatars* (Figure 3b) that are visual counterparts of the YouTube3D users and of their networks of friends. The golden avatars of a specific zone represent their owners and can be queried for retrieving the user identity just passing the mouse over their head. Copies of these avatars, identified by the silver color, may be placed by friends in their private zones, to visually represent a network of distinguished friends. It is possible to click on the arrows placed on the body of the avatars for lifting all the copies associated to a specific user over the level of the other objects that populate the world; such functionality is useful for having an idea of the extension and of the spatial distribution of a specific network of friends. The *mill* (Figure 3c) is a metaphorical object that can be seen at a distance and that can be associated both to public and private areas. The mill may rotate at three different speed levels, depending on the date of the last update of the area content, giving therefore to the users an immediate feeling of the most *fresh* zones of the YouTube3D world.

The next steps of the project will include an extensive testing with a group of users that will exercise the system, comparing it with the classic 2D interface offered by YouTube, in order to receive feedback and direct future research.

References

1. Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Difino, A., Nigro, B.: User modeling and recommendation techniques for personalized electronic program guides. In: Kobsa, A., Ardissono, L., Maybury, M. (eds.) Personalized Digital Television. Targeting Programs to Individual Viewers, pp. 3–26. Kluwer, Dordrecht (2004)
2. Cotter, P., Smyth, B.: Ptv: Intelligent personalised tv guides. In: Proc. of the 12th Innovative Applications of Artificial Intelligence (IAAI-2000) Conference, Austin, TX, USA, pp. 957–964 (2000)
3. Parisi, T.: Ajax3d:the open platform for rich 3d web applications (2006), `http://www.ajax3d.org/whitepaper/`
4. Pittarello, F.: The time-pillars world. a 3d paradigm for the new enlarged tv information domain. In: Kobsa, A., Ardissono, L., Maybury, M. (eds.) Personalized Digital Television. Targeting Programs to Individual Viewers., pp. 287–320. Kluwer, Dordrecht (2004)
5. Robertson, G.G., MacKinlay, J.D., Card, S.K.: Cone trees: Animated 3d visualizations of hierarchical information. In: Proc. CHI 1991, New Orleans, Louisiana, USA, pp. 189–194 (1991)
6. Rosenfeld, L., Morville, P.: Information Architecture for the World Wide Web. O'Reilly, Sebastopol (2006)
7. Vinson, N.G.: Design guidelines for landmarks to support navigation in virtual environment. In: Proceedings of Conference on Human Factors in Computing Systems, pp. 278–285 (1999)
8. Yates, F.A.: The Art of Memory. University of Chicago Press (1966)

ZapTV: Personalized User-Generated Content for Handheld Devices in DVB-H Mobile Networks ⋆

Yolanda Blanco-Fernández, José J. Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, and Martín López-Nores

Department of Telematics Engineering, University of Vigo, 36310, Spain
{yolanda,jose,agil,mramos,mlnores}@det.uvigo.es

Abstract. During the last years, we had witnessed the boom of the digital market due to proliferation of emergent audiovisual services and increasing number of broadband networks. In this scenario, users insistently demand for innovative services for exchanging and sharing their own audiovisual contents and productions. In order to meet these needs, in this paper we propose the ZapTV system. Broadly speaking, this tool broadcasts user-generated audiovisual contents for handheld devices in a mobile network based on the DVB-H broadcasting standard. ZapTV offers diverse added-value services to these new active users, such as: (i) multi modal access (via Web and by handheld devices) to digital contents anywhere and anytime, (ii) availability of a return channel to transmit interactive contents that enhance the user's viewing experience, and (iii) annotation, sharing and personalized distribution of audiovisual contents. To achieve these goals, our system adopts both well-known technologies for broadcasting and semantic annotation of audiovisual contents (such as DVB-H and TV-Anytime), and emergent technology from the so-called Web 2.0, which permit the users to actively cooperate in tasks of generation, annotation and classification of digital contents.

1 Introduction

In the last years, we had witnessed the boom of the digital market, an effect derived from proliferation of emergent audiovisual services and increasing number of broadband networks. In this scenario, users demand for a more and more active and collaborative role, which is encouraged from the powerful new ways of communicating arose in the Web.

This new scenario has allowed to create an enhanced networked effect among Internet users that is changing the way people communicate, where and how they congregate, and the way they seek and share information [2]. This focus on a social Web does lay the foundations of the so-called Web 2.0, a new initiative that boosts active participation and collaboration among users [8]. In fact, Internet users are increasingly visiting social networking sites —sites that promote networks of relationships around

⋆ Work funded by the Ministerio de Educación y Ciencia (Gobierno de España) research project TSI2007-61599, by the Consellería de Educación e Ordenación Universitaria (Xunta de Galicia) incentives file 2007/000016-0, and by the Programa de Promoción Xeral da Investigación de la Consellería de Innovación, Industria e Comercio (Xunta de Galicia) PGIDIT05PXIC32204PN.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 193–203, 2008.

users and information [6]— for entertainment and news, business relationships, consumer product reviews, connecting with friends, and more. But they are doing more than just visiting; instead, as members contribute site content in the form of journal entries, photos, videos, and weblogs, becoming producers and authors. Definitively, the value of these sites is derived not from content provided by the site's owners, but from the emergent relationships among users and the content they create and consume. This process of user-driven content generation is firmly supported by the development of consumer electronics, which has given rise to new generation handheld devices that enable to easily create contents, and to access them from any location.

Bearing in mind the current demand for this kind of anytime and anywhere services, along with the boom of digital contents and the technological advances related to creation and sharing of contents, we have developed the project described in this paper. The main goal of our system, named *ZapTV*, is to broadcast *personalized channels of mobile TV* in a network compliant with DVB-H, a technical specification for accessing DVB services on handheld devices [4]. As main novelty, note that these channels include both *interactive services* and *user-generated audiovisual contents*, which are shared and annotated to favor advanced searching tasks in the system.

According to these functionalities, ZapTV must adopt well-known technologies from a wide variety of fields, such as: (i) broadcasting for handheld terminals, (ii) semantic annotation of audiovisual contents, (iii) use of a return channel to deliver interactive capabilities, (iv) acquisition and publication of user-generated content from diverse sources and with different ways of access (via Web and by DVB-H handheld devices), and (v) emergent technologies bound to Web 2.0, which provides mechanisms for collaborative annotation and classification of user-generated content.

This paper is organized as follows: Sect. 2 is focused on the technologies gathered together in our ZapTV project. Next, Sect. 3 detail the main functionalities and the architecture of our broadcasting system for mobile environments. Finally, Sect. 4 concludes the paper.

2 Technological Landscape

Multiple emergent technologies from diverse areas of information and communications technology are especially useful in our project. Firstly, in order to meet our broadcast requirements, we put the focus on the technical specifications of the Digital Video Broadcasting Project[1] (DVB), which is an industry-led consortium of over 250 broadcasters, manufacturers, network operators, software developers, and regulatory bodies in over 35 countries who are committed to designing global standards for the delivery of digital television and data services. The DVB standards cover all aspects of digital television from transmission through interfacing, conditional access and interactivity for digital video, audio and data. The consortium came together in 1993 to create unity in the march towards global standardisation, interoperability and future proofing.

Regarding the transmission of the audiovisual contents for domestic receivers, the DVB consortium has proposed successful initiatives for terrestrial, cable and satellite forms of broadcast (DVB-T, DVB-C, and DVB-S, respectively). As for mobile

[1] http://www.dvb.org

devices, the consortium developed the initiative DVB-H (Handheld), a standard that allows the handheld terminals to receive data from the network while a return channel (e.g. modem, ADSL, 3G) is required for data and interactive applications. For that purpose, DVB-H resorts to IP DataCast (IPDC), a system for delivery of any types of digital content and services using IP-based mechanisms, which comprises of a unidirectional DVB broadcast path and a bidirectional mobile path for interactive purposes.

Although DVB-H builds on existing DVB-T infrastructure to reach wide coverage with limited incremental investments, it introduces some technical changes to meet the specific requirements of the handheld devices. The most significant change in this regard is the so-called *time-slicing* technology, which is employed to reduce power consumption for small handheld terminals. This technology is based on transmitting the information as data bursts in small time slots, so that the front-end of the receiver switches on only for the time interval when the data burst of a selected service is on air. Within this short period of time a high data rate is received and stored in a buffer, which can either maintain the downloaded applications or playout live streams. Besides, DVB-H includes other technical improvements in order to increase the Signal to Noise Ratio (SNR) and mitigate Doppler effect in the mobile channels.

While DVB-H brings broadcast services to mobile handsets, we have chosen TV-Anytime to describe the attributes (e.g. topics, intended audience, intention, involved credits) of the audiovisual contents available in ZapTV, thus enabling enhanced searching tasks in the system. These specifications have been proposed by the so-called TV Anytime Forum[2], which standardizes multimedia services based on digital storage in consumer platforms, combining the immediacy of television with the flexibility of the Internet. In TV-Anytime Phase 1, XML applications are defined to describe: (i) generic audiovisual contents, (ii) specific instances of these contents, (iii) users' profiles, (iv) information concerning segmentation, rights management and protection of contents, and (v) mechanisms to reference contents regardless of their location and broadcast time. The TV-Anytime semantic descriptions are provided by metadata specifications aimed at endowing the digital receivers with capability of locating and storing TV programs that match user-specified characteristics. This way, for instance, TV-Anytime allows the user to find, navigate and manage contents from a wide variety of sources, including traditional broadcasting, interactive TV, Internet, and local storage in a PDR (Personal Digital Recorder).

After Phase 1 in 2003, TV-Anytime Phase 2 was released in 2006. Its key areas include sharing and distribution of diverse types of multimedia content (e,g. games, enhanced TV, graphics, music files) through home networks, and automatic matching and delivery of relevant content to profiled consumers. To this aim, TV-Anytime Phase 2 must enable the secure management of content within and among compliant devices, as well as ensure the privacy and security of data related to user identity and interaction. Nowadays, media personalization using TV-Anytime is gaining momentum as a promising mechanism that lays the foundations of novel business models in the field of provision of personalized services for IDTV, whose utility is undoubted in view of the content overload available in the digital stream.

[2] http://www.tv-anytime.org

Although TV-Anytime standardizes a common format to describe audiovisual contents, we also permit the users to categorize and annotate contents in a free and collaborative way. For that reason, ZapTV adopts emergent technologies from the Web 2.0, a new generation of services that stresses on-line cooperation and information sharing among users. The foundation of the Web 2.0 is the creation of virtual communities establishing links among users who share common characteristics or interests. A virtual community or social network represents a structure of the relationships (edges) existing among the individuals (nodes). The language commonly used to model social networks is FOAF[3], a simple technology that permits to share and use information about people and their activities (e.g. photos, weblogs) to transfer information between Web sites, and to automatically extend, merge and re-use it online.

In a social network, the annotation process gathers the efforts of the community of users to create a shared collection of metadata, since its expressive richness and quality improves progressively. As each individual chooses freely a set of labels for a resource, the collection evidences both the social aptitudes of the community and a shared organization of the target annotation space. This community-managed classification process was coined by Thomas Vander Wal with the term *folksonomy*, referring to a progressive and collaborative definition of the categorization and organization of the content.

However, the use of a vocabulary without a common semantic substratum hampers to carry out automatic processes in the applications involved in the Web 2.0. To mitigate these drawbacks, it is possible to employ the technologies developed in the Semantic Web, which improve the automatic access and sharing of resources. Among these technologies, note the *ontologies*, which are defined as "'formal, explicit, shared conceptualizations of a domain" [5]. In this definition, *conceptualization* describes and abstract, simplified model of concepts in a particular domain of interest, by resorting to classes, properties and relationships among them. The term *explicit* indicates that the domain concepts and the constraints imposed on their use are explicitly defined. Finally, the terms *formal* and *shared* mean that the ontology specification must be machine readable, and that the represented knowledge must be agreed by consensus, respectively. One of the main strengths of ontologies is related to the possibility of carrying out logic-based deductive processes by reasoning about the semantics formalized by a specific language. In this regard, note that several languages have been proposed in literature, ranging from RDF and RDFS to OWL, which is the most expressive format to date. These reasoning processes lead to discovering relationships among the semantically annotated resources. As we well explain in the next section, these relationships provide new knowledge, thus enabling to improve the accuracy of ZapTV in content searching and personalization processes.

3 The ZapTV System

3.1 Main Functionalities

Once the technological scope of our project has been delimited, we highlight the most appealing functionalities provided by ZapTV:

[3] http://www.foaf-project.org

– **User-driven content generation in social networks**. The users in ZapTV congregate in social networks where they take advantage of the last generation consumer devices (e.g. digital video cameras and mobile phones) to generate their own audiovisual contents in a flexible and easy way. These contents are shared with the remaining users of the social network by accessing a website where they can be uploaded, published, viewed, rated, and even suggested to other individuals in the system. As these contents can come from a wide variety of codification technologies, sources, and formats, it is necessary to resort to transcodification processes so that they can be broadcast through the DVB-H channels in our system.

  In the context of the social network, the user-generated contents are also annotated progressively by means of folksonomies. Starting from these shared annotations, the system creates automatically a domain ontology where the contents are formalized and semantically described by employing both TV-Anytime metadata specification and the user-defined folksonomies. This ontology will be a key element to achieve the personalization goals of our system.

– **Broadcasting of personalized contents**. ZapTV offers two modes of personalized broadcasting, which we refer to henceforth as *explicit* and *implicit recommendations*. The contents suggested by the system to the users can be either stored in their terminals or accessed on-line via the Web site. In the first case, the recommendations are sent through a 2.5G/3G network, along with the alerts employed to notify the users of the availability of the suggested contents.

  • *Explicit recommendations*. In this scenario, the users take the initiative of suggesting contents they liked to other individuals in the system. These contents can be recommended from either the system Web site or a client application running in the DVB-H receiver available in the user terminal.
  • *Implicit recommendation* In the implicit mode, ZapTV resorts to a personalization engine that automatically selects contents of interest for each user by considering the TV-Anytime metadata describing the contents, their audience levels, and the ratings and viewing history of the user. Specifically, this engine employs a two-phase hybrid personalization strategy that combines the two filtering techniques most used in literature, named *content-based methods* and *collaborative filtering* (see [1] for details). Both techniques suffer from severe drawbacks due to the syntactic mechanisms used to match the users' preferences against the available contents. To alleviate these limitations, the two phases of our hybrid strategy reason about the domain ontology in ZapTV to discover semantic relationships that provide additional knowledge about the users' interests (see [3] for details), as we explain next.

    The content-based phase selects the appropriate programs for a user by maximally matching her preferences against the semantic descriptions of contents formalized in the domain ontology. For instance, if a user has rated as interesting contents about potholing, our reasoning-based engine could suggest other programs about applied sciences strongly related to the former one, such as archaeology (specifically, *potholing* and *archaeology* are hierarchically related in our domain ontology as both are instances of the *Applied Sciences* class). Next, in the collaborative stage, the users with similar preferences (named *neighbors*) assist each other in finding contents of interest. Our semantic reasoning processes help to create the users' neighborhood, permitting to detect that the preferences of two users are similar when the contents they have rated belong to a common category in the domain ontology. For instance, the neighbors of a user who has enjoyed the highlights of a soccer match are other users who have rated highly contents about tennis and basketball (as all these contents are classified under specific subcategories of the *Sports* class in the ontology).

- **Planning of thematic broadcast channels**. ZapTV arranges the available contents into thematic channels to be broadcast though the DVB-H channel. For that purpose, the system resorts to an intelligent mechanism that tackles these planning tasks by considering different factors, such as the ratings assigned by the users, the metadata formalized in the domain ontology, and the kind of target audience, among others. This way, for instance, ZapTV provides channels with diverse contents: the favorite programs for most of the users, contents classified under a specific genre (e.g. fiction, sports, music, leisure), programs aimed at specific audiences (e.g. children under 12, teenagers, young couples without kids), contents about a particular topic (e.g. natural disasters, politics).
- **Viewing of contents and interactive capabilities in the users' receivers**. The client application running in the DVB-H receiver permits the user to view the contents she has generated, contents she has downloaded from the system Web site, and those that have been suggested to her by either other users in the system or the smart personalization engine included in ZapTV. This client application allows also to endow the user's receiver with interactive capabilities leading to a much more rich viewing experience for the users logged into our system. This way, by a return channel based on a 2G/3G network, the user can rate contents, send her viewing histories (i.e. list of contents viewed in the past), suggest programs to other users, and even, access online interactive trading services.
- **Broadcasting to handheld devices in mobile networks**. ZapTV includes a DVB-H headend in order to: (i) transmit the broadcast channels to the handheld devices, (ii) send the Electronic Guide of Services (ESG) with the available channels and the schedule of their contents, (iii) manage the interactive capabilities, and (iv) send alerts to users to notify them of the availability of appealing contents.

3.2 Architecture

As shown in Fig. 1, the architecture of ZapTV is divided into the five functional blocks enumerated below, which are described in detail in the following sections: (i) Provision and adaptation of contents, (ii) Web interface for publication, rating and recommendation of contents, (ii) Personalization and planning of channels and contents, (iv) Broadcast and interactivity management, and (v) User terminal.

3.2.1 Provision and Adaptation of Contents

Broadly speaking, this block takes charge of: (i) acquiring and processing the audiovisual contents, and (ii) managing the interactive services delivered by the system. As shown in Fig. 1, these tasks are distributed among the modules described next:

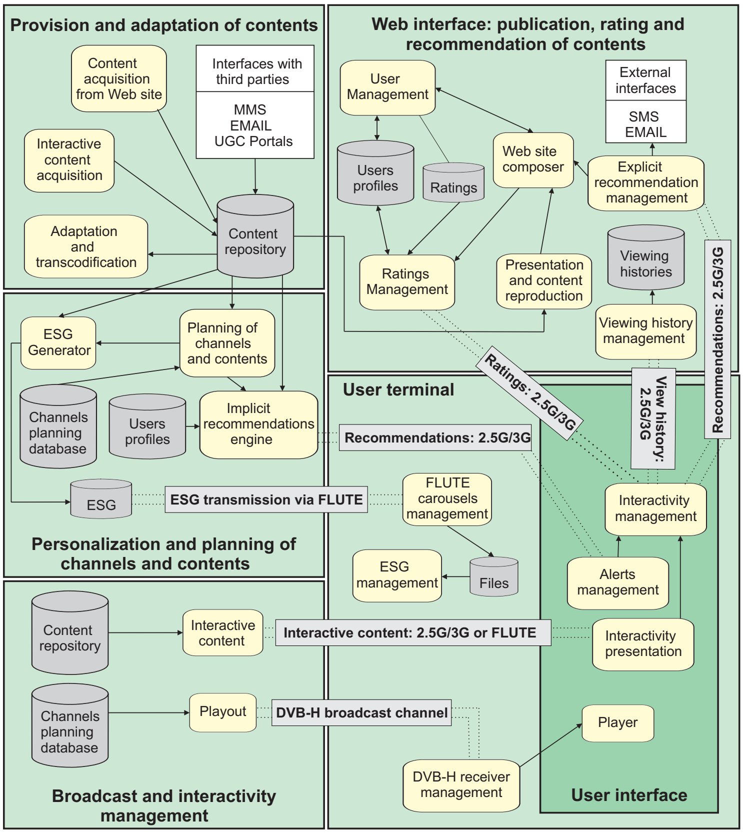


Fig. 1. Architecture of ZapTV system

– **Content acquisition from Web site**. One of the key elements in the Web site is the module that develops the mechanisms to upload the user-generated contents from a Web page and to store them in a repository.
– **Acquisition interfaces to third parties**. Other contents available in ZapTV come from some well-known UGC portals, and even are sent via multimedia messages (MMS) and e-mail. This acquisition process requires interfaces that are developed and managed by the module accessing the content repository in Fig. 1.
– **Adaptation and transcodification of contents**. The module that converts the contents originating in different sources to the formats and codifications handled by the DVB-H broadcast channel is also included in this first functional block of the architecture. To fulfill the handheld devices requirements, our system uses the video compression standard H.264 and the resolution 320x240 set in QVGA (Quarter Video Graphics Array), typically employed for displays of mobile terminals.
– **Interactive content acquisition**. The interfaces to receive interactive contents in ZapTV permit to download music, ringtones, access on-line trading services, and

even to download advertisements that are synchronized with the audiovisual contents without hampering the user's viewing experience (see [7] for details).
- **Content repository**. The available programs and the interactive contents are stored in the system repository, along with the domain ontology where the semantic descriptions of audiovisual contents are formalized. As these descriptions are extracted from the user-generated folksonomies, the contents are annotated freely, being the system which translates these annotations to a common format (TV-Anytime) to tackle personalization and searching tasks. In fact, the domain ontology in ZapTV enables advanced searching of user-requested specific contents, and also it permits to reason about the semantics of contents leading to the enhanced automatic recommendations mentioned in Sect. 3.1.

3.2.2 Web Interface for Publication, Rating and Recommendation of Contents

The functionalities of the Web site are carried out by the modules described below, which are mainly focused on publishing contents and developing the mechanisms the users need to rate and recommend contents to other individuals, as depicted in Fig. 1.

- **Presentation and contents reproduction**. This component shows an interface with common reproduction options (e.g. pause, forward, rewind), which is adapted to the formats and codifications considered by the transcodification module.
- **User management**. The interfaces for logging into the system, logging out and modification of the users' personal information are developed by the *user management module* depicted in Fig. 1.
- **Rating management**. The ZapTV architecture includes a separate module to manage and store the ratings the users assign to the contents both from the Web site and from their handheld devices through the return channel.
- **Explicit recommendation management**. The contents a user recommends to other individuals are managed by the homonymous module depicted in Fig. 1. Such explicit suggestions can be ordered from both the Web site and the mobile device, hence the fact that this module includes interfaces with external notification systems (based on e-mail and SMS) to report the users of the availability of the suggested contents (which are either received via a 2.5G/3G network or accessed on-line).
- **Viewing history management**. ZapTV includes also a component that communicates with the client application running in the user's DVB-H receiver to obtain information about her viewing history, from which times with a large number of viewers can be inferred. These audience levels will be exploited by the module in charge of channel planning, so that the contents with the highest ratings will be broadcast in times that are in great demand by the users.
- **Web site composer**. Lastly, all the components included in the Web site are coordinated by a module that also composes the interface shown to the user.

3.2.3 Personalization and Planning of Channels and Contents

This block delegates to three modules both the generation and planning of the DVB-H broadcast channels, and the implicit recommendations automatically selected for the users of our system (see Sect. 3.1).

- **Planning of channels and contents**. The module that generates the thematic channels described in the previous section also develops interfaces to access the databases of ratings and viewing histories. So, the favorite contents for most of the users are broadcast in times with maximum levels of audience, as we commented before.
- **Generation of ESG**. After the planning tasks, it is necessary to generate the ESG including the DVB-H channels available in the digital stream, along with information about broadcasting time and date of each content. As shown in Fig. 1, the module that manages this process sends the ESG by a FLUTE carousel, which is a solution considered in DVB-H for broadcasting data that must be available anytime.
- **Implicit recommendation engine**. The personalization capabilities provided by ZapTV are distributed between the Web site –where the explicit recommendations are managed– and the module *implicit recommendation engine* depicted in Fig. 1. As shown in this figure, the personalization engine resorts to a 2.5G (GPRS) or 3G (UMTS) data channel to send the contents that our two-phase strategy suggests.

3.2.4 Broadcast and Interactivity Management

The two modules included in this block centre on the transmission of contents through the broadcast channel. The first component (named *playout*) encapsulates as IP datagrams the information stored in the database used by the module of channel planning. Next, these data are modulated and finally broadcast though the DVB-H channel. The second module focuses on the *interactive contents*, which are typically transmitted via a 2.5G/3G network, although it is also possible to send them through the FLUTE carousel available in the system.

3.2.5 User's Terminal

The last block in the proposed architecture includes the modules required to compose the client application that allows the user to access the system functionalities by handheld devices. The three former modules described below develop access interfaces to the internal resources of the mobile terminal, whereas the remaining components take charge of the user interface, as shown in Fig. 1.

- **FLUTE carousel management**. The user's terminal includes a module to manage the reception and storage of information about the ESG (and interactive contents) by means of the FLUTE carousel.
- **ESG management**. From this information, other module obtains and presents the ESG information on the user's device the list of available channels and contents.
- **DVB-H receiver management**. The browsing of channels and selection of specific contents are controlled by a manager module which is focused on the DVB-H receiver available in the user's handheld device.
- **Interactivity management**. The return channel available in ZapTV, as well as its associated processes (e.g. sending of ratings, viewing histories and explicit recommendations) are controlled by a component inside the user's interface, which manages the wide variety of interactive services offered to the user.
- **Interactivity presentation**. The interactive options provided by ZapTV are shown to the users by interfaces developed by the *interactive presentation module*. As seen

in Fig. 1, this module communicates with the component of interactivity management in order to agree the use of the return channel and to synchronize the interactive contents with the programs received through the DVB-H broadcast channel.
– **Player**. Lastly, the user terminal includes a player that decodes the DVB-H stream, allowing to view the received audiovisual contents.

As conclusion, highlight that the openness and modularity of the proposed architecture make it easy to incorporate into ZapTV both new services appealing to the users of mobile TV, and future technologies that promote an extended use of our system.

4 Conclusions

In this paper, we have described the main functionalities and architecture of ZapTV, a system that broadcasts channels of TV for handheld devices compliant with DVB-H. Our system is inspired by the principles of participation and sharing between individuals promoted by the so-called Web 2.0. For that reason, the users in ZapTV gather together in mobile social networks where they generate and share audiovisual contents; also, the user rate and annotate the contents in a collaborative way in order to facilitate advanced searching tasks. Along with the user-driven content generation and the ubiquitous access via handheld devices, our broadcast system for mobile environments is supplied with extra functionalities, such as unobtrusive synchronization of interactive services with the DVB-H channels, and broadcasting of personalized audiovisual contents. In this regard, note that, although the users can explicitly recommend contents to other individuals, ZapTV also includes a smart personalization engine that automatically selects contents appealing to each viewer (according to her preferences and viewing history), by combining traditional filtering mechanisms with semantic reasoning processes that lead to improved suggestions.

Thanks to the mentioned functionalities, ZapTV offers an enhanced viewing experience to the new users of TV, who are demanding insistently for more and more active roles that permit them to generate and share audiovisual contents.

References

1. Adomavicius, G., Tuzhilin, A.: Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 739–749 (2005)
2. Akamai White Paper. Successful Social Networking and User-Generated Content: What You Need to Know (2006), `http://www.akamai.com/dl/whitepapers/social-networking-wp.pdf`
3. Blanco Fernández, Y., Pazos Arias, J.J., López Nores, M., Gil Solla, A., Ramos Cabrer, M.: AVATAR: An improved solution for personalized TV based on semantic inference. IEEE Transactions on Consumer Electronics 52(1), 223–232 (2006)
4. Digital Video Broadcasting (DVB); Transmision system for handheld terminals (DVB-H). ETSI EN 302 304 V1.1.1 (2004)
5. Gruber, T.R.: A translation approach to portable ontologies. Knowledge Acquisition 5(2), 199–220 (1993)

6. Hales, D., Arteconi, S.: Friends for Free: Self-Organizing Artificial Social Networks for Trust and Cooperation. Technical Report UBLCS-2005-20, Department of Computer Science (University of Bologna) (2005)
7. López Nores, M., Pazos Arias, J., Blanco Fernández, Y., García Duque, J., Tubío Pardavila, R., Rey López, M.: The MiSPOT System: Personalized Publicity and Marketing over Interactive Digital TV. In: E-Business and Telecommunication Networks. Springer, Heidelberg (2007)
8. O'Reilly, T.: What Is Web 2.0? Design Patterns for the Next Generation of Software (2005), http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Prototyping Interactive and Personalized IPTV-Services on Top of Open IMS Infrastructures

Oliver Friedrich, Robert Seeliger, Adel Al-Hezmi, Christian Riede, and Stefan Arbanowski

Fraunhofer Institute for Open Communication System, Kaiserin-Augusta-Allee 31
10589 Berlin, Germany
{Oliver.Friedrich, Robert.Seeliger, Adel.Al-Hezmi, Christian.Riede, Stefan.Arbanowski}@fokus.fraunhofer.de

Abstract. In this paper a proof of concept infrastructure for the delivery of standardized IMS-based IPTV services driven by the IPTV enabled Open IMS Playground at Fraunhofer FOKUS will be presented. The so called Media Interoperability Lab represents one of the world's first real world test beds for these kind of services. This has been achieved by actively guiding the different SDOs work with a focus on ETSI TISPAN. On top of this framework different personalized and interactive services have been implemented to show the predominance and advantages of IMS-based IPTV solutions compared to other mainly black box approaches.

Keywords: IPTV, Standardization, Interactivity, IMS.

1 Introduction

With the advent of the first IPTV deployments the question for a more generic solution based on worldwide standards - as they do exist for Digital Television - has been raised throughout industry and academia.

The year 2007 was driven by different Standard Development Organizations (SDOs) working on guidelines, frameworks and standards for Internet Protocol Television (IPTV). With respect to their different scope all of them are working on or recognized the need to integrate solutions following the approach of Next Generation Networks (NGN) and the IP Multimedia Subsystem (IMS) [1] respectively. Mainly this includes ITU-T Focus Group IPTV [2], ETSI TISPAN [2], Digital Video Broadcasting Group (DVB), ATIS IIF and the Home Gateway Initiative (HGI).

With the availability of ETSI TISPAN Release 2 a framework not just dealing with architectural issues but also details on signaling and used data formats will be available to built IMS-based IPTV end-to-end solutions.

This paper presents Fraunhofer FOKUS' activities within the so called Media Interoperability Lab (MIL) [7] setting up a real world implementation for IMS-based IPTV combined with interactive IPTV services as Targeted Advertisement and Interactive Shopping or a Remote Parental Control Service.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 204–208, 2008.

2 Media Interoperability Lab – A Prototype Infrastructure for IMS-Based IPTV

To enable the IMS for IPTV and interactive services a set of common service functions has been identified and integrated into the FOKUS Labs on top of the basic Open IMS architecture [4]. The key components of Open IMS are depicted in the center of **Fehler! Verweisquelle konnte nicht gefunden werden.** and include:

(1) The **IPTV Service Provisioning Function** offering information on Service Discovery by so called entry points enabling the connected UEs to find adequate Service Selection entities. These entities provide the Electronic Program Guides (EPG) including information on available channels and Video on Demand feeds.

(2) The **IPTV Service Control Function** that is in charge of the whole life cycle of a service based on a defined business model. This includes session management, access control, service composition, personalization, charging, processing and managing of the interactive application;

(3) The **IPTV Service Personalization Function** which actively gathers information on user behavior and generates (content) recommendations taking also into account the user profile and manual user rating. Recommendations will be calculated by integrating also collaborative approaches as considering other user's behavior. More information on the techniques behind and also engineered at FOKUS can be found in [5].

(4) A set of **IPTV Media Functions** responsible for content delivery to the end user as well as media processing and an efficient utilization of available network resources including QoS and mobility support.

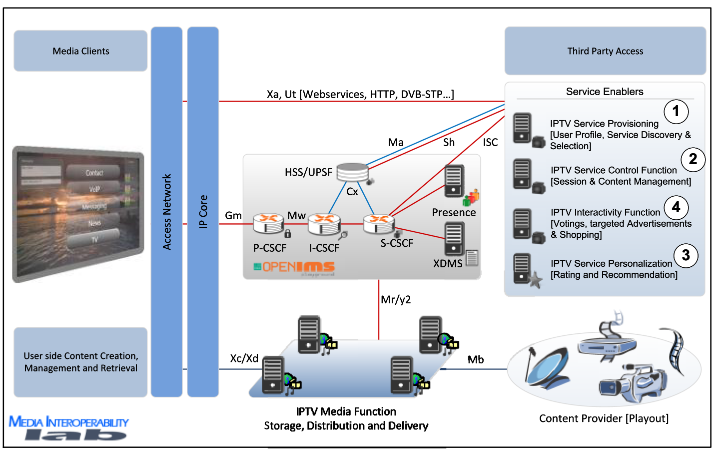


Fig. 1. High level architecture for IMS based IPTV

2.1 Architecture Enhancements for Supporting Interactive Scenarios

This section shall give a short overview on the different types of services provided by the proposed infrastructure and necessary architectural enhancements. As all scenarios have already been described in detail in a previous paper [6] just the main three categories and some examples are presented here which are **IPTV features** as Plain streaming services as Live TV and Video on Demand controlled through the IMS infrastructure, **NGN Telecommunication Services** provided by the NGN infrastructure including presence, VoIP Telephony and Messaging and **IPTV-NGN interaction** showing the combination of common NGN services with new IPTV services as CallerID on TV screen, Time Shift Mode for Video on Demand and Live TV on accepted calls and call forking to multiple devices including the TV set**.**

To enable interactive services three different types of modifications were applied to the lab's infrastructure: This includes modifications to the client middleware, application logic and GUI to support incoming and outgoing interactive event handling as well as enhancements to the basic IPTV signaling and extensions to the Service Control Function to forward messages to the Interactivity Function providing access to active user sessions.

To host the different types of interactive applications a Service Interactivity Function as depicted in Figure 1 and marked as (4) has been added to the architecture acting as a generic AS to process the server side logic for the scenarios presented in the next section. In future revisions these services will be provided by different Application Servers in the Service Providers Domain or at the Content Provider side which implies the need for open interfaces in this direction.

These interfaces are currently out of scope and will be part of future work.

3 Interactive and Personalized Service Scenarios

The basic hooks for the scenarios described are on the one hand given by the underlying NGN infrastructure or on the other by adding Application Server and client logic to the architecture already being integrated into the lab's components.

By making available the AS infrastructure together with the NGN infrastructure the vision being able to address single users or groups of them with the same interest with personalized or so called targeted advertisements becomes true. Scenarios as interactive voting and personalized advertisement have been realized and can be enriched with advanced interactivity to the advertisements by providing access to the products presented in the different spots. This allows the consumer to buy the presented or related products through his IPTV interface and getting charged either through his Service Provider on his monthly bill or by credit card credentials also stored in the user profile.

3.1 Evaluation and Proof of Concept

To evaluate the proposed solution of the IMS-based interactive multimedia framework a scenario for the injection of personalized advertisements has been selected to demonstrate the advantages of IMS-based media delivery. Figure 2 depicts a simplified message flow for this scenario. Details on how the ad-injection is triggered from the outside and details on RTSP signaling are out of scope of this description and will be analyzed in one of our following papers just focusing on the different personalized and interactive enhancements of our architecture. The described scenario enables the end-user to watch different live channels where personalized advertisements will be inserted by a managing entity based on the user profile. To simplify the tasks also for the Service Provider each user has been assigned to a specific group matching his behavior and target group. After selecting a specific channel from the EPG and the setup of the multicast delivery for this service (see messages 1-7 in figure 6) the Session Control Function (SCF) initiates or receives a command for injecting a personalized advertisement. The SCF selects an appropriate Media Function for the advertisement play out (MF #2 in this case) and forwards this information to the UE (see messages 8-12).The UE's application logic interprets the received SIP INFO (12) message and switches to the signaled RTSP connection. After the advertisement has been played out by the client the application logic switches back to the live stream. It may be noted that one major drawback of this scenario is that the user is missing content while receiving the personalized advertisement. As this is just an example how the MIL infrastructure is used to signal this service various scenario modifications could solve this problem as using a modified signaling approach to enable picture-in-picture advertisements blended on the UEs screen or in Video on Demand scenarios where the (paid) content is paused during the advertisement.

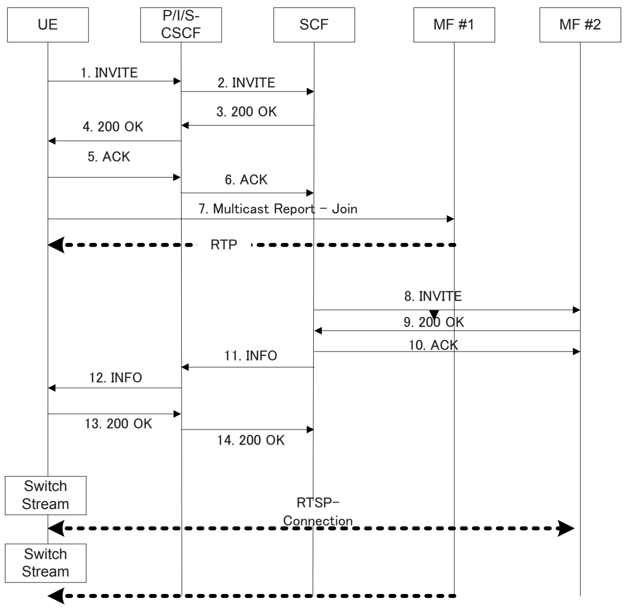


Fig. 2. Interactive Service Signaling

4 Summary and Future Work

With this paper we presented Fraunhofer FOKUS approach for the delivery of IMS-based IPTV services over NGN networks incorporating the well known Open IMS Playground. With the presentation of more advanced IPTV services going beyond plain broadcasting we showed that the IP Multimedia Subsystem is a perfect architecture to deliver personalized and interactive streaming services. Future work will concentrate on improving the implemented solution and adding new services and components. This includes work on issues like Third Party Openness e.g. for the connection to advertisement content providers and interfaces to upload user generated content (UGC). Also Application Servers and signaling issues to support Networked based Personal Video Recorders and Home Network issues (e.g. bridging between the IMS and DLNA world) will be targeted in future projects.

References

1. 3GPP, TS 23.228. IP Multimedia Subsystem (stage 2), http://www.3gpp.org
2. ETSI ES 282 001: Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN), NGN Functional Architecture Release 2
3. International Telecommunication Union; Focus Group On Iptv Telecommunication Standardization SectorFG IPTV-DOC-0115, Working Document: IPTV Architecture
4. Vingarzan, D., Weik, P., Magedanz, T.: Development of an open source IMS core for emerging IMS testbeds. Special Issue on IMS, Journal on Mobile Multimedia (JMM) 2 (2006), http://www.rintonpress.com/journals/jmm/
5. Räck, C., Arbanowski, S., Steglich, S.: A Generic Multipurpose Recommender System for Contextual Recommendations. In: 8th Intern. Symposium on Autonomous Decentralized Systems (ISADS 2007), Sedona, Arizona (March 2007)
6. Al-Hezmi, A., Friedrich, O., Magedanz, T.: Requirements for an IMS-based Quadruple Play Service Architecture. Network 21(2), 28–33 (2007)
7. Fraunhofer FOKUS Media Interoperability Lab, http://www.mediainteroperabilitylab.org

A Software Component for Content Management and Delivery to Mobile Phones and Digital Television

Gaetanino Paolone and Eliseo Clementini

Department of Electrical and Information Engineering - University of L'Aquila
gpaolone@ing.univaq.it, eliseo@ing.univaq.it

Abstract. The proposed software component is intended to facilitate the development of interactive software for the Digital Terrestrial Television (DTT), mobile phones and any other digital platform. The component implements and solves some software standard issues and is the basis for the integrated development of cross-platform software. We also illustrate a real process [1, 4, 5] that uses the software component for the development of an advertisements management application. The requirements that are needed for software development with the component are the knowledge of the "layer" pattern and the specific digital platform [4, 6].

1 Introduction

Nowadays, one of the most important challenges in computer science is the definition of production processes for software that allow us to reuse components and services towards the development of software systems on different platforms. The "content management" component implements a software layer capable of providing and receiving information from different clients (DTT, mobile phones and any other digital platform). This is possible through the definition of a set of services that the component is granting and an XML based communication. The component permits the implementation of an application that adheres to the guidelines of the Service Oriented Architecture (SOA) [7], where the main concept is the "service" that is provided by the component and used by the client. The content management has its roots on the studies made on the deployment of DTT applications, and on our belief that the content must be kept independent from the communication system. This work extends and generalises the content management for the DTT [4] through services for the mobile phone. Three are the relevant factors on which we base the whole proposal:

1. the interface design to manage the information on the specific platform;
2. the need to have a unitary processing for the business logic;
3. the automation of the required computations to find the data from a generic source.

The new technology and the lack of interactive multimedia applications suggests the research and experimental character of our product. In Section 2, we show the architecture of the component and how it works. In Section 3, we describe a real project that uses the content management. In Section 4, we draw short conclusions.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 209 – 213, 2008.

2 The Content Management Component

The key elements of the component architecture are the communication system at disposal of the user to interact with the application (notebook, television, telephone) and the supplier of the contents to be published. The applications that use the content management component implement only the presentation layers for the specific platform, while the business layer is unique. Applications are written in Java [1, 6]. In Fig. 1, it is possible to observe the break-up of the necessary elements to implement an interactive software with related connections.

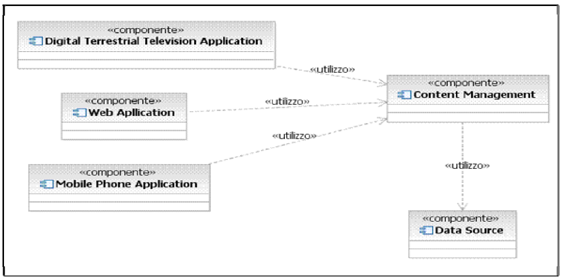


Fig. 1. Component Diagram

The content management design has been made to allow a simple and fast management of any source without having to modify the inner structure; it is independent from the kind of offered service and from the kind of processed data. The designed and implemented content management is a software module that can be re-used in many applications or software systems.

Re-use is a typical approach of many engineering fields and, in order to obtain the expected results, it requires a specific organizational and methodological formulation [4, 5]. The component design takes into account a context where it is necessary to supply services towards generic applications and the information may originate from any data source. The typical distribution is illustrated in Fig. 1: different resident applications on different media can be connected to each server component and, in turn, every server can be connected to different data sources, which are themselves on separate machines. The structure of the component provides a presentation layer, a control layer, a domain layer and a resource layer. The presentation layer and the control layer are developed on the terminal system (client) and it is specific for one platform. The content management implements a service to facilitate the application development. Also the component server has a user interface for the management of data sources and the activation service. The use cases of the content management are: “source management” and “service management”. The source management is used by the administrator to configure the content management and set up the sources on which retrieving data. The service management is used by any DTT or mobile phone application, which offers retrieval and manipulation services. The content management application displays the interface to the user and allows the activation of the two use cases [1, 4].

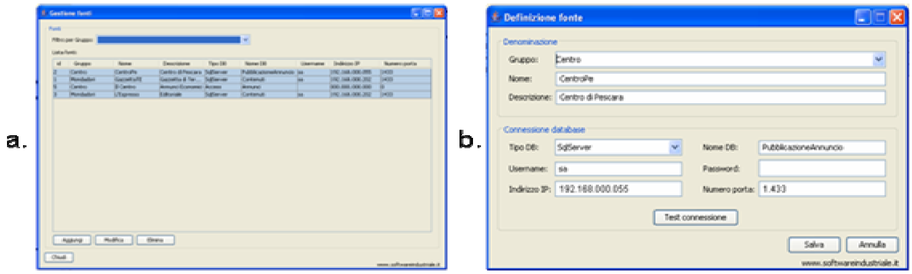


Fig. 2. The source management use case

The source management use case (Fig. 2.a) allows us to manage the required information for the publication of contents of a generic source.

Fig. 2.b shows the data source detail with all the information that are needed by the content management. To connect to a different data source, with the same database structure, is as simple as adding inside a new source, without changing the software component.

The service management use case (Fig. 3.a) allows us to coordinate all the interaction phases that can occur between the system and the applications on the decoder or on the mobile phone. The domain includes all object classes that are dedicated to the manipulation of the necessary information for all applications. The domain classes interact with external resources through a dedicated layer of Java classes. Specifically, three resource layers for the treatment of permanent data are included. To dialogue with a generic data source the content management implements the “abstract factory” pattern, which provides an interface to create families of objects and make the client independent from the implementation of real objects.

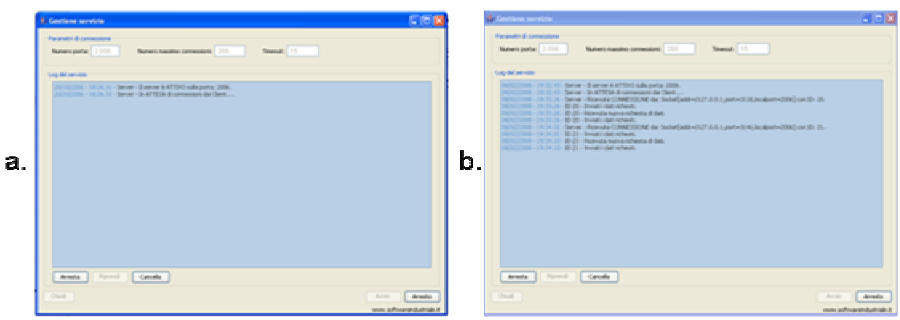


Fig. 3. The service management use case

Fig. 4 illustrates the architecture [4] of a digital interactive application that uses the content management. The architecture is very powerful: it is sufficient to design and develop just the client of any application and implement the abstract classes of the “creational” pattern of the resource layer to have a software product running on multiple digital platforms. Therefore, the content management is a component to develop a digital application providing that a software system is already available that handles data manipulation and user interaction [2]. The designer focuses on the interfaces for user interaction [1, 5, 6] on the specific platform and on the query system of the specific database being made available by the data source.

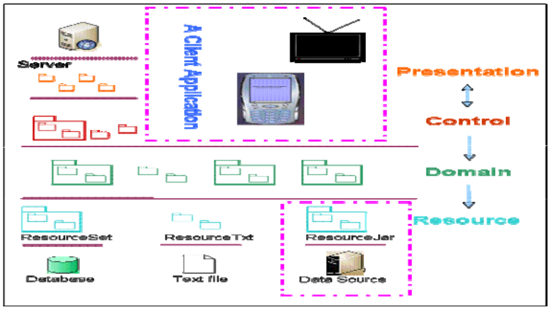


Fig. 4. The content management complete architecture

3 The Example Application

The application allows the look up of press articles, advertisements, and bank accounts and was built with the content management component. We have two applications: one on the mobile phone and one on the DTT. When an application connects to the content management server we can have different requests. If a DTT application sends a request to the server, we have a request identifier "ID 20"; if at the same time a mobile phone application sends a request to the server, we have a request identifier "ID 21" (Fig. 3 b). When the mobile phone or the DTT is sending a request, the content management will process it and send back the information. If a user requests a list of advertisements through a mobile phone, he/she will get the result shown in Fig. 5 a. In the same way, the user will get the same results if sending the request through the DTT (Fig. 5 b).

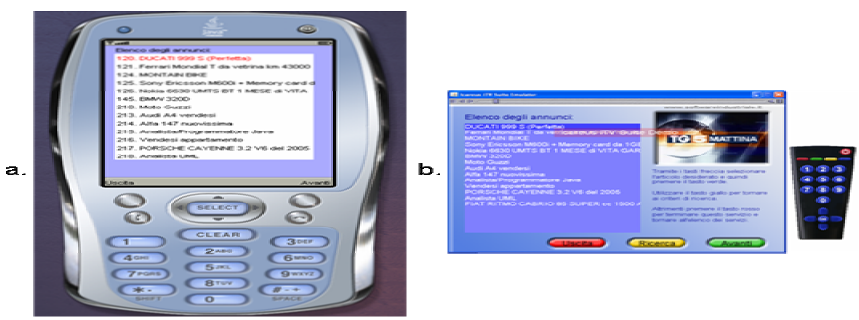


Fig. 5. A list of advertisements

We have different navigation systems, but the same content on different media. The human interaction can be very different on various communication systems and the designer should propose a solution with specific interfaces using the content management component. For example to search the advertisements on the mobile phone we have an additional menu (Fig. 6 a), while on the DTT we simply have a green button to press on the remote control (Fig. 6 b).

Fig. 6. The advertisement research

4 Conclusions

This paper presented a content management capable of offering services for a generic digital client. One of the future goals is to build the interface by using the content management for any digital media. So far, we have successfully developed a demonstration of the user interface for interactive television and mobile phone services, as shown in the article. The component "Content Management" is an innovative way to make multiplatform software, based on the total reuse of the component itself and its services. The idea to design such a component moves by the need to have a single methodological approach to develop integrated systems with multiple interfaces on multiple media. Our next objective will be to develop a methodological process for developing digital applications based on our component.

References

1. Flower, M.: UML 2 Distilled. Addison Wesley, Pearson (2004)
2. Iatrino, A., Modeo, S.: Text Editing in Digital Terrestrial Television: a comparison of three interfaces. In: Proc. of conf. EuroiTV 2006, Athens, Greece, May 25-26 (2006)
3. MacKenzie, I.: I.S.E. (2002); Special Issues on Text Entry for Mobile Devices. Human-Computer Interaction 17(2,3)
4. Paolone, G., Clementini, E.: A Methodology for building enterprise software for digital terrestrial television. In: Proc. Conf. MITIP 2007, Florence Italy, September 6-7, pp. 420–425 (2007)
5. Sewall, S.J.: Executive Justification for Adopting Model Driven Architecture (MDA) (November 2003)
6. SUN Microsystem, JavaFX Mobile Architecture (February 15, 2008), `http://java.sun.com/software/javafx/mobile/index.jsp`
7. Wikipedia: Service-oriented architecture (February 15, 2008), `http://it.wikipedia.org/wiki/Service-oriented_architecture`

Development of a Generic XML Personality Metadata Handler for Distributed Entertainment Services

Simon Reymann, Jakub Rachwalski, Stefan Kemper, and Artur Lugmayr

NAMU Lab., Tampere University of Technology
http://namu.cs.tut.fi/
FIN-33101 Tampere, +358 (0) 50 935 6668
{simon.reymann, stefan.kemper}@tut.fi,
jakub.rachwalski@gmail.com,
lartur@acm.org

Abstract. Especially in the case of entertainment services, the provision of a general valid personality profile is gaining in importance. The personality profile shall include information about the usage-context of media services. These media services can be arbitrary and range from simple playlist compilations of audio-visual content to complex feedback profiles in interactive television. This research paper provides an overview of a metadata based solution for merging personalized content. Multiple applications, such as audio or video players contribute with user-context information enabling future personalization scenarios. An XML metadata specification is presented providing a generic container format for representing a personality user profile. To enable generality, the personality profile simply extends existing XML standards by adding attributes and thus keeping the existing structure intact. This also allows the creation of slim and resource efficient personality schemas which are especially beneficial when deployed on resource-sensitive client devices like mobile phones.

Keywords: Profiling, Personalization, Personal Content, Metadata Handling.

1 Introduction

Nowadays people are surrounded by multiple electronic devices which, with an increasing probability, hold a personality based profile of a user. This particular piece of information could be used by various applications creating new and customized services. Due to the variety of sources, incompatibilities need to be resolved and a smooth combination needs to be created.

Profiling has been in development for years and is considered as the antidote for the rapid growth of information overrun of e-services. Profiling technologies have been widely described, e.g. in [1] or [2] and also profile merging algorithms from multiple sources have been discussed [3]. However, problems in gathering information from multiple sources and producing personal content [10] of unknown structures are not solved yet. Additional and comparable works are dealing with very application specific solutions as e.g. personalization in the context of broadcasting, applications of specific standards such as MPEG-7 for personalization, specific personalization algorithms, or agent based approaches for personalization. This work is based on the

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 214–218, 2008.

architectural approach proposed by the Portable Personality (P2) project [4] embedded in the New Ambient MUltimedia Lab. at the Tampere University of Technology [5]. The key-feature of P2 is its general applicability in many application scenarios, especially on the audio-visual sector. The goal of the project is to bridge multiple personal content providers with contemplable consumers evolving into a distributed profiling network with many nodes. In the P2 context, nodes are every-day portable devices like laptops and mobile phones, or stationary systems like desktop PCs and terminals. The following presented solution describes how P2 manages personal information collected from multiple contributing sources and what kind of extension mechanisms are needed to be able to identify and merge personal content from multiple providers.

2 System Architecture

The P2 network consists of three types of nodes, namely providers, consumers and daemons. The structure of the system together with the flow of the information is presented in Fig. 1.

Due to its features and popularity the XML format is the only accepted format in P2. The elements of the system which are responsible for data acquiring are called providers. They send personal content together with P2 management functions to a specific daemon. The daemons store the data, apply the functions provided and send the acquired information to contemplable consumers.

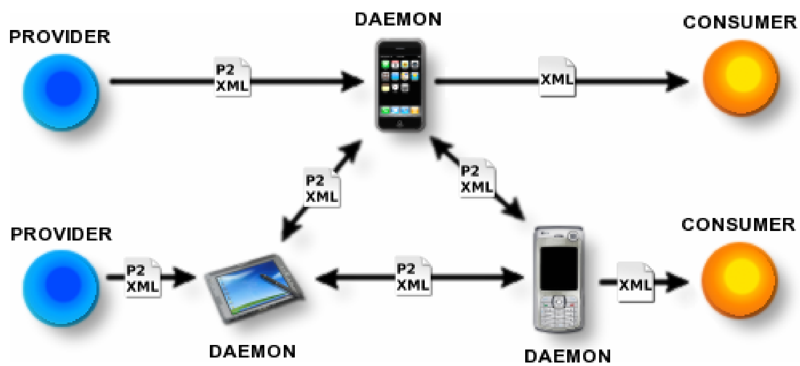


Fig. 1. Overview of the P2 system architecture with multiple provider and consumer nodes interconnected through the distributive P2 daemon network

The P2 architecture could be easily applied on a home entertainment environment, consisting of a digital TV, a Bluetooth enabled set-top-box and multiple Bluetooth enabled mobile phones representing the audience. The mobile devices carry information about the general audio-visual behaviour of its user, such as genre, theme or consumption. That means that each phone represents a P2 daemon with a P2 provider. The set-top-box instead detects all mobile phones within its vicinity and plays a central role in this scenario. On the one hand it behaves like a P2 daemon and collects available data from surrounding mobile phones via Bluetooth, i.e. it gathers audio-visual behaviour patterns from the audience and merges each individual profile to one representative audience profile. On the other hand the set-top-box is a P2 hybrid being provider and consumer at the same time. Based on the generated audience profile an integrated recommender system could provide a personalized selection of available audio-visual content (P2 consumer). Then, based on the performed decisions of the audience the existing profile can be updated (P2 provider). Additionally, individual profile updates are distributed to each mobile phone of the audience. By making personal content identifiable and mergable a set-top-box is able to pro-actively establish a personalized entertainment environment.

Every XML package received by a daemon may be very distant in structure, as it could come from different providers. What is more, each contained personal content item needs to be identified before it is processed by the daemon according to given management rules. In order to process an unknown standard the personal content provider needs to identify a personal item by extending it with specific P2 attributes (from the xmlns:p2="http://www.portable-personality.org/") before sending it to the daemon. If the daemon knows a standard it is able to identify items and extend them automatically by default rules. As the provider or the daemon knows the structure and extensions he can easily create a complaint version of the XML schema accordingly. Furthermore, when sending personal content to a consumer the daemon deletes all P2 attributes generating valid XML content. Known standards, like Friend of a Friend [6], Description of a Career [7] or Media RSS [8] could be processed automatically.

This is an example of XML content with a specific P2 attribute extension.

```
  <favorites p2:item="1">
    <show p2:item="1" p2:limit="title,1,9">
      <title p2:merge="title,1">Found</title>
      <director
p2:merge="director,1">Spielberg</director>
      <value p2:merge="value,3">2.5</value>
    </show>
    <show p2:item="1" p2:limit="title,1,9">
      <title p2:merge="title,1">Spiderwoman</title>
      <director
p2:merge="director,1">Spielberg</director>
      <value p2:merge="value,3">3.3</value>
    </show>
  </favorites>
```

Each required management function has its own P2 attribute defined. The attributes' values consist of coma separated fields (CSF) defining options for a single personal content item of the original XML content. If the attribute refers to a couple of elements each field is separated by a semicolon. Following attributes types are valid:

- Single (S SYNTAX), e.g. p2:item- with 1 VALUE FIELD or V FIELD;
- Comma separated pairs (CSP SYNTAX), e.g. p2:merge- with 2 fields in CSF;
- Comma separated thirds (CST SYNTAX), e.g. p2:limit- with 3 fields in CSF;
- Comma separated quads (CSQ SYNTAX), e.g. p2:delete- with 4 fields in CSF.

The first item in a CSF is called REFERENCE (R) FIELD, the second MODE (M) FIELD, the third ADDITIONAL (A) FIELD and the fourth EXTRA (E) FIELD. A P2 attribute placed in an element with an R FIELD may refer to the element, its value or a child element. However, if it is an attribute it contains the name of the attribute proceeded by $, e.g. $<attribute_name>. What is more, the default value of P2 attributes may be changed for the element and all its values, attributes and the child nodes. In this case the element is preceded by @, e.g. @<element_name>.

3 Personality Profile Handling

This section describes the changes of the metadata structure and possible operations on the personality profile, especially merging.

In XML content some of the elements together with their child nodes should be treated as one piece of information (further referred as ITEM). In the exemplary code each “show” element represents an ITEM and is then described by several child nodes like “director” or “title”. The element must be marked with the p2:item="1" attribute which is of S SYNTAX (using p2:item="0" ignores the element and is considered as default). ITEMs represent an actual personal content item and hold all required rules for handling contained data. In some XML structures, like iTunes XML [9] in which the parent-child relationship is not defined by its structure the usage of p2:key, for parent nodes, and p2:parentkey, for child nodes was invented. With this, parent-child relationships are uniquely identified.

3.1 Merging

The profile merging is one of the most important features in the P2 system. In order to merge two ITEMs the system needs to know what ITEMs are of the same nature, i.e. what ITEMs representing the very same personal content but holding different values. It is important to identify the type of an ITEM before merging it with an existing ITEM. Due to the fact that two ITEMs can be of a different XML structure, an identification and merging attribute is implemented. The identification of an ITEM merge key (identification key) is done by using the p2:merge=”<REFERENCE>,1” attribute which defines <REFERENCE> as an identifier of the ITEM. All values of all defined identifiers and their location within the structure must be equal before declaring two ITEMs to be of the same type. After characterization through the identifiers the ITEM’s child elements get merged according to specified merging modes, each of them representing an individual merging strategy. These in fact can be of any type, e.g. sum, multiply, take the newer, concatenate etc.

For merging, a daemon must hold information about the provider source and details about its type. The source is provided as a Globally Unique Identifier (GUID) and the type is represented by its namespace. Content of multiple sources (multiple GUIDs) embedded in the same namespace specification is merged into one XML representation. Merging of personal content originating from two different namespaces, instead, is not supported. The final result is a collection of multiple namespaces holding personal content of multiple profiling sources.

3.2 Sorting, Limiting and Deleting

Due to the fact that the storage capacity of a daemon is finite, e.g. on mobiles, the provider may set rules when to remove an ITEM using additional P2 attributes. The first option could be deleting on limit in which the number of ITEMs contained in the profile could be limited up to a certain value using the p2:limit attribute (CST SYNTAX). The limitation takes effect on every ITEM which lies at the same location within the XML structure holding the same ITEM element name and having the same p2:limit attribute defined. The M FIELD defines the sorting mode, e.g. ascending or descending. The R FIELD refers to the element on which the sorting is based on while the A FIELD defines the limitation value, e.g. 100. A second option is deleting on a value by using the p2:delete attribute (CSQ SYNTAX). An ITEM may be deleted when the value of a certain attribute or child element reaches a limit. The M FIELD holds the condition to be met, e.g. =, <, >, <= or >=, the A FIELD states the limit value and the E FIELD provides grouping functionalities meaning that the conditions of all group members need to be met before deleting. With this, logical AND and OR operations are implemented.

4 Conclusions

The proposed solution for handling personal content based on different XML standards is straight forward and resource-effective introducing small overhead. Based on simple attributes, XML content can be easily extended enabling features like personal content identification and also merging, sorting, limiting and deleting.

As being part of the P2 public domain forum "*www.portable-personality.org/*" an open source library implementation using C# .NET will be available.

References

1. Crossley, M., Kings, N.J., Scott, J.R.: Profiles — Analysis and Behaviour. BT Technology J. 21, 56–66 (2003)
2. Newbould, R., Collingridge, R.: Profiling Technology. BT Technology J. 21, 44–55 (2003)
3. Zhiwen, Y., Xingshe, Z., Yanbin, H., Jianhua, G.: User Profile Merging Based on Total Distance Minimization. In: Proceedings of the 2nd International Conference On Smart homes and health Telematics (ICOST 2004), pp. 25–32. IOS Press, Singapore (2004)
4. Bruns, V., Reymann, S.: Development of a Middleware Solution for Consumer Profiling and Advanced Profile Distribution. Master of Science Thesis, Tampere University of Technology, Tampere (2007)
5. Namu Research Group, http://namu.cs.tut.fi/
6. Friend of a Friend Project, `http://www.foaf-project.org/`
7. Description of a Career, http://ramonantonio.net/doac/
8. Media RSS, http://search.yahoo.com/mrss
9. iTunes Library, http://www.apple.com/DTDs/PropertyList-1.0.dtd
10. Lehikoinen, J., Aaltonen, A., Huuskonen, P., Salminen, I.: Personal Content Experience - Managing Digital Life in the Mobile Age. Wiley-Interscience, Chichester (2007)

EMTV – A Component-Based DTV Middleware Extension for Educational Purposes

Juliano Rodrigues Costa[1,2] and Vicente Ferreira de Lucena Junior[2]

[1] Genius Institute of Technology, Av. Dr. F. Coelho, 64 São Paulo – SP – Brazil 05423-911
jcosta@genius.org.br
[2] Federal University of Amazonas; Ceteli – Electronics and Information Technology R&D Center - Av. Gen. Rodrigo Otávio, 3000 Manaus – AM – Brazil 69065-190
vicente@ufam.edu.br

Abstract. This work introduces a software framework called "Extended Middleware for Digital TV (EMTV)" which is suitable for the generation of interactive applications executed over digital television systems. Its concept was developed focusing on the Brazilian technological options for Digital TV. Technically, EMTV is a procedural GEM compliant application which, from the programmer's point of view, acts as a declarative middleware extension. The framework was developed to be component-based in order to minimize the need for programming knowledge to deploy the digital TV applications using EMTV. The main goal of the platform is to facilitate the construction of interactive educational applications, a crucial field for the Brazilian population. The concept is tested and validated by the construction of a Quiz application presented at the end of the paper.

Keywords: Interactive Digital TV; Educational Applications; Digital TV Middleware; Component-Based Software Development; Quiz.

1 Overview of the Brazilian DTV System

Brazil is at a point in time where an important technological decision will affect the life of 90% of its 184 million citizens who consider television as one of the most important sources of information and entertainment. This decision refers to the use of digital technology in the current process of transmitting and receiving open TV signals in the country, which has started in December 2007.

It is only in 1998 that Brazil started to research DTV technology and initially decided to develop its own standard which used to be called SBTV (Brazilian TV System) and whose main characteristic is the use of an OFDM (Orthogonal Frequency-Division Multiplexing) modulator system equipped with an artificial intelligence module on the reception side to make the receiver multipath-noise robust. Despite the good results achieved a few years ago, the Brazilian government showed clearly that it was giving in to the pressure of the TV content providers when it decided to adopt the Japanese ISDB (Integrated Services Digital Broadcasting) standard. At that time, the Brazilian committee renamed the standard to ISDTV [1] (International System for

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 219 – 228, 2008.

Digital TV) on account of some local contributions. As a matter of fact this model, already in use in Brazil since last December 2007, consolidates important aspects of the Japanese technology such as the use of the BST-OFDM (Band Side Transmission – Orthogonal Frequency-Division Multiplexing) modulation which is very effective against multipath noise even with fast-moving mobile receivers[1]. The Brazilian contributions are related to the adaptation of ISDB to use the MPEG-4 AVC standard, also known as MPEG-4 “Part 10” or H.264, which is an evolution of the MPEG-2 standard because it achieves compressed audio and video rates from 40% to 70%[2] higher than those of the MPEG-2. With such characteristics, and considering that, the Japanese standard committee is aggregating the proposed changes, the Brazilian model was recently defined to be referred to internationally as International ISDB.

Another Brazilian contribution to ISDB refers to the development of its DTV middleware specification called GINGA[2]. Just like other DTV middleware, it has a procedural part, the GINGA-J (Ginga–Java) [1], and a declarative part known as GINGA-NCL[1] (Ginga–Nested Context Language).

Figure 1 illustrates that the most important middleware specifications available nowadays has a procedural part, represented by the grey boxes, and a declarative part, represented by the white-dashed boxes. It also shows a tendency, the procedural middleware has to be GEM [3] (Globally Executable MHP) compatible. Despite of the fact that GINGA-J has not yet been officially released it will probably be a GEM implementation. GINGA-NCL is the declarative part of GINGA and it is considered to be very powerful and flexible as it not only controls appearance and the positioning of media objects but also considers the temporal relationships among them [**Error! Reference source not found.**].

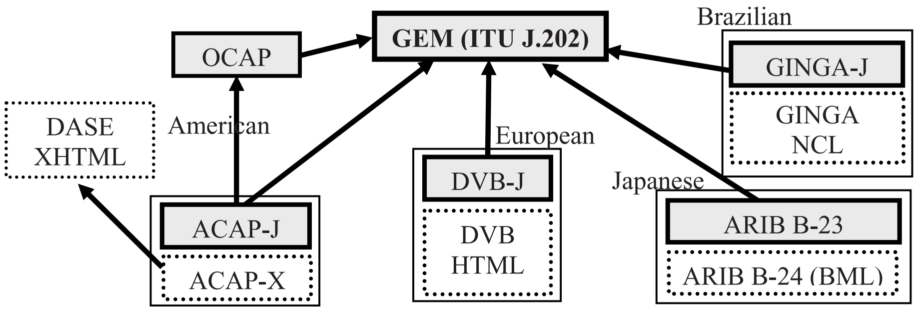


Fig. 1. Relationship between DTV Middleware Specifications

Terrestrial open DTV transmissions started in Brazil without any middleware support. According to the local industry, this was due to the need to the middleware to support the H.264 standard and also due to the wait for official GINGA specifications. Anyway, even without the support to any type of middleware, the technology used by

[1] HDTV (1920x1080i 16:9) in fixed terrestrial TV receivers. LDTV (320x240 4:3) in fast moving mobile receivers.
[2] Depending on the MPEG-4 profile and on the nature of the images.

the Brazilian system made both, the signal transmitter equipment and the set-up boxes very expensive, a fact likely to be a problem for the popularization of the technology in the country.

Nevertheless, the new interactive features of the digital system are expected to have great impacts on the population, the most significant being the use of this technology as a tool to contribute to educational processes. This paper introduces a component-based framework, named "Extended Middleware for Digital TV (EMTV)" [4], which was developed at a time when the GINGA was not even available for public download and whose main target is to help TV-content providers with no advanced programming skills to deploy DTV applications for educational purposes more specifically. The EMTV is free for any use and does not require any expensive tool since it thereby hopes to contribute in democratizing DTV technology in Brazil.

2 Used Concepts on the Extended Middleware

The development of complex DTV applications requiring the procedural approach is a relatively difficult task. Besides the logical concepts, the programmer must have borad knowledge of several software interfaces and must be able to build a very efficient software code due to the hardware memory and processing limitations [5]. The programmer also has to predict all necessary software responses to any user actions and system errors. A DTV application cannot, in any circumstances, force the user to reset the television set. This is why, although the procedural approach is powerful, it demands the professional services of an experienced programmer with software engineering capabilities.

The development of applications using declarative middleware [6], on the other hand, is simpler than the procedural applications, since its main functionalities are internally programmed. And so, programmers won't have to concern with most of exception handling once it's already treated by the middleware itself. The limited number of functionalities has the advantage of guarantying the simplicity despite of being less powerful. Declarative middleware only becomes more complex as the offered number of functionalities and flexibility increase. That is a problem most declarative middleware have nowadays.

Figure 2 shows that EMTV has an extra software layer to run over the available middleware in the STB, thereby extending its abilities and allowing it to be able to generate a specific family of DTV applications. In addition, EMTV is designed to be user-friendly, as the knowledge required on software libraries and the executor system

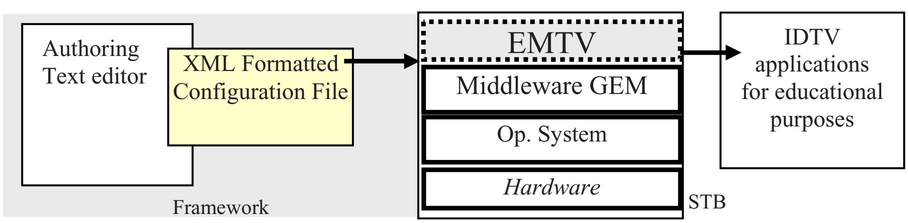


Fig. 2. EMTV framework

is reduced to a minimum. In terms of applications, this additional software layer assumes the middleware attributions once it keeps converting data information into interactive applications. The EMTV framework uses an external file which format is defined in the framework. The handling of this file demands the use of any third-part text editor referred to as authoring text editor in Figure 2.

2.1 Declarative Approach

To minimize the need of high qualified specialists on DTV software development, allowing program providers, more concerned with the content to be presented than to software programming, EMTV offers a declarative approach. It follows the same idea as GINGA-NCL but, is much more simplified and targets a very specific kind of application so that EMTV is able to provide just enough flexibility to deal with the main demands of a specific type of DTV application.

Although the programmer will view the platform proposal as declarative, it was developed using a GEM compatible procedural middleware. This choice is very convenient not only because GEM offers all the necessary library software [3] to develop any desirable feature but also because it allows EMTV to run over any compatible middleware, including MHP, ACAP, ARIB B-23 and especially the Brazilian GINGA-J which is the purpose of this work.

The decision to develop EMTV through the resources offered by GEM specification defines that the platform is in fact a Java XLet [14], meaning that its execution processes can be done just like any other GEM compatible application: The application will first be transported via the data channel or via the return channel[3] before being loaded and managed by the GEM middleware, and controlled by the DTV system or any user event through the remote control. At this point, the EMTV assumes its task as a middleware extension, loading text and image content.

2.2 Educational Purposes

Interactive applications have already been used in several countries where DTV technology has been implemented. As the range of applications increases, new terms like T-commerce, T-government and T-learning and others have been created to classify it according to its main purpose. The term "T-learning" [7] refers to the use of DTV technology in educational processes, one of the main purposes of EMTV. T-learning is viewed as the convergence between DTV technology and E-learning which basically refers to the use of computational technology for training or any other educational activity. It comes from "E-learning" the popularization of Quizzes as a very effective educational tool[4] [12]. A Quiz is a kind of test containing a series of

---

[3] The MHP 1.0.3 profile enables the XLet applications to be loaded directly from the return channel if available, which is connected to a TCP/IP network.

[4] Walldén and Soronen [11] wrote a paper about relationship between e-learning and t-learning which classifies education processes in four classes: Formal learning (leads to acknowledge diplomas), Non-formal learning (education from formal institutions but that does not provide official diplomas), Informal learning (education which comes from social activities) and Accidental learning also known as edutainment (education + entertainment: the knowledge or skill is acquired not intentionally).

questions along with some alternatives which the users can select according to their skills, knowledge or personal opinions. This principle allows Quizzes to be used not only in educational processes, typically in edutainment and informal processes, but also in other non T-Learning applications.

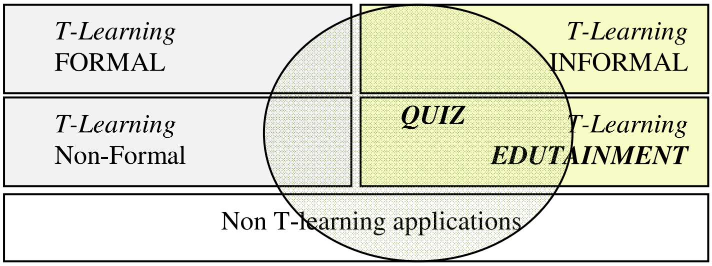


Fig. 3. Proportionality of Quiz Applications in the Universe of T-learning Applications

2.3 Designed to Be Configurable and Easy to Reuse

The need of an external file is already part of the declarative approach. It allows EMTV to be easily reused, as different configuration files generate different applications with a different content and behaviour. The EMTV configuration file is based on the XML format which furthers three main advantages at least:

- XML is easily read and written by both humans and algorithms.
- XML can be validated through DTD files.
- Several free XML editors can be used for edition and validation.

Another advantage of the use of the XML format for the configuration file is that it is very convenient to describe the software components properties. Software components are artifacts constituted by one or more instances of classes, independent enough to provide some interest functionality such as visual, behavioural or both. The behaviour and characteristics of a software component are defined by the properties it makes available through public methods which constitute the interface of the component. There are many good software-engineering proven reasons [13] to design components-based software. EMTV is component-based mainly because it contributes to an easy system reuse, maintainance and expansion.

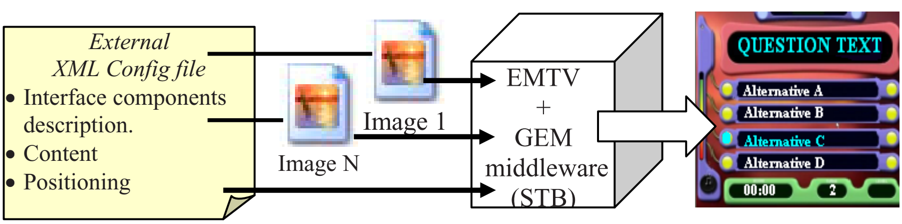


Fig. 4. The EMTV configuration files provides enough information to generate the applications

3 EMTV Components

The first version of EMTV provides only the minimal set of components to allow the easy deployment of DTV Quiz applications with the possibility to use or not a return channel. Those components were defined from a study of similar systems, such as MOODLE [9] which has a module to generate Quiz applications typically for the Internet, and commercial products for DTV such as Sofia Digital Voting and Quiz [10].

Fig. 5. Quiz application interface that helps to identify graphical components for EMTV

Figure 5 helps to identify some of the basic components to build a Quiz application. It leads to construct 5 basic visual components identified as: Application Screen, Application Text, Application Image, Application Image Button and Application Quiz. There is also the need of a non visual component called "Application Communication" to perform communications between EMTV and an external server. The user indicates the components in the XML configuration file which is read[4] by an EMTV Application Manager class that creates and controls the components according to the configuration file.

3.1 Application Screen

It is a singleton graphical component to represent the screen on which all other graphical components are placed. It is built using HAVI, DVB and JMF libraries and allows the programmer to control the size, position and appearance of the application. One of its attributes indicates if the screen background should be filled with an external picture, giving to the application an enhanced appearance.

3.2 Application Text

It is a graphical component to represent texts over the screen. It is made up by HAVI and DVB libraries but mainly by the instantiation of the org.havi.ui.HText HAVI Class. The attributes allow the programmer to control the text, position, font, font

[4] EMTV uses nanoXML library to be able to read the XML file attributes. This library is a very small and efficient XML parser for Java and can be found at http://nanoxml.cyberelf.be/

size, foreground colour, background colour and also has a field to indicate some basic animations as blinking and scrolling. Another field also allows the programmer to use some keywords to build a Boolean expression to controls the visibility of the component. The same principle is applied in the Text-content field as EMTV interprets some keywords which are related to information of the running application like the date and time, information about Questions Group application navigation and about the sending status of an Application Communication component. The programmer can place as many Application Text components as needed.

3.3 Application Image

It is a graphical component that is used to represent static images over the screen. It is made up by HAVI, DVB and JMF libraries but mainly by the instantiation of the Java Image Class. The attributes allow the programmer to indicate an external image and controls its position on the screen. The loading method of the image is done by JMF through the classes java.awt.MediaTracker and java.awt.Toolkit. Just like the Application Text, there is a field to control the visibility of the component. The programmer can place as many Application-Image components as needed.

3.4 Application Image Button

This is just a specialization of the Application Image component. It quickly switches between two defined images when the user presses on a key that is also defined on the remote control. This component is just useful to provide a graphical effect like a button if the proper pair of images is used.

3.5 Application Questions Group

This is a singleton component which enables EMTV to build Quiz applications. Its fields provide all the information to create a multi-page Quiz. Each page contains one question, represented by an HText Havi instance, and multiple alternatives for this question. The programmer can define the position, the foreground colour, the background colour of the question whose text is updated every time the user changes the question page through the remote control left and right arrows. The programmer can indicate a different number of alternatives for each question and if each question has a single or multiple choices answer. Fields control the initial position for the first alternative of all questions as well as the relative increment on the x and y axis for the next available alternatives. Other fields indicate the foreground colour for the alternatives, the foreground colour used as the user navigates through the alternatives, using the remote control up and down arrows, as well as a field to indicate an external picture which will be placed close to the chosen alternatives by pressing on the ENTER key of the remote control. Each alternative has special attributes which allows the programmer to define if the alternative is in fact an edit box where the user can input alpha-numeric, numeric or password characters when the alternative, by default represented by an org.havi.ui.HText is replaced by an org.havi.ui.HSinglelineEntry instance. This last situation is useful for the Quiz to get specific user information, like the username to be registered on an external server if an Application Communication component is used. In this case, an alternative can be converted into a text instruction for the edit box if the selectable attribute for an alternative is indicated as false. The component always inserts an additional configurable question which allows the user to indicate he/she has given the answers. This event disable the possibility of user change its responses and also starts the sending of the collected information in case of the presence of an Application Communication component. After this event the component changes the background color of all alternatives marked with the "isAnswer" attribute allowing the user to check his score.

3.6 Application Communication

It is a non-graphical singleton component which tries to establish a TCP/IP communication through a permanent or dial-up interface, if available, to send the information to an external TCP/IP server. Its fields provide all necessary information to connect EMTV to an external server. The data field replaces several reserved keywords into application specific information including the answers captured in an Application Question Group component and other useful pieces of information. This component is built based on org.davic.resources, java.net.Socket, java.io.DataOutputStream and java.io.DataInputStream classes.

4 EMTV Tests and Validation

EMTV was tested and validated at the Ceteli DTV laboratory of UFAM[5] as shown in Figure 6. The process used an external Apache server connected to the same TCP/IP network as an interactive profile MHP Set-top box (STB). The server is used to provide the EMTV packet application as requested by the STB. Once loaded in the STB, the EMTV requests the server the configuration file and all necessary picture files to generate the Quiz to be presented on the HDTV monitor. The user interacts with the application

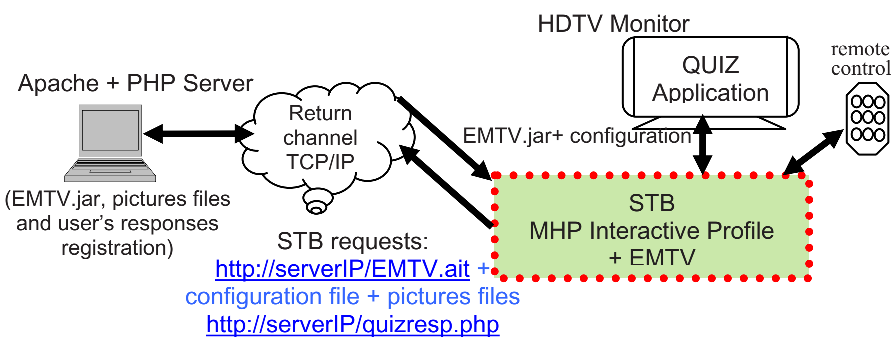


Fig. 6. EMTV tests schema

[5] Ceteli is a R&D center of Federal University of Amazonas (UFAM) in Brazil.

though the STB remote control and once he/she has finished, the information gathered is sent to an external Apache/PHP server. In the case of this test, the server used to provide the application was also used to register the users' responses.

Figure 7 illustrates a Quiz application generated by EMTV:

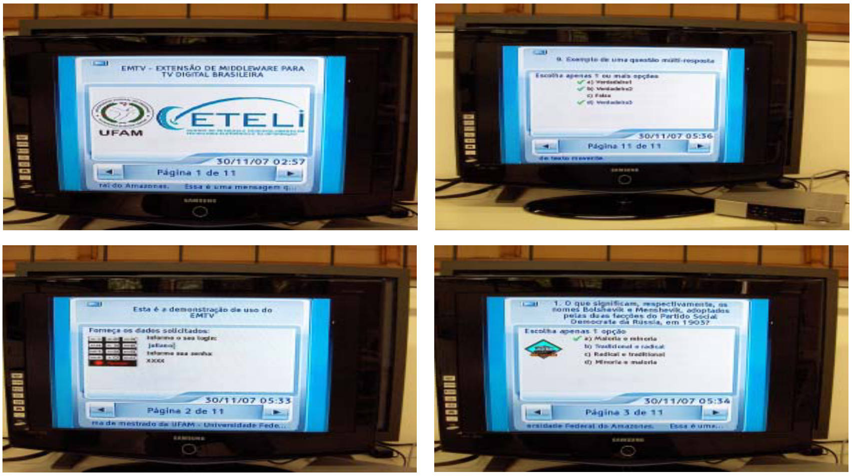


Fig. 7. Pictures of a Quiz application generated by EMTV running in a real DTV system

The same results could be achieved loading the EMTV application directly from the DTV data channel. To validate EMTV on this case would be necessary, besides the proper hardware equipment to generate the MPEG-2 Transport Stream signal and the DSMCC (Digital Storage Media Command and Control) system, to correctly setup the PMT (Program Map Table) and the PAT (Program Association Table) so the STB can detect and load EMTV. This test would also require a minimal change on the software so EMTV would be able to download any external files through the DSMCC synchronously or asynchronously.

5 Conclusion

Although the EMTV framework is still under development, it is very simple and already capable to generate a series of useful applications for educational purposes as Quizzes. It has the advantage of neither requiring any broad knowledge on DTV systems, nor the purchase of expensive proprietary tools, and thereby encourages the popularization of a generation and use of interactive DTV software.

This first EMTV version is only about 70 Kbytes in size since it's based on the use of GEM classes and it is available for free, for any use, by sending an e-mail request to the authors. Future work will include the development of a graphical authoring tool and the development of new graphical components besides of the improvement of the existing ones.

References

1. Soares, L.F.G., Souza, G.L.: Interactive Television in Brazil: System Software and the Digital Divide. In: EuroiTV (2007)
2. Middleware GINGA (2008), http://www.ginga.org.br/
3. MHP Organization (2008), http://www.mhp.org
4. Costa, J.R.: EMTV – Extensão de Middleware para TV Digital com Componentes de Software para Educação (2008)
5. Interactive TV (2008), http://www.interactivetvweb.org
6. César, P.: A Graphics Software Architecture for High-End Interactive TV Terminals, Helsinki University of Technology (2005) ISBN 951-22-7888-X
7. Bates, P., Atwere, D.: Interactive TV: A learning platform with potential, Learning and Skills Development Agency (2003)
8. Jones, J.: DVB/MHP Java TV™ Data Transports Mechanisms (2002)
9. Moodle Course Management System Home Page (2008), http://moodle.org/
10. Sofia Digital Voting and Quiz (2008), http://www.sofiadigital.com
11. WalldÉN, S., Soronen, A.: Edutainment. From Television and Computers to Digital TV (2004)
12. Johnson, K., Hall, T., O'Keeffe, D.: Generation of Quiz Objects (QO) with Quiz Engine Developer (QED) (2005) ISBN: 0-7695-2385-4
13. Heineman, G.T., Councill, Book, W.T.: Component-Based Software Engineering, 1st edn. Addison-Wesley, Reading (2001)
14. Java TV™ (2008), http://java.sun.com/products/javatv/index.html

DVB Service Security—A Problem in Development of Multi-service Television

Zbigniew Hulicki

AGH University of Science and Technology, al. Mickiewicza 30,
30-059 Kraków, Poland
hulicki@kt.agh.edu.pl

Abstract. The paper aims to discuss the issues those concern secure service delivery on DVB platform. Although the infrastructure exist, the end-users are still not fully satisfied with development and proliferation of new TV services. Because traditional DVB network security model has a partial nature and provides only a certain level of security, it is difficult to offer new multimedia services with assurances that the required security level will be held. This restrains successful deployment of new content delivery services on digital TV platform. To solve the problem, technical capabilities of the DVB based infrastructure have been explored in terms of service implementation aspects and potential evolution of services in the future. The paper does propose to use comprehensive, multi-dimensional measures necessary to secure the whole distribution way of a content. Such approach, assuring a given security level, seems to be essential for the successful service deployment in a new, personalized multimedia environment.

Keywords: DVB, content delivery services, network and service security.

1 Introduction

During the last decade the TV (television) infrastructure entered a new era in service to the consumers [17]. DVB (Digital Video Broadcasting) – one of the most perspective technologies – can be used both to provide efficient communication and support creation and delivery of various multimedia applications [11]. Simultaneously, because of the explosive growth of Internet and service demands of the users, bound up with the development of transmission and switching technologies, the new diffusive and interactive services could be offered to the customers. Unfortunately, worried about the number of customers, the market players (network operators and service provides) are afraid of investments and for lack of attractive, useful business models they keep themselves from offering new services to prospective users [18]. In consequence, the customers could be disaffected from the service offering.

Nowadays, it seems to be obvious that development and proliferation of services is strictly related to the involved security measures [9], [12]. One of the major issues in the provision of the content-based services is the protection of copyright [3], [14]. Depending on the dissemination infrastructure different approaches are today available providing a certain level of security. Although a lot of work has been done

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 229 – 237, 2008.

in implementing such mechanisms as watermarking and fingerprinting, there are still a lot of gaps in the field of copyright protection [1], [16]. In addition, the convergence of various applications and availability of interconnection technologies enable a migration towards digitally-networked devices in users' homes. This is not only a new opportunity but it can be also considered as a possible threat to service security [2], [4], [13].

In case of existing DVB systems security has been provided at the data link layer without recourse to additional measures [5], [6]. Because network operators are under constant pressure to develop and launch both innovative and profitable new services, while faced with deploying new technologies to support those services, in 2000/2002 the DVB consortium standardized an open multifunction platform for interactive DTV (digital TV) and multimedia services [7]. This tool, called the MHP (Multimedia Home Platform), enables users to move seamlessly between these domains. The MHP security model ensures however, that functions raising security concerns can only be used by authenticated MHP applications [10].

Various requirements imposed by multimedia communications can have consequences on communication system design and its development [8], [15], [22]. The requirements will include not only *who* has access to *what* content, but also *when* and from *where* the content is accessed [19]. Hence, the evolving market of multimedia communications is in need of a system which will allow content distribution with assurance the producer's rights, collection of revenue, and which will be flexible enough to accommodate new marketing policies and value added services.

Because traditional DVB network security model does not go far enough to address the requirements of content delivery services and become obsolete in a personalized multimedia service environment, the objective of this paper is to consider the issues those concern secure content delivery on DVB platform and seem to be crucial for development of the comprehensive security framework. It has been proposed to use multi-dimensional measures in order to secure the whole distribution way of a content, what is essential for successful deployment of the interactive multimedia services on DVB platform.

2 Infrastructure for the New Multimedia Services

One of the most important and insightful technologies for providing a personalized service environment is DVB. Its technical capabilities can be used to support the integration of the interactive multimedia and DTV services and meet user demands in the future.

The indispensable infrastructures for transporting DTV and/or multimedia services include both wire and wireless broadcast networks, and not only traditional TV network carriers (terrestrial, satellite, and cable) try to be able to offer such services [10], [18]. Currently, the major way for distributing complex interactive multimedia services, including DTV, is the broadcast of radio signals. The satellite TV broadcasting is a relatively simple arrangement, but the reception of DTV services calls for the installation of subscriber decoders (set-top-boxes). In some countries however, inclusion of a digital tuner in all new TV sets became mandatory for the manufactures [10]. On the other hand, one can use PC to select and receive TV programs, even though it has a smaller screen and its life is short [8]. Although computer and TV worlds are still deeply different, one can expect that in future TV sets will access to some computer services and a universal and cheap PC architecture will enable provision of TV and high speed multimedia services.

Aware of a successful deployment of interactive services and stimulating content distribution, DVB has specified the DVB-RCT (Return Channel Terrestrial) system [7]. That system offers a wireless interaction channel for interactive TV and multimedia services, even in the congested UHF/VHF bands.

Until recently, a satellite version of the return channel seemed to be not attractive. However, the standardization process of DVB-RCS (Return Channel Satellite) system [7], its practical approval, and decreasing costs of satellite technologies allow rapid implementation of DTV and multimedia services in a cost-effective way. Now, DVB-RCS system becomes a strong competitor in the satellite coverage area.

Today, the use of telecommunication networks for transactions and data interchange is very common and widely accepted. On the other hand, Internet has become a real multimedia network which is used for real time teleconferencing and broadcasting of radio and video. Hence, the combination of television and the Internet leads now to powerful new applications built around two-way interactive technologies, with limitless possibilities of great commercial potential.

3 Implementation Aspects of Service Providing

The best solution to solve the above mentioned problems seems to be the broadband communication via satellite and/or terrestrial TV channels. It enables broadcast transmissions of large amount of data and it has a number of features (e.g. simple network topology and management etc.) [17].

Based on the object and content of services aimed at a residential audience, the palette of interactive multimedia services involves: information and education services, services for households, and leisure services. In [18] all these categories of the interactive services have been discussed in details, i.e. from the bandwidth requirement point of view, some services can be seen as strictly asymmetrical (broadcast) while others as symmetrical (conversation). Besides, according to the needs for interactivity we have also uni- and bi-directional services [11], e.g. an access to the resources of the Internet on DVB platform means a possibility of tuning TV receiver into the "*WWW* channel" and individual browsing the network resources in real time. In such system, transaction server is responsible for service of user demands and moves user URL addresses to the Internet gateway. The gateway fetches appropriate content, converts its own format (e.g. HTML) to the interactive TV format (i.e. MPEG), and then the application server puts it into the sending server, which multiplexes received data streams into the MPEG Transport Stream. Such architecture enables network operator to provide (in a similar way) also other type multimedia services.

A number of Internet applications, those require credibility of data, can be transmitted over a satellite and/or terrestrial medium without any additional changes. However, the basic problem in satellite communication is a delay, i.e. the price for credibility is paid by a high sensitivity for delays. Many of such applications require the response time as short as possible and their performance depends on implementation of TCP protocol [11].

With the rapid adoption of digital video technology in the cable, satellite, and terrestrial broadcasting, the foundation is set for the creation of an interactive TV and multimedia segment that introduces to a mass consumer market a whole new range of possibilities in sound economic principles. Available technologies support interactive features for advertising, e-mail, game shows, sports and customized channels filled with favorite movies and/or content driven or value added services. Rather than concentrating just on *Web* services, the goal is to deliver *enhanced* television experience [18]. An important component of such services is the hard-drive-based PVR (Personal Video Recorder) at the user's receiver.

The digital content distributed in a compressed format (MPEG-2) can be accessible at the customer premises via diverse electronic devices with SIM (Subscriber Identification Module) Smart Card for service management. The functions required for multimedia content distribution are variable but can be addressed in accordance with bandwidth, interactivity, and subscriber management [12]. All devices can be connected by an IEEE 1394 high speed home network bus or similar interconnection system.

CA (Conditional Access) methods and DRM (Digital Rights Management) systems allow service providers and network operators to control who uses the content and how much services customer is entitled to, i.e. provide the services according to their business paradigms. Nevertheless, the era of digital services introduces new challenges to content developers and service providers, i.e. any scheme protecting the service or content must also take into account the perceived rights of the users and the ease with which they obtain, view, store, and continue to use the services or media they purchase. Therefore, the trade-offs between system's flexibility on one hand and complexity and cost on the other are always difficult to decide.

4 Security on DVB Platform — Threats and Challenges

In a framework of the future broadband communication scenario specific emphasis should be placed on critical issues dealing with a security of the whole distribution way (i.e. from the information source to the destination) of a content from unauthorized access. One can assume that only a content will be kept private, however in some cases one needs to secure both, i.e. the distribution list of some content and content itself. Therefore, the issues those concern secure content delivery on a given platform (e.g. DVB) should be analyzed together with an impact of service attributes on system architecture.

In case of DVB platform security is intended to protect the user identity including its exact location, the signalling traffic to and from the user, the data traffic to and from the user and the operator/user against use of the network without appropriate authority and subscription. Although DVB standards include three levels of security those can be applied to the different layers [7], i.e. DVB common scrambling in the forward link, required by the service provider; individual user scrambling implemented at the section level in the forward and return link; IP or higher layer (application specific) security mechanisms can be used optionally by the service and/or content providers. Besides, even though the user/service provider could use its own security systems above the data link layer, existing DVB systems provide a security at the data link layer only so that the system is inherently secure on the satellite section without recourse to additional measures. Also, since the satellite interactive network forward link is based on the DVB/MPEG-TS standard, the DVB common scrambling mechanism could be applied. Service providers however, do not consider it necessary because such mechanism would just add an additional protection to the entire control stream for non-subscribers [7]. However, it makes one wonder if such approach is reasonable from the service provision point of view, i.e. in future one can not exclude that increasing user demands and a necessity to provide various services (or service bundles) at the given security level will impose severe discipline on the use of security mechanisms.

DVB security is defined at the link level using classic cryptographic handshake to prevent eavesdropping on the communication between two secure devices (units). Because it is the PKC (Public Key Cryptography) system, the signature not only provides validity but it also uniquely identifies the signer. Unfortunately, there are some weaknesses of a public key system those seem to be crucial to DVB, e.g. because cryptographic handshake is performed at the link level, secrets might be easier to find, and it might be also easier to find other interfaces where content is unprotected (e.g. between the content and service providers [16]). Besides, an overhead introduced by PKC is of great importance in DVB, at least from the user devices and a return channel point of view. Hence, in the last two years, there has been much discussion in that research area. On the other hand, it seems to be obvious that the guaranteed security is prerequisite and essential in public networks, however it will be insufficient if we consider the whole distribution way of the content, i.e. also in home networks and its possible redistribution.

Except of providing authentication and data encryption for the digital connection of consumer electronics devices (DVD, DTV, STB), the content and service providers need a more general approach that protects content during transmission over public networks, in customer premises networks, and while stored on removable or fixed media. Existing solutions do not completely address that problem.

One of the main issues in successful provision of TV and multimedia services is also security of smart cards used for service subscription, customer management and other functions in finance-related services. As data security is provided by the card itself, off-line operations and various transactions can be safely carried out. This opens up the number of applications, with the well-known example of pay TV being the most important.

If security features are not an integral part of system infrastructure, various threats can occur, showing the points (targets) of potential attacks. Depending on a type of DVB system and services offered to the customers, these threats can include [12], [20]: servers, network equipment, and end-user devices (TV receiver, PC, etc.). Because a PC can easily emulate the operation of any equipment, and both the MAC and IP address can be modified '*on the fly* ', most of attacks can be mounted from a PC platform. Such attacks can exploit weaknesses in the IP stack and could result in a control over the remote equipment. Then, a *denial of service* attacks against the service delivery infrastructure can be issued from the compromised device [21]. Moreover, if the compromised TV receiver (IRD) has sufficient local storage, it can be also used as part of a peer-to-peer infrastructure for distribution of illegal versions of the premium content. Therefore, the key problem with the end-user equipment is the client software, which, in most cases, has not been designed to defend against intentional misuse or clever malicious attacks that try to make the software fail.

The business model of DVB services and the associated delivery infrastructure will also subject to a number of business threats, which can occur due to attacks to the delivery infrastructure. For the network operator and service provider these attacks can result in various losses, including a loss of service availability (for some time), its perceived value (if the service could be obtained illegally), a loss of customer's respectability (if the theft of viewing patterns and user's account details would occur), and as a result a loss of revenue. The risk of such threats is very real [21]. In future, most of enhanced TV services will be based on a consumption related business model (user charged on the service usage). Therefore, the removal of such model (e.g. due to non-repudiation attacks) will significantly reduce the revenue generated from new multimedia and TV services. Hence, a set of procedures and mechanisms is needed to secure DTV infrastructure in delivery of enhanced services.

5 Issues and Mechanisms for System Security

The exposure or alteration of sensitive information, together with financial theft, are particularly serious for businesses based on communication and content delivery networks. After the security breaches, network systems may have limited or reduced service or be temporarily unavailable. In many cases it can be very difficult or impossible to determine precisely what subtle damage of sensitive information the attacker left behind [20]. Therefore, content and service providers as well as network operators are highly susceptible to a loss-of-confidence crisis. Hence, it becomes necessary to adjust security architecture decisions in accordance with unforeseen changes in technical attack methods or technology advances.

Entities involved in the provision of services and their consumption (end users) can perceive security in a quite different way, e.g. confidentiality and integrity of a content together with non-repudiation are of great importance for service providers. On the other hand, network operators attach more importance to both authorization and encryption (confidentiality), whereas the customers consider privacy as the most significant security feature. Hence, the need for protection from various losses associated with mentioned security breaches has been fanning a demand for more secure environments in many different areas [2].

The increased worldwide interest in individual privacy has motivated researchers to search for the new cryptographic tools. One of many possible solutions that provides just as much protection as PKC system seems to be broadcast encryption [16]. It hides device keys deeper in the software, near the point of content consumption. Broadcast encryption is very fast and content protection is well suited to it, but it is impractical to use it for transaction services such as electronic fund transfers.

Given the tradeoffs between the PKC and broadcast encryption, it is obvious that the latter method is not the optimal choice for all applications but in some is reasonable and well suited (e.g. content driven and enhanced TV services).

If we consider any service scenario in a system implemented on DVB platform, it seems to be obvious that the broadcast delivery media network can suffer mainly from passive attacks such as interception and release of the message contents. Whereas passive attacks are difficult to detect, measures are available to prevent their success [20]. On the other hand, both attack types, i.e. passive and active, can be launched in the interaction network of that system. These active attacks can involve some modification of the information that is being transmitted, message interruption and /or its fabrication (creation of a false stream). Unfortunately, it is quite difficult to prevent active attacks absolutely [15], since to do so would require physical protection of all communication facilities and paths in the interaction network at all times, what is practically unfeasible. Such vulnerability could have a crucial impact on DVB system operation because all users' information and management data are transmitted over interaction channels.

Any effort to breach network security can be viewed from the perspective of means, motives, and opportunity to commit the crime. However, weaknesses of traditional single-dimensional approaches to the network security consist in their piecemeal defensive nature, that based primarily on preventive mechanisms, such as firewalls, anti-virus software, encryption, or other access control mechanisms. It is obvious that properly secured servers and networks make content information available to the customers and business partners (i.e. co-operating service providers), and good security now is better than perfect security never.

Some mechanisms and tools (e.g. *service monitoring* or *Web access management*) can be used in future to provide security of the interactive services on DVB platform. Unfortunately, the network security requirements have consequences on communication system design, i.e. network security and system usability are often inversely proportional [15]. Therefore, it is important to concentrate on real and probable threats, reducing and managing risk, in order to achieve fairly good security while still allowing system productivity.

A single-layer security system contains only one type of defense and it does not ensure comprehensive security. Hence, in order to create as comprehensive a security system as possible one should use multidimensional security mechanisms and multilayer defenses. Such approach should protect information from unauthorized modification, destruction, or disclosure, whether accidental or intentional. However, perfect security of the network seems to be a dream, and even if it would exist, usability of the networking systems would be questionable [12]. Nevertheless, one should try to create as comprehensive a security system as possible. It is not a trivial task because a single-layer security based on preventive mechanisms is not effective [2], and if complex, multidimensional, multi-layer approach is used, the trade-offs between maximum security and flexibility on one hand, and usability and costs on the other, are always difficult to decide, since they have an impact not only on the initial deployment of a system, but also on its future evolution and market acceptance. A network might contain internal and external security domains, i.e. in terms of the physical layer the internal domains reside within a physically protected area that can only be accessed by authorized personnel, whereas an external security domain may reside in publicly accessible areas. The access-control mechanism constitutes a desirable foundation on which higher security layers can be built or added on DVB platform. Hence, regardless of the hard problem to solve, properly arranged multidimensional security mechanisms, together with multi-layer defenses ought to be used, because it is only the way to prevent, detect, and respond to sophisticated breaches of network security.

Network security seems to be an ongoing, iterative, technical, and business process that include a drawing up of security policy, an analysis of security risks, development of the target security architecture, purchase of security products, and monitored network operations. Therefore, the multidimensional security measures will comprise at least three sets of mechanisms, i.e. preventive, detective, and responsive measures. All those defenses have been discussed in detail in [12].

In order to ensure a given security level of the system, in addition to the mentioned security mechanisms one should also deploy multi-layer security devices throughout a network. Because network systems share information with other network-connected devices, all defense layers also require security protection, detection, and response procedures. Moreover, the same principles can be applied to both wireless and wire networks, including DVB and other broadcasting systems.

6 Conclusions

The issues those concern secure content delivery on DVB platform have been discussed in the paper. They have been examined in terms of service provision and their potential evolution in the future. It has been stated that traditional DVB network security model has a partial nature and, depending on the dissemination infrastructure, DVB platform provides only a certain level of security. Hence, it does not go far enough to address the requirements of new content delivery services. Because the traditional single-dimensional security approaches are no longer adequate and they can create a false sense of security, one has to apply the multi-layer framework, comprising multidimensional security mechanisms to be arranged in an incremental and continuous fashion. Multidimensional approach seems to be mandatory because different services corresponding to various user profiles may require diverse security levels. The solutions (i.e. new services to be offered) without guarantee of the required security level might not be accepted by the customers (end users), and hence, they will not produce a revenue for the network operators and service providers, what is essential to cover capital expenditure in the infrastructure as well as for further investments in the development of new services. This appears the main reason that restrains the audiovisual market players from doing investments in the infrastructure and development of both the existing and new services. Therefore, multidimensional security approach seems to be indispensable for successful and profitable implementation of new multimedia services.

References

1. Bancroft, J.: Fingerprinting: monitoring the use of media assets. In: Proc. IBC 2001, Amsterdam, pp. 55–63 (2001)
2. Cass, S., Riezenman, M.: Improving security, preserving privacy. IEEE Spectrum, 45–49 (January 2002)
3. DAVIC. Digital Audio-Visual Council 1.4 specification (1998), `http://www.davic.org/`

4. Desmond, J.: eSecurity: a high priority. BusinessWeek (February 2002)
5. DVB/European Standards Institute. Support for use of scrambling and Conditional Access (CA) within digital broadcasting systems. ETR 189 (1996)
6. EBU: Functional model of a conditional access system. EBU (European Broadcast Union) Technical Review, winter 1995, pp. 64-77 (1995)
7. ETSI: ETSI EN 301 790 V1.2.2 (2000-12): Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems (2000)
8. Furt, B., et al.: Multimedia broadcasting over the Internet: part II — video compression. IEEE Multimedia, 85–89 (January-March 1999)
9. Ghosh, A.K.: Addressing new security and privacy challenges. IT Pro, 10–11 (May-June 2002)
10. Huffman, F.: Content distribution and delivery. In: Tutorial, Proc. 56th Ann. NAB Broadcast Eng. Conf., Las Vegas, NV (2002)
11. Hulicki, Z.: Integration of the interactive multimedia and digital TV services for the purpose of distance learning. In: Proc. IBC 2001, Amsterdam, pp. 40–43 (2001)
12. Hulicki, Z.: Security aspects in content delivery networks. In: Proc. 6th World Multi conf. SCI 2002 / ISAS 2002, Orlando, FL, pp. 239–243 (2002)
13. Kamperman, F., et al.: Digital rights management in home networks. In: Proc. IBC 2001, Amsterdam, pp. 70–77 (2001)
14. Keating, S., et al.: The use of watermarking for managing metadata in content production. In: Proc. IBC 2001, Amsterdam, pp. 46–54 (2001)
15. Liu, S., et al.: A practical approach to enterprise IT security. IT Pro., 35–42 (September-October 2001)
16. Lotspiech, J., et al.: Broadcast encryption's bright future. IEEE Computer 35(8), 57–63 (2002)
17. Pagani, M.: Multimedia and Interactive Digital TV: Managing the opportunities created by digital convergence. In: IPG, Hershey, PA (2003)
18. Pagani, M. (ed.): Encyclopedia of Multimedia Technology and Networking. Idea Group Inc, Hershey (2005)
19. Rodriguez, A., et al.: File security and rights management in a network content server system. In: Proc. IBC 2001, Amsterdam, pp. 78–82 (2001)
20. Stallings, W.: Network security essentials: applications and standards. Prentice Hall, Upper Saddle River (2000)
21. Thanos, D., et al.: A model for the commercial dissemination of video over open networks. In: Proc. IBC 2001, Amsterdam, pp. 83–94 (2001)

Where Have You Ended Up Today? Dynamic TV and the Inter-tainment Paradigm

Rossana Simeoni[1], Marina Geymonat[1], Elena Guercio[1], Monica Perrero[2], Amon Rapp[3], Francesco Tesauri[4], and Roberto Montanari[4]

[1] Telecom Italia, Technology, Research and Trends department
Via G. Reiss Romoli, 274, 10148 Torino, Italy
{rossana.simeoni, marina.geymonat, elena.guercio}@telecomitalia.it
[2] Department of Computer Science – University of Torino C.so Svizzera, 185, Torino, Italy
perrero@di.unito.it
[3] Department of Computer Science – University of Torino C.so Svizzera, 185, Torino, Italy
Telecom Italia, Technology, Research and Trends department
"Progetto Lagrange – Fondazione C.R.T"
amon.rapp@guest.telecomitalia.it
[4] Department of Science and Methods for Engineering, Human-Machine Interaction Group, University of Modena and Reggio Emilia, Via Amendola, 2, 42100 Reggio Emilia, Italy
{montanari.roberto, tesauri.francesco}@unimore.it

Abstract. Traditional TV was based on a time-dependent and passive paradigm of use: the availability of audiovisual contents was pre-defined by a rigid scheduling, and user's role was to choose among limited alternatives. A potential for active TV experience is now available: users should be allowed to access contents that fit their attitudes at any time, while being encouraged to discover new domains of interest. This paper outlines the basic elements of the new Inter-tainment paradigm, which builds on an active role of the user and on attitude-centred fruition. Successful applications of the new paradigm require flexible and reconfigurable structures of navigation: the development stages of an Inter-tainment system are reported here.

Keywords: Interaction Design, Social TV, Long Tail Theory, User Experience, Usability.

1 Introduction

New broadcast technologies as IPTV, DVB-H and DTT have modified the market of television services under several sides. Changes could be roughly divided between qualitative and quantitative. On the one hand, the number of possible contexts for watching TV is increased, expanding the user experience beyond the TV set: contents are available through PCs and mobile phones, shifting the TV experience towards personalized services. On the other hand, the amount of available contents has dramatically increased, mostly addressing market segmentation through thematic channels. Moreover, and most notably, TV consumption is losing most of its original

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 238–247, 2008.

bound, that is time-dependency: as new contents add up to existing and ever available archives and are equally available at any time (besides live ones, of course), the most important question is no longer "when and on which channel?", but "what and why?". In spite of this evolutionary rush, the navigation structure of current TV systems appears to disincentive and hinder the discovery of new contents of interest: studies (e.g. [1, 2]) suggest that the traditional navigation on current interactive TV systems do not improve user's experience within a multimedia environment. According to some authors (e.g. [3].) today's "interactive television" doesn't show a real interactivity, and should more properly be defined "enhanced TV – e-TV". Several attempts were made to improve the user experience: for instance, web TVs propose new paradigms of navigation that combine television zapping with web surfing, as showed by Babelgum [4] and Joost [5] – These innovations apply to PC only and are not adequate for a relaxed TV experience.

Other studies combine the opportunity of interactive television systems to realize a sort of social television. On the one hand, they use television as a simple media that supports remote communication, allowing users to talk and share contents (pictures, movies) with relatives and friends that live faraway [6], [7]. On the other hand television contents become drivers for a new form of sociability: on one side interactive television is a "virtual couch" that allows distant people to communicate during the vision of a program [8], on the other side it supports the creation of some kind of social networks [9].

A complete solution to the problem to have benefits of the previously described features in the TV context seems still to be faced and studied in depth.

Taking into account the above issues this paper proposes a new Interactive TV paradigm allowing users to navigate among large amount of contents and encounter unpredicted and probably interesting content. This result could be obtained by developing a dynamically-reconfigurable navigation structure able to provide navigation paths and suggestions to discover previously unexplored contents.

Some general design issues had been dealed with in a previous study [10], in which one of the key feature of this new entertainment paradigm appeared to be the capability to encourage serendipity, i.e. finding valuable content while performing an apparently unrelated search. In the following the first steps towards the definition of the paradigm are reported, aiming to stress in which way navigation interfaces could favor an active and enjoyable user experience.

2 Today and Tomorrow's TV: The Inter-tainment Paradigm

Television watching is traditionally thought of as passive fruition of pre-selected and pre-organized contents. TV spectators' role is typically limited to selecting sources of information and entertainment contents (i.e. zapping TV channels), following pre-determined content policies and time scheduling. This consumption habit still dominates large part of the audience [11]. Novel ways for selecting and consuming multimedia contents have become available with the Internet age, through systems allowing users to select and acquire contents independently of third parties' scheduling and content policies (e.g.: file sharing software); in this case, users are focused on actively extracting contents of interest from undistinguished complexity. In latest years, a huge number of technologies, generically referred to as Recommenders, have provided more and more refined ways of suggesting contents of interest to users, by inferring their tastes from previous behaviour and explicit preferences. Two well known examples are the Amazon-like (people who bought this, also bought…) and the Netflix-like (if you liked these movies, you will also like ….).

As a whole, a substantial dichotomy can be detected in today's domain of multimedia consumption. On the one hand, traditional TV audiences still choose what to watch from pre-defined schedules, being able to choose anything that meets their interest. Their fruition is *limited by contents*. On the other hand, internet users have way of collecting any kind of content at any time, but they have to find interesting items among unlimited and ever increasing amounts of irrelevant ones; recommenders help them finding what they already like, but not what they may like and differs from their previous choices. Their fruition is therefore *limited by search tools*.

A consumer technographic survery carried on by Forrester Research [11], revealed how european population (adult on-line consumers) uses New Media: it appears that non active users are about 54%.

A recent ethnografic study carried on in our labs shows that part of the pleasure related to these novel ways of media selection and consumption resides in the huge quantity of contents to choose from, stimulating users' curiosity and appetite.

These evidences suggest there is room for a more rewarding form of television, that starting from the current channel zapping paradigm may allow the 46% of non-inactive users to enrich their viewing experience.

So far, content overload doesn't seem to be a problem: for the business coming from the Long Tail Theory [12], direct access to content through PC oriented metaphors seems to be enough. Nevertheless, if we focus on TV, some big questions are still unanswered: what about using a keyboard while relaxing on a sofa? What about zapping through data popping out as lists and text menus on TV? Some of these issues were covered by [13] and have now been more deeply investigated.

Three main attitudes related to TV usage were drawn from the previous evidences, describing users' needs and desires in specific moments of their lives:

1) Relaxed attitude: when a user is interested in watching a content
2) Curious attitude: when a wants to "surf on the surface", watching a few minutes of many contents, possibly related one another;
3) Participant/relational attitude: when a user is willing to be part of the system, sharing his/her knowledge about the contents, rating or reviewing it.

This paper introduces a new paradigm for Multimedia content fruition, aiming to join the benefits and overcome the limits of each approach, offering a rewarding experience for users in all the three attitudes, with particular attention to the first two.

That means to free traditional TV spectators from content limitations and to allow active media users to fully exploit the potential of unlimited content and powerful exploration tools. We called this paradigm Inter-tainment (Interaction + Entertainment), as a statement of an active and enjoyable role of the user in choosing and consuming TV contents, still allowing traditional users to keep their long-acquire watching habits.

The Inter-tainment paradigm encompasses and merges three different concepts, i.e. enhanced TV, Recommendation Technologies and Serendipity. Enhanced TV provides interaction models derived from traditional TV, particularly as to control devices (e.g.: multifunctional remote controls), allowing any users to easily interact. A number of different recommendation technologies can be used and merged in order to guide both traditional and active users in retrieving contents that match their preferences, though taking into account serendipity. The latter represents the strongest driver for the success of the Inter-tainment approach: by definition, it is the condition by which one accidentally discovers something worthy, especially while looking for something else[1]. Applied to the Multimedia content domain, Serendipity means providing users with Search/Navigation, Visualization and Fruition strategies attracting users towards the exploration of unknown content. According to the Inter-tainment paradigm, not only watching, but navigating among content previews is part of the fun. The Inter-tainment paradigm is defined by three basic features:

-navigation among contents can be engaged at any point of the viewing experience;

-navigation is conceived as exploring galaxies of contents and the connections among them;

-exploration, choice and consumption are not isolated actions, but a contextualized moment in the overall flow.

3 Dynamic TV: An Interaction Model for Inter-Tainment

Dynamic TV is a concrete attempt of implementing the Inter-tainment paradigm approach. This interaction model has to be appealing to both traditional and active users and be shaped in an intuitive navigation metaphor, tailored around users' fruition habits and familiar interaction patterns. A state of the art analysis of eTV and p2p software applications was led, defining a set of interface design drivers. The aim was to merge the interaction modalities of traditional and widespread entertainment systems with the new approach to content access adopted in new media. As a whole, a new navigation metaphor needed to be defined, to allow different navigation strategies. Possible approaches range between two extremes: on the one hand, the traditional categorical approach recalls the structure of contents arranged in thematic channels; on the other hand, a new serendipitous approach requires contents to reconfigure dynamically whenever the user asks to be advised according to his/her taste, or wants to discover new contents. Dynamic TV's navigation metaphor is designed to provide users with ways of following both approaches,

Guidelines for developing such flexible navigation metaphor:

- The interface design should encourage the "serendipity"[14], i.e. users should be provided with ways to find unpredicted, yet interesting available contents starting from previous search results;
- Users should be provided with suggestions about the available contents, according to a user-targeted recommendation engine;
- As counterbalance to the previous guideline, the system should be perceived as non-intrusive, especially as to the provision of recommenddations and suggestions;

[1] Wikipedia reference: http://en.wikipedia.org/wiki/Serendipity

4 The User Interface Design

The analysis of existing systems [14] provided evidence that some interaction solutions appear to be universally adopted and accepted in TV systems. For instance, navigation is generally allowed in four directions (up/down/left/right), suggested by a compass-like arrow keys configuration on the remote control.

The guidelines for the navigation metaphor presented in the previous chapter were interpreted according to a full User-Centered approach. The basic requirement of this interface is that different aims of the user are allowed, as well as different criteria for organizing the user experience; that means that he/she must be able to easily modify strategies for accessing the content. The most important transitions to be allowed appear to be:

- from leading the navigation to being led by the navigation (and back)
- from targeted search to serendipitous search (and back)

These requirements were interpreted as a need for the user to be in control of the main high-level decisions in navigation, i.e. defining the points of navigation engagement, setting the target of the contents navigation and reconfiguring the visualization strategies. Each of these decisions impact on the overall user interface configuration, and therefore users must be in full control of them.

Main decisions turn into reconfiguration of the user interface structure, that has to be managed by the system according to the general user inputs. Once the interface structure has been reconfigured to allow a given navigation strategy, users may proceed in micro-navigation decisions by selecting specific contents.

As a first step towards the specification of the user interface, a set of alternative metaphors were proposed, sharing a visualization strategy that places the point of observation in the eyes of a user moving through the contents. This approach differs from existing e-TV interfaces in that contents are differently organized into the space depending of the “level” of navigation the user has entered (thus adding a visual information about the level of category currently navigating), though maintaining the same meaning for all the commands: as a whole, this approach overcomes the traditional “browsing” strategy, replacing it with a full yet simple navigation solution.

The basic navigation features were implemented in a prototype, whereas advanced functions were drafted through a storyboard (an animated slide show) depicting how the most relevant actions would be performed. The aim of this decision was twofold: on the one hand, the usability of the most traditional functions (e.g.: control device, menu structure, content details) needed to be assessed in fine details; on the other hand, reconfiguration and recommendation strategies represented a novelty for most users, so that acceptance and interest needed to be investigated from scratch – as no previous example in the TV domain could be used as a blueprint for prototyping.

5 Enhancing the User Experience

Based on the attitudes described at paragraph 2, we faced the need to imagine and describe in enough detail a set of typical behaviors of the users when interacting with this enhanced TV. This is necessary to define use cases and, most important, the tasks for performing usability tests. Some examples of use cases are reported.

As shown in Figure 2, contents are classified in a two level taxonomy, whose structure was defined by a group of media experts that outlined categories (e.g. movies, tv series) and subcategories (thriller/romantic, soap operas/series) necessary to classify all the possible television contents, taking into account standardization bodies, best practice of content providers as well as the national cultural context.

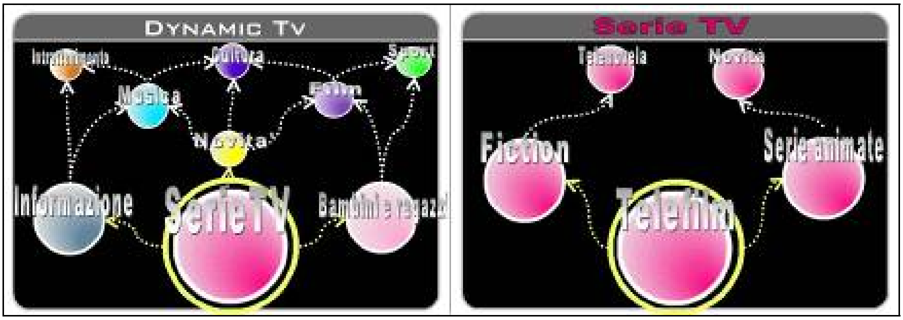


Fig. 2. a. All the television contents Categories – b. An example of subcategory

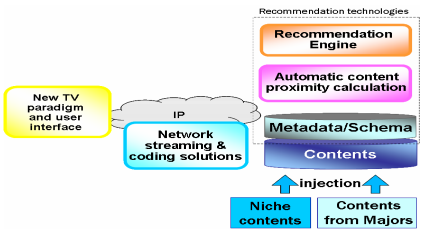
As shown in figure 3, each subcategory contains a set of contents so that when the user enters that level of navigation, he/she can easily look at the whole offer just moving left or right. Even if it cannot be shown in the figure, all pictures of contents are actually streaming videos, with audio only for the one in the first place. For every content it is also possible, by simply pressing the OK button, to see a “content profile box” (3.b) that contains additional information (e.g. director, actors and synopsis for a movie), the trailer and some suggestions on other contents that the user can follow to discover new and unexpected paths.

The metaphor described so far allows to exemplify use cases that match with all the user attitudes described above. For instance, if we suppose that the user is in a relaxed attitude we can guess that he/she moves through the categories just to find the one matching his/her interest, then chooses a content and watches it in the full-screen

Fig. 3. a. Contents belonging to a specific subcategory - b. A content card

mode, from the beginning to the end. Differently, a user in a curious attitude might not be interested in watching a single content but rather to have a "taste" of many of them, so he/she will explore different categories just to watch only few minutes of each content. He/she can also enter the "content profile box" and follow suggestions proposed to discover new and unexpected contents.

Finally, if a user is in a participant attitude he/she probably wants to share his/her knowledge with the other users, for instance, giving a rating or inserting a review.



Architecture and technical issues

Since the whole study exposed in this paper has the final aim of actually implementing an end-to-end prototype, all of the choices and the assumptions made about the user interface has to be dealt with from the feasibility point of view. The main component studied and implemented to support the metaphore are the depicted in the figure above and briefly described hereafter:

- **Metadata structure** defined taking into account both the national culture and its "translations" into the various schemas of the different content providers.
- **Semantic proximity:** Automatically detects the level of similarity between contents, based on their textual desciptions. .
- **Audiovisual proximity:** If the content has no text description, audio/visual low level features analysis will allow detecting similarities between contents
- **Recommendation engine:** Powerful and flexible algorithms are needed, in order to maximise the probability that a user that chooses to follow a path actually encounters something that he/she might like.

These 4 modules can be used to determine the way contents are presented to the user, overcoming the current use of the alphabetical or "newest-oldest" order, neither of which is satisfactory to match the chosen drivers, as well as to propose appropriate and acceptable suggestions.

- **Network:** Since one of the important and innovative aspects of the Dynamic TV paradigm is its ability to satisfy the most curious and bulimic attitudes towards media, immediate contents' preview is a very important requirement. While MPEG-4 AVC is a stable coding, new techniques for coping with the bandwidth fluctuation are emerging. We investigated SVC (Scalable Video Coding, extension of MPEG4) and MDC (Multiple Description Coding).

6 User Study

In order to evaluate the appeal and the appropriateness of the user experience with the new TV paradigm, a usability and acceptability test was conducted on the prototype and the storyboard. The test is just a preliminary one and some features, like serendipity, were not tested yet. Anyhow, they will be performed as soon as the prototype will achieve a more mature development phase. Before going on it was important to verify that the direction we were following was a right one. Three main issues about navigation and interaction design were investigated :

1) General feeling of potential users for the Dynamic TV paradigm. Does it meet users' actual interests, including the "no TV" and the most technology oriented?
2) General usability (efficacy, efficiency and using satisfaction) and navigability of the prototype. Users were asked to use it through specific usability tasks, i.e. reaching a category starting from a specific point, looking for a content, etc.
3) Suitability of controls. Is a common Tv input device suitable?

6.1 Method

According to the three described main aims some experimental tasks were defined, that every participant performed individually. The test driver gave him/her the task without suggestions and observed him/her in all its operations on the prototype, writing down the number and the typology of errors. The test setting were located in the Usability Lab of Telecom Italia and in the Interaction Design Lab of the University of Modena &Reggio Emilia. The sample of users was equally distributed in the two sites. To reduce the setting experimental effect people tested Dynamic Tv in a residential room inside the labs. The test driver was inside the room, besides the user, to follow directly his behaviour. From a control room, through a one side mirror, the test was monitored also by two external test drivers. Every user tested the system for about an hour and half carrying on 8 tasks. The first one was a free navigation task. All the others had a specific aim focused on critical points of the navigation flow (i.e. reaching a category, reaching a content, moving across categories, etc.

At the end of every task a short interview was carried on to investigate user feeling about the interface and the simplicity/difficulty in interaction, navigation and content search. Then, the participants had to fill in a self-filling questionnaire that allowed us collecting subjective and quantitative data to be analyzed in comparison with direct observation data and interview ones. In particular questionnaire data were collected on 5 point Likert and semantic differential scales ("1" is lowest, "5" the highest). We defined the sample of potential users we were interested in, according to previous market analysis that identified specific "stereotypes" of TV users and the hypotesis

Table 1. Users categories for usability and acceptability tests

TV fruition	Videogame use	Pc use
YES	NO	NO
YES	NO	YES (also Internet)
NO	YES	Indifferent
YES	YES	Indifferent
NO	NO	YES (also internet)

we made about favourite targets of Dynamic TV. Five user targets were defined on the basis of their habits and capabilities in TV fruition, PC and videogame use. The sample was composed by 30 participants equally distributed among targets (6x5). For each target, users were equally distributed on three age ranges (<35, 35-65, > 65).

6.2 Main Results

The main results are optimistic. Qualitative and quantitative data are coherent.

In particular qualitative data come from the short interviews and users' thinking aloud analysis. Quantitative ones come from direct observation of behaviour, number and kind of errors and number of attempts to reach the aim but also the final selfcompilative questionnaires. In the following, the main results:

in general the prototype was easy to use for every target of users (mean = 3,88over 5). Also over 65 could use it in a simple way because of high learnability and easiness in error recovery. The appeal of the system was good (mean= 3,86 over 5) as all the different sample's typologies, namely less interested in tv, generalist and heavy TV users really appreciated the system. The users proved to be able to navigate through the control device in a simple way (mean usability = 3,86 over 5) even if not always efficiently, especially for heavy pc/videogame users. For this reasons, part of the re-design phase, in a way which is still in course, has been pointed in this direction.

Interesting data came out from the system perception on easyness to use and appeal before and after using: after using the system, most evaluations were more positive.

7 Conclusions

This paper is a first description of the interaction design process, results and findings related to a new interactive TV paradigm under definition in the DynamicTV project. Guided by design drivers coming from previous studies, the new paradigm merges interactive and entertainment elements with the aim to be suitable for different attitudes of users. The paradigm is based on standard de-facto solutions combined with new interaction metaphors. Storyboards and prototypes were developed to enable usability and acceptability tests. User tests showed good results and, on the base of these findings, the DynamicTV project is now focusing on the design of new interactive features and on the implementation of an enhanced prototype. The ongoing interaction design activities continue to be guided by a user-centred approach, with the final goal to simplify and enhance the TV user-experience by exploiting the technical opportunities of the new media landscape.

References

1. Swearingen, K., Sinha, R.: Beyond Algorithms: An HCI Perspective on Recommender Systems. In: Proceedings of 2001 Workshop on Recommender Systems at SIGIR (2001)
2. Taylor, A.S., Harper, R.: Switching on to switch off. In Inside the Smart Home, pp. 115–126. Springer, London; Van Dijk, J., de Vos, L.: Searching for the holy grail – Images of interactive television, in New Media & Society, vol. 3. SAGE (2001)

3. Van Dijk, J., Heuvelman, A., Peters, O.: Interactive Television or Enhanced Television? The Dutch users interest in applications of ITV via set-top boxes. In: 2003 Annual Conference Of the International Communication Association, San Diego, USA (2003)
4. (last checked on Jan 18th 2008), http://www.babelgum.com
5. (last checked on Jan 18th 2008), http://www.joost.com
6. Ghittino, A., et al.: Living@Room: a Support for Direct Sociability through Interactive Tv. In: Proceedings of the EuroITV 2007, Interactive TV: a shared experience (2007)
7. Hemmeryckx-Deleersnijder, B., Thorne, J.M.: Awareness and Conversational Context Sharing to Enrich Tv Based Communication. In: Proceedings of the EuroITV 2007, Interactive TV: a shared experience (2007)
8. Boertjes, E.: ConnecTV: Share the Experience. In: Proceedings of the EuroITV 2007, Interactive TV: a shared experience (2007)
9. Harrison, C., Amento, B.: CollaboraTV – Making Tv Social Again. In: Proceedings of the EuroITV 2007, Interactive TV: a shared experience (2007)
10. Simeoni, R., Etzler, L., Guercio, E., Perrero, M., Rapp, A., Montanari, R., Tesauri, F.: Innovative TV: from an old standard to a new concept of Interactive TV – an Italian job. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 971–980. Springer, Heidelberg (2007)
11. Schönbach, K.: The Hyperactive Audience – Still An Illusion. An essay, revisited. In: Roessler, P., Krotz, F. (Hrsg.) Mythen der Mediengesellschaft – the media society and its myths, Constance, Germany, UVK, pp. 267–278 (2005)
12. Forrester's 2007 European Consumer Technographics Survey. Forrester Research (2008)
13. DynamicTV: The Long Tail applied to broadband-broadcast integration. In: Euro ITV 2006, Andrea Belli, Marina Geymonat, Monica Perrero, Rossana Simeoni - Telecom Italia, Monica Badella - Politecnico di Torino
14. Anderson, C.: The Long Tail: Why the Future of Business is Selling Less of More. Hyperion (2006)
15. Toms, E.G.: Serendipitous Information Retrieval. In: Toms, E.G. (ed.) Proceedings of the First DELOS Network of Excellence Workshop on"Information Seeking, Searching and Querying in Digital Libraries", Zurich, ERCIM Workshop Proceedings - No. 01/W001 (2000) (lastchecked on January 18th, 2008), Retrieved: http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf

Media Space Navigator: Navigating Video Content on IPTV Portals

Simon Vogl[1], Peter Halbmayer[1], Christoph Lichtenberger[1], Helmut Rauscha[2], Doris Rodler[2], and Waltraud Müllner[2]

[1] Austrian Research Centers Gmbh - Research Studios Austria
`first.lastname@researchstudio.at`,
`http://www.researchstudio.at/`
[2] Telekom Austria TA AG,
`first.lastname@telekom.at`,
`http://www.telekom.at/`

Abstract. With the introduction of IPTV services, today broadcast and on demand services may be offered seamlessly. The growing amount of content creates a significant navigation challenge for the user that we try to resolve by providing a uniform way to access both services simultaneously.

The Media Space Navigator combines entries from Video on Demand sources with data about life shows from EPG feeds into a unified graph structure that links entries by similarity. Users can browse the graph via an interface that has been designed for casual browsing with a remote control: For each entry, it provides a ranked view of the most relevant neighbours, which are accessible using cursor buttons. By selecting an entry, users can either select a node, or directly jump to the content (either the broadcast channel or a payment dialog to access a stored movie).

The similarity measure is computed in a hybrid approach by merging different meta-data domains (film genre, actors, etc.), graph distance metrics and user preferences that can be entered casually during navigation: Rating entries with one to three plusses influences genre and actor weights in the similarity calculation.

1 Motivation

With the introduction of IPTV services, today broadcast and on demand services may be offered seamlessly. In the current navigation scheme, once one of the two services is selected via the main portal, two different types of navigation are implemented for the two domains; a tabular EPG in the broadcast domain and static linear navigation structure for the VOD content. As the number of broadcast channels increases and the VoD library grows significantly, the searches for the desired movie may become long and tiring, whereas the motivation of the user decreases.

A way to support the user in finding his movie, and to speed things up, is to include contextual information about movies into the search, like linking together

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 248–252, 2008.

movies of the same genre or with the same actors. An additional part would be to include user preferences and context into the search. This means to prefer search results where the genre or the actors are to the users liking.

The system consists of two components: The backend analyses video metadata gathered from different sources and user preferences to create a similarity graph, linking all entries. The frontend enables navigation by visualizing a suitable subset of this graph.

2 Frontend

The Frontend application enables navigation of the similarity graph is intended for use on set-top boxes. As they use IR remote controls as user input, we have adapted the navigation principle to arrow keys and the OK button, in contrast to PC-based graph navigation mechanisms (Herman et.al. provide a nice overview), which rely on mouse input.

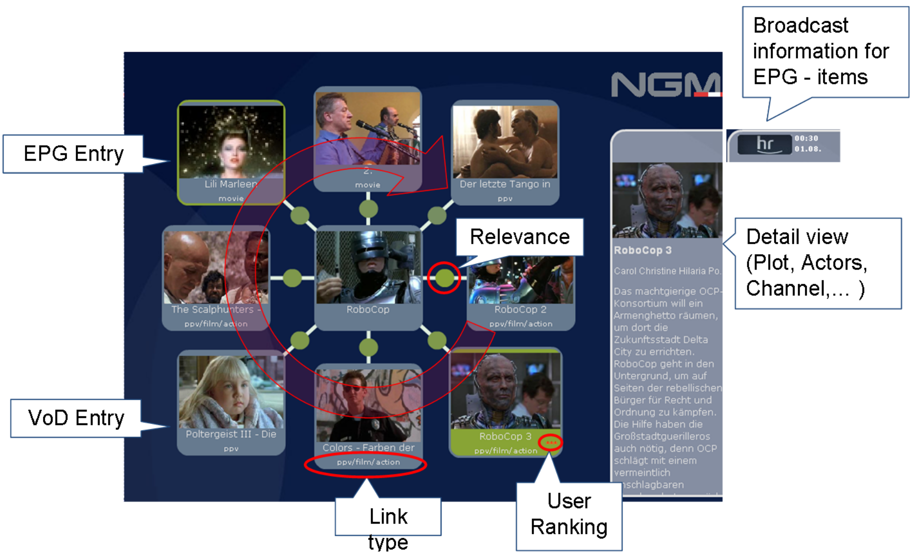


Fig. 1. Media Space Navigator Frontend

Fig. 2 shows the frontend with the most important features: The starting entry is located in the center, surrounded by up to eight of the most relevant neighbours, entries coming from the video on demand (VoD) data set or the electronic program guide (EPG). The relevance of a movie item is mapped to a color range between green (most relevant) and blue (least relevant). Each entry is displayed with its preview icon, movie title and the linking genre or actor; in addition EPG data is marked with a green border around the movie item. Besides the the graph a detail area on the right side provides information about the selected item like actors, creators and plot summary. For EPG entries this is enriched with information about start time and date as well as the broadcast channel.

When the user selects an item by pressing the OK button, this movie is centered and the most relevant neighbours are displayed. Pressing the OK button on a centered item leads to the broadcast channel for EPG items and for VoD items the payment dialog is opened. After insertion of the PIN the VOD movie is started.

3 Backend

Delivering the necessary information to the frontend, the functionality of the backend is depicted in figure 2. Starting from the top different data sources are read into a data pool. This way electronic program guide (EPG) data and video on demand (VoD) data can be handled in a uniform way in following processing stages. The metadata about movie entries is contained in XML files where the format for EPG entries is different to the format of VOD entries. Therefore it is necessary to convert entries from the two data sources to an object format that can be handled the same way in succeeding processing steps.

The categories an entry belongs to and the persons that appear in the movie are extracted too. They are used for linking entries together in the upcoming stage. If for example a person appears in two entries they are potential candidates for linking over this person. All existing categories and persons are extracted

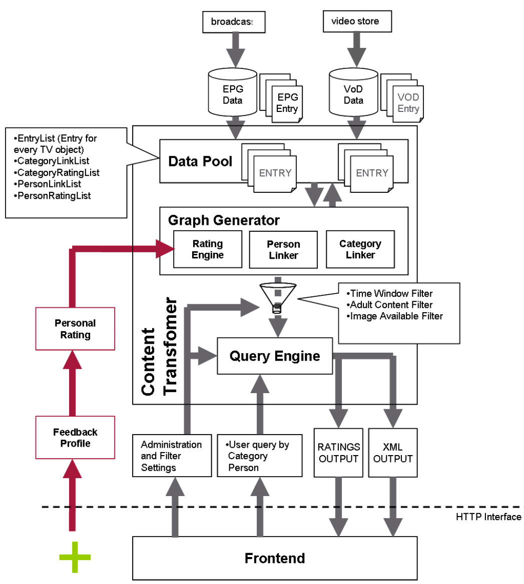


Fig. 2. Content Analysis Backend

and the number of entries that belong to each are calculated. These serve as the weights for an initial linkage. All the data gathered until now is kept in a Data Pool that is the central structure for all subsequent user queries. Two entries can be linked by more than one relation. This happens for example if two entries belong to the same categories and the same actors appear in both entries. This is often the case for series or sequels of movies.

Initial weights are determined by summing up the number of entries within a category or the number of entries a person appears in. This category or person weight is then taken to give the initial connection weight between two entries linked over this category or person. Then the inverse weight sum is taken which means that categories or persons with a lower weight have higher importance for linking. This is because for example categories with a high weight are too general to give a reliable picture about what is tightly related and what not. This is caused by the metadata as for example a lot of entries are tagged with the category "movie" where at least two third of the whole database will fall into. It is clear that a more meaningful link description is desired that substantiates the relation between two entries at a more concrete level. Therefore categories with many entries shall be considered only if more special categories do not result in enough items in the search of neighbors for an entry. The Graph Generator is used in the next step by the Query Engine to process user requests received from the frontend. It uses the Data Pool to access the entry items and creates a list of entries that are linked by categories and persons. The graph generator applies the set filters that reject items that do not meet the specified filter criteria.

As the frontend displays the central movie item surrounded by the eight most prominent neighbors it is obvious that eight related entries must be found. Under normal conditions much more than eight related items will be found for an entry due to the vast amount of data items and the first eight will be returned as they are the most important ones. In some cases when an item lies in a very special category it is possible that less than eight neighbors are found. In this case the actually found entries are returned. A possibility would be to use the empty slots to take some general movie items that should be specially advertised like new movies that were inserted in the database recently.

If a user likes an item more than others he/she is able to rate an entry on a scale from one to three (with one being the lowest and three being the highest rate). When the user sets the maximum rating to, for example, an action movie, this leads to a higher ranking of all items in the action category which again promotes action movies in subsequent queries of this user. Additional features include an adult content filter and a time window for EPG entries that states how far in the future EPG entries are considered. These filters can be set over an administration interface.

4 Conclusions

With the Media Space Navigator, we have successfully transferred a graph based navigation interaction metaphor from the PC into the TV interaction domain.

The feedback from an initial set of test users gathered during a three-day multimedia exhibition showed that the presented approach is easily understood by a demographic diverse community and is a viable extension of traditional browsing mechanisms that enforce a more static ordering of contents in genre/sub-genre lists.

The extraction and analysis of meta-data resulted in a usable graph structure that we will extend with an user-activity based recommender system as soon as an initial set of test-users has access to the interface.

References

[AG07] Ahmed, A., Hong, S.-H.: Navigation techniques for 2.5D graph layout. In: 6th International Asia-Pacific Symposium on Visualization APVIS 2007, February 5-7, 2007, pp. 81–84 (2007)

[GK06] Gstrein, E., Krenn, B.: Mobile Personalization at Work. In: Workshop on Recommender Systems, ECAI 2006, Riva del Garda, August 18-19 (2006)

[HMM00] Herman, I., Melancon, G., Marshall, M.S.: Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics 6(1), 24–43 (2000)

An Adaptive Cartography of DTV Programs

Jean-Gabriel Ganascia, Charles Madeira, and Karan Fouladi

Université Pierre et Marie Curie (Paris 6)
Laboratoire d'Informatique de Paris 6 (LIP6)
104, avenue du Président Kennedy, F-75016, Paris, France
{Jean-Gabriel.Ganascia,Charles.Madeira,Karan.Fouladi}@lip6.fr

Abstract. With the development of information technologies and more precisely, the so-called digital convergence, the number of digital TV (DTV) programs available on any kind of DTV receiver (satellite, cable, terrestrial or IP) has increased dramatically. This has made the task of selecting a DTV program very complicated and time consuming. This paper proposes an intelligent graphical user interface (GUI) tool that gives a new appearance to the on-air and recorded DTV contents in order to enable DTV viewers to, quickly and in a user-friendly manner, select a program. The GUI tool is based on a synthetic cartography of DTV programs in which the shape generated reflects the DTV viewers' preferences over a period of time. These preferences are obtained by using any method for program clustering and recommendation. Results are demonstrated on the task of recommending programs in the context of the French terrestrial DTV by using a content-based approach.

Keywords: Intelligent Graphical User Interface, Digital TV Recommendation System.

1 Introduction

With the development of information processing, Internet, digital TV (DTV) and telecommunications, the so-called digital convergence becomes a reality [10]. As a consequence, DTV viewers face the difficulty of quickly finding programs that best fits their preferences.

In order to aid DTV viewers for selecting programs more easily, several TV recommender systems have been proposed and developed in the last years. These systems take advantage of different approaches such as content-based filtering [1,16], collaborative filtering [2,14] and hybrid filtering [4,3,12] for creating personalized Electronic Program Guides (EPG).

Nevertheless, even if these recommender systems restrict a large number of possibilities of choice, many of them remain. So, the graphical user interface (GUI) appearance used for presenting the personal EPG seems to be also a key point for the success of the DTV viewer choice. So far, the GUIs implemented on DTV receivers are usually awkward. Some of them only provide a textual list of DTV programs, while others are built on small views of current programs.

This paper constitutes an attempt to simplify visual access to DTV programs by using a novel GUI inspired by the technique of "Memory Islands" [6]. This

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 253–262, 2008.

novel GUI is based on a dynamic 2D cartography that represents personalized EPGs by clusters of DTV programs. As we shall see, personalization of EPGs can be generated by any existing recommender system approach.

In the next section, we recall what Memory Islands are, how they can be used to describe DTV programs and what the ontology on which program clustering is based on. Following this, we present the process for building Memory Islands in the context of the French terrestrial DTV programs, we briefly describe a content-based recommender system that we have developed to demonstrate Memory Islands, and we illustrate it with examples of some of the generated shapes. Finally, we conclude the paper and discuss ongoing work.

2 Representing Personal EPG with Memory Islands

2.1 Memory Islands

Memory Islands [6] were initially designed to ease browsing in e-books and more generally in digital textual documents. The original idea consists in transforming the linear structure of a textual document into a 2D space that maps its logical structure. It is inspired by the ancient "Arts of Memory" which relate to us how our memory capacity can be considerably increased by placing things into "locis" belonging to architectures that were previously learned [15]. In other words, to "embody" the virtual content of e-books, Memory Islands anchor it on an automatically generated territory. In a way, it represents on a 2D space a digital content that appears linear at first sight. This representation corresponds to an increase of dimensions, which is quite unusual in information visualization, since the general aim is to reduce data dimensions [13]. Note that the main goal here is not to focus attention on a particular item, but to represent a wide variety of contents and to stimulate human memories with an easy to remember picture, which facilitates user interactions with the contents.

The adopted solution consists in mapping each unit of virtual content onto a small deserted island, since uninhabited islands are known by sailors who simply sail around it, and therefore only know and name the coast (1D space) without worrying about the surface. These fictitious territories are designed to strike imagination and to remain anchored in our memory. Consequently, they must be distinguished the ones from the others as much as possible in order to reflect the structure of the digital content and to ease navigation. Therefore, the map generation program endeavors to produce forms that are as diversified as possible in order to build a realistic map of a territory that offers an overview of the content. This cartography takes advantage of historical studies of old maps by using legends, texts, colors, and more amazingly for us, relevant icons [8].

As presented in [6], Memory Islands may also represent any cultural/multimedia contents such as digital audio, digital video or DTV programs. In this case, the "islands" map the multimedia contents by some clustering. For example, clustering can be made with respect to the media theme, e.g. movies, news, etc. This allows to generate a *weighted ordered tree* that is transformed into a shape enriched by colors and icons correlated to each theme.

2.2 Shape Generation Algorithm

From a practical point of view, the logical structure of books corresponds to their table of contents. It may be transformed into a *weighted ordered tree*, of which node weights are computed according to the proportion of the book covered by the corresponding part. For instance, if a book containing 300 pages is divided in three parts containing respectively 50 pages, 100 pages and 150 pages, then the weights are respectively 1/6, 1/3 and 1/2 for the parts 1, 2 and 3. This allows to generate a 2D shape from the *weighted ordered tree.*

More generally, any e-document structure can be assimilated to an ordered weighed tree, i.e. to an ordered tree of which nodes are associated to numbers corresponding to the weight of the corresponding parts. For instance, the number of characters for texts or the required space or time for some digital multimedia contents such as audio or video. A shape is generated from this tree.

The algorithm for generating shapes considers each weighted node of the structure of the e-document, i.e. each set of contents, as a sector of a circular disk, i.e. an angle associated with a radius. The angle corresponds to the proportion of the section in the whole book while the radius is computed taking into account the size of the section and its level. Moreover, some blank contents are virtually considered to separate blocs, which cut out the coast of the artificially generated island.

The algorithm takes as input an ordered labeled tree derived from the multimedia content organization. For instance, the content organization may cover five meta-themes – *Society and Politics*, *Information*, *Culture*, *Youth* and *Fiction* – while each one covers from two to four themes. It corresponds to the following multimedia organization that is voluntarily oversimplified, for the sake of the presentation: [[*Society and politics*, [*Religion*, *Talk show*, *Society*, *War*]], [*Information*, [*News*, *Newsine*]], [*Culture*, [*History*, *Music*, *Theater*]], [*Youth*, [*Cartoon*, *Manga*, *Movies*, *Games*]], [*Fiction*, [*Thriller*, *Spies*, *Fantastic*, *SF*]]].

Once this multimedia content organization ontology is given, the programs (e.g. next hour, next day or next week programs) and coefficients weighting their importance (e.g. program duration combined with its interest), are added to each leave of the tree. It gives a weighted ordered tree. More precisely, each node N of the weighted tree corresponds to a level of a hierarchy characterized by:

– A label (called *label(N)*), that is the characteristic associated to the level of the hierarchy. In the case of DTV programs, we restrict it to meta-themes and themes;
– A weight, which is the size of content covered by the nodes, e.g. a combination of the running time and preferences of DTV viewers;
– The depth of the node, i.e. the level of the node. For example, it may be 0 for the root of the hierarchy, 1 for meta-themes, 2 for themes and 3 for programs.

Moreover, other parameters are involved in order to accentuate singularities of the generated island through their variations. Since our goal is to remember a form, we have to make each form as particular as possible. Here are all existing parameters:

- An internal size ρ. It may be predefined or computed as a function of the global running time or of the available space for digital audio and video.
- An eccentricity, ϵ. It accentuates the length and the extension of capes, shaping them like islands.
- The number of blank contents between blocs. It is usually an input parameter, since it is difficult to get it by just reading the table of content. It helps to increase the size of gulfs in order to differentiate shapes. In the DTV program case, we use theme to build a separate section for each metadata.

The output of the algorithm generates a sequence of triplets $< \theta_1, \theta_2, \rho >$, each one corresponding to an arc in which the radius is ρ on the sector $< \theta_1, \theta_2 >$. This sequence is automatically translated in Scalable Vector Graphics (SVG), generating a path passing by these generated points.

For instance, being given a set of programs described with the themes and the meta-themes given in the above described content organization, the previously referred algorithm generates the shape presented in figure 1.

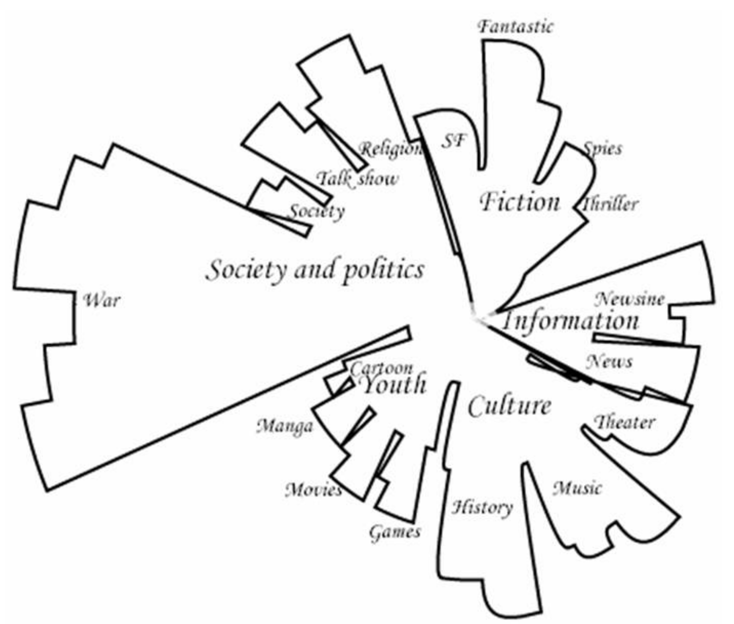


Fig. 1. A shape generated with a 5 meta-theme content organization

2.3 Ontology of Themes and Meta-themes

As we have previously seen, the Memory Islands are built from a *weighted ordered tree*, which corresponds, in the case of cultural/multimedia contents, to the hierarchical organization of those contents. In the case of DTV programs, we identify the clusters related to a predefined organization. Three hierarchical levels are used to distinguish them:

- The first level is constituted by sub-themes. A group of homogeneous sub-themes are assembled to build a theme.
- The second level corresponds to 39 classes that are characterized as themes, e.g. *thriller, adventure, funny, drama, fantasy, cartoons, etc.* They are built of the sub-themes groups. Each theme is associated with a meta-theme, which means that the ontology of theme consists in a hierarchy.
- The third level gives a broad definition of 12 classes as meta-themes, e.g. *charm, fiction, society and politics, entertainment, youth, sport, sciences and nature, TV and shopping, information, do-it-yourself, culture, other.*

Figure 2 contains themes associated with their corresponding meta-theme. Our first experiments using this setup demonstrate that it works very well in the case TV programs for at least two reasons. On the one hand, it can automatically determine themes and meta-themes from the metadata associated to each program. On the other hand, it corresponds to distinguishable sets of preferences associated to the social profile of families, which can facilitate a learning task.

3 Applying Memory Islands to the French Terrestrial DTV

In this section, we show a 2D cartography built from a set of DTV programs and automatically adapted to viewers' preferences. The first subsection explains the process of identifying themes and meta-themes for DTV Programs. The second subsection shows the building of a Memory Island from a DTV program set. The third subsection describes the process of adjusting DTV viewers' preferences by taking their behaviors into account. The fourth subsection shows how viewers' preferences shape the Memory Islands.

3.1 Identifying Themes and Meta-themes of DTV Programs

Once meta-themes and themes have been defined in the ontology introduced in section 2.3, we need to associate them with DTV programs. The goal consists in dynamically classifying EPG programs for a period of time in order to obtain the information required to build Memory Islands. This procedure is crucial for the usefulness of the Memory Islands generated in the context of the DTV.

Considering that program metadata obtained from the French Terrestrial DTV EPG are expressed in natural language, we propose an abstraction method that takes the metadata as input in order to identify their themes and meta-themes.

Meta-themes	**Themes**	**Meta-themes**	**Themes**
charm	pornography	youth	youth
charm	eroticism	youth	cartoons
fiction	fiction	sport	sports
fiction	martial arts	sport	soccer
fiction	western	sport	motorsport
fiction	thriller	sport	swim sports
fiction	adventure	sciences&nature	wildlife
fiction	funny	sciences&nature	sciences
fiction	drama	sciences&nature	hunting&fishing
fiction	horror	tv&shopping	tv shopping
fiction	science fiction	information	tv news
fiction	fantasy	do-it-yourself	do-it-yourself
society	society	do-it-yourself	culinary arts
society	politic	culture	culture
society	religion	culture	theater&dance
society	VIP	culture	literature
society	history	culture	arts
entertainment	games&gambling	other	diverses
entertainment	popular music	other	not announced
entertainment	show biz		

Fig. 2. List of 12 meta-themes and 39 themes defined in the ontology

This consists in simplifying the level of detail of the information available [7]. The input in here is composed of eight fields: *Time, Channel, Title, Sub-title, Description, Category, Authors* and *Review*. The procedure is performed as following:

– A parser uses a list of stop words to clean up the program metadata. This reduces each field to a bag of words;
– A list of keywords is selected by the TF-IDF (term frequency / inverse document frequency) information retrieval technique [11];
– The bag of words and the list of keywords are asserted into a knowledge-based system implemented with the *NASA's C Language Integrated Production System* (CLIPS V6-24) in order to identify themes and meta-themes of each program.

3.2 Generating a DTV Program Memory Island

Let us now suppose that we have a DTV program set composed of clustered programs. The first step required to generate a Memory Island consists in transforming this set into a weighted ordered tree. This is achieved by using the main theme identified by the clustering process, i.e. *"Theme"* attribute, which belongs to the theme ontology. The main theme associated with each program allows to cluster the program set into a three-level hierarchy. As previously said, the first level corresponds to the meta-theme, the second to the theme and the third to the programs themselves.

Therefore, we use the algorithm presented in section 2.2 to build the Memory Island corresponding to the weighted ordered tree. An example is given in Figure 3.

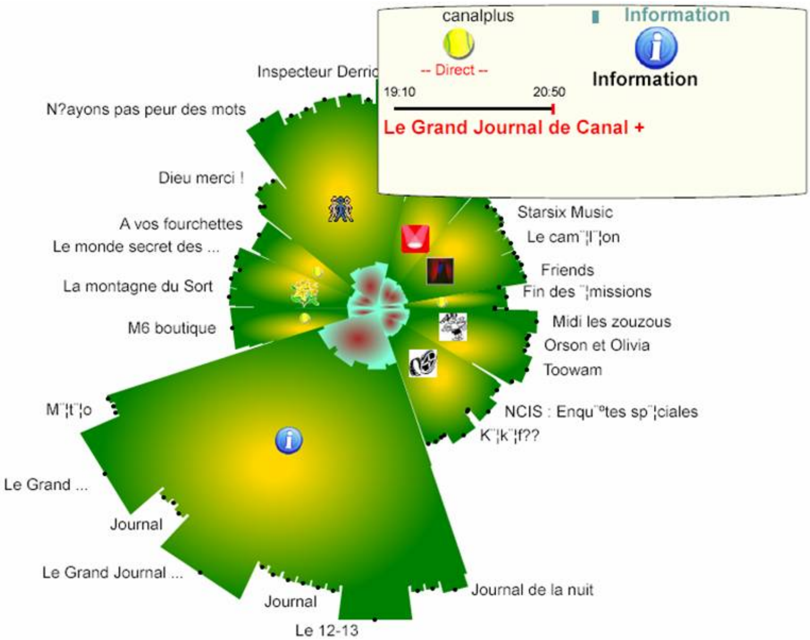


Fig. 3. A Memory Island with a pop-up for a focused program

We can see that the islands can be enriched with colors and icons corresponding to the most developed meta-themes and the main theme. Moreover, pop-up windows describing program characteristics can be automatically showed when a program is focused on. User interactions (selecting a theme, selecting a program, zooming a theme) are based on basic buttons of the remote control (up/down, left/right and +/-) that can be managed easily by anyone, without any requirement for learning.

3.3 Adjusting DTV Viewers' Preferences

In order to build personalized EPGs for testing the recommendation of programs by Memory Islands, we have been implementing a content-based recommender system that creates dynamic profiles reflecting viewers' behaviors.

Understanding Viewers' Behaviors. To understand how viewers behave, we interpret the events on the DTV receiver. First of all, we need to know which programs are available (on-air). This information is obtained by the EPG. Secondly, we need to monitor viewers' interactions with the DTV receiver. Viewers' interactions are described by a pair $< b, s >$, where b is a button on the remote control that is pressed at a time t and s is the current screen state on the receiver. The EPG and the viewers' interactions are then associated together, referring to the time scale matching. The matched information is then analyzed in order to learn the viewers' behaviors.

This analysis is carried out in two stages: a *syntax validation* and a *semantic validation*. The syntax validation consists in combining interactions that operate on the collected data during a specified period of time. It generates an outline perspective on the viewers' interactions. This combination procedure is based on a set of interactions that lead to changing a channel. They are called *main interactions*. All programs watched on this specific channel are recorded in a devoted list. Other interactions such as volume increase or decrease are associated to this list as *secondary interactions*.

The semantic validation consists in evaluating the programs watched by the viewer at each period of time identified through syntax validation. The goal here is to rank the programs according to the time that the viewer spent on watching them and the reason of eventually changing the channel. In order to dynamically express the viewers' preferences, the program's rank contributes to reward or punish the associated abstract representation of the program introduced by the clustering method (see section 3.1). This plays a very important role in the performance of machine learning techniques [9].

Learning Viewers' Preferences. The abstract representation of programs is then matched to the viewers' profile in order to get their ranks. Using these ranks, a filtering process is carried out to arrange programs by decreasing order of relevance. Programs ranked higher occupy bigger regions on the generated Memory Island in order to reflect the viewers' preferences on screen. Machine learning techniques are advocated in this context to dynamically learn the viewers' preferences. The task consists in learning a function that ranks all programs according to user preferences. The goal of this task is to automatically update a pair $< r, w >$, where r is an abstract program representation and w is a weight value associated with the abstract representation. We can use different learning strategies to recommend programs (short-term and/or long-term).

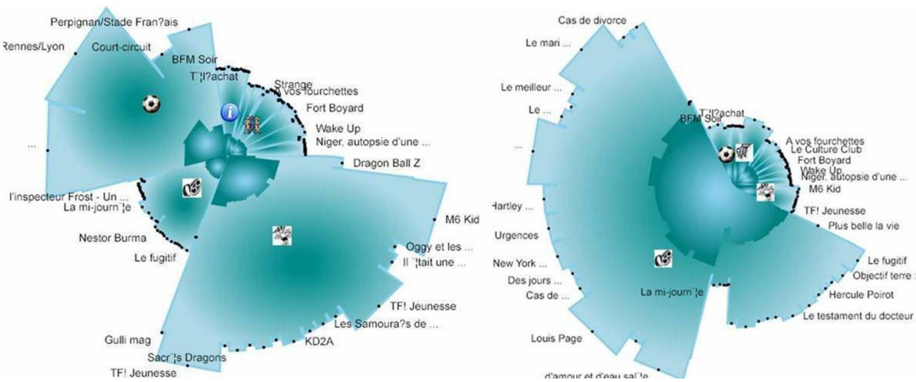


Fig. 4. Two Memory Islands showing the preferences for two kind of group of users

3.4 Adapting the Shape of Memory Islands through Viewers' Preferences

In this section, we take into account a DTV program set of roughly two hundred programs and show them through two Memory Islands that reflect the preferences of two kind of group of users. The Memory Island presented on the left in Figure 4 corresponds to a group of users interested in watch football matches and youth programs. The Memory Island on the right corresponds to another group of users that is a film-loving.

4 Conclusion and Future Work

Intelligent GUIs for DTV receivers are a very interesting and original problem little investigated until now. Designing GUIs that facilitate the DTV viewer choice seems to be a key point for the next generation of DTV receivers. Moreover, they can significantly aid the recommendation of DTV programs made by automated systems.

This paper proposed a novel intelligent GUI tool that constitutes an alternative to the GUIs usually employed on DTV receivers. This novel GUI tool is based on a synthetic 2D cartography of DTV programs, inspired by the technique of Memory Islands, to enable DTV viewers to quickly select programs. It can be associated with any method for program clustering and recommendation in order to generate personalized EPGs.

The generation of shapes through Memory Islands has been illustrated for some personalized EPGs in the context of the French terrestrial DTV by using a content-based recommender system. The first results have demonstrated that desired programs can be easily located by using our system. In cooperation with a team of the IWEDIA Company and another of the LIMSI Laboratory, we are carrying out experiments with our system in the context of several groups of users. We will be evaluating, quantitatively and qualitatively, the behavior changes and the level of satisfaction of the different groups when using our system. More results will be available in the next months.

References

1. Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Chiarotto, A., Difino, A., Negro, B.: User Modeling and Recommendation Techniques for Personalized Electronic Program Guides. In: Ardissono, L., Kobsa, A., Maybury, M. (eds.) Personalized Digital Television: Targeting Programs to Individual Viewers, pp. 3–26. Kluwer, Dordrecht (2004)
2. Baudisch, P., Brueckner, L.: TV Scout: Guiding Users from Printed TV Program Guides to Personalized TV Recommendation. In: Proceedings of the 2nd Workshop on Personalization in Future TV, Malaga, Spain, pp. 157–166 (2002)
3. Buczak, A., Zimmerman, J., Kurapati, K.: Personalization: Improving Ease-of-Use, Trust and Accuracy of a TV Show Recommender. In: Proceedings of the 2nd Workshop on Personalization in Future TV, Malaga, Spain (2002)

4. Hsu, S., Wen, M.-H., Lin, H.-C., Lee, C.-C., Lee, C.-H.: AIMED - A Personalized TV Recommendation System. In: Proceedings of the 5th European Interactive TV Conference, Amsterdam, The Netherlands, pp. 166–174 (2007)
5. Ganascia, J.-G.: RECIT: reprsentation cartographique et insulaire de texts. In: Proceedings of the CIFT 2004 (Colloque International sur la Fouille de Texte), La Rochelle, France (2004), http://archivesic.ccsd.cnrs.fr/sic_00001258/en/
6. Ganascia, J.-G.: AC3 - Automatic Cartography of Cultural Contents. In: Lévy, P.P., Le Grand, B., Poulet, F., Soto, M., Darago, L., Toubiana, L., Vibert, J.-F. (eds.) VIEW 2006. LNCS, vol. 4370, pp. 253–263. Springer, Heidelberg (2007)
7. Giunchiglia, F., Walsh, T.: A Theory of Abstraction. Artificial Intelligence 56(2-3), 323–390 (1992)
8. Jacob, C.: L'empire des cartes. Albin Michel (1992)
9. Mitchell, T.: Machine Learning. McGraw-Hill, Sing (1997)
10. O'Driscoll, G.: The essential guide to digital set-top-boxes and interactive DTV. Prentice Hall, Englewood Cliffs (1999)
11. Singhal, A.: Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24(4), 35–43 (2001)
12. Tsunoda, T., Hoshino, M.: Automatic metadata expansion and indirect collaborative filtering for TV program recommendation system. Multimedia Tools and Applications 36(1-2), 37–54 (2008)
13. Tufte, E.: The Visual Display of Quantitative Information, 2nd edn. Graphics Press (2001)
14. Wang, J., Polwelse, J., Fokker, J., Reinders, M.: Personalization of a Peer-to-peer Television System. In: Proceedings of the 4th European Interactive TV Conference, Athens, Greece, pp. 147–155 (2006)
15. Yates, F.: The Art of Memory. Penguin Books (1969)
16. Zimmerman, J., Kurapati, K., Buczak, A., Schaffer, D., Gutta, S., Martino, J.: TV Personalization System. Design of a TV Show Recommender Engine and Interface. In: Ardissono, L., Kobsa, A., Maybury, M. (eds.) Personalized Digital Television: Targeting Programs to Individual Viewers, pp. 27–52. Kluwer, Dordrecht (2004)

Electronic Programme Guide Design for Preschool Children

Ana Vitoria Joly, Lyn Pemberton, and Richard Griffiths

University of Brighton, Lewes Road, BN2 4GJ,
Brighton, United Kingdom
{A.V.B.Joly, Lyn.Pemberton, R.N.Griffiths}@brighton.ac.uk

Abstract. The proliferation of audiovisual content available for young children brings with it the need for an appropriate Electronic Programme Guide (EPG) that meets needs and preferences of pre-schoolers. Two approaches to the design of such a service are discussed. Firstly, a range of existing guidelines on interactive television applications, personalized recommendation systems and interaction design for children are reviewed, in the context of theories of child development. Secondly, we explore the potential for children's collaboration during the categorization process to create a user-appropriate information architecture and interface.

Keywords: Electronic Programme Guide, Children, Interactive Television.

1 Introduction

The Electronic Programme Guide (EPG) is a vital component of interactive television, enabling viewers to find their way in an ever-increasing landscape of audiovisual content. The design of EPG's has been addressed by a number of studies, focussing variously on efficient EPG design [1,2] and personalized recommendation systems [3-5], yet these researches tend to assume a homogeneous adult audience. However, it may be that significant sub-groups within the television audience might benefit from specialised EPGs. For pre-school children, for instance, television viewing is part of the daily routine and is often an activity performed unaccompanied. Despite a general concern about TV's negative impact on children's development several studies underline that programmes designed specially for young children benefit their cognitive and social development [6]. As part of a larger research project analyzing iTV interfaces for young users, in this paper we discuss the design issues raised by the development of an EPG for (and with) pre-school children, which could be used independently and which would provide access to adult-moderated content.

2 Guidelines for EPG Design

Guidelines developed for interaction and interface aspects of general EPG's need to be supplemented with child-oriented approaches. While adults can manipulate complex remote controls, children have small hands and less developed motor skills,

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 263 – 267, 2008.

suggesting that a limited number of keys should be used in the remote and paging should replace scrolling [7, 10]. Given limited reading skills, text on the EPG menu could be replaced by real world metaphors associated with simple words [9]. Animation and audio on rollovers can help indicating where to find functionality [9] while clarifying what to find on the selected button.

For navigation, Bakhshi (2007), for instance, proposes thumbnails, which allow viewers to perform tasks with drop-down menus. Children up to the age of five, however, appear to be able to hold in memory only one chunk at a time, suggesting the use of a flattened hierarchy instead of sub-menus [7].On the other hand, studies of adult users demonstrated that navigation by colour buttons works well, with colours and labels used carefully to match buttons of the remote only when this correspondence is intended [1, 2]. Since pre-school children can match and name four primary colours [8] the use of colour buttons should also be explored on applications for young users. As in applications for adults the core functionality needs to be presented consistently and be available throughout the application [10] with icons always visible and placed in the same position [9]. Children usually focus on the middle of the screen so important icons should be placed in the middle of the page [7]. Kindergarten children are imperfect in the course of their actions and tend to make errors [12]. A proposed solution to help with error prevention, recognition and recovery is to map children's action directly to actions on the screen [9].

Studies on personalized recommendation systems could also offer some inspiration but will need adaptation. The EPG-Board, for instance, integrates a to-watch planner with a message board and a rating and tagging system [5]. The SenSee Framework combines the context with user profile to improve the multimedia consumption experience [3]. And AIMED is another personalized TV recommendation system based on user properties such as activities, interests, moods, experience and demographic information [4].Young children don't read or are not experts readers, so text should be avoided to reduce cognitive load [9]. In this case, a message board could be hard to be incorporated; but rating could be appropriate and recommendations based on user profile and preferences would also be relevant and useful, this could be set by children or by parents with what they think best suits their children's needs.

3 Methodology: Closed Card Sorting with Pre-school Children

Search and classification present a particularly difficult problem for this audience. Bakhshi (2007) recommends two complementary methods in the context of searching for video. The 'bag of words' approach where viewers can search for content by entering keywords and the 'Subject/Genre/Channel' approach where viewers can choose videos within a set of pre-defined categories [13].

The first approach suggested would be hard to be implemented for pre-school children. Even with a speech interface such as that developed in the VISTA project [14], the children's limited vocabulary and possibly idiosyncratic speech patterns would make the task complex for both system and user. Browsing via established categories is more viable and adequate for this age group. However, conceptual categories should not be merely characterized in terms of objective properties of categories members. Human conceptual categories have properties determined by the nature of the people categorizing and have properties that are a result of imaginative processes [15]. We therefore felt the need to involve children in the process to define categories for the EPG: in this way, categories appropriate to the user group would be developed, making their browsing tasks easier to accomplish.

According to Piaget, children from 2 to 7 years old are in the preoperational stage in which they begin to understand the classification of objects. About the age of four years, children begin to classify in a more systematic way, however they are unaware of the principles of this operation, cannot explain why they have done it and cannot carry the out the activity of classification in a fully satisfactory way [16]. For this reason, an open card sorting task, process that involves asking participants to sort items into meaningful groups would not be appropriate. However, children could be asked to sort cards into pre-determined categories (closed card sorting). This procedure can be used to check how well the pre-assigned categories fit children's expectations.

We designed a card sorting activity inspired on the Dimensional Change Card Sorting (DCCS) task frequently used to determine extradimensional shifting abilities in preschool children [17]. In current UK EPG's, children's channels are found grouped under a simple "Kids" heading, with no further sub-categorisation. This indicates a need for further categories to be labelled. For the study, suggested categories were pre-determined based on children's channels websites such as CBeebies, CBBC, CITV and Disney, children's films categories from Amazon and books categories from the International Children's Digital Library. This resulted in seventeen categories: movies, cartoons, music and songs, make and do, animals and nature, fairy tales, super heroes, around the world, TV shows, Cbeebies, CBBC, CITV, Jetix, Disney Channel, Boomerang, Cartoon Network and Nick Jr.

Four children aged three to four years old (two girls and two boys) participated in a pilot experiment. Each child was tested individually in one session that lasted for approximately ten minutes. First the researcher introduced herself and asked the child if s/he watched television and what was hers/his favourite programme. The researcher then explained that she was working on a TV guide for children but she did not know many children programmes so she needed children's help. She asked if the child could help her: if s/he decided not to help or to stop helping at anytime it was acceptable. Following this introduction the researcher described the card sorting activity. Two side-by-side shoeboxes were designed, each with a plastic sheet to display a predefined category plus a slot through which the child posted a laminated card showing a thumbnail screenshot from a programme. Two predefined categories were displayed at a time and the child's task was to post the card in the box they found more appropriate. First the researcher showed the thumbnail and asked if the child recognized it. If they did, the child was asked to post the card into a box. There were icons representing each category but the researcher read out the names of the categories for clarity. If the child did not recognize the screenshot, the researcher would show another picture of the same category. If the child did not recognize any screenshot available for a determined category, they were asked to choose a screenshot and place it on the box it would be most appropriate. Two new categories would then be displayed. In all, twelve screenshots would have been sorted.

This card sorting activity was a pilot test conducted with a very small sample. Further studies will need to be carried out to produce results that are statistically significant, but there were interesting outcomes that are worth reporting. Some screenshots such as the cartoon "Pingu" were identified instantly and inserted immediately in the correct box, with the child stating that it should go in the "CBeebies" not in the "CITV" box. The four children also identified the "Club House Disney" picture and put it in the "Disney" box instead of the "CBBC" box. This is evidence that young children do associate audiovisual content with the channels on which they appear. The "Wild Show" was a screenshot with two deer: no participant recognized the show but they all said it belonged to the "Animals and Nature" not to the "Cartoons" box. Thus these children could manipulate abstract categories.

No single screenshot was assigned to the wrong category by every child but the three last screenshots to be presented were mis-classified by all but one child. This may be due to tiredness at the end of the session, but it could also be due to wrongly pre-determined categories. In the case of "Teletubbies", for instance, the choice was between "Super Heroes" or "TV Shows" and three children promptly chose the "Super Heroes" box. If children really consider "Teletubbies" super heroes they would certainly expect it to be in the super hero's category and the EPG should display that, reflecting children's views and choices. However, the most important result is that preschoolers are capable of categorising videos. As we would predict from psychological studies [16, 17] the four year olds performed the card sorting task better than the three year old children. The two four year olds in the pilot test placed nine out of twelve screenshot in the correct box. Given that children of this age are capable of categorisation, they could benefit from an EPG that take into account their developmental level, preferences and concept of categories.

4 Future Work

As the amount of video available for children increases the EPG will have to be modified to meet preschooler's needs and allow them to find content according to their preferences. Guidelines suggested for iTV applications, personalized recommendation systems, interaction design for children along with children developmental theories could inform an appropriate design. But they do not replace user involvement during the design: the most important elements of the Electronic Programme Guide, such as structure and categorization, should have direct input from the end-users. The method described here is one way of eliciting this design information and will be refined for application in further work.

Acknowledgments. We would like to thank the collaboration of children and staff from One World Nursery.

References

1. Daly-Jones, O.: Case Study: The usability of electronic programme guides. In: Gawlinski, M. (ed.) Interactive Television Production. Focal Press, Oxford (2003)
2. Eronen, L., Vuorimaa, P.: User interfaces for digital television: a navigator case study. In: Proceedings of the working conference on Advanced visual interfaces. ACM Press, Palermo, Italy (2000)

3. Aroyo, L., Bellekens, P., Bjorkman, M., Houben, G.-J., Akkermans, P., Kaptein, A.: Sensee Framework for Personalized Access to TV Content. In: Cesar, P., Chrorianopoulos, K., Jensen, J.F. (eds.) Euro iTV, pp. 156–165. Springer, Amsterdam (2007)
4. Hsu, S.H., Wen, M.-H., Lin, H.-C., Lee, C.-C., Lee, C.-H.: AIMED- A Personalized TV Recommendation System. In: Cesar, P., Chrorianopoulos, K., Jensen, J.F. (eds.) Euro iTV, pp. 166–174. Springer, Amsterdam (2007)
5. Iatrino, A., Modeo, S.: EPG-Board a Social Application for the OmegaBox Media Center. In: Cesar, P., Chrorianopoulos, K., Jensen, J.F. (eds.) Euro iTV, pp. 31–36. Springer, Amsterdam (2007)
6. Hynd, A.: Evaluating four and five-year old children's responses to interactive television programs, Vol. Doctor of Philosophy. Murdoch University, p. 313 (2006)
7. Hutchinson, H.B., Bederson, B.B., Druin, A.: The evolution of the international children's digital library searching and browsing interface. In: Proceeding of the 2006 conference on Interaction design and children. ACM Press, Tampere, Finland (2006)
8. Sheridan, M.D., Frost, M., Sharma, A.: From Birth to Five Years: Children's Development Progress. Routledge, London (1997)
9. Chiasson, S., Gutwin, C.: Design Principles for Children's Software. Computer Science Department, University of Saskatchewan (2005)
10. Gawlinski, M.: Interactive Television Production. Focal Press, Oxford (2003)
11. Gilutz, S., Nielsen, J.: Usability of Websites for Children: 70 Design Guidelines. Nielsen Norman Group, Fremont (2002)
12. Mansor, E.I.: 'My world(s)': a tabletop environment to support fantasy play for kindergarten children. In: Proceeding of the 2007 conference on Interaction design and children, pp. 193–196. ACM Press, Aalborg, Denmark (2007)
13. Bakhshi, O.: The Implications of Convergence on EPG Design: Enabling the Mediation of Content through Converged Technologies. In: Lugmayr, A., Golebiowski, P. (eds.) Euro iTV, pp. 9–16. TICSP, Amsterdam, The Netherlands (2007)
14. Carmichael, A., Petrie, H., Hamilton, F., Freeman, J.: The Vista Project: Broadening Access To Digital TV Electronic Programme Guides. PsychNology Journal 1, 229–241 (2003)
15. Lakoff, G.: Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. The University of Chicago Press, London (1987)
16. Smith, P.K., Cowie, H., Blades, M.: Understanding children's development. Blackwell, Oxford (2003)
17. Kloo, D., Perner, J., Kerschhuber, A., Dabernig, S., Aichhorn, M.: Sorting between dimensions: Conditions of cognitive flexibility in preschoolers. Journal of Experimental Child Psychology (2008) doi:10.1016

Interaction Design in Television Voting: A Usability Study on Music TV and Input Devices

Ralph Riecke[1], Alex Juers[1], and Konstantinos Chorianopoulos[2]

[1] University of the Arts Berlin, Germany
[2] Ionian University, Greece
alex_juers@t-online.de, ralph.riecke@gmx.de, choko@ionio.gr

Abstract. The aim of this work is to study the usability of voting on music TV channels. We asked subjects to perform a voting-task on two different music TV shows. The results indicate, that 1) there are small differences in acceptance and understanding of the voting-instructions between users and non-users, 2) the mobile phone is a familiar and the most preferred voting-device and 3) sociability features is a way to support the pricing model of voting services for entertainment applications in TV.

Keywords: Usability, voting, interactive TV, music TV, mobile phone.

1 Introduction

Voting over TV shows has been a popular [3], but controversial kind of TV interactivity [4]. Call-In shows like “Who wants to be a millionaire” and voting shows like TRL are only two out of hundreds different formats of interaction between the TV channels and the viewers. Especially music television channels recognized the potential of those formats for their program very early. They started to produce some new TV shows that are based on voting – Get the Clip on VIVA and TRL on MTV are the most popular ones for German viewers.

Previous works have studied user interaction with music TV, such as information overlays and animated characters [2]. Researchers have also studied voting in the context of e-goverment systems [1], but there is no research on the entertainment aspects of voting in the context of a television show (e.g., Music TV, reality show).

Interactive voting-shows offer lots of possibilities to companies and viewers. Usability plays a major role in interactivity. It took the producers some time to develop useful formats and voting instructions to motivates the audience to use this modern way of interacting – and also spending money. As a matter of fact, music channels and telephone companies earn millions of dollars with voting and some viewers more and more incur debts. So is this topic all about making money?

Besides the ethical considerations, how easy is the voting, what are the main functions and how clear is the pricing presented in the shows? Do all of the developed and existing instructions help the user in praxis to make voting easy or Interactive TV Voting-Shows on TV music channels in terms of usability spending money easily? Are the users able to handle the tasks of voting and realize the costs with the offered instructions? Or are they useless? All of those questions are still to answer.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 268 – 272, 2008.

2 Method

We employed usability testing to explore the above research questions.

2.1 Test Material

Viva – Get The Clip VIVA was the first TV Music Station to broadcast in German VIVA focuses on its target group of feminine teenagers. Get the Clip is an interactive Music-Format on VIVA which empowers the viewers to act like program directors. It works like a kind of a playlist which is build together by the viewer. The Show is aired seven days a week for about three hours a day. The viewer has the possibility to vote by telephone call and SMS. The clip-list with all the video-numbers is shown on teletext and channel-website. In addition, users can send an instant message while voting with SMS. On the TV screen the viewer can see the messages, the playlist und some Information about the running top three of the voting in real-time with the music video-clips. The top three are placed with percentages on the upper part of the screen. Short messages send by the viewer and voting instructions are shown at the bottom of the screen.

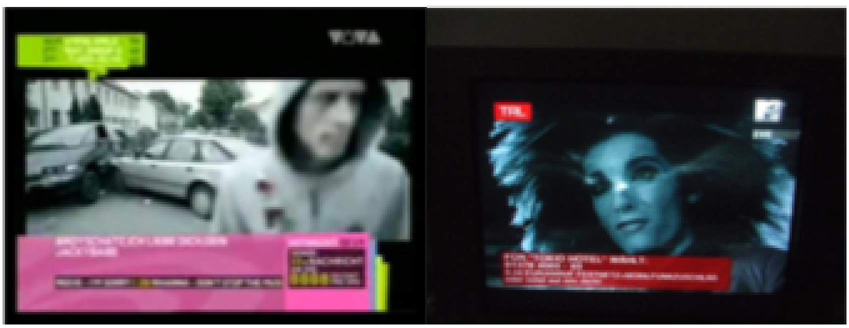


Fig. 1. Get the Clip voting show on VIVA and TRL voting show on MTV (Germany)

MTV – TRL MTV stands for music television owned by MTV Networks (Viacom). In Germany it is on air since 1997. MTV focuses on its target group of more edgy boys. TRL is a live show hosted by a 'VJ' (Video-Jockey) on MTV, where the viewer can select from a sample of the newest clips by voting. Before the actual show starts, each viewer can vote 5 out of 40 songs online. This only works on the MTV Website and is not featured on TV. By the beginning of the TV show the top ten are created out of the pre-voting. During the show the viewer has the possibility to change the top ten chart-list ranking online or per call. At the time of the study a phone-call counts 3 times more than an online vote. The viewer can see the ranked chart-list position of the actual played song in the left corner of the TV screen. All the available videos for voting and the voting instructions are shown in a red box in the bottom area of the screen.

One significant difference between the two shows (Figure 1) is that the videos played on VIVA are overlaid all the time by percentages, messages, voting instructions and the playlist whereas the clips on TRL are shown mostly without numbers and text-boxes, which only appear occasionally.

Both shows were recorded a few days in advance to make sure that every user received the same content. Therefore the text written by the users couldn't be displayed on the TV Screen. Also a few options weren't able to take into consideration because they are only available when the show is on air.

2.2 Study Set-Up

The study took place in an relaxed setting (Figure 2). The same qualitative environment and the same technological set-up for every subject ensured a controlled environment. The voting devices telephone (traditional) and notebook with internet connection (Windows software package) were provided. On the Notebook the websites of the two music channel were already opened to avoid mistakes in getting to the websites. The mobile phone belonged to the subject because there can be some differences between the usability of mobile phones (Figure 2). All the arising costs were taken over by study organizers. The main limitation of this study is that there was no remote control interactivity for the music TV channels.

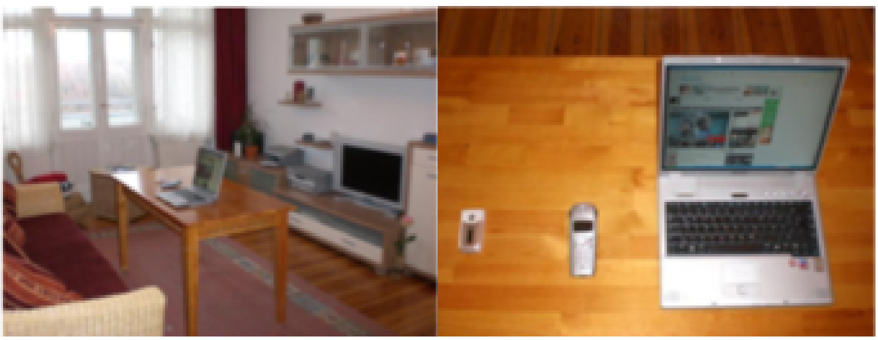


Fig. 2. Experimental set-up and input devices employed in the study

2.3 Subjects

The recruitment of subjects was facilitated by a screening procedure. All subjects were recruited from a school class consisting of 31 pupils. 29 of them confirmed that they know about voting. 4 girls and 5 boys out of the 29 pupils have voted at least once. In the end, 3 boys and 4 girls agreed to take part in our exploration. One of that girls didn’t have any experiences concerning music channels, so we decided to exclude her, in order to keep a balance between genders. The ages were between 13 and 20. Overall, six users with voting experience and two without were recruited.

2.4 Testing Procedure

Each subject had to perform the following task (translated from German): “In the next 10 minutes you will see two voting shows from two different music channels. Each one of the shows will be presented over a duration of 5 minutes. Your goal is to vote for one clip of your choice in each show. You can freely choose one or more out of three voting devices: mobile phone, or telephone, or a notebook with an internet connection."

The study organizers payed for the cost of voting, in order to ensure that users would not hesitate to do the voting. In terms of acceptance, we found users would accept a pricing model which lies between 9-17,5 Cent for each SMS. Nevertheless, the economic aspects of voting is outside the scope of this study and more tests should be done to predict acceptable price range.

Task completion was observed with the help of an observation checklist. We measured time to watch a clip completely (measured average time of a clip: 2 minutes 48), time to read the instructions, and time to vote with any of the given devices.

Further we considered that the subject may start again after an incomplete attempt of voting which is defined as missing out of one of the major steps for voting – wrong chosen clip number, wrong usage of the voting tool and wrong sending procedure.

After the subject accomplished the task, a questionnaire-based interview followed. The interview provided more detailed information about facts like voting-instructions, device usage and preference, screen design and the pricing.

3 Results

All subjects solved the voting task on each TV music channel within the allocated time. Average time of the users was 3:24 minutes and of the non-users 4:24 minutes. The preferred voting tool for all subjects was the mobile phone. The notebook connected to the internet was used only two times by the subjects to look for voting instructions whereas the telephone wasn't used at all. The voting instructions provided by the TV music channels (on screen and online) were mostly used by the non-users.

The pricing model of both TV music channels was clearly visible for 5 subjects. 3 subjects weren't able to remember the costs for their voting procedure. More than 80% of the subjects would vote again and even more often, if a lower price was offered. The average suggested price for the two shows was: 9 Cent on TRL and 17,5 Cent on "Get the Clip."

Although, the task was easy to solve, non-users had a few problems and therefore reviewed the task as medium-hard to solve. One non-user criticized that "the voting instructions on both music channels were not clear and therefore confused" him whereas the other non-user liked the way TRL formulated their voting instructions but disliked the confusing screen design with "too much information and too many symbols and colors" of "Get the Clip."

Although the subjects had the possibility to choose out of three voting devices all of them used the mobile phone for voting. In general, all subjects explained their decision towards the mobile phone with following statement: "It's is my daily companion and I'm used to it."

4 Discussion

Based on the results of the study we draw some implications for theory and practice. Furthermore, we provide suggestions for further research in this direction:

Television voting is an established practice: It was interesting to see that all subjects solved the task; even the non-users did so.

Mobile phone as a multipurpose tool (remote control): All subjects preferred their mobile phones for voting because in everyday life “it is always close to” them. Indeed, always and everywhere teenagers communicate with their mobile phones However, our study did not include the option to interact directly with the TV through a remote control, which might have been a close competitor to the mobile phone.

Online voting instructions are not enough to ensure usability: There are two different kinds of voting instructions – on TV screen and on the website of the two TV music channels. We found out that most users didn’t look for the online instructions, but further research (e.g., eye tracking) is needed, in order to examine what captures the attention of the users during voting.

Perceived economic value of social TV: There was a major difference between the two music shows concerning the desired price for voting. Users would spend nearly double of the amount for the Clip’ as they would spend on ‘TRL’ – 9 Cent for ‘TRL’ and 17,5 Cent for ‘Get the Clip’. The main difference between the two formats is the messaging option provided by ‘Get the Clip’. This outcome could be essential for further development and improvement of voting shows especially in terms of customer satisfaction and business profit maximization.

Future research should study the differences between the remote control and other alternatives for interaction with TV and in particular with voting shows. Especially, how the television remote control measures up against the mobile phone? Moreover, further research should consider the differences between voting for entertainment and voting for other aspects (e.g., public issues). Finally, changes in technology and consumer needs should be taken in consideration by improving the existing or developing completely new formats, like mobile TV voting shows.

References

1. Bederson, B.B., Lee, B., Sherman, R.M., Herrnson, P.S., Niemi, R.G.: Electronic voting system usability issues. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2003, pp. 145–152. ACM, New York (2003)
2. Chorianopoulos, K., Spinellis, D.: Affective usability evaluation for an interactive music television channel. Computers in Entertainment 2(3), 14 (2004)
3. Slot, M.: Changing user roles in ICT developments; the case of digital television. Telematics and Informatics 24(4), 303–314 (2007)
4. informitv, Big Brother drops text votes and cuts call costs, `http://informitv.com/articles/2007/06/01/bigbrotherdrops/`

What You Expect Is What You See

Dirkjan Joor, Wilco Beekhuizen, Lidwien van de Wijngaert, and Pascal Ijegalu

Utrecht University, Dept. of Information and Computing Sciences,
Utrecht University, P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
d.joor2@chello.nl, wbeekhui@cs.uu.nl, l.vandewijngaert@cs.uu.nl, pjeijega@cs.uu.nl

Abstract. In this paper an experiment was conducted to measure the effect of framing a high definition television (HDTV) clip. One group of participants was told they were watching a brand new HDTV clip while the other group was told they were watching a digital DVD clip. Both groups were in fact watching the same (low) quality DVD clip. After watching this clip the beliefs of participants and their viewing experience was measured through a questionnaire. A significant more positive viewing experience for people framed to watch the HDTV clip was found. This means participants were unable to discriminate properly between digital and high definition signals and were influenced by the frame set for them. This effect has been shown in many different situations and now it has been established for watching HD-quality television as well. The results still indicate that the HDTV-frame is already associated with a high quality viewing experience, which may influence the selling strategy and/or adoption speed of this technology.

Keywords: Framing, value-expectancy theory, HDTV, Diffusion of innovation.

1 Introduction

High Definition Television (HDTV) is aimed to provide a noticeable higher quality level of display for observers. This is because HDTV offers a higher resolution signal compared to standard television, Furthermore, HDTV is watched on a flat panel and is available in various high resolutions. HDTV in general should result in a higher quality and a better viewing experience for observers (Belgacem, Bizzocchi and Bowes, 2005) [1]. This research does not test if there is a noticeable difference between a high definition television signal and standard digital signals, nor does it look for a noticeable difference in quality between different HDTV formats. This research is of a more psychological nature as it investigates the relation between expectations, viewing experience and beliefs. It will find an answer to this question: “Will telling viewers that they are watching a HDTV signal lead to a higher quality of their viewing experience?” To answer this question we use framing theory as described by Druckman [4] and Nelson, Oxley and Clawson [6,7]. The assumptions of framing theory state that to influence the attitude, the frame is a more direct mechanism than belief change. The attitude is a result of both the frame and ones individual beliefs (Nelson et al. [6,7]). Intuitively this experiment can be grasped by making the parallel with the tale of the

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 273–282, 2008.

new clothes of the emperor by Hans Christian Andersen. Setting a frame is highlighting certain aspects of reality to direct the attention of the information receiver. The scientific relevance of this research is that framing theory is used to direct the participants in their attitude toward a viewing experience. Previous research has focussed on the opinion of participants towards a topic such as abortion (Nelson et al. [6,7]) or free speech (Druckman [4]). The outcome of this research will show if experience itself, or at least the assessment of this experience, is moderated by framing. Within the viewpoint of HDTV from a user and consumer perspective, the outcome of the experiment may also reflect on the ability of the term "HDTV" to influence consumers into expecting higher quality viewer experience and is therefore relevant in studying the individual decision process (Rogers [9]) as the ability to change the attitude of the consumer plays a role the persuasion process for users that frame HDTV as a pleasant viewing experience, whether this is true or not. Framing a viewing experience as HDTV reveals the current ability of the HDTV- concept to affect the attitude in of possible consumers in the persuasion phase. When the framing effect can be proven, this would also emphasize the role for mass media to frame an innovation and thereby increase expectation to stimulate the adoption of an innovation (Le Book and Barnett [5]; Rogers [7] p. 170). Finally it is interesting to see from a provider perspective whether image improvement from a technical perspective is actually fully responsible for a perceived experience, or whether the expectation alone already influences the experience. In that case other features of digital television may be more interesting to provide first than the improved quality offered by HDTV.

2 Theoretical Framework

While the quality of experiencing television is measured relatively often (Rohali et al., 2000), [10] in product reviews by either supplier or consumer and can be readily measured using quality attributes (Péchard, Sylvain, Callet, Le Carnec and Barba, 2006) [8], the concept of expectancy is not as direct to grasp as it is of a more psychological nature. To form a conceptual model for measuring expectancy, the concepts of "Beliefs", "Framing" and "Attitude" are used as the basis for the theoretical framework. These concepts are described by the research from Nelson et al. [6,7]. The process of framing is well researched in fields such as psychology, political science, and communications studies but definitions differ. For the purpose of providing clarity, they use an algebraic model to explain the relation between the underlying concepts of expectancy:

$$A = \Sigma v_i * w_i \tag{1}$$

In this formula the A stands for attitude, while the concept of attitude is not explicitly defined in Nelson et al. [6,7], within the context of this experiment "attitude" is assumed to be the same as "expectancy". Using this assumption, expectancy can be seen as the total sum of v (value beliefs) and w (weight) of each i (attitude object). They use the framing of a message as the factor that is telling information receivers how to weight their beliefs. Following this reasoning, respondents that view an experience under the HDTV-frame should also increase the weight of their beliefs of that the image is high quality and thereby have a better expectation of the movie experience. This leads to the H1 hypothesis:

– H1: Individuals that have been framed that they are watching a HDTV signal, will have a higher expectation and therefore a better quality viewing experience than those that have been framed to belief they watch a standard digital signal.

The first hypothesis measures only the effect of increasing the weight (W) of beliefs by the frame. Following the formula, respondents are likely to differ in their beliefs about HDTV however. One respondent may believe that the quality of colour is increased, while another one believes colour does not differ too much, if anything at all from earlier television. When a respondent has a low belief of the quality increase of HDTV, this should then cancel out the framing effect. When a respondent does believe HDTV has a better quality, this should then be noticeable in their experience evaluation. This leads to the H2 hypothesis:

– H2: A higher belief in the quality increase of HDTV compared to standard digital television results in a higher appreciation of the viewing experience for the group that is framed to watch HDTV, while this result is not visible in the group that is framed to watch standard digital television.

The second hypothesis does take in account both the beliefs V and frame W from the formula. It does not relate this to the individual attitude objects (i) however. Attitude objects in the context of HDTV are single quality attributes such as the colour. According to the formula, the added sum of each attitude object (belief * frame) will result in the final attitude. This results in the third and final hypothesis:

– H3: The beliefs that are weighted by the HDTV-frame should have a positive relation with their identical experience item.

An example of H3 can be illustrated in the formula (2) by using the attitude object of “image colour”.

$$A_{colour} = v_{color} * w_{color} \tag{2}$$

In this research, the attitude (A) is measured indirectly through the viewing experience, the weight (w) is assumed to be a direct result of the HDTV-frame and the belief (v) is measured directly. Because attitude objects require measuring both the experience and belief about the object, which might be repetitive for participants and thus only measuring the consistency in answering questions, this research limits itself to only three matching beliefs and experience items: colour, motion and realism that were asked among other beliefs.

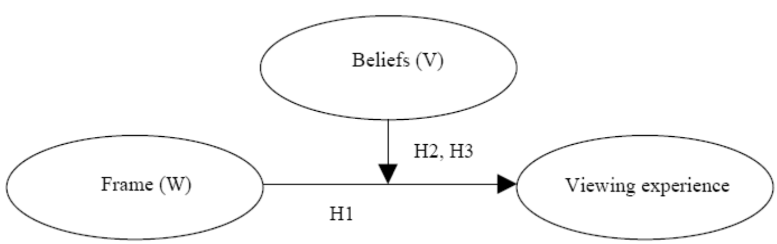


Fig. 1. Conceptual model for the relation of expectation and experience

Combining the previous three hypotheses with the underlying variables results in the conceptual model of Figure 1. The concept of attitude and the synonym concept of expectation are a latent variable that results from the combination of the belief and frame combination that influences the viewing experience.

Staples, Wong and Seddon [11] elaborate on expectations (related to information systems) using the disconfirmation theory. In short, this theory measures perception compared to expectations: (unrealistically) high expectations will result in lower perceived benefit. This theory offers a possible explanation when setting a high expectation by framing backfires. However, to formulate the hypotheses, a positive effect of framing is assumed.

3 Method

The viewing experience and belief of participants was studied by participating them to watch a DVD clip on a high definition television (Full HD). One group of participants (N=30) was told they were watching a HDTV clip, others (also N=30) were told they were watching a standard digital DVD signal. Both groups were in fact watching the same quality DVD signal. More information on the video clip is provided in appendix B.30 participants were used for each group. All participants were informed they were watching a clip on a flat panel device. Both groups had the same flat panel and the same signal, but the groups were told they were watching either a HD signal or a DVD signal. All participants are Dutch media and communication students approached at the hallways of the Dutch “Hogeschool” in Utrecht, Netherlands. The participants in this convenience sample are estimated between 19 and 25 years of age. Participants were taken into a room individually, or with at most 3 other students. The students in groups were requested to refrain from commenting during the movie play to prevent influencing the other participants during their session, at least verbally to prevent participants influencing each other. All participants were sitting about 4 meters of the screen. In the room we placed four posters indicating which signal they were framed to watch. The group framed to watch HDTV had the addition of a thick cable going from the television to a computer, as if the signal was coming from a source other than the DVD player. A strong determinant of frame effectiveness is the credibility of the one setting the frame (Druckman, 2001) [4]. Credibility is defined

Table 1. The two frames created for the experiment

Frame	Text	Accessories
HDTV	“Welcome. You are about to watch a HDTV clip on a brand new HDTV. HDTV is the future of television as it promises a better television experience. We research the quality HDTV offers by letting you watch a new HDTV clip followed by a short questionnaire in which you can indicate how you experienced this clip.”	Flyers, fake cable connected to pc
Standard digital frame	“Welcome. You are about to watch a standard DVD signal on a brand new HDTV. HDTV is the future of television as it promises a better television experience. We research if the old DVD signal is still good enough on a new HDTV, or if you really need the newest HDTV signal for a pleasurable viewing experience. After watching a short DVD clip you will receive a questionnaire on which you can indicate how you experienced this clip.”	Flyers

with two dimensions 1) that the speakers target must believe that the speaker has relevant knowledge and 2) that it is believed that the speaker reveals everything relevant of what he or she knows. Therefore a credible story must be constructed and understood by the performers of this experiment beforehand to answer questions for the inquiring participants.

Table 1 illustrates both frames. Both were communicated by spoken language, in Dutch as all participants were Dutch. The questionnaire was available in both Dutch and English; all participants used the Dutch version. The full questionnaire is included in Appendix B. The questionnaire shows the items that were used to develop the constructs that were used. Cronbachs' alpha reveals that by leaving out sound and realism experience, a construct of 4 items can be formed for experience (alpha = 0,74). The data items for beliefs will be used separately as no satisfying alpha is found. The DVD clip shown to both groups is a 90 second clip of "Life in the Blue: HD" [2]. The first 90 seconds of this clip were shown to all participants. The clip shows the text "Live in the blue in HD". The second group was told the truth: the original version was converted to DVD for this research; the first group (framed to watch HDTV) was not made aware of this. The clip showed an underwater footage, the demo was obtained through the procedure described in Appendix A.

4 Results

From the descriptive data it is found that 20 of a total of 60 participants owned a flat screen themselves and 11 participants were expecting to buy one in the coming twelve months. Analysis shows that the experience does not differ significantly between owners of a flat screen television and non owners, although owners did have a higher extend of belief that HDTV was worth paying extra for (t = 2,69, significant at 0,01 level). To analyze the difference in experience submitted by people that were told to watch HDTV and standard digital signal, the H1 hypothesis uses an independent sample T-test. The standard digital group valued the experience construct with an average 4,98, while the HDTV group valued this a full point higher at 5,98. The T-value resulting is 4,56 which is significant at 0,001 level. For the individual experience items the difference the groups is significant at the 0,05 level for colour (t = 2,55), 0,01 for both movement (t= 2,99) and sharpness (t= 3,26) and significant at 0,001 for brightness (t=4,57). For the experience items of sound (t= 1,78) and realism (t= -0,40), the difference between groups is not significant. The alpha value already indicated that sound and realism did not respond to the other experience items and are not used to measure experience. Using the first four experience items, the hypothesis 1 can be accepted.

Hypothesis 2 is tested by measuring the correlation value of the individual beliefs with the experience of each group. For the HDTV this is expected to be positive to confirm H2, while there is expected to be no effect on the correlation at the standard digital group for H3. The results of the analysis are shown in Table 2. Although there is a significant correlation between belief and experience in the HDTV-group, the other values in the table reveal that the values between groups are too similar to give a positive validation on the framing effect and belief combination. This may be explained because beliefs were asked about the signal and the screen of HDTV, while the frame only focused on the signal.

Table 2. Correlation between belief and experience for the differently framed groups

Frame	Sharpness of HDTV compared to normal	Colour of HDTV compared to normal	Realism of HDTV compared to normal	Belief of quality caused by signal	Belief of sound improvement from HDTV	Belief of worth paying extra
HDTV	0,129	0,379(*)	0,325	-0,122	0,016	-0,020
Standard	0,235	0,311	0,237	-0,172	0,122	0,146

N=30 per group, *Significant at the 0,05 level.

Hypothesis 3 is another attempt to find evidence for the relation between of frame and belief, the value-expectancy formula is applied directly for the attitude objects of colour, sharpness and realism. For these attitude objects a corresponding experience items and belief items are measured in the survey. The H3 hypothesis does therefore not measure the effects of the experience items of movement, sound and brightness, as there are no corresponding beliefs in the survey to prevent repetition for respondents. The attitudes consist of multiplying the beliefs of each object with the frame. To create a scale based on the nominal variable of frame, the HDTV frame is given an arbitrary value of 1,5 (value of 2,0 has similar results) and that of standard digital a value of 1. To quantify the "correlation" of HDTV and standard digital for illustration purposes, the HDTV cases are also given the value of 1,5 and standard digital of 1. Table 3 shows the prediction value of attitude (=expectancy) objects compared to the frame. As can be seen in Table 3, the correlations slightly increased when frame and belief are combined. The exception is the correlation of realism, although this question that measure the experience item was reported to be ambiguous by several participants however and was already taken out of the experience construct variable.

Table 3. Correlation value of frames, beliefs and attitude for each experience item

Experience item / Independent variable	Colour experience	Sharpness experience	Realism experience
Corresponding belief item (color, sharpness, realism)	,148	,186	,226
Frame: HDTV = 1,5 SDTV = 1,0	,317*	,393**	,052
Expectancy (belief * frame)	,344**	,413**	,203

N=60, * Significant at the 0,05 level, ** significant at the 0,01 level (2-tailed).

Although there is only a slightly stronger correlation between the expectancy product (belief * frame) compared to just the frame, the attitude objects do have a reasonable correlation. The H3 hypothesis can therefore be accepted.

4 Conclusion

The group of participants that was told that they are watching HDTV indicated a higher appreciation of their viewing experience than the group that was told that they watch a standard digital signal. At the basis of this research was the value-expectancy formula as the hypotheses were formed around the concepts in this formula. Using only the concept of framing, the H1 hypothesis proves the guiding effect of these frames in the expectation of the participants. The H2 hypothesis brought the beliefs of the participants concerning HDTV in the equation and predicted that participants with a higher belief in the quality increase from HDTV compared to standard television appreciated the viewing experience better when told it was HDTV. Although a correlation was found with the belief of "better colour than normal", the correlating values were most similar to the group that was told that they watch standard digital television. For this reason the H2 hypothesis was rejected. An explanation for that the belief and frame combination did not have a strong prediction effect is that the beliefs questions measured the whole HDTV (signal + television) and the frame was only about the signal. The H3 hypothesis also measured the predicting effect of combining beliefs with the frame, but did so using individual attitude objects. Being a cross product of both frame and beliefs, these attitude objects did show a stronger correlation toward the viewing experience than the frame or beliefs alone. The limitation of H3 is that of H2; the beliefs may not correspond with the frame and therefore limits the combining effect. Another limitation of H3 was that there were only three attitude objects for which the beliefs matched the experience items. This was done to prevent measuring the "consistency" of participants which may result in that similar looking questions get a similar answer. This could have been prevented by using a pre-experiment questionnaire for beliefs, but was not done due to time constraint of the experiment. A final limitation of this research is the limited sample size and the measuring of demographic data. One aspect that could have been measured was the distribution of gender among groups during experimentation; this is estimated to be slightly higher for females in both groups but not significant in either one. For future experiments the addition of gender is suggested as this could influence framing effect.

The relevance of this experiment for framing theory is that the attitude or expectation is shown to be strongly effected by framing. This was proven before (Nelson et al. 1997, Druckman, 2001), but this research showed that framing can be used in the wide sense as interpreting information in the outside world instead of just interpreting a communicated message. The information in this case was the movie fragment that was interpreted as being high quality because of the "HDTV-frame". Although framing- theory was chosen for the strong potential for influencing people, this research did not apply it in the traditional way of presenting a message with a certain emphasis on one aspect. The experiment used the "HDTV-frame" as a collection of aspects that are associated with a viewing experience of high quality. Such a collection of emphasis on reality might have been described by an "illusion" or similar term. Because of the time constraint, the exploration of literature stayed within the field of communication and information. Therefore there may have been other theories from the field of psychology or sociology that can been an addition to framing theory. One outcome of this research is that it is difficult for participants to objectively judge television quality. This research provides evidence that participants are generally unable to see the whether there is really a high definition or standard digital signals without reference. One application of framing for cable providers may be to frame people into believing digital cable is almost high definition or at least of high quality in its own right. Cable providers can consider to provide features such as interactivity to their users before delivering high definition content through their set top boxes. That participants are unable to discriminate properly between HDTV and standard digital without having a reference, does not deduct from any real technical improvements from HDTV that improve a movie quality. It does however illustrate the role of framing on expectation, which puts a powerful role on the mass media in the persuasion phase of the decision process from consumers (Rogers, 2003). Advertisement may therefore stimulate the early adoption process by providing a clear focus on the aspects that are of interest to the consumer and stimulate the positive-association with the new technology. The group of respondents were all students from a higher vocation education. When drawing on the H1 hypothesis it can be said that this positive-association for HDTV is already strongly present among this group and individual decision making for them is likely to focus on later phases of the process. For consumers with different demographs such as age and education level, this may differ however. Another interesting addition is the effect of framing backfire on the outcome. Although we mentioned that HDTV offers a major improvement in quality we have not exaggerated this too much. It would be interesting to measure a possible backfire effect by setting unrealistically high expectations. Finally, this research can be repeated after HDTV is more fully diffused. Because the participants will get more used to HDTV it can be expected that the frame becomes less effective and it then becomes easier to discriminate between the quality of an HD signal and a standard digital DVD signal.

References

1. Belgacem, B.Y., Bizzocchi, J., Bowes, J.: The future of video: user experience in a large-scale, High definition video display environment. In: ACM International Conference Proceeding: International Conference on Advances in computer entertainment technology, Valencia, Spain, vol. 265, pp. 204–208. ACM Press, New York (2005)
2. Life in the blue in HD. Retrieved, from the World Wide Web, `http://www.apple.com/quicktime/guide/hd/bbc-blue.html`
3. Belgacem, B.Y., Bizzocchi, J., Bowes, J.: The future of video: user experience in a large-scale, High definition video display environment. ACM Press, New York (2005)
4. Druckman, J.N.: On the Limits of Framing Effects: Who Can Frame? The Journal of Politics 63(4), 1041–1066 (2001)
5. Le Book, C., Barnett, B.: PCTV: Consumers, Expectancy-Value and Likely Adoption. The International Journal of Research into New Media Technologies 12(3), 325–339 (2006)
6. Nelson, T.E., Oxley, Z.M., Clawson, R.A.: Toward a Psychology of Framing Effects. Political behavior 19(3), 221–246 (1997)
7. Nelson, T.E., Oxley, Z.M.: Issue Framing Effects on Belief Importance and Opinion. The Journal of Politics 61(4), 1040–1067 (1999)
8. Péchard, S., Sylvain, S., Callet, P., Le Carnec, M., Barba, D.: Towards video quality metric for HDTV. In: Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, United States, pp. 1–6 (2006)
9. Rogers, E.M.: Diffusion of Innovations, 5th edn. Free Press, New York (2003)

10. Rohaly, A.M., Corriveau, P., Libert, J., Webster, A., Baroncini, V., Beerends, J., Blin, J.-L., Contin, L., Hamada, T., Harrison, D., Hekstraa, A., Lubin, J., Nishida, Y., Nishihara, R., Pearson, J., Pessoa, A.F., Pickford, N., Schertz, A., Visca, M., Watson, A., Winkler, S., Ngan King, N., Sikora, T., Ming-Ting, S.: Video Quality Experts Group: Current Results and Future Directions. In: Proceedings of SPIE Visual Communications and Image Processing, Perth, Australia, vol. 4067, pp. 742–753 (2000)
11. Staples, D.S., Wong, I., Seddon, P.B.: Having expectations of information systems benefits that match received benefits: does it really matter? Information & Management 40, 115–131 (2001)

Append ix A - Video Clip

The original video used is the 720p HD version. This video uses the H.264 codec with a framerate of 16.216 per second. The video was converted to a standard PAL DVD format using open source programs. The following programs and versions were used for converting: MPlayer (1.0rc2), transcode (1.0.2), ffmpeg (0.4.9-pre1), mjpegtools (1.9.0rc2), dvdauthor (0.6.11). The following commands converted the original HD movie to a DVD format:

```
mencoder bbc-blue_m720p.mov -ovc raw -oac pcm -MC 0 -o
bbcblue_m720p.mov.new

transcode -i bbc-blue_m720p.mov.new -y ffmpeg --export_prof dvd-pal -
-export_asr 2 -o bbc-blue_m720p -D0 -E 48000,16,2 -b 224 -s2 –m bbc-
blue_m720p.ac3 --export_fps 25 -J modfps

mplex -f8 -o bbc-blue_m720p.mpg bbc-blue_m720p.m2v bbcblue_ m720p.ac3

dvdauthor -o DVD/ bbc-blue_m720p.mpg

dvdauthor -T -o DVD/
```

The conversion to DVD caused the framerate to increase from 16.216 to 25 fps. To maintain the original video speed we used the modfps pre-processing filter of the transcode program. This filter shows some frames for a slightly longer period to fill up to 25 frames per second. This filter causes the video to become slightly shaking or choppy. We found that this choppiness causes no significant negative viewing experience when framed to watch HDTV, as elaborated on in the results.

Appendix B - Questionnaire

The Future of Television

Thank you for participating in our research! The following questions will help us understand more about the quality of television. To reward you for your participation you can win a free cinema ticket!

Viewing experience
Please indicate the most appropriate answer.

1. How were the colors?	Faint	O O O O O O O	Clear
2. How did you experience the motion of the scene?	Shaking	O O O O O O O	Fluent
3. How did you experience the sharpness?	Blurry	O O O O O O O	Sharp

4. How would you judge the brightness?	Vague	O O O O O O O	Clear
5. What do you think about the sound quality?	Very bad	O O O O O O O	Very good
6. How real did the scene look to you?	Artificial	O O O O O O O	Natural

Your future of television

7. Do you have a flat screen at home?	Yes /No
8. Do you think you will switch to a flat screen in the next 12 months?	Yes /No

What do you think HDTV has to offer?

9. To what extent do you think HDTV has a sharper image than a regular television?	Little difference	O O O O O O O	Much sharper
10. To what extent do you think HDTV has better colors than a regular television?	Little improvement	O O O O O O O	Lot better
11. To what extent do you think HDTV has a more realistic image than a regular television?	Little extra realism	O O O O O O O	Lots of extra realism
12. To what extent do you think image and sound quality depend on the signal?	Little influence	O O O O O O O	Lots of influence
13. To what extent does HDTV influence sound quality?	Little influence	O O O O O O O	Lots of influence
14. Are you willing to pay extra for the possible quality improvement offered by HDTV?	Little	O O O O O O O	Lot

Thank you for filling in this questionnaire. To qualify for the free cinema ticket please leave your phone numer or e-mail address below.

‘I Want My HDTV’? Underlying Factors of Perceived Usefulness for High Definition Television

Eva Baaren, Lidwien van de Wijngaert, and Erik Huizer

Utrecht University, Department of Information and Computing Sciences, Padualaan 14, 3584 CH Utrecht, The Netherlands
Baaren@cs.uu.nl, L.vandeWijngaert@cs.uu.nl, Huizer@cs.uu.nl

Abstract. While traditional technology acceptance models concentrate on relationships between usefulness and acceptance, they leave unresolved the questions about why a certain technology is found useful, in which contexts, and by whom. This paper presents an empirical study of some of the factors behind perceived usefulness of HDTV in the Netherlands, where the adoption process has just begun. Results show that HDTV awareness, knowledge about HDTV, daily television viewing habits, household innovativeness and the presence of digital television contribute to perceived usefulness of HDTV characteristics. They also demonstrate that these characteristics accord with Dutch viewer demands for future television.

Keywords: HDTV, Adoption, Perceived Usefulness, Diffusion of Innovations.

1 Theories of Technology Adoption

When examining the history of technology adoption research, many scholars in the social sciences have concentrated on the relationship between personal attitudes towards a new technology and the actual behaviour that derives from these attitudes. They often employ the Technology Acceptance Model or ‘TAM’ [1], which builds upon the theories of reasoned action and planned behaviour, both of which argue that an individual’s attitude towards a certain behaviour and preceptions about the individual’s own performance will determine the actual execution of this behaviour [2, 3]. TAM posits that the intention to use a new technology and its actual acceptance depend upon the manner in which people perceive a technology to be useful and easy to adopt. While this statement has been proven empirically correct for many different situations and technologies, the model itself is limited in its claims when researchers seek to understand technology acceptance or adoption. The model suggests that intention to use a technology is equivalent to actual usage acceptance, which is not necessarily true. For example, lack of accessibility and competing intentions regarding the use of other technologies can interfere. Another point of critique arises because TAM only regards adoption choices as outcomes of personal attitudes towards the technology and does not acknowledge external social influences on these attitudes. Both points of critique were addressed by Venkatesh, Morris, Davis and Davis in their creation of a Unified Theory of Acceptance

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 283 – 292, 2008.

and Use of Technology (UTAUT) model. This model employs TAM as a basic principle, but acknowledges intervening factors, such as social influence, facilitating conditions and user demographics, which can influence intention to use and usage [4]. The UTAUT model, while an improvement on TAM, still leaves one important limitation found in TAM unresolved. TAM and UTAUT both concentrate on the relationship between perceptions of a technology, on the one hand, and the intention to use the technology, on the other; however, they do not address the question as to which general and contextual factors lie behind perceptions of a new technology, specifically in terms of 'perceived usefulness' and 'perceived ease of use'. In order to obtain more insight into these underlying factors, they will be examined in the case of High Definition Television (HDTV). In the Netherlands, HDTV adoption is still in an early stage and hardly any research has been done on its adoption process. The focus in this paper will be on *why* HDTV characteristics are found (un)useful, instead of merely measuring *if* they are (un)useful. Due to limitations on paper length, TAMs perceived ease of use will not be examined.

1.1 Perceptions of Usefulness

With 'perceived usefulness' as the central concept of this paper, the question arises as to *when* HDTV is perceived of as being useful. Seen from a social constructivist perspective, the manner in which people make use of any technology does not (only) depend upon its intrinsic characteristics, but also on the function(s) they create for the technology in their daily lives. Theories of media domestication and appropriation that focus on the manner in which end users adopt, use and reshape media technology, according to their own specific desires and contexts, also assume this perspective [5, 6]. Yet notions about this contextual usage and shaping do not imply that a technology has unlimited possibilities for use. Hutchby argues that each technology has 'affordances'. This term refers both to the different uses people create for a technology, and also to the fact that these uses are limited by specific technical boundaries [7]. From this perspective, it can be said that perceived usefulness of a technology depends upon the manner in which people regard its affordances as useful. For the HDTV standard, current technical affordances that entail usefulness can be divided into superior image quality and better sound quality. However, perceived affordances are also inextricably linked to related technologies and also to devices that share affordances with HDTV, such as the size and shape of televisions. New widescreen TVs have evolved as part of a new television and 'home cinema' aesthetic, and are prominently displayed in the living room. As these sets are also HD-compatible, the perceived usefulness of the size and aesthetics of a new TV (along with image and sound quality) can thus influence adoption of HDTV units.

1.2 Factors That Influence Perceptions of Usefulness

Having specified perceived usefulness of HDTV in terms of image sharpness, sound quality, and screen size, the conditions that influence the manner in which people rate these three characteristics remain open for evaluation. At the time this article was written, HDTV in the Netherlands stood at the beginning of the diffusion process: although HD compatible televisions were already being sold, prospects for viewing broadcast HD content were still limited and peripheral equipment was not yet standardised. According to Rogers, the adoption stage of an innovation's diffusion process consists of chronological sub-processes, including attainment of knowledge, persuasion in favour of or against the innovation, the decision to adopt, actual adoption, and confirmation of the 'correctness' of the decisions [8]. Understanding the factors behind the usefulness of HDTV in the Netherlands therefore indicates that the focus should examine the stages of knowledge and persuasion prior to the point at which the adoption decision is made. The persuasion stage itself depends upon five criteria of personal technology perception: 'relative advantage', 'compatibility', 'trialability', 'observability' and 'complexity' [8]. In contrast, Lin offers a non-chronological overview of the categories whose factors influence technology adoption [9], according to which system factors, such as social influence, post-adoption situations, personal (psychological), and technological characteristics, play a role in and exert influence on the adoption process. As research into all of the factors mentioned by Rogers and Lin would be too complex and unnecessary (not every factor applies to each technology and to any situation), the next paragraphs will explore and combine some of these factors and their possible importance within the context of HDTV usefulness.

Awareness and actual knowledge

According to Rogers, prior to the formation of an opinion about the usefulness of an innovation, a process of uncertainty reduction has already occurred. In this process, the potential adopter becomes aware of the existence and possibilities of the innovation [8]; however, knowledge is not only a precondition to the formation of an opinion, it also influences the opinion. In their research on the adoption of digital television in the USA, Chan-Olmstead and Chang concluded that misconceptions about the medium diminished the perceived value of DTV [10]. In addition to factual information about the new medium, the necessity of HDTV features can also be provoked by the presence or 'observability' [8] of HDTV as a concept and artefact in society. Advertisers and the direct social environment have the ability to emphasize the importance of HDTV, or they can confront a person with the medium itself [11]. This may cause individuals to reconsider the sharpness, size, and sound of their own television, which may have been satisfactory before. In addition to awareness of the medium's existence, actual knowledge about its capabilites, and which requirements are neccesary for its use, Rogers' concept of 'trialability' may be of importance here as well. Because HDTV has been specifically and explicitly linked with the concept that 'seeing is believing', the essence of image sharpness can hardly be demonstrated without its display. Seeing the medium in action can, therefore, be considered to comprise part of the knowledge about the medium.

Personal compatibility: uses and gratifications of television

In understanding attitudes towards media use, Uses and Gratifications and Media Choice theories argue that the day-to-day use of a medium emerges from people's needs [12, 13]. From this basic assumption, different theories have emerged that address choices in specific content, exposure to a medium, and choices between media. A well known categorization for television content choice appears in the choices made by the ritualized/habitual user (passive, entertainment seeking, frequent) versus those made by the instrumental user (active, information seeking, less frequent) [14. 15]. While the Uses and Gratifications theory has been criticized, primarily due to its premise of conscious user choices and the argument that users do not have to belong to either individual group [15], the separation of user groups may still be plausible for HDTVs perceived usefulness. However, while movies may be of interest, due to the sense of 'being there' or of sitting in a movie theatre, watching a quiz show may not. Also, sports viewers, who sit at the intersection of information and entertainment, may also have special preferences for HD due to perceptions of immediateness. Along with film and television productions, frequent (console) gamers should, for the same reason, be considered as a separate but important group as well.

In addition to content choice, it can be argued that the importance people attach to HDTV features also depends upon the importance they attach to television in general. A study of leisure activities in Dutch households over the last ten years documents that traditional television viewing hours have decreased [16]. People who do not attach much importance to traditional television in their daily live, may, therefore, be in less need of image, sound, and screen size improvements than those who do.

Household compatibility: life cycles and innovativeness

Another factor that can influence perceived usefulness may derive from lifestyle differences. While Rogers speaks of 'compatibility' in a broad socio-cultural sense, in which a technology has to accomodate existing cultures and norms, this can also be examined from a household pount-of-view. According to the (extended) 'model of adoption of technology in households', differences in individual attitudes towards technology adoption depend, not only upon personal demographics, but also upon entire household demographics, or 'life cycle stages' [17]. These stages are determined by a combination of the demographics of the individual household members. In the case of HDTV usefulness, the household income, the number of household members, their ages, and the presence of children may influence attitudes towards HDTV, due to underlying differences in eyesight, variations in the sizes of the respective living rooms, what is watched, and by whom.

Another factor related to household compatibility is that of household innovativeness. Past adoption research has documented that a person's general interest in and speed of adoption regarding other innovations in the past had a positive influence on the adoption speed of a new innovation [18]. For HDTV, it can be argued that innovativeness with regard to television is linked to increased knowledge about new developments in this area, and, with that, to perceptions of usefulness in these developments. Innovativeness with regard to television is also a household matter instead of merely an individual preference. A change of televisions and subscriptions affects all members of a household, and is likely to be discussed within that household at some point. Therefore, innovativeness on a household level can influence personal HDTV perceptions.

Relative advantage and adoption satisfaction: analogue and digital television

Closely related to innovativeness is the manner in which households already watch television and the 'relative advantage' [8] they perceive in switching from an old technology to a new one. An old technology may be abandoned when its performance is dissatisfactory, or it may be replaced when a better alternative appears [8]. As HDTV in the Netherlands has emerged at a time when standard digital television is in the midst (40%) of its own diffusion [19], the existence of groups of analogue and digital (non-HD) viewers can affect perceived usefulness of HDTV. Like HDTV, digital television has been promoted as an improvement over analogue television subscriptions in image quality. Satisfaction with the continued performance of analogue television or that of digital television may decrease HDTV's usefulness. Satisfaction is not guaranteed, however, as households are also switching from CRT screens to flatscreens, which have different sizes and resolutions. As a result, people who own a flatscreen TV, yet still have an analogue subscription, risk having a *lower* image quality than they had when watching on a CRT unit. Also, viewers with digital subscriptions may have had expectations that were not met. In this case, the change to digital television can trigger the need for a further improvement in image quality.

In summary, from the framework sketched above, it is apparent that possible factors that can determine perceived usefulness of HDTV (in terms of a sharper image, better sound and wider television screen size) derive from knowledge (awareness and actual knowledge), uses and gratifications (in groups of viewers, specific genres and the role of television in the daily life), household situations (life cycles and innovativeness), and relative advantage (current analogue or digital subscription and screen use). The next section will explain how these factors were tested.

3 Method

To test the relationship between perceived usefulness of HDTV and the factors mentioned above, we conducted a survey of a select sample of 435 Dutch households in different geographical regions and with different television service providers in the spring of 2007. The five main elements in the questionnaire appeared as follows:

Perceived usefulness: respondents were asked for their opinions on the importance of image and sound quality improvement for the future (Likert scale). To establish a perception of their value, we also asked the same questions about other digital television features, such as more channels, video on demand and interactivity. In addition, opinions about the importance of screen size were sought.

Knowledge: respondents were asked if they had heard of HDTV and whether they had seen it in practice. Also, they were presented with eight statements about HDTV and its practical use. Options for answering included 'true', 'false', and 'I don't know'.

TV Viewing: to measure genre preferences, respondents were asked how often they watched movies and sports (amongst other genres) or played games on a game console. After the survey was conducted, responses about genre and channel viewing frequencies were combined into different groups using hierarchical cluster analysis. The first group consisted of 'frequent omniviewers'; the second of 'less frequent information seekers' (news, sports, documentaries); and the third of 'less frequent mostly entertainment viewers' (films, soaps, and children's programs). Finally, respondents were asked to respond to statements about the importance of television in their daily lives. Three statements (Cronbachs Alpha 0.7) measured the importance attached to television.

Households: to determine household type, respondents filled out information pertaining to household income, and the ages and sex of the household members. From analysing this data, a set of six household types was established by age: young people from 18-40, families with children, and seniors aged 40+; and by income: high and low income brackets. Finally, to measure household innovativeness, respondents were asked to agree or disagree with statements about their household behaviour. This resulted in a three-item scale with a Cronbachs Alpha of 0.8.

Current adoption: respondents were asked about their television subscription and living room television. Afterwards, four groups were constructed, comprising analogue or digital viewers with CRT or flatscreens.

4 Results

The outcome of the analysis demonstrated that HDTV features themselves were considered important in future television wants, even before the data for correlations between the different variables and HDTV characteristics were analyzed. Among other features, a sharper image was considered of most importance, programs on demand appeared second and sound quality was third. Table 1 (App. I) shows that the importance of individual HDTV characteristics demonstrate significant correlatation with each other. This may indicate that the concept of HDTV has already reached these respondents, or, simultaneously, that the current television viewing experience requires an 'upgrade' of image, sound, and/or screen size in order to remain satisfactory.

Awareness and actual knowledge

Beginning with the concept of awareness: Table 2 (App. I) shows that there are no significant relationships between having heard of HDTV and the need for any of its features. Significant relationships did however, appear in the responses about the importance of image quality, sound quality, and screen size, when the respondents had seen HDTV in practice. With the adoption of HDTV still at an early stage, confrontations with the medium are likely to have occurred primarily in electronics stores. Respondents that have experienced HDTV in these stores can be influenced, not only by what they have seen, but also by the context in which they have seen it, and the actual knowledge they gained from this context. When comparing the results about actual knowledge, 41 percent of the answers given by respondents to all of the questions were correct, 18 percent incorrect, and 42 percent said that they did not know the answers. Percentages were similar for most individual questions. The knowledge that HDTV supposedly offered improvement in image sharpness, and that an HD-ready label was required to view HD signals, appeared to be most common. Less well known were possibilities for sound improvement, and also subscription- and decoder requirements for watching broadcast television.

In summary, the presence of HDTV in society, in terms of trialability and knowledge about its features, contributes to the perceived importance of its image and sound quality improvements, and the size of the television screen. However, knowledge about HDTV remains only partly correct, and some information is simply lacking, especially when it comes to requirements for actually receiving (broadcast) HDTV images. This can be explained by the fact that relatively little HD-related advertising from distributors and broadcasters has appeared, while TV-set related issues are advertised by manufacturers and in electronics stores. The transition from wanting HDTV features, to actually adopting HDTV, therefore remains uncertain.

Television uses and gratifications

For television uses, there appears no difference in perceived usefulness between omniviewers (mean on a five point scale = 3.9), entertainment viewers (mean = 3.7), and information seekers (mean = 3.8). However, Table 1 shows that there are some significant relationships between viewing movies and perceived usefulness, and between viewing sports and perceived usefulness: the frequency of sports viewing is related to preferences for a better image quality, better sound quality, and increased screen size; frequent television movie viewing only demonstrates a significant relationship with a better image quality. However, when it comes to watching movies, either on *DVD* or from Internet downloads*,* sound quality and screen size also matter. More conscious movie viewers may desire a 'cinematic' experience, which may require not only a sharper image, but also improved sound and screen size. In addition, the poor quality that can appear in downloaded movies, particularly when compared to regular DVD's can also help to trigger the need for a better image. For gaming, significant relationships exist in regards to image quality and screen size. This must, however, be asserted with caution: gaming on consoles is still a niche market in television use and the low number of respondents who are gamers may be responsible for these results.

Finally, the importance of television in people's daily life, as demonstrated in Table 1 of the appendix, accounts for 20 to 29 percent of the correlation with the importance of a better image quality, sound quality, and screen size. While it appears only logical that people who attach little value to their television may also be less interested in its features, this remains an important fact to remember: if (traditional broadcast TV viewing or using continues to decline, so may interest in its innovations.

Household life cycles and household innovativeness

In regard to household life cycles, perceived usefulness for HDTV was the highest for high-income families with children (means between 3.7 and 4.0). Low-income senior respondents (40+) and high-income young respondents (18-40) without children both appear in second place, with means ranging from 3.6 to 3.9. As Table 2 (App. I) shows, none of the differences between these groups was large enough to be significant. Yet the second factor, household innovativeness, does demonstrate significant results by positively correlating with importance of image quality, sound quality and, most significantly, screen size (Table 1). The importance respondents attached to screen size can be explained by interest in new flatscreen televisions as a general type of aesthetic technical 'gadget'. The importance of image and sound quality may be linked with awareness and knowledge about HDTV as a new medium that offers the possibility of sound and image improvement. It is also probable that innovative households are simply more likely to have digital television. Why this can be uses as an argument will become clear in the next section.

Current analogue and digital adoption satisfaction

When examining the influence of current subscriptions and screen possession, both analogue and digital viewers have high scores (means between 3.5 and 4) where image quality, sound quality, and screen size importance are concerned. However the importance of image quality improvement as well as screen size was rated significantly higher by digital flatscreen viewers (mean = 4.0) when compared with analogue CRT viewers (mean = 3.8). Other categories (analogue subscriptions with flatscreen and digital subscriptions with CRT devices) showed no significant differences. Thus, instead of high perceived usefulness amongst analogue flatscreen users, this usefulness is found amongst those who have already adopted *digital* television and have flatscreen TV's. Dissatisfaction with the digital signal may offer a partial explanation (technical difficulties with set top boxes or the simultaneous change to a bigger television screen may have resulted in disappointed viewers). A further possibility may be that current digital TV viewers are not necessarily dissatisfied, but instead belong to a category of 'early adopters', who are already interested in 'what comes next'. Further research is needed on this topic, and also in regard to actual HDTV adoption. While HDTV perceived usefulness is high, investments already made in digital television and 'yet another promise' of better image quality may well become a stumbling block to adoption, rather than a stimulus.

5 Conclusions

In examining perceptions of usefulness in the case of HDTV, knowledge and confrontation with the medium, specific television uses, household innovativeness and relative advantage appear as some of the factors that significantly contribute to perceived usefulness. There appears no significant relationship between those who have only heard, but not seen, HDTV, and household demographics in perceptions of usefulness. Based on these results, a few remarks can be made. First, the actual existence and promoted functions of digital television and HDTV together appear to create their own demand; more knowledge about HDTV (sought by innovative households or found for other reasons) results in higher perceived usefulness. Also, the need for a sharper image appears higher amongst digital television adopters rather than analogue viewers. Second, perceived usefulness for HDTV is explained, not though classifications of viewers (frequent entertainment viewers or less frequent information seekers), but rather by preferences for specific, unrelated genres that involve the concept of 'being there' or infer a cinematic experience. Third, in terms of adoption, while larger screens are already being purchased, both image quality and sound quality are high on the list of improvements for television viewers. However, as was stated in the first section, the status of HDTV's perceived usefulness still reveals little about its actual adoption. Social, psychological and practical factors can intervene between the point of desire for HDTV's features and the actual adoption of HDTV. If time spent watching television declines further, so may interest in HDTV features. Even with persistent interest, adoption of HDTV devices and subscriptions may be too expensive, or may compete with earlier investments in digital television. Also, TAMs other independent factor, 'perceived ease of use' and Rogers' complexity have not yet been researched. Increasing knowledge about costly subscriptions, different set top boxes and standards may influence the adoption decision negatively. Finally, it is important to realize that the meaning of perceived usefulness employed in this paper has an expiry date. While technology and contexts change over time, so do the possibilities for their use. For HDTV, this change may come when Dutch broadcasters and distributors treat HDTV as *the* new digital format. Usefulness of HDTV will then be expanded into the current usefulness of standard digital television, including video-on-demand, interactivity and electronic program guides. If this happens, these digital services may also become powerful drivers for HDTV adoption.

References

1. Davis, F.D.: Perceived Usefulness, Perceives Ease of Use and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–340 (1989)
2. Fishbein, M., Ajzen, I.: Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley, Reading (1975)
3. Ajzen, I.: From intentions to actions: A theory of planned behavior. In: Kuhi, J., Beckmann, J. (eds.) Action-control: From cognition to behavior, Heidelberg, pp. 11–39 (1985)
4. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27(3), 425–478 (2003)
5. Silverstone, R., Hirsch, E. (eds.): Consuming technologies: media and information in domestic spaces. Routledge, London (1992)
6. Lie, M., Sørensen, K. (eds.): Making technology our own? Domesticating technology into everyday life. Scandinavian University Press, Oslo (1996)
7. Hutchby, I.: Conversation and Technology; From the telephone to the Internet. Polity Press, Cambridge (2001)
8. Rogers, E.: Diffusion of Innovations. Free Press, New York (1995)
9. Lin, C.A.: An interactive Communication Technology Adoption Model. Communication Theory 13(4), 345–365 (2003)
10. Chan-Olmstead, S.M., Chang, B.: Audience knowledge, perceptions and factors affecting the adoption intent of terrestrial digital television. New Media & Society 8(5), 773–800 (2006)
11. Quiring, O.: Social and situational influences on the acceptance and adoption of interactive technologies. In: The social shaping of new media: studies on media appropriation, ICA Convention Dresden, pp. 25–35 (2006)
12. McQuail, D., Blumler, J.G., Brown, J.: The television audience: A revised perspective. In: McQuail, D. (ed.) Sociology of mass communication, pp. 135–165. Penguin, Harmondsworth (1972)
13. Katz, E., Blumler, J.G., Gurevitch, M.: Uses and Gratifications Reseach. Public Opinion Quarterly 37(4), 509–523 (1974)
14. Rubin, A.M.: Ritualized and Instrumental Television Viewing. J. of Communication 34(3), 67–77 (1984)
15. Lin, C.A.: Looking Back: The contribution of Blumler and Katz' Uses of Mass Commmunication To Communication Research. J. of Broadcasting & Electronic Media 40, 574–581 (1996)
16. Sociaal Cultureel Planbureau TBO (October 2006), `http://www.tijdsbesteding.nl`
17. Brown, S., Venkatesh, V.: Model of Adoption of Technology in Households: A Baseline Model Test and Extention Incorporating Household Life Cycle. MIS Quarterly 29(3), 399–426 (2005)
18. Lin, C.A.: Webcasting adoption: technology fluidity, user innovativeness, and media substitution. J. of Broadcasting & Electronic Media 48(3), 446–465 (2004)
19. Monitor Digitale TV in Nederland (September 2007), `http://www.immovator.nl`

Appendix

Table 1. Means and Pearson Correlations for Perceived Usefulness of HDTV

		Pearsons Correlations (r) Importance of…		
	mean	image quality	sound quality	screen size
Importance of a sharper image	3.80	-	0.73*	0.41**
Importance of better sound quality	3.69	0.73**	-	0.38**
Importance of screen size/aspect ratio	3.61	0.41**	0.38*	-
Amount of correct knowlegde	3.28	0.24**	0.21**	0.25**
Amount of incorrect knowledge	1.43	0.01	- 0.01	0.05
Lack of knowledge	3.28	- 0.18**	- 0.15**	- 0.22**
Frequency of television movie watching	3.84	0.10 *	0.03	0.06
Frequency of movie downloading	1.75	0.13**	0.11*	014**
Frequency of viewing movie DVDs	2.57	0.15**	0.09*	0.08
Frequency of sports watching	3.27	0.12**	0.10*	0.08*
Frequency of game console gamers	1.49	0.13**	0.08	0.12**
Importance of TV in daily life	3.41	0.21**	0.22**	0.29**
Household Innovativeness	2.80	0.18**	0.17*	0.23*

*Significant at 0.05 level, (1-tailed). ** Significant at 0.01 level (1-tailed).

Table 2. Compared Means for Perceived Usefulness of HDTV

	Compared means by Anova (F) or T-test (t) Importance of…		
	sharper image	better sound	screen size
Had heard of HDTV (t)	1.41	1.09	0.61
Had seen HDTV work (t)	3.30*	3.25*	3.39*
Groups of television viewers (F)	1.12	1.38	0.88
Subscriptions and Screens (F)	2.91*	2.58	3.25*
Household type (F)	1.77	1.50	1.57

*Significant at 0.05 level

Prospecting the Appropriation of the Digital TV in a Brazilian Project

Elizabeth Furtado[1], Thais Kampf[2], Lara Piccolo[3], and Maria Cecília Calani Baranauskas[3]

[1] Mestrado em Informática Aplicada – UNIFOR
CEP 60811-905 – Fortaleza – CE – Brasil
[2] Universidade Estadual do Piauí– PI – Brasil
[3] Institute of Computing (IC) – UNICAMP
13084-971 Campinas - SP, Brazil
elizabet@unifor.br, thaiskampf@ig.com.br,
lpiccolo@cpqd.com.br, cecilia@ic.unicamp.br

Abstract. A good practice for developers of a new technology, which has great influence on the life of interested users (as in the case of the Digital TV), would be to: inform them about the benefits of the technology and challenges for its appropriation. Usually the focus of works that investigate the appropriation of Digital TV has been on avoiding the non acceptance of users towards usability and accessibility issues. This work presents other risks related to the non acceptance, influenced by the way the technology will be owned, by the topology of the city and by the life style of the population. Human and contextual factors help to overcome the non acceptance of users and give rise to some recommendations useful for the developers of this technology to mitigate such risks.

Keywords: Technology Acceptance, community-oriented TV.

1 Introduction

The deployment of a new technology requires developers to have background knowledge about the needs of the users inclined towards this technology. These professionals apply various practices, such as: the identification of what users expect and the analysis of the viability of such expectations. User expectation refers to the founded hope towards supposed rights, probabilities or promises, [1] cited in [2], which can influence in the way he/she will interact with the technology. Many times an individual does not have any hope left when his/her experiences are of repeated failures.

However, many times these practices do not happen because the professionals do not believe that the expectations of some users may be representative of a community. Bodker and Buur [3] affirm that it is not recommended to focus on the needs of a particular community, but the inputs from users should be regarded as a source of inspiration to better design the products. In this text, such inputs refer to expectations, characteristics of a sample of users, residents of a small Brazilian municipality, and which were inputs for an awareness process of the acceptance of the Digital TV. It was thought that the inputs from users, representative of a small group, would only

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 293–302, 2008.

serve as source of inspirations to better design products. However, the participation of users revealed a series of human and contextual factors involving the community as a whole. More specifically, the analysis of the expectations of these users through the acceptance of the Digital TV described in this article was influenced by the following factors: the knowledge level of users of the proposed technology, friendship, thrust and solidarity among people of the community, involvement and support of local authorities of the municipality, quality of the interaction established with the researchers, among others. Such factors have been useful to help professionals in understanding the risks that would lead to the unsuccessful appropriation of the technology (that is, its non use and/or non acceptance by the users involved). Our aim at this paper is to show how people would be prepared to accept a new technology by analyzing the risks that they face on and their strategy applied to mitigate such risks.

This article is organized as follows: contextualization of this research, preliminary elicitation of risks that may compromise the acceptance of the technology, description of the identified human and contextual factors and of some recommendations from lessons learned.

2 Context of the Project in Brazil

The experiments that guided the analysis in this article are part of a research project, which aims at developing and evaluating applications for Digital TV, called *System for Advanced interactive digital television and Mobile services in BrAzil* – SAMBA [4]. SAMBA, financed by the European Union, started in January of 2007 and will take place until December of 2008. SAMBA aims at creating a computational environment that allows citizens to have access to contents produced by the own population through the WEB. In the Brazilian context, these citizens are users residing in the city of Barreirinhas, a small municipality (47.728 inhabitants [5] in the state of Maranhão). TV Mirante in the state of Maranhão will be the channel responsible for the transmission of content generated through this environment. The Set-Top-Box (STB) to be built in the project, called SAMBOX, will be compatible with the European *middleware* MHP (Multimedia Home Platform). An Interactive STB consists of a device connected to the television, which, among other things, can convert digital signal into analogical signal. In SAMBA, the built STB will allow sending and receiving data through a return channel, established via PLC (*Power Line Communication*) [6].

The reader can then ask if this System that is being financed will be compatible with the Brazilian Digital TV System (SBTVD) [7], which has started in December of 2007. Explaining all the differences between the European standard (DVB - Digital Video Broadcasting) and the Japanese standard (ISDB - Integrated Services Digital Broadcasting), adopted in Brazil, is out of the scope of this article. However, two aspects are worth mentioning: the time to deploy the SBTVD for the whole country and the benefits that this project can bring to the research on Digital TV in Brazil.

The deployment of the SBTVD requires the preparation of broadcasters of the national digital TV for transmission, as well as of viewers for buying converters (STB) following the Brazilian Middleware, called GINGA [8], which was integrated to the Japanese standard. Deadlines were established to start works on transmission of TV content in this technology in the country. The expected date to access this technology in small municipalities in the northeast, as in the case of the study, is for 2010. In the Samba project, it is expected that the transmission starts in 2008.

Concerning the benefits with the results of this project, we can say that the SAMBA environment developed to access content through the digital TV can be compatible with the SBTVD. It is based on some characteristics, such as the adoption of JAVA-TV of the HAVi standard, and GEM (Globally Executable MHP) specifications [9], which are common to both the Japanese and European standard. .

Another benefit to contribute for the quality of the SBTVD is related to the fact that SAMBA is being built considering the socio-cultural diversity of users, their difficulties to use this new technology and their strategies to overcome them.

Until now, we have not found works in this direction, despite some efforts from the SBTVD. This is the case of Brazilian works [10] and [11], which have demonstrated initiatives to develop Digital TV applications focusing on usability and accessibility by suggesting providing users with, for instance, opportunities to personalize such applications, increasing font size, changing colors, providing alternatives to write and read, etc. Other works have presented studies about the purchasing power of Brazilians to acquire this technology [12].

Summarizing, accepting that the appropriation of a new technology is dependent on human and contextual factors was the hypothesis that made the authors of this article investigate the following questions: What risks people face to own this technology? How can the relationship between people in a community decrease risks for the coverage of this technology? Following, we present two experiments that guided us in answering these questions.

3 Method

The team of researchers (authors of this article) has worked closely with the residents of the mentioned city (the interested users) to guarantee a good level of satisfaction of the services and contents presented through applications that access the generated contents. The team performed two experiments in the municipality in January and in October of 2007, using as meeting place with participants, an institution that has a good reputation in the municipality and which acted as a gateway to the initial contact with the population.

3.1 First Experiment

The main goal of the team was to know the needs and expectations of the participants of the study via the Digital TV. Therefore, with instruments (questionnaires, prototypes of applications for Digital TV) and techniques (interviews and focal groups) it was possible to elicit data, thus, providing the team with material to advance in their findings about Digital TV services that are more appropriate for that community [13]. The study involved participants with different ages (between 8 and 85 years old), levels of education (2nd grade of elementary school to post-graduate) and socio-economic (between 1 and 10 minimum-wage) and professional groups (teachers, entrepreneurs, public servants, merchants, freelancers). 117 questionnaires were applied with occasional interviews. For a good understanding of the emotional level

Fig. 1a. Participants watching TV in their home

Fig. 1b. Workplace of the participant

and real situations experienced by the participants, we conducted interviews in their residence or workplace. We observed their behavior with the technological resources in their natural environment during the use of television and of other common technological resources (see Fig. 1a and 1b).

Following, 26 participants used some applications for Digital TV in various situations (voting, communication). In the focus groups, we registered the opinions of the participants of the study about the used resources. The acquired data were consolidated and the results were analyzed in a quantitative and qualitative manner [14]. The quantitative analysis referred to the presentation of statistics on socio-demographic data (such as: number of residences that have TV, Internet access, use of the remote control, etc.) [2]. The qualitative analysis enabled the definition of the needs of users in Barreirinhas, which were to: communicate, obtain information, participate in educational programs, and have more alternatives for leisure. Such needs were related to the limitations of the city on the basis of factors, such as: restrictions on access to the Internet, few schools of average level, and little choice of entertainment offered to the community. With this information, the team of designers developed several interaction scenarios about how TV services would address the elicited needs. We generated a video of these interaction scenarios illustrating future users interacting with the new technology, using TV, computer, mobile phone, etc.

3.2 Second Experiment

With the goals of validating the obtained results of the first study and discovering the maximum number of new user characteristics and expectations, we conducted a second meeting with the participants of the first experiment. This time, we included new users, but still keeping the diversity.

In this meeting, we performed 5 workshops with thirty participants, 15 male and 15 female, grouped mainly by age, such as: Group 1: 6 children; Group 2: 4 adolescents; Group 3: 1 youth and 5 adults, community leader profile; Group 4: 5 adults, including 2 low-literacy people; and Group 5: adults and elderly (4 teachers, 4 retired, 1 pastor).

In each workshop, after an explanation about the goals of the workshop, the users had the opportunity to use an existing Digital TV application [15]. This application invites the user to express his/her opinion regarding a municipal achievement by voting and by sending a free-text message to the city mayor, while watching its announcement. This moment also occurred in the first experiment, but this time, we prioritized group activities (that is, the participants helped each other as they interacted with the application using remote control) (see Figure 2a). Then, we presented the video with the possible solutions of interaction that SAMBA could offer. To improve the understanding about what was possible to do, the screens (still as low fidelity prototypes) were inserted in the video. Figure 2b illustrates the participants of a workshop watching a video of a professor that prepares his classes to be viewed by students through the TV. The video presents a screen with fields to be filled out by the professor about class slides.

Fig. 2a. Participants interacting with the TV

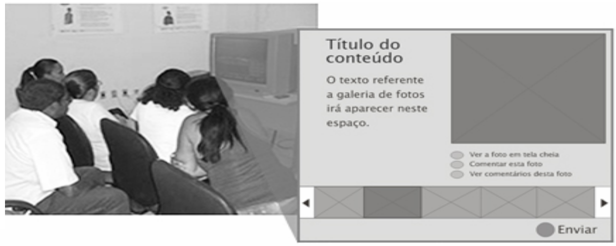


Fig. 2b. Participants watching the video with prototypes

Each workshop finished with a moment for considerations about the study, electing as a target point of the study, the acceptance factor of the new technology to be deployed in the municipality. The subjective analysis to be described as follows involves perceptions about this factor for the participants involved in the study.

4 Risks Organized on the Basis of Organizational Semiotic

The level of user satisfaction called the attention of the researchers to consider some possibilities that could risk the appropriation of the interactive Digital TV in the community. The notion of risk refers to possible problems that might cause obstructions to the normal flow in the performance of any plan. In the context of the development and usage of the technology in this research, there are technical, human, and contextual risks, which are interconnected.

An organizational semiotic tool was applied to classify the identified risks that could influence the user satisfaction towards the appropriation of Digital TV. The "semiotic onion" [16], shown in Figure 3 represents the risks in technical, formal and informal levels. The technical level is under the influence of both formal and informal levels and, at the same time, it has an impact on them. The technical level comprises technical aspects of the project, such as coverage area. The identified risk depends on the topology of the city, mentioning: Limitation of the scope of the Digital TV. The formal level refers to possible or existing contracts or agreements that regulate the project. The identified risk was influenced by the support that can be offered to people, mentioning: Digital TV from SAMBA is different from the Brazilian Digital TV.

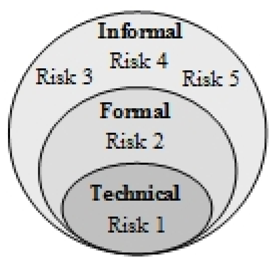


Fig. 3. The semiotic onion

Informal Level, which refers to the relationships with the community, with the social world. In this level, we find risks influenced by the following aspects: works that are similar and previous to the ongoing research and those that also happened in the same city; profile of the population, such as diversity of education and urban migration. They are: Disassembly of the structure after the project; Digital TV not being adapted to the profile of the population, and change of users participating on the studies.

The identified risks will be described as follows, based on the comments of participants during the meetings with the research team. The participants demonstrated behaviors that induced us to suppose levels between satisfied, partially satisfied, and even unsatisfied.

Risk 1: Limitation of the Scope of the Digital TV. 50 SAMBOX will be distributed free of charge in places, situated on the streets where the PLC is currently installed. Despite the efforts of the electric engineers of the municipality towards the expansion of the scope of the Digital TV, many residences of the participants of the study will not receive the STB because they are out of the area benefited by the electrified area.

To address this risk, the investigating team requested the participants to make a list of addresses of the possible benefited houses in which they could have easy access to watch TV. The list soon had various addresses. When we asked the reason why there were so many possibilities, they said: “The house of Joe, (a resident who will be benefited), is always open. Anyone can come in... ”. Another important factor, as proved during the application of the questionnaire (N=117, 68%), people usually watch TV in groups – with friends or with their family. This was once again evidenced on the occasion of testimonies in which they proposed public places for the installation of SAMBOX, since they are not available in enough quantity for all residences. Despite the fact that the people knew each other and demonstrated the feeling of cooperation, we observed, in many moments, the wish of all participants to have a SAMBOX in their residences. This was evidenced when they said: “What I would really like is to receive a SAMBOX, but since it won’t be possible…”. “This line is very small, it won’t get nearly anyone. Only the commercial area will have the Digital TV”.

These quotes and behaviors signaled a risk that leaded users to a behavior of **partially satisfied**.

Risk 2: DIGITAL TV from SAMBA is different from the Brazilian DIGITAL TV. As previously mentioned, there are differences between the Digital TV made possible by Samba and by SBTVD. In the specific case of this research, when the TV Mirante sends the content (Digital TV applications along with the normal program, such as soap operas and commercials) via broadcast, all the content needs to be understood by the STB installed in the residences. When the STB is Brazilian, maybe in 2010, the applications developed for the SAMBOX that are compatible with the functions of GEM should not be re-written.

However, some adjustments (such as: having to buy another STB in the future) should be needed. We stress the fact that the community has not given importance to this for some reasons: i) the delay of this municipality to be benefited by the program of deployment of the SBTVD; ii) the fact that the municipality was elected to be the first Brazilian municipality to use the Digital TV with interactivity was reason of pride for all; and iii) including also its gratuity, during the period of the project. These reasons are evidenced in the following quotes: “It doesn’t matter; this study has already brought good results for our children considering the future". "We are proud and interested to disclose Barreirinhas for the world." "[...] Better yet the fact of being free (Digital TV)"…

Such acceptance behaviors, in the view of the researchers, point to strong evidence of **user satisfaction** related to the appropriation of the new technology.

Risk 3: Disassembly of the structure after the project. Two years ago an electric structure was installed in the municipality in order to allow the installation of internet in schools. There was a pilot project, located on the most famous educational institution in the city, where many people were involved direct and indirectly. But the project was not self-sustaining: the computers did not have maintenance or were destined for another purpose, the installations to support cabling were withdrawn, and the technicians involved did not provide any justifications to the community. The team, in fear of a similar situation informed the participants that they did not know yet what will happen with the SAMBOX when the project is over. They might be returned to the Italian company that will build them. The team explained that they were searching for alternatives to continue the project.

Given this possibility, the behavior of the users with respect to this risk was ranked as **not satisfied**. This classification is based on the testimonies from the first trip, related to its discontinuity and the lack of explanation on the suspension of the project. To illustrate, two testimonies are mentioned as follows: "I only hope that it does not happen with this (SAMBA project) what happened with the other one: nothing". (Speech made by one of the participants during the first visit of the team to the municipality); "Will you really come back?". (Speech from one of the participants during the first visit of the team to the municipality).

Risk 4: Digital TV not being adapted to the profile of the population. The education level is one of the most relevant factors to put at risk the use of a technology by users with little familiarity with the technology, mainly to: type texts using remote control, read long texts on the TV screen, etc. Certain care during design must be taken, such as the ones already mentioned previously. In this research, to have a better understanding of this risk, we analyzed users interacting with Digital TV applications in simulation situations. The behavior manifested by users enabled us to believe that they are able to overcome certain difficulties. For instance, they are willing to interact, when followed by one or more people that can slowly demonstrate the steps. That was registered in the following quotes from two users: "We think that there is a dog that will bite us, but after we see another person using it, we feel like doing it too". It also leads them to request help from someone: "If I can't use it, I will ask my son to help me".

These manifested behaviors, which were motivated by the presence of collaboration of a more experienced user, and which might also have been motivated by the interesting possibilities of interaction with the technology, led users to say that they were willing to overcome difficulties, such as typing texts using a remote control. Therefore, this risk was considered as **partially satisfied**.

Risk 5: Change of users participating on the studies. In the second experiment, which happened 10 months after the first one, we contacted again the same participants. However, some unexpected situations avoided the participation of all of them: change in residence, transfer of the workplace, lack of telephone, work trip. Only 9 people participated in both experiments and other participants (11) were present only during the second meeting. This led to a new awareness process about the project, as well as in the performance of sessions to use existing TV applications. Although the change of users participating on the studies is usual in users' field studies, in this particular project, it compromises the progress of the work of researchers, who have to define the places where the SAMBOX will be installed. It is difficult to choose people who are committed in a meeting to create content for digital TV and/or to use digital TV and then in the following meeting do not even attend it.

To avoid that this risk compromises the project, the team personally visited some residences and/or workplaces to investigate the interest of the proprietary. In the public places that did not have TV, the responsible for the place said that they would be able to install it without any problem. The team returned to a store in which during the first trip had a TV, but that was substituted by a computer. After acknowledging the evolution of the project, the proprietary said that he would bring the TV back. As some actions need to be taken in function of the selected users, and as they are always moving, the team identified this risk as inconvenient, resulting in the behavior **partially satisfied**.

5 Results

Facing the facts and data analyzed, we believed in the possibility of establishing associations between the behavior of users and their acceptance towards Digital TV. We therefore understand that the most satisfied the user is, more easily it will be to overcome the risks already mentioned, and therefore more the user will be willing to accept the new technology. Consequently, we established human and contextual factors that were grouped in the following manner:

Factors that influence the acceptance of the new technology and that will help to overcome the risks, such as:

- Thrust in the project, which was influenced by the fact of being financed by the European Community, and by the credibility of the work performed;
- Feeling of friendship, collaboration between the population and the pride for being from Barreirinhas;
- Benefits for the collectivity through the use of various functions in the Digital TV and for their children in a future perspective, where the people responsible for the kids and teenagers see a better life for their children; and
- Gratuity of the project, bringing the Digital TV to public spaces and to some residences.

Factors that partially influenced the acceptance and that may help in overcome the risks, such as:

- Lack of regularity of participants in the meetings with the research team because of difficulties in communication, change of address, transfer to other cities;
- Non comprehension of the use of the technology, but the availability and cooperation of other people, who were more skilful with the technology; and
- Limited scope, in the municipality, of the Digital TV, thus, restricting the access to a large part of the community.

Factors that influenced the non acceptance, such as:

- Disbelief with regard to the continuity of the project, because of events prior to it, was characterized as a factor that influenced some of the contacted people to avoid getting involved with the study, thus, immediately rejecting to participate in the first phase; and
- Lack of time of some of the merchants, not demonstrating interest in being available nor to make the commercial shop available.

In general, the risks were considered of little importance by the researcher team for this context. The users in Barreirinhas were more enthusiastic about the potentials of iDTV for providing mechanisms to interact more within the local community and even for obtaining useful content while also producing it for others. In addition, they found alternatives to overcome their own difficulties, based on a community-oriented life style, instead to an individualistic one [14].

6 Future Works and Conclusion

In this section, we briefly describe some recommendations to mitigate the risks, before describing the idea for future works and the conclusion.

The recommendations, based on human factors, are the following: i) search for people that have influence in the society and that are concerned with the well being of the community; ii) involve the participants, using techniques that are collaborative and directed to their present and future daily life; iii) define roles and commitments between the participants so they disseminate the purpose of the project to the community; and iv) commit to agreements with the community: “When will you return?”. In a small municipality, where the routine of events is a constant, any event that changes with the established situation is a matter of curiosity and expectation. It is believed that in a filed study, which interferes with the community, we should take the utmost care in providing answers.

Future work will consist in a deeper analysis of the preliminary recommendations. It is intended to make a deep analysis of the presented risks considering the organizational semiotic point of view. The idea will be to think about recommendations according to the positions of the risk in the onion. For instance: bring Risks 1 and 4 to the formal level, that is, establish formal commitments between the participants of the project and the population, could be one of the recommendations. This way, it is possible to think about the other risks, thus, obtaining recommendations that consider not only human factors, but also contextual and organizational ones.

Concluding, Digital TV has been characterized as a very important resource in the digital inclusion in which the population involved lacks information resources, and where the TV is one of the only means of communication. In this article, we focused on analyzing the expectations of users inclined towards this technology (such as: will I know how to use it? Will the whole community have access to the Digital TV? How could the information from the TV help us?). The registered and observed quotes and behaviors signaled risks that are making users become satisfied, partially satisfied and unsatisfied. Human and contextual factors were identified aiming at analyzing how it is possible to mitigate such risks, when some recommendations are elaborated.

We believe that the preliminary findings of this work are useful, in general, to other international technological projects as hints for dealing with cross-cultural issues. They are also particularly useful for the SAMBA European project as a mean for knowing the adequate expectation of the community and for dealing with their frustrations.

Acknowledgements. This work was funded by the IST EC-project SAMBA.

References

1. Holanda, A.B.: Mini-Dicionário Aurélio da Língua Portuguesa, 3rd edn. Nova Fronteira, Rio de Janeiro (1999)
2. Vasconcelos, P.: USE – user experience uma metodologia de conhecimento das experiências dos usuários com base na etnografia. Dissertação de Mestrado, Unifor. (2007)
3. Bodker, S., Buur, J.: The Design Collaboratorium – A Place for Usability Design. In: ACM TOCHI (2002)
4. SAMBA Project. System for Advanced interactive digital television and Mobile services in BrAzil (2007), http://www.ist-samba.eu/
5. IBGE Instituto Brasileiro de Geografia e Estatística (accessed in 2007), http://dtr2002.saude.gov.br/caadab/indicadores/maranhao/BARREIRINHAS.pdf
6. Open PLC European Research Alliance (OPERA) (2007), http://www.ist-opera.org/
7. SBTVD. Sistema Brasileiro de TV Digital (2007), http://sbtvd.cpqd.com.br
8. GINGA. Middleware Brasileiro Ginga (accessed in 03.12.2007), http://www.ginga.org.br/
9. ETSI. ETSI TS 102 812 - Digital Video Broadcasting (DVB) – Multimedia Home Platform (MHP) Specification 1.1.1. ETSI Standard, (2003)
10. Piccolo, L.S.G., Melo, A.M., Baranauskas, M.C.C.: Accessibility and Interactive TV: Design Recommendations for the Brazilian Scenario. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 361–374. Springer, Heidelberg (2007)
11. Furtado, E., Carvalho, F., Sousa, K., Schilling, A., Falcão, D., Fava, F.: Interatividade na Televisão Digital Brasileira: Estratégias de Desenvolvimento das Interfaces. In: SBRT (2005)
12. Holanda, G., Ávila, I., Martins, R.: Mapping users' perspectives and outlining social impacts from digitalization of terrestrial TV in Brazil. Telematics and Informatics. Elsevier, Amsterdam (2006)
13. Vasconcelos, P., Fava, F., Kampf, T., Schilling, A., Furtado, M.E.S.: Ethnographic Investigational Methodology and Evaluation on Local Television Channel Creation that Allows Interaction with the Community. In: EUROiTV (2007)
14. Furtado, E., Mayora, O., Anastassova, M., Kampf, T., Vasconcelos, P.: An Investigation of iDTV User Needs in Brazilian and Italian Communities: Preliminary Cross-Cultural Findings. In: CLIHC (2007)
15. Piccolo, L.S.G., Baranauskas, M.C.C.: Understanding iDTV in a Developing Country and Designing a T-gov Application Prototype. In: Proceeding of DIS. ACM Press, New York (in press)
16. Liu, K.: Semiotics in Information Systems Engineering. Cambridge University Press, Cambridge (2000)

Usability & User Experience: Preliminary Results from Evaluating an IPTV Community Platform

Marianna Obrist[1], Sara Kepplinger[1], Elke Beck[1], Manfred Tscheligi[1], and Paul Muzak[2]

[1] ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
{marianna.obrist, sara.kepplinger, elke.beck, manfred.tscheligi}@sbg.ac.at
[2] Telekom Austria TA AG, Lassallestraße 9a, 1020 Wien, Austria
paul.muzak@telekom.at

Abstract. The promises of IPTV and related changes of the traditional TV environment are broadly discussed in the EuroiTV community. In this paper we present preliminary, qualitative results from a user evaluation conducted for an IPTV community platform focusing on usability and user experience issues. Our results show that major obstacles in the first phase of developing an IPTV community application are the usability of the system as well as the motivation of people to share personal audiovisual content.

Keywords: IPTV, user evaluation, community TV, user-generated content.

1 Introduction and Background

"We're entering an era in which people are participating rather than just receiving information" (Jonathan Swartz, in [8]). IPTV could bring a breath of fresh air into local communities by providing new ideas for sharing and creating content, exploring new ways of doing things, and enriching the perspective of community members on local topics and social interaction. Strengthening the relation building among people via the television is one of the possibilities for the IPTV technology and earns attention [7], in particular considering trends such as user-generated content [3].

If we trust prognoses, personalization, personal TV as well as personal or local content will play a bigger role in the near future [8]. Field tests, like for example within the Lommel TV project [10] in Belgium or SAMBA in Brazil [11] have already collected experiences on community related TV services and applications. Leading players are already stepping into the market by offering services related to community or social topics [6] (e.g. Amigo TV Application [12]). Although, usability and user experience research in this area is still in its infancy. Within this paper, we try to fill this gap by presenting the results from a user evaluation of an IPTV community platform.

2 User Evaluation

A first user evaluation of the IPTV platform (which was developed as part of the CITIZEN MEDIA project) was conducted in November 2007. The focus of the user

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 303 – 307, 2008.

evaluation was to discover usability problems as well as to gather preliminary, qualitative feedback from users' experience with the community platform.

2.1 Tested IPTV Application

The evaluated application was developed for supporting a local community in creating, uploading and sharing audio-visual content. The application, based on the DSL broadband IP technology, is an integrated part (walled garden – see Figure 1) of the service range of a local telecommunication provider. In a long-term field trial, which started in January 2008, a limited number of test households have access to the IPTV platform and can create and share videos and pictures with other community members.

Fig. 1. Start Page of the IPTV Application

On one hand, the evaluated IPTV platform (see Figure 2) supports the traditional viewers' role in front of the TV. On the television itself the user can watch the uploaded audiovisual content (either his own or of other users) and rate the content as well (e.g. rate the quality of a picture, rate the friendship with other users, rate the quality of a video). Another feature is the possibility of searching for certain other users and adding them as a friend. On the other hand, users can take an active role by using their PC or mobile phone. Through the PC people can upload videos, pictures and edit their profile. People are able to use their mobile phone to take pictures and display them on the IPTV platform in real time.

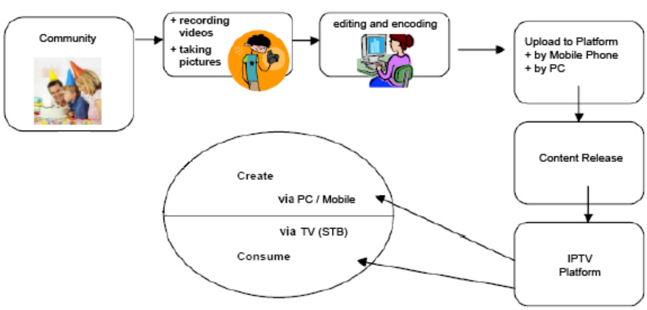


Fig. 2. Multiple IPTV Platform Environments

The decision of having multiple media in the IPTV environment builds on the idea to support the pre-experiences of people with each device. Existing media experiences with each device are supported instead of replaced by new interaction concepts.

2.2 Participants

A broad distribution of age, household structure and media experience was used for the recruitment of the test participants. The intention was to have a good representation of the demographics of the local community addressed within the CITIZEN MEDIA project. Eleven test participants – five female and six male – at the age between 23 years and 64 years participated in the evaluation of the IPTV platform. The average age was 40,73 years (SD 14,227). Two single households participated, four participants live in pairs, and four participants live in a three-person household and one participant counts four household members. All of the participants had Internet access. Three participants had three TV sets at home, four participants had two TV sets, two participants had one device and one participant had no TV at home anymore. Every participant used (at least) one mobile phone. In sum, the majority of the participants had basic experience with the TV, Internet, and mobile phones. Additionally, each participant was a member in at least one offline or online community.

2.3 Test Set Up

The study was conducted in a laboratory environment using video and audio recording as well as eye-tracking (see Figure 3).

Each participant was tested alone in a one-way mirror room, equipped with a classical IPTV set top box and remote control, connected to a flat screen TV (CRT-TV).

Fig. 3. Laboratory Usability Test Environment

2.4 Procedure

The test participants took part in a one and a half hour evaluation. Every participant carried out six realistic and typical tasks with the IPTV platform in front of the TV. The functionalities available through the PC (video upload) and through the mobile phone (picture upload) couldn't be included due to technical constraints.

A main focus of the tasks performed was on searching and navigational issues. For each task, common data such as success rate, time needed and errors occurred were collected. Apart from a questionnaire addressing the socio-demographic data, two additional usability and user experience questionnaires were used (i.e. SUS – System Usability Scale [2] and AttrakDiff [3]). Moreover, two qualitative interviews (a pre- and post-interview) were conducted. Additional data was collected through eye-tracking in order to underline identified usability problems. Due to the limited space available, we will not present the eye-tracking results in this paper. The main results are summarized towards identified usability problems and first impressions and experiences users had with the platform.

2.5 Results

Several problems, mostly related to the design of the user interface, but also related to the labeling and the general navigation on the IPTV platform were identified in this first user evaluation.

2.5.1 Success Rate and Usability

The following table gives an overview on the number of tasks successful solved, the tasks solved with help, not solved tasks and the average rating of the degree of difficulty of each task on a scale from 1 to 5 (where 1 = very easy and 5 = very difficult).

Table 1. Task / Success Overview

Task	Solved	Solved with Help	Not Solved	Task Rating
T1: Search for a User Profile	8	2	1	2,42
T2: Find and watch a Video	5	3	3	3,00
T3: Find the first Task of the Game	1	1	9	4,25
T4: Add a New Friend	6	1	4	2,92
T5: Find and rate a Picture	7	4	0	2,67
T6: Look at the Game High Score	5	4	2	2,83

Participants had the most problems to solve task 3 due to design and labeling problems (mainly related to navigation and orientation problems), as they couldn't find the task for the game offered within the platform. The SUS questionnaire revealed a score of 55,7 points (from 100) for the usability of the IPTV platform. Consequently there is room for improvements of the system and to support the users' requirements.

2.5.2 Overall Impression and User Experience

The results from the AttrakDiff Questionnaire [3] (indicating how users experience the system on a pragmatic and hedonic quality level) revealed that the user interface of the IPTV platform was generally rated as neutral. There is still a need for improvements to make the system more attractive to the users. In the post-interview, the majority of the participants were critical about the IPTV platform, but had also a lot of sympathy for the offered features and the provided content. Following statements give some insights in the overall impression participants had of the platform:

"As I am not a fan of pictures I would rather use that for the exchange of messages like in Skype in earlier days. Perhaps like for example: Ask a question to the community and expect an answer" (female, 28) or *"Well, when I can show pictures to my friends, I would use it, … if I am sure, that just my friends can see them"* (male, 62).

Overall it can be summarized, that features or elements motivating the participants to use the platform were mostly community related. Participants mostly liked the possibility of getting to know other people and to look at their pictures. The friends list and picture gallery were mentioned in this case as motivating features for the usage. Features of the platform, which discourage the participants to use the platform, are mainly concerned with the usability of the platform. The participants most mentioned the fussiness of the platform and a lack of meaningful labeling of links, as well as lack of transparency. On the question what shouldn't be changed on the platform, the participants most often mentioned the possibility to watch videos or pictures of friends and adding friends to the personal space at the platform.

3 Conclusions and Outlook

In this paper an IPTV community platform was evaluated by focusing on usability and user experience aspects, which are influencing the acceptance of a new system. From previous research we know, that aspects related to "ease of use" community design both related to features, content variation and usability are important. Seventeen percent of the people that stop using communities or use communities less than before explain this fact by usability issues and 24 % by content issues [1]. Therefore it is important to take usability issues into account throughout the whole design and development process.

As part of a one-year field trial – just started in the project – selected test households will be involved in future in situ evaluation activities. We will address user experience issues and in particular also sociability of the IPTV platform, as well as the question how the deployed IPTV application could enhance the community experience and social interaction in a rural environment.

References

1. Brandtzæg, P.B., Heim, J.: Initial context, user and social requirements for the CITIZEN MEDIA applications. Participation and motivations in off- and online communities. P6-2005-IST-41, CITIZEN MEDIA, D.1.2.1 (2007), http://www.ist-citizenmedia.org
2. Brooke, J.: SUS: A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor and Francis, London (1996)
3. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Teubner, B.G. (ed.) Mensch & Computer 2003: Interaktion in Bewegung, pp. 187–196 (2003)
4. Jensen, J.F.: User-generated Content – a Mega-trend in the New Media Landscape. In: Lugmayr, A., Golebiowski, P. (eds.) Interactive TV: Shared Experience, TICSP Adjunct Proceedings of EuroITV2007, pp. 29–30 (2007)
5. Nielsen, J.: Participation Inequality: Encouraging More Users to Contribute (2006), http://www.useit.com/alertbox/participation_inequality.html
6. Schmitt, D.: European IPTV: Market assessment and forecast. Screen Digest, UK (2005)
7. Weber, J., Newberry, T.: IPTV Crash Course. Mc Craw Hill Companies, NY (2007)
8. Wellman, B., Hogan, B.: Connected Lives: The Project. In: Purcell, P. (ed.) Networked Neighbourhoods. The Connected Community in Context, pp. 161–216. Springer, London (2006)
9. Zu Salm, C.: Zaubermaschine interaktives Fernsehen? TV-Zukunft zwischen Blütenträumen und Businessmodellen. Gabler Fachverlage, Wiesbaden (2004)
10. http://www.lommel.tv/
11. http://www.create-net.it/samba/
12. http://www1.alcatel-lucent.com/

iTV as a Platform for Rich Multimedia Reminders for People with Dementia

Alex Carmichael[1], Mark Rice[1], Stephen Lindsay[2], and Patrick Olivier[2]

[1] School of Computing, University of Dundee, Dundee, DD1 4HN, UK
{acarmichael, mrice}@computing.dundee.ac.uk
[2] School of Computing Science, Newcastle University, Newcastle upon Tyne, NE17RU, UK
{s.c.lindsay, p.l.olivier}@ncl.ac.uk

Abstract. This paper gives an overview of a reminder system, based on digital television technology, which can present rich multi-media prompting messages in order to help maintain a regular daily schedule for people with mild to moderate dementia. The ongoing development of this application has the aim of supporting busy carers by making it easy to create, adapt and maintain a schedule of reminders to help keep a regular daily routine, relieving some of the burden of care and freeing more time for better human contact and interaction.

Keywords: Dementia, DTV, Older People, Carers, Functional Prototypes, Prompting & Reminder Systems, Activities of Daily Living, Schedule Control.

1 Introduction

It is well known that the population aged 65 years and over is expanding both absolutely and as a proportion of the population as a whole. Of this group, about 5-6% are estimated to suffer from some form of dementia. Prevalence at the age of 65 is about 2% and this percentage approximately doubles for each subsequent 5 year age band, i.e. 4% at 70 yrs., 16% at 80 yrs. [1]. Alzheimer's disease (or Dementia of the Alzheimer's Type; DAT) is the most common form, and accounts for about 60% of all dementia cases, making it a major (and increasing) healthcare issue in later life.

A progressive disease, people with DAT experience a range of cognitive impairments that typically start with relatively minor working memory problems as is common among many older adults. Symptoms can vary from very mild cognitive decline or memory lapses, such as forgetting familiar words or names (which again can be difficult to distinguish from the effects of 'normal' ageing), to very severe cognitive decline including loss of the capacity to understand (and occasionally even recognise) speech and to communicate more generally. As mentioned, over time the extent of these declines is generally progressive, however, over the shorter term the impacts can undergo relatively unpredictable cycles of remission and relapse. Further, these complicated trajectories of change can be different for different people and for different aspects of cognitive functioning, including the individual's degree of awareness of their current cognitive state. In addition to declines in cognitive ability *per se*, DAT can also have negative impacts on the sufferers' affective state which can manifest as relatively generalized depression or as a wide range of more idiosyncratic 'mood-swings' which in turn can be accompanied by various

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 308–317, 2008.

challenging behaviours [2]. There can also be physical impacts which can compromise musculoskeletal functioning including the ability to walk without assistance and even to sit without support [3]. Consequently, many people with DAT can feel scared, frustrated and/or ashamed by their increasing frailty, memory and cognitive difficulties, and the accompanying dependence on others that this can bring.

A very important issue for people with DAT and their carers is the maintenance of a stable and regular daily routine [2]. This is far from being a trivial issue given the wide range of other care requirements the individual may have (often exacerbated by a general paucity of resources to meet them) which in the main are dealt with by over-stretched care workers or informal carers such as their spouse or other family member. Although a wide range of 'memory aids' have been researched and developed for people with mild cognitive impairment or memory problems, most become of limited utility when the user's cognitive limitations are severe enough to *need* them (e.g. they forget that a 'beep' is a reminder prompt), other issues also limit the utility of such devices intended for operation by the person with DAT [4, 5].

This paper describes the use of interactive digital television as a platform to provide rich, personalised multi-media prompts and reminders for people with DAT, primarily to relieve carers of some of the burden of this scheduling activity, particularly 'nagging' (i.e. repetitively reminding) their charge and the emotional challenges associated with this. The main focus is on reminders for activities of daily living such as meals, but there is also clear potential for its role to be extended to include 'encouragement' for events the patient may resist, 'calming' for potentially agitating events (or if the patient becomes agitated without any connection to a scheduled event *per se*), and more general 'reality orientation' (day, time, etc.). It is not intended for this application to replace human care, but rather to support carers in achieving a beneficial, stable routine for their charge, while alleviating some of the burden on their time and emotional resources. Helping to expand and improve the quality of the human contact between carer and patient is a key goal. It is worth noting that this application is a work-in-progress and is currently at a point where it can act as a functional prototype suitable for hands-on evaluations with carers and patients and can be readily adapted in light of feedback from such sessions.

2 Interactive Systems for Cognitive Support

Information and communication technology (ICT) is increasingly being employed as a prosthetic for 'information and communication' functions which have been affected by illness, accident, or aging. To date, the increasing capacity and power of computational systems has been recognised as having the potential to provide 'scaffolding for cognitive tasks' [6]. Similarly, recent years have seen digital technology become an increasingly important part of the 'mainstream' domestic environment, a development that has been identified as having both positive and negative impacts on home life. This 'domestication' of technology has highlighted its potential to support people with non-mainstream needs in their own homes. In particular, attention has been paid to the use of 'smart home' technologies to provide functional support for independent living in areas of safety, security and social interaction for older and/or disabled people. Some of these systems primarily monitor the patient and 'report' to others, while some interact with the patient to provide support. For example, Hoey et al [7] developed an intelligent prompting system that monitors a user's progress and uses visual and verbal prompts to assist them during the (relatively constrained) activity of hand washing.

Another area of development has brought the capabilities of digital technology to bear on fostering and potentially enhancing human interaction between carer and patient [8]. For example, *CIRCA* a touch-screen, PC based support for conversation and reminiscence [9] has been found to be successful and more efficient than non-technological equivalents (e.g. boxes of photos and memorabilia) which serve their purpose but tend to distract the carer's attention from the interaction itself in order to deal with the material supporting it. The success of *CIRCA* has led the same team to develop *LIM* (Living In the Moment) [10] which utilises a similar interaction approach for a system that DAT patients can operate themselves in order to engage in mentally stimulating (and fun) activities without the need for assistance.

Another aspect of the domestication of ICT relates to the emergence of digital television which is currently in the process of replacing analogue broadcast in the UK (with many other countries following close behind). Beyond the potential to simply provide more television channels, the processing requirements for receiving digital broadcasts has also made it possible to support interactive services of various kinds. A wide variety of such services are available, some related to programme and channel content and others related to the provision of government and other commercial services. This development from 'TV' to 'ICT' also supports the potential for 'local' interactive applications that reside on the user's home equipment (Set-Top-Box) and can be configured to their own requirements, either in relation to 'TV' or as more generalised 'ICT'.

In the UK, it has been identified that older people represent the largest proportion of television viewers, particularly in day-time audience groups, averaging around 36 hours per week for those aged 65-75 years [11]. It seems reasonable to assume that a similar pattern relates to older people who are relatively frail and/or have minor or moderate cognitive limitations, either as a consequence of 'normal' ageing or of entering the early stages of DAT (or other dementia).

These various factors make TV a feasible platform for a personalisable application that can present rich multi-media prompts and reminders, using the TV as an output channel. The utility of such an approach is clearly limited to circumstances where the recipient is in the proximity of the TV and/or engaged with it. Despite this, the potential benefits of supporting a regular routine and relieving even a small amount of the care burden could have significant value. Further, the adaptability of digital TV means that a 'basic' reminder application could well be extended to incorporate (or be connected with) other interactive 'modules' that are emerging for this platform. Examples of this include the development of iTV systems that support social communication either as a stand-alone activity, or in association with TV viewing. Researchers have explored, for example, the conceptual design of audio-based applications to communicate with family and friends [12], including the use of graphical overlays indicating availability and status of possible interlocutors, [13] and 'instant-messaging' type features aimed at enriching engagement in such communication [14]. However, mixed reactions to two-way communication via TV have been reported

[12], suggesting possible limitations (currently at least) on people's willingness to breakaway from relatively passive 'TV viewing'.

3 Designing Reminders for Everyday Routines

A fundamental requirement of the reminder system is that it can reside in the context of the relationship between the carer(s) and the person with dementia in a meaningful manner. That is, not only should the system provide effective reminders that relate to routines of daily life, but that the carer can set and modify these in accordance with their daily routines. A number of issues arise in the design of the prompts themselves. Such as, the delivery of prompts through an embodied agent, the potential qualities of the prompts (e.g. 'tone', layout, frequency, etc.) including the provision of supporting information in different modalities (text, graphics, sound/music), and issues of prompt presentation and integration into existing programming (or not). There are also a host of issues surrounding the design of the carer's 'schedule control' interface. These issues will be discussed in the following sections.

3.1 Prompting through Embodied Agents

Embodied conversation agents (ECAs) intended to 'persuade' humans, or effect behavioural change, raise important ethical issues which are beyond the scope of our inquiry. However, Bickmore [15, p.727] cites a range of analyses of these issues. Existing studies evaluating the effectiveness of ECAs have used limited empirical approaches that do not identify objectively measurable variables, nor do they adequately explore the impact of the low level of functionality of current agents. For example, Bernsen and Dybkjær [16] gather subjective data of users' perceptions of interactions with an ECA through structured interviews and present the results of conversational analysis of the transcripts. The challenges of designing and building ECAs has meant that research has focused more on "specific problems which are prerequisites for developing full-fledged multimodal ECAs" [17, p. xv], rather than focusing directly on evaluating the full systems [17, p. xvi]. More fully-fledged ECAs have been developed but still the focus of even 'full' system development has been on relatively limited aspects of the ECA's behaviour.

The potential of ECAs to affect behaviour has however been demonstrated in previous work on persuasion and social influence [18,19,20] which primarily use metrics based on self-reports of attitudes and belief, although a limited number of empirical studies have measured behaviour change directly. Bickmore et al. [15] used a *relational agent*, ("computational artefacts that build and maintain long-term social-emotional relationships with users", p711), in the role of an exercise advisor to encourage older adults to meet the minimum level of physical activity currently recommended, and used a combination of both questionnaires and direct behavioural measures to indicate effectiveness. One direct behavioural measure was the number of steps taken each day, recorded from a pedometer. The results demonstrated that relational agents significantly increased the amount of physical activity (i.e. number of steps) faster over the duration of the study than the control group (who used a conventional text interface). This suggests using embodied agents for presenting the prompts should be feasible and even beneficial, although this cannot be taken for granted. It is possible that some people with DAT may fail to 'engage' with a virtual person or may be unsettled by it, thus it will be beneficial to incorporate additional 'communication channels' (e.g. voice-only, text, etc.) in the prompts so that the carer can 'mix and match' to suit their (and their charge's) requirements and preferences.

3.2 Prompt Qualities

The use of ICT means that the breadth of possible qualities of a prompt is very wide. However, the practicalities of allowing the carer to easily create and adapt the prompts make it vital that the range of possible customizations is suitably constrained. As described, a major element will be the embodied agent or avatar with utterances derived from text input which therefore can also be presented without the avatar. This also allows the possibility of shaping the emotional tone (and other qualities) of the voice, although this has not been included in the current implementation (this and other technical limitations are indicated below). Given the known difficulties older people, particularly those with DAT, can have with perceiving and understanding speech (and which go beyond 'deafness', *per se* [21]), there will also be benefits in allowing the spoken message to be supported by the presentation of the text equivalent (i.e. 'captioning'), which has previously been shown to be of benefit to older adults [22]. Less directly, the 'message' of a prompt can also be supplemented by the use of suitable photographs and other graphics. Some of the current implementations of these possible layouts are illustrated in Figure 1.

The use of text input and the possibility to include either generic or personal pictures gives the carer a great deal of choice and flexibility for tailoring the prompts in a way that they feel will maximize their efficacy. The relative limitations of the current layout style are intended to have two related, but distinct benefits. First, it should allow a prompt 'message' to contain sufficient information to serve its purpose (i.e. 'engage' and 'inform') while avoiding possible 'information overload' for the person with DAT. Second, it makes it feasible to develop a control interface (used by the carer to input/update the content for prompts and to schedule them) that can be used easily and effectively by someone who may have limited capabilities themselves, and/or little time, and/or many distractions. The current implementation of the control interface is described further in section 4.2, below and loosely constrains the user through selection and creation of content for each prompt and its scheduling and currently supports text input via a virtual (or possibly real) keyboard and handwriting recognition. This layout is intended as a suitable basis to initiate formative evaluations with a range of carers in order to inform further development.

Beyond the use of individual prompts for particular events, there is potential for further benefits in the presentation of messages that 'lead-into' and/or 'follow-up' a main prompt. For example, there could be circumstances where a series of messages with an 'encouraging' or 'comforting' tone could better prepare the recipient for an event that is prone to agitate or upset them. In other circumstances post-prompt messages may be helpful, and are particularly likely to require instigation on an ad hoc basis, such that, the appropriate follow-up will be different if a reminder was simply missed compared to one that was 'ignored'.

Such follow-up messages will also benefit from the system's endless patience, meaning that (suitably designed) follow-up messages should minimize, for all involved, the emotional impact usually related to 'nagging'. The current system has the capability to support such pre- and post-prompts, but these will not be implemented until the 'basic' control interface has been refined and a clearer picture has emerged about the feasibility of including control for these extra elements.

Fig. 1. Example layouts of the different elements of a prompt

These variations on the theme of reminders also open the possibility of the system incorporating other elements aimed at 'calming' or 'reality orientation' unrelated to reminders *per se*, although these will not be introduced until the current system has been further evaluated. Similarly, there is potential for the current system to be connected with associated, but effectively stand-alone, applications aimed at mental stimulation and engagement, such as *CIRCA* and *LIM* described above.

3.3 Prompt Introduction and Integration

Another major issue for the presentation of reminder messages on the TV is how the prompt is integrated with the surrounding content. An initial issue here is whether a prompt is scheduled to appear with regard to 'clock' time or 'TV' time. Using 'clock' time may be suitable in certain circumstances, although because TV schedules are rarely precise regarding 'clock' time, the possible negative impact of interrupting the current TV program cannot be overlooked. On the other hand, prompts coordinated with the 'breaks' in scheduled TV content may be less intrusive but to such an extent that the challenge becomes one of ensuring the prompts 'stand out' from TV content, so that the personal relevance of the reminder can be recognised by the 'remindee'. Both of these approaches to scheduling are possible with the current prototype, but initial evaluations will focus on 'clock' time. The current approach to 'announcing' a prompt is for a pair of theatre style curtains to close over the current content and re-open to reveal the prompt content, with a similar approach to 'closing' the prompt and returning to scheduled TV. This is based on the idea of theatre curtains as a familiar visual metaphor for 'something is about to start' from an 'audience/viewer' perspective, but which currently is rarely, if ever, used as such on TV. However this is currently the 'default' approach and there are many other possible graphical approaches to may prove effective, including the possible inclusion of music (possibly a 'favourite' tune) and/or some other accompanying audio 'announcement'.

4 Prototype Reminder System

4.1 Architecture

The TV prompting interface was built to run on a 3.0 GHz Intel Pentium PC, while the carer interface was developed on a Samsung Q1 7" 900 MHz Intel Celeron M touch-screen display. The development environment used was Microsoft Visual Studio 2005 and the user interfaces were implemented and written in Visual C#. Creation of a 3D animated character is provided by the third party Haptek Player[1] plug-in, which is embedded as an Active-X control in the TV prompting application.

The PC is responsible for simulating a TV broadcast, using short clips of pre-recorded videos to give the appearance of watching a small set of TV channels (while the system is compatible with digital broadcast, the simulation makes the prototype more potable and robust). Simulated TV control is currently implemented by mapping software *event listeners* to key press functions for 'channel switching'. When the system is running it checks an XML data file before the reminder is given, and identifies the reminder to be played. Updates to the data file are checked for regularly to see if they will be played on the day they are added. Once the prompt is initiated, the prompt (including avatar if chosen) appears with a transitional effect and presents the appropriate prompt message (with lip-synching if needed). The spoken message can also optionally be displayed in text form (i.e. 'captioning').

To set a prompting message, an XML file (which is stored locally on the Q1 touch-screen unit), is automatically updated to the TV prompting system via the schedule control interface. In the schedule control (Q1 touch-screen) unit, prompts are presented within forms, which once set are stored in the XML file. Software in the TV prompting interface then remotely checks for updates in this file via Wi-Fi.

[1] http://www.haptek.com

4.2 Schedule Control Interface

The touch-screen interface consists of four separate but related screens and includes (a) a video demonstration facility of the system (not yet implemented), (b) a simple data table list of available prompts (initially pre-loaded but 'editable'), (c) basic options to set up a new prompt (i.e. date, time, frequency, title and message) and (d) additional features to customise the prompt (i.e. avatar, text, image, clock, etc.).

Given the limited real estate of the 7" touch-screen, careful consideration was given to the visual design of the interface to ensure users could adequately navigate and touch-activate available controls. This included appropriate visual feedback by using 'tick icons' to indicate selected menu options, and to provide visible confirmation when all aspects of the prompt were set (see Figure 2).

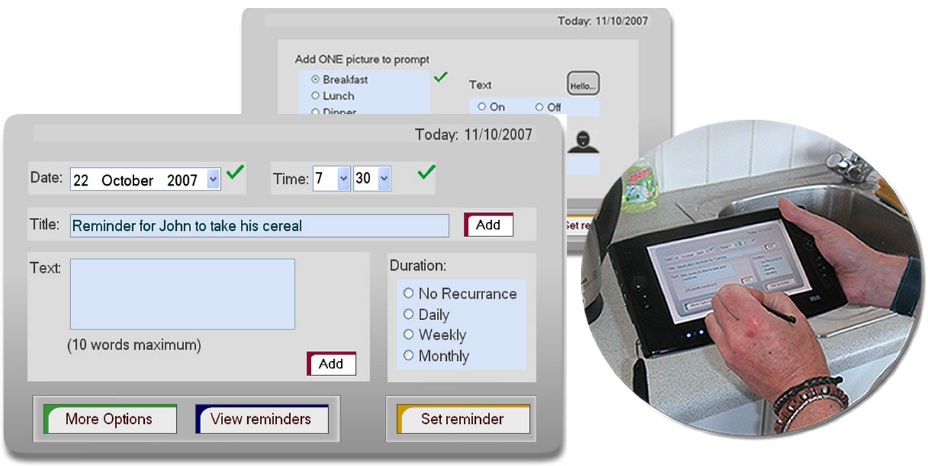


Fig. 2. Elements of the carer's schedule control interface

On the 'new reminder' screen, three simple drop-down menu boxes are used to initially set the date and time of a new prompt, followed by the display of two text boxes, one for the title of the prompt (which is then listed in the database table for the carers own future reference), the other, is for the message text (which can subsequently be verbally and/or textually displayed on the patient's TV screen). Users can either hand-write the message using a stylus, or type it using an on-screen keyboard (a separate keyboard is also possible). Final options on this screen allow the user to indicate the frequency of the message (e.g. no recurrence, daily, weekly or monthly). As previously mentioned, additional aspects of the prompt layout are available via the 'more options' button and include a selection of different photographic and illustrative images (e.g. breakfast, lunch, new visitor, etc.) to accompany the prompt. A list of radio buttons also allow users to turn the avatar 'on' or 'off' and/or the accompanying text and/or image, to further personalise the prompt.

4.3 Implementation Issues

Limitations have been identified with the use of the Haptek player related to the availability and control of facial expressions, gestures, body movements and emotional vocal tone. Technical limitations also emerged for displaying animated graphics over the avatar plug-in for the transition effects between prompts and (simulated) broadcast content. Further investigation is also needed for the optimal mechanism to trigger the avatar prompt through available MPEG video streams. Questions about the development of open standards and an API to render 3D graphics, most notably on MHP (Multimedia Home Platform) also need addressing [23]. In particular, one likely constraint will be the necessity for higher-end home media and set-top box systems (similar to high-spec game consoles) with adequate memory, processing power and built-in graphics card to run the prompting system. Currently, this excludes many people, particularly those who might benefit most from using this system. However, the recent history of commercial ICT suggests this barrier will reduce in the near future, particularly when digital broadcast has replaced analogue. Similarly, when the final requirements of the schedule control interface are known it can be implemented on more affordable hardware than the current prototype.

5 Conclusion

The main challenge in developing this prototype has been achieving the balance between including enough functionality to accommodate suitably adaptable and rich multi-media reminders, and ensuring that the control interface is easily useable by carers. This will remain the focus during the planned series of formative evaluations with a wide range of carers and subsequently in conjunction with people with DAT.

Acknowledgments. This work was supported by a grant from the ETAC program (Everyday Technology for Alzheimer's Care - Grant# ETAC-04-1027) funded by the Alzheimer's Association of America and the Intel Corporation.

References

1. Bäckman, L., Small, B.J., Wahlin, Å., Larsson, M.: Cognitive Functioning in Very Old Age. In: Craik, F.I.M., Salthouse, T.A. (eds.) The Handbook of Aging and Cognition. Lawrence Erlbaum Associates, Mahwah (2000)
2. Koenig Coste, J.: Learning to Speak Alzheimer's. Vermillion, Random House (2003)
3. Alzheimer's Association: Stages of Alzheimer's (2007), `http://www.alz.org`
4. Inglis, E., Szymkowiak, A., Gregor, P., Newell, A.F., Hine, N., Shah, P., Wilson, B.A., Evans, J.: Issues surrounding the user-centred development of a new interactive memory aid. Universal Access in the Information Society 2(3), 226–234 (2003)
5. Riederer, E., Pieper, R.: Adaptable, adaptive and extendable Information systems for occupation, activation and stimulation of people with dementia. In: Computers and Assistive Technology: Proceedings of the XV. IFIP World Computer Congress, pp. 265–272 (1998)

6. Newell, A.F., Carmichael, A., Gregor, P., Alm, N.: Information technology for cognitive support. In: The Human-Computer Interaction Handbook, 2nd edn., pp. 811–828. Lawrence Erlbaum Associates, Mahwah (2007)
7. Hoey, J., von Bertoldi, A., Poupart, P., Mihailidis, A.: Assisting Persons with Dementia during Handwashing Using a Partially Observable Markov Decision Process. In: International Conference on Vision Systems, Bielefeld, Germany (2007)
8. Spector, A., Thorgrimsen, L., Woods, R.T., Royan, L., Davies, S., Butterworth, M., Orrell, M.: Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia: randomised controlled trial. Br. J. Psychiatry. 183, 248–254 (2003)
9. Alm, N., Astell, A., Ellis, M., Dye, R., Gowans, G., Campbell, J.: A cognitive prosthesis & communication support for people with dementia. Neuropsych. Rehab. 14, 117–134 (2004)
10. Alm, N., Dye, R., Gowans, G., Vaughan, P., Astell, A., Ellis, M.: Living in the moment: An interactive entertainment system for people with dementia. In: International Workshop on Cognitive Prostheses and Assisted Communication, Sydney, Australia, pp. 16–19 (2006)
11. BBC (2007), http://www.bbc.co.uk/commissioning/marketresearch/audiencegroup5.shtml
12. Harboe, G., Massey, N., Metcalf, C., Wheatley, D., Romano, G.: Perceptions of value: The uses of social television. In: 5th European Conference of EuroiTV, Amsterdam, The Netherlands, pp. 116–125 (2007)
13. Vanparijs, F., Coppens, T., Godon, M., Bouwen, J.: Social Television: enabling rich communication and community support with AmigoTV, `http://www.ist-ipmedianet.org/Alcatel_AmigoTV_paper_ICIN_2004_final-1.pdf`
14. Luyten, K., Thys, K., Huypens, S., Coninx, K.: Telebuddies: social stitching with interactive television. In: CHI 2006 extended abstracts on Human factors in computing systems, pp. 1049–1054. ACM Press, New York (2006)
15. Bickmore, T., Caruso, L., Clough-Gorr, K., Heeren, T.: 'It's just like you talk to a friend' - Relational agents for older adults. Interacting with Computers 17(6), 711–735 (2005)
16. Bernsen, N.O., Dybkjær, L.: Evaluation of spoken multimodal conversation. In: 6th International Conference on Multimodal Interfaces, State College, PA, USA (2004)
17. Ruttkay, Z., Pelachaud, C.: From brows to trust, vol. 7. Kluwer Academic Publishers, Dordrecht (2004)
18. Bailenson, J.N., Yee, N.: Digital chameleons. Psychological Science (2005)
19. Baylor, A.L.: Interface agents as social models: The impact of appearance on females' attitude toward engineering. In: CHI, Montreal, Canada (2006)
20. Blascovich, J.: A theoretical model of social influence for increasing the utility of collaborative virtual environments. Paper presented at the Collaborative virtual environments, Bonn, Germany (2002)
21. Helfer, K.S., Wilber, L.A.: Speech understanding and aging. Journal of the Acoustical Society of America 83, 859–893 (1988)
22. Sancho-Aldridge, J., Davis, A.: The impact of hearing impairment on television viewing in the UK. British Journal of Audiology 27, 163–173 (1993)
23. Ugarte, A., Garćia, I., Orliz, A., Oyarzun, D.: User interfaces based on 3D avatars for interactive television. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 107–115. Springer, Heidelberg (2007)

A Comparative Study of Remote Controls for Digital TV Receivers

Jane Lessiter, Jonathan Freeman, Andrea Miotto, and Eva Ferrari

i2 media research ltd
Psychology Department Goldsmiths
University of London New Cross
London SE14 6NW UK
{J.Lessiter, J.Freeman, A.Miotto, E.Ferrari}@gold.ac.uk

Abstract. This study was designed to explore the usability of three remote controls that operate a specific digital set top box (Logik LDR V3) amongst consumers who may have more difficulty than most in accessing and using digital television equipment. Participants were UK consumers (a) aged over 75 years with various sensory, physical and/or cognitive impairment, (b) aged under 45 years with learning difficulties, and (c) aged under 45 years without any sensory, physical or cognitive impairment. Using a repeated measures design, participants were asked to perform a series of everyday tasks using remote controls with digital television equipment. Subjective and objective data were collected to explore how intuitive and desirable the remote controls were, and how well subjective preferences related to objective performance data. The study provides evidence that user interfaces that meet the UK Digital TV receiver recommendations (V1.3) for digital terrestrial television (section 5: remote controls) better meet the needs of consumers likely to face difficulty using digital television equipment.

Keywords: usability, user interface, digital television, digital set top box, remote control, elderly, impairment.

1 Introduction

The process of digital television switchover (DSO) started in the Copeland region at the end of last year and is scheduled to end in 2012. For many, digital television switchover will involve learning to use a new remote control for everyday television use. Some consumers with the biggest DSO related support needs are entitled to apply for assistance via a Government established Digital Switchover Help Scheme (DSHS).

Research into good practice of remote control design features has been reported previously [1,2,3,4,5]. Lab-based research into button labeling and other usability research has informed the UK Digital TV receiver recommendations (V1.3) for digital terrestrial television [1,2,4]. Recommendations include variation in button size, shape, texture and spacing, functional groupings, and positioning.

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 318–322, 2008.

The study reported here was commissioned by the UK Government Department for Business, Enterprise and Regulatory Reform (BERR) to address specific items under the Usability Action Plan [9]. It provides BERR with an independent evaluation of the market in this area to inform their work. The subjective and objective performance benefits of 3 remote controls that operate the DSHS set top box used in the Copeland switchover, the Logik LDR V3, were tested in this multi-phase research study (see Figure 1). Each remote control was compared against the remote control features recommended in the UK Digital TV Receiver Recommendations (V1.3). Results showed that remote control 1 (DSHS in Copeland) best matched the recommendations (90%), followed by remote control 2 (manufactured by TW Electronics; 85%), then remote control 3 (manufactured by Tvonics; 63%). Next, each was subjected to a detailed expert evaluation. And finally, the comparative usability (intuitiveness) and aesthetic appeal of the three remote controls was tested by consumers likely to have more difficulty than others in getting and using digital television equipment, and a control group of consumers. This paper focuses on the final experimental phase of this research.

Fig. 1. (From left to right) Remote control 1 (DSHS in Copeland), 2 (TW Electronics) and 3 (Tvonics)

2 Method

Thirty-five people were recruited comprising 3 sub-samples of participants: (a) 12 adults aged over 75 years ('75+') including people with various combinations of mild visual and dexterity impairments, and age related cognitive decline; (b) 11 adults aged under 45 years, with various learning ('cognitive') difficulties including Autism, Asperger's Syndrome, and Attention Deficit Disorder; and (c) a control group of 12 adults aged under 45 years, without any sensory, cognitive or physical impairment ('young').

A repeated measures design was employed. The trial order with each remote control was counterbalanced across participants. The independent variable was remote control for which there were three types: (i) the remote control that was supplied with the DSHS set top box in Copeland, designed to meet remote control recommendations (rc1); (ii) the 'Echo' remote control, manufactured by TW Electronics, also designed to meet remote control design recommendations and able to control the DSHS digital box as well as basic functions of the TV to which the DSHS digital box is connected (power-standby on/off, volume up and down, mute) (rc2); and (iii) a thin, curved edged remote control that uses the same infra-red code-set as that receivable by the DSHS digital box, which had not been specifically designed to meet the remote control recommendations and has a younger target market than catered for by the Help Scheme (rc3).

Participants were asked to perform a series of 12 typical TV-use tasks, which relate to basic use, access service use or interactive service use, with each remote control. The tasks were presented in fixed presentation order in a logical sequence that might be typical with TV use, ranging from 'change channel' (basic use), through 'access' and 'exit' 'text' (interactive service use), 'access subtitles' (access service use); to 'switch off set top box' (basic use). Time to identify the correct button was recorded by two independent observers. Time was recoded into one of five categories [A=Pass <10s; B=Pass 11-20s; C=Pass 21-30s; D=Fail >30s (time out); E=Fail, participant gives up]. At the end of each trial simple 4 point Likert scale ease of use and attractiveness ratings were collected. At the end of the three trials, participants were asked to directly compare the remote controls (judge their favourite, most expensive, most similar to home TV remote control). Participants were thanked, de-briefed and received a small payment for their participation.

3 Results and Conclusions

Participants performed better with the remote controls that more closely met Core Receiver Requirements. Across all groups participants performed fastest with remote control 1, then remote control 2, then 3. Statistically this difference (irrespective of user group) approached statistical significance ($F(2,62) = 3.04$; $p.< 0.06$). This effect was largely due to faster performance with remote control 1 compared with remote control 3; no other remote control comparisons (rc1 vs. rc2; rc2 vs. rc3) approached significance.

Adults aged over 75 years, with a range of impairments had the most difficulty using all 3 remote controls. Participants in the ('young') control group were able to complete all the usage tasks with each of the remote controls, with limited differences in time taken to complete the tasks with each of the remote controls. The difference between user groups (irrespective of remote control) was highly significant ($F(2,31) = 17.78$; $p.<0.01$). For all tasks taken together, each user group was significantly different from each other ($p < 0.05$); the 'young' group was significantly faster than the 'cognitive' group which in turn was significantly faster than the '75+' group. The '75+' group found that the remote control which least met design guidelines and relied most on button labels to identify button functionality was the most difficult to use. For this remote control, button layout was indistinct and uniform and all button shapes and sizes were identical. In contrast, the '75+' group found that the remote controls which best met design guidelines were the easiest to use; they had larger handsets with variation in button size, shape, layout and height.

Similar proportions of adult participants aged over 75 years, with a range of impairments, successfully completed usage tasks within the designated time limit using remote controls 1 and 2; they were less able to complete usage tasks with remote control 3. For the tasks of switching the digital set top box on (at the start of each trial) and switching it off (at the end of each trial) successful completion within the designated time limit of 30 seconds was more likely with remote control 1. This result is not surprising given that remote control 2 included two power standby on/off buttons, one for the digital box, the other for the television.

Participants aged over 75 years were more likely to 'change channel' (numeric channel entry, and channel up and down) in under 10 seconds with remote control 2 than with remote control 1, a result that is likely to have arisen from the atypical location of the "0" (zero) key on remote control 1.

Ease of use ratings were consistent with the extent to which the remote controls meet Core Receiver Requirements. For all user groups as a whole, there was a significant main effect of remote control ($F(2,64) = 5.24$, $p < 0.01$): remote control 1 was rated significantly more easy to use than remote control 3 ($p<0.01$), and remote control 2 was rated more easy to use than remote control 3 although this difference only approached significance ($p = 0.07$).

The difference between user groups' ease of use ratings were significant ($F(2,32) = 13.8$; $p < 0.01$): the '75+' group gave significantly lower ease of use ratings to all the remote controls compared with both the 'young' and 'cognitive' groups, but there were no significant differences between the ease of use ratings of these latter two groups.

There was a significant user group*remote control interaction ($F(4,64)=2.72$, $P<0.05$). Whilst there was little difference in ease of use ratings across remote controls for the 'young' sample, for the 'cognitive' and '75+' user groups remote control 1 was perceived as easier to use than remote control 2, which in turn was perceived as easier to use than remote control 3.

Attractiveness ratings. Overall, statistically there were no significant main effects of remote control or user group on attractiveness ratings, though the latter approached significance. However there was a significant user group*remote control interaction ($F(4,64) = 2.51$, $p = 0.05$). Whilst there was little aesthetic preference for any remote control amongst the '75+' group (all remote controls were rated similarly), for the 'cognitive' group there was a preference for remote control 1, then 2, then 3; and for the 'young' group, the reverse was found (preference for 3, then 2, then 1).

Other ratings. Results for other ratings are beyond the scope of the current paper, and are published elsewhere.

Finally, significant proportions of the '75+' participants in the experiment did not find the design of any of the trial remote controls sufficiently intuitive to enable them to complete a variety of usage tasks. This finding suggests a need for the provision of additional tools to help, be it step-by-step usage instructions or a remote control that accesses reduced functionality. Indeed the performance results suggest that the inclusion of interactive navigation functions on remote controls can complicate usage for some older consumers without providing the user benefits for which they are designed to provide access.

Acknowledgements

This work was commissioned by the UK Government Department for Business Enterprise and Regulatory Reform. i2 media research limited at Goldsmiths would like to thank the following organisations for their assistance with recruitment of participants in this study: The Darby & Joan Club, Croydon; Toucan Employment, London; and Goldsmiths, University of London.

References

1. Summary of Research on the Ease of Use of Domestic Digital Television Equipment (March 2006)
2. Lessiter, J., Freeman, J., Davis, R., Dumbreck, A.: Understanding DTT remote control button labelling: a multimethod approach. In: Second European Conference on Interactive Television (EuroiTV 2004), University of Brighton (2004)
3. Freeman, J., Lessiter, J., Williams, A., Harrison, D.: Easy TV 2002 Research Report. ITC and Consumer's Association (2003)
4. Lessiter, J., Freeman, J., Davis, R., Dumbreck, A.: Helping viewers press the right buttons: Generating intuitive labels for digital terrestrial TV remote controls. Psychology 1(3), 355–377 (2003)
5. Freeman, J., Lessiter, J.: Easy to use digital television receivers: remote control buttons and functions used by different types of consumer. Report for Ofcom (2007)
6. Freeman, J., Lessiter, J., Beattie, E.: Digital Television Switchover and Disabled, Older, Isolated and Low Income consumers. Report commissioned jointly by Digital UK and Ofcom's Advisory Committee on Older and Disabled people (ACOD) (2007)
7. Freeman, J., Lessiter, J., Pugh, K.: Equipment needs of consumers facing the most difficulty switching to digital television. Report for the UK Department of Trade and Industry (2006)
8. Klein, J., Scott, N., Sinclair, K., Gale, S., Clarkson, J.: Equipment needs of consumers facing the most difficulty switching to digital television. Report for the UK Department of Trade and Industry (April 2006)
9. Government Department for Business Enterprise and Regulatory Reform, BERR has updated its Action Plan on usability of digital TV receivers (October 25, 2007)
10. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human–computer interaction. In: Karoulis, A., Demetriades, S., Pombortsis, A.: Comparison of Expert-Based and Empirical Evaluation Methodologies in the Case of a CBL Environment: The Orestis Experience (2006); Computers and Education 47(2), 172–185 (1994)

Author Index