Introduction to Special Issue: Human-Centered Television—Directions in Interactive Digital Television Research

PABLO CESAR
Centrum Wiskunde & Informatica
DICK C. A. BULTERMAN
Centrum Wiskunde & Informatica and Vrije Universiteit
and
LUIZ FERNANDO GOMES SOARES
Pontificia Universidade Católica do Rio de Janeiro

Introduction to Special Issue: Human-Centered Television—Directions in Interactive Digital Television Research

PABLO CESAR
Centrum Wiskunde & Informatica
DICK C. A. BULTERMAN
Centrum Wiskunde & Informatica and Vrije Universiteit
and
LUIZ FERNANDO GOMES SOARES
Pontificia Universidade Católica do Rio de Janeiro

The research area of interactive digital TV is in the midst of a significant revival. Unlike the first generation of digital TV, which focused on producer concerns that effectively limited (re)distribution, the current generation of research is closely linked to the role of the user in selecting, producing, and distributing content. The research field of interactive digital television is being transformed into a study of human-centered television. Our guest editorial reviews relevant aspects of this transformation in the three main stages of the content lifecycle: content production, content delivery, and content consumption. While past research on content production tools focused on full-fledged authoring tools for professional editors, current research studies lightweight, often informal end-user authoring systems. In terms of content delivery, user-oriented infrastructures such as peer-to-peer are being seen as alternatives to more traditional broadcast solutions. Moreover, end-user interaction is no longer limited to content selection, but now facilitates nonlinear participatory television productions. Finally, user-to-user communication technologies have allowed television to become a central component of an interconnected social experience. The background context given in this article provides a framework for appreciating the significance of four detailed contributions that highlight important directions in transforming interactive television research.

Categories and Subject Descriptors: H.1.2. [**Models and Principles**]: User/Machine Systems—*Human factors*; H.5.2 [**Information Interfaces and Presentation**] User Interfaces; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Video*; J.7 [**Computers in Other Systems**] Consumer products

General Terms: Design, Human Factors, Standardization

Additional Key Words and Phrases: Interactive television, shared experiences, standards, survey

ACM Reference Format:

Cesar, P., Bulterman, D.C.A., and Soares, L.F.G. 2008. Introduction to special issue: Human-centered television—directions in interactive digital television research. ACM Trans. Multimed. Comput. Comm. Appl. 4, 4, Article 24 (October 2008), 07 pages. DOI = 10.1145/1412196.1412197 http://doi.acm.org/10.1145/1412196.1412197

This research was supported by the NWO BRICKS PDC3 project, the FP7 IP project TA2, and the FP7 STREP project iNEM4U.
Authors' addresses: P. Cesar and D.C.A. Bulterman, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands, email: p.s.cesar@cwi.nl; dick.bulterman@cwi.nl; L.F. Gomes Soares, Pontificia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil, email: lfgs@inf.puc-rio.br.

1. INTRODUCTION

The past decade has seen a fundamental evolution in the attitude of the end-user towards networked media. Beginning with the World Wide Web (WWW), the content infrastructure has changed from a static global library to a dynamic meeting place in which users not only consume but share rich media content. Similar developments are happening in the television domain, where person-to-person communication mechanisms are (re)shaping the way people consume media at home and the way they interact with each other.

This special issue provides an insight into emergent research across the television content production lifecycle. Before introducing the four articles included in this issue, we start with a review of current research challenges that the interactive TV community is facing. We base this discussion on the traditional television content flow, identifying basic building blocks from a functional viewpoint. We then summarize research directions that will impact future work. We conclude by introducing each of the articles in this special issue.

2. TRADITIONAL TELEVISION CONTENT FLOW

Traditionally, the outward flow of television content starts when raw media is captured. After acquisition, a variety of media elements are encoded and aggregated into a coherent program, possibly including handlers for user interaction. The content package is sent to end users via a distribution infrastructure, in which assets (and economic value) are added incrementally. Finally, the content is consumed by the end user, using both a primary delivery channel and occasionally an interactive return channel. Together, these activities form the three major stages of the content lifecycle: content production, content delivery, and content consumption. They are illustrated in Figure 1.

During the production/composition stage, extensive studio settings are used. Major research in this area includes the provision of efficient video-encoding mechanisms for effective video-streaming rendering. This led to important solutions that include the Moving Picture Experts Group (MPEG)-2[1] and the MPEG-4[2] video formats. Even though these formats provide an efficient packing solution for linear content, they provide only limited capabilities for interaction. This need has fostered higher-level formats to allow the integration and synchronization of individual media elements and user interaction capabilities. Some examples include Synchronized Multimedia Integration Language (SMIL) [Bulterman and Rutledge 2004], Flash,[3] and Nested Context Language (NCL) [Soares et al. 2006]. These integration languages allow multimedia presentations to be created by defining flexible spatial and temporal relationships among media elements. Although popular within the research community, these solutions have been largely ignored in television standards. Instead, more restrictive frameworks have been used, such as Multimedia Home Platform (MHP)[4] in Europe [Morris and Smith-Chaigneau 2005; Cesar et al. 2006] and the Open Cable Platform (OCAP)[5] in the USA [Morris and Smith-Chaigneau 2005]. An exception has been *Ginga* in Brazil [Soares et al. 2007], which includes support for NCL. The lack of acceptance for innovation is also present for semantic modeling metadata annotations: in spite of significant work performed by the TV-Anytime[6] Forum, the majority of digital television systems continue to use MPEG-2 description tables [Lugmayr et al. 2004].

---

[1] http://www.chiariglione.org/mpeg/standards/mpeg-2/mpeg-2.htm
[2] http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm
[3] http://www.adobe.com/products/flash/
[4] http://www.mhp.org/
[5] http://www.opencable.com/
[6] http://www.tv-anytime.org/

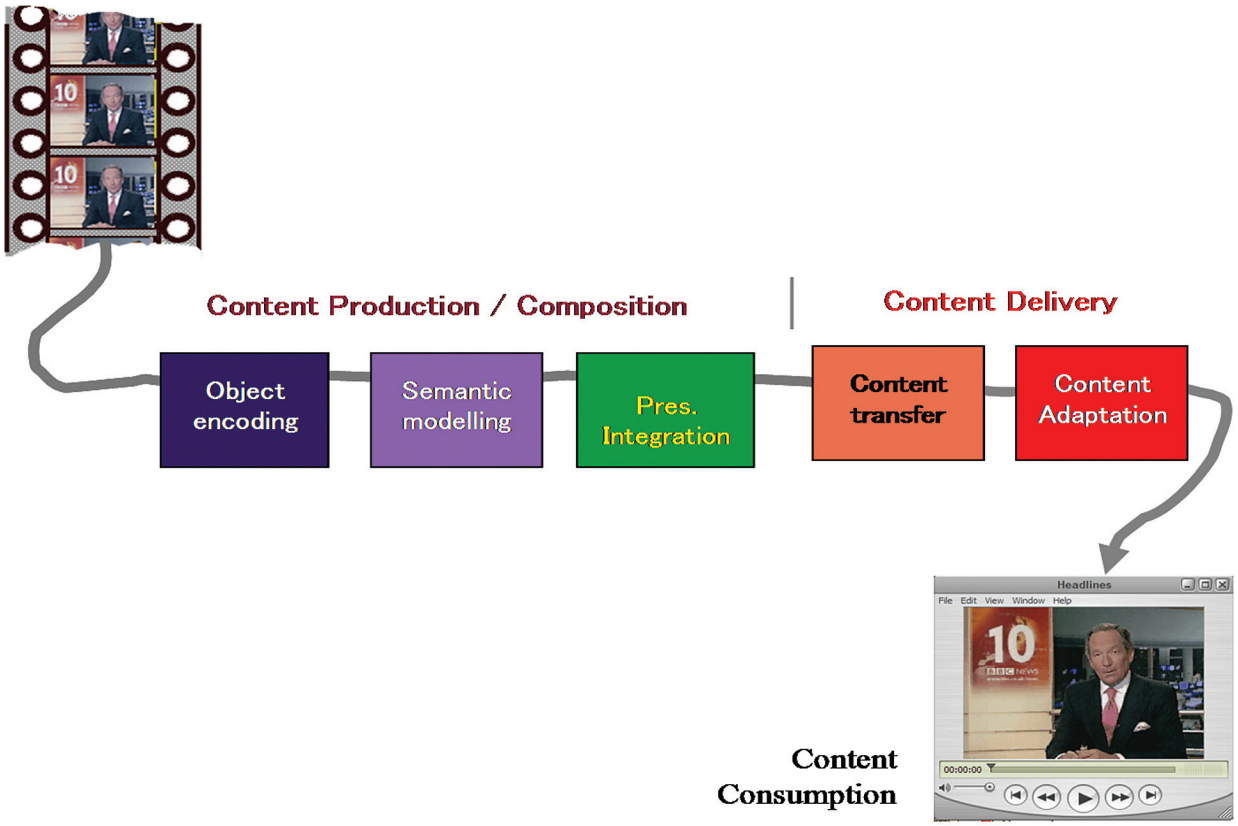

Fig. 1. The content production lifecycle.

The complexities of the national and international content delivery infrastructure focused much of the early work on digital television on two main topics: the efficient distribution of the massive amounts of data involved in a digital content stream and the commercial packaging and protection of highly licensed TV content. This wave of research concluded in digital television systems where three major regional body standards were created: Advanced Television Systems Committee Advanced (ATSC) in North America, International Standard for Digital Broadcasting (ISDB) in Japan and Brazil, and Digital Video Broadcasting (DVB) in Europe. While the focus of these organizations has been to manage the transition to digital distribution of content, they have also considered ways to make the relatively passive television environment more interactive.

There are three main areas of end-user interaction that have been addressed by first-generation interactive television standards: interacting with (and within) a particular piece of content, interacting across multiple streams of content, and interacting with a single stream on multiple devices. We consider each of these briefly in turn.

The principal end-user interaction innovation in first-generation standards has been the provision of a degree of fine-grained interactive content encoded in a collection of on-demand asynchronous media components that augment a given television program. These components are transmitted cyclically in elementary data streams, using the data carousel in the MPEG-2 Digital Storage Media–Command and Control (DSM-CC) protocol [Reimers 2005]. Nevertheless, efficient carousel management

procedures and resource identification are still issues that deserve more attention regarding pushed and pulled data [Costa et al. 2008]. The data carousel allows lightweight applications to be downloaded and executed on end-user devices such as set-top boxes. For the first time, interactive applications, synchronized to the video content could be launched. Games, enhanced information about the viewed program, or quizzes are common services for interactive television systems. These systems are still under development, such as providing a means to insert, in real-time, enhanced services when, for example, a goal is scored in a soccer match [Costa et al. 2006]. Fine-grained interactive applications were also originally seen as the interface for commercial applications, such as home shopping. Because of the one-way nature of content delivery, however, such applications required a separate return infrastructure. Ad hoc solutions, including separate low-bandwidth telephone back-channels, lacked a substantial degree of fluid interaction with the downstream video content. This lack of significant interaction capabilities severely limited the potential of interactive television. At the same time, advances in the storage structure of set-top boxes allowed end users to obtain an unexpected degree of interactive freedom: personal video recording systems such as TiVo[7] provided users with the ability to alter distribution schedules (time-shifted consumption). This was augmented with linear navigation capabilities such as play, pause, forward, and rewind within the context of a single program. The significance of these developments was that, for the first time, the user was able to gain control of the flow of content in the home rather than having this flow dictated by the infrastructure.

The growth of the content offering brought about by new transmission technologies, coupled with the ability for time-shifted consumption, provided another dimension of interaction; interfaces for managing the vast collection of programs that could be saved from the large number of additional television channels. Active systems, such as the electronic program guide, and more passive interfaces, such as personalization systems, were devised to support content selection in the midst of a video-data explosion [Adomavicius and Tuzhilin 2005, Ardissono et al. 2004; Smyth and Cotter 2000]. At the same time, a diversity in the types of end-user rendering equipment required new solutions for content adaptation that had previously not existed in the television infrastructure. The adaptation could take place locally at the terminal, but more often, it was managed by a recoding prior to content delivery. This increased the complexity of the distribution infrastructure without significantly increasing the degree of end-user interaction.

Placed against the context of an interactive Internet infrastructure, the first generation of interactive television seems exceptionally parochial; the original model was top-down and proscriptive and only open to incidental end-user interaction at the content fringes. The walled garden approach was evident across the entire infrastructure from inflexible content packaging to restrictive content distribution. The end-user's role was one of the passive consumer with little more control than content selection.

3. CURRENT AND FUTURE DIRECTIONS

The second generation of interactive television has not departed significantly from the traditional content lifecycle. Content still needs to be gathered, aggregated, modeled, and distributed. The key change is that the end user, not an individual content stream, has become the central focus of the interaction infrastructure. Initially, this change has been visible in the early stages of acceptance of user-generated content. Rather than this being the end of a process of transformation, we are convinced that it is simply the beginning.

The increased role of the end user across the content lifecycle can be illustrated by three examples. First, we consider a shift in production technology, resulting in new lightweight authoring systems.

[7]http://www.tivo.com/

Second, we consider a shift in the infrastructures for media sharing from centralized to peer-level distributed. Third, we consider a shift in content paradigms from the isolated to the social.

3.1 Example 1. Production: Lightweight Authoring

There have been several events that have reshaped multimedia authoring for television-like applications. One is the ubiquity of raw media capture devices, most notably the wide proliferation of mobile devices with video cameras, which has had a major impact on how we understand media production [Kirk et al. 2007]. The second is the demonstrated ability of end users to create (relatively) compelling content beyond the genre of stupid pet tricks. The popularity of YouTube[8] and its derivatives has demonstrated that individuals can create content that interests some portion of the masses (assuming, of course, that it can be found). A third is that an increasing number of interfaces have been developed that allow end users to remix and to repurpose multimedia content [Shaw and Schmitz 2006]. The main idea behind these systems is to reuse existing material on the Web or on the television channel repositories in order to create new media content ready for consumption. Such remixing is not restricted to directed content authoring; other systems consider the authoring process as an incidental and spontaneous or a lightweight authoring task that can be performed during content consumption [Bulterman 2007]. The functionality required for such systems include video fragmentation, video annotation, and video enrichment (e.g., audio commentaries) [Cesar et al. 2008]. The nature of remixing (semi-)private media assets requires that content enhancements and compilations not be encoded as closed media object but rather as a layered collection of enrichments [Bulterman 2003]. This way legal rights can be protected while still allowing media enrichments to be shared. This also allows subsequent generations of media asset use to be traced, searched, and nondestructively annotated.

3.2 Example 2. Delivery: Media Sharing

Interactive video is becoming a popular Web content type not so much because of the conventional broadcasting infrastructure but in spite of it. Major newspapers and local television outlets are making their video content available online, and video sharing systems (such as YouTube or MySpace[9]) provide interfaces for uploading, searching, viewing, and rating videos. Distribution has been expanded from an anonymous infrastructure to a recommendation-based direct distribution model. At the same time, traditional communication systems are starting to integrate rich media capabilities into immediate messaging systems. For example, enriched instant messaging solutions such as Messenger TV[10] from Microsoft and Zync! [Shamma et al. 2008] from Yahoo! provide the option of sharing synchronized videos while chatting. From a human-centered perspective, such direct recommendation interfaces represent an important research direction for further exploration. Although the value of centralized delivery of significant temporal events will remain high, it can be expected that there will be an increased recognition of the value of direct recommendation as a delivery paradigm [Gill et al. 2007].

3.3 Example 3. Consumption: Social Television

While lightweight authoring and directed media sharing are important indicators of change, one of the most significant developments for interactive television has been the broader acceptance of a merging of personal content consumption and social sharing. Social television tries to enrich the human connectedness between individuals consuming media content in different locations. This line of research, which is enjoying a renewal of interest, primarily focuses on providing one-to-one or one-to-many communication

[8]http://www.youtube.com
[9]http://www.myspace.com/
[10]http://messengertv.msn.com/

mechanisms between peers watching media content. Such mechanisms are currently synchronous (using an audio link or a text chat), although asynchronous solutions are beginning to appear. Luckily, a number of technical challenges remain, including scalability, synchronization of media streams across multiple locations, noise reduction, and audio-video signal processing among the available peers. This should provide a source of interesting results in the coming decade. Other essential challenges include presence awareness [Harboe et al. 2008], design guidelines [Ducheneaut et al. 2008], privacy concerns, and social network integration. One interesting example service is the torcida virtual[11] or virtual fans where groups of disjoint users can select virtual seats together in a remote stadium. Each user receives the video and hears the sound from his seat location together with the sounds of his friends.

4. FOUR VIEWS OF HUMAN-CENTERED TELEVISION

The examples presented in the previous section point to a specific direction: human-centered television. In this paradigm, the end user becomes an active node in the multimedia content chain as producer and distributor, even as a contributor of computing resources. Moreover, the end user attains increased interactive capabilities, both within and across content streams, for himself and his social network.

This special issue provides a collection of views of the human-centered television paradigm. The underlying argument behind this special issue is that a more human-centered approach is required in all the stages of the television content lifecycle: production, delivery, and consumption. The content lifecycle no longer ends when television content is broadcast.

This issue, contains the following contributions.

—The first article, Interactive TV Narratives: Opportunities, Progress, and Challenges, presents from the content producers perspective a new authoring tool for interactive narratives. Such tools are the basis of innovative use cases in which television viewers can actively select and decide how television programs will evolve.

—The second article, GridCast: Improving Peer Sharing for P2P VoD, introduces P2P technology as a viable solution for digital television, in this case Video on Demand services, transmission. The authors introduce a deployed working system, including user tests and performance evaluations.

—The third article, Examining Presence and Lightweight Messaging in a Social Television Experience, represents another step in the social interactive television sphere. The work investigates human connectedness around the television experience, focusing on issues such as nonobtrusive presence awareness and instant messaging capabilities across different locations.

—The fourth and final article, Watch-and-Comment as a Paradigm toward Ubiquitous Interactive Video Editing, introduces an architecture that allows end users to enrich fragments of television content using ink or audio commentaries. Such enrichments then can be shared with other people via a P2P infrastructure.

All of these articles, in addition to being interesting individual contributions, provide significant markers in the transition to a new generation of interactive television systems. In this generation, the end user, once a faceless, nameless entity without any direct influence on content and quality, becomes the centerpiece of a new human-centered interactive television infrastructure.

ACKNOWLEDGMENTS

We are thankful to the external reviewers who helped us in the article selection process for this special issue: Konstantinos Chorianopoulos, Ionian University, Greece, David Geerts, Katholieke Universiteit

[11]http://www.hxd.com.br/site/index.php?option=com_content&task=view&id=35&Itemid=80

Leuven, Belgium, Gunnar Harboe, Motorola Labs, USA, Hendrik Knoche, UCL, UK, Artur Lugmayr, Tampere University of Technology, Finland, Marianna Obrist, ICT&S Center, Austria, Jose Pazos, University of Vigo, Spain, and Marian F. Ursu, Goldsmiths College, University of London, UK.

REFERENCES

ADOMAVICIUS, G. AND TUZHILIN, A. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Engin. 17*, 6, 734–749.

ARDISSONO, L., KOBSA, A., AND MAYBURY, M. T. (Eds.). 2004. *Personalized Digital Television Targeting Programs to Individual Viewers*. Kluwer Academic Publishers.

BULTERMAN, D. C. A. 2003. Using SMIL to encode interactive, peer-level multimedia annotations. In *Proceedings of ACM Symposium on Document Engineering*. 32–41.

BULTERMAN, D. C. A. AND RUTLEDGE, L. 2004. *SMIL 2.0: Interactive Multimedia for Web and Mobile Devices*. Springer-Verlag, Berlin, Germany.

BULTERMAN, D. C. A. 2007. User-centered control within multimedia presentations. *Multimed. Syst. J. 12*, 4–5, 423–438.

CESAR, P., VUORIMAA, P., AND VIERINEN, J. 2006. A graphics architecture for high-end interactive television terminals. *ACM Trans. Multimed. Comput. Comm. Appl. 2*, 4, 343–357.

CESAR, P., BULTERMAN, D. C. A., GEERTS, D., JANSEN, A. J., KNOCHE, H., AND SEAGER, W. 2008. Enhancing social sharing of videos: Fragment, annotate, enrich, and share. In *Proceedings of ACM Multimedia Conference*.

COSTA, R. M. R., MORENO, M. F., RODRIGUES, R. F., SOARES, L. F. G. 2006. Live editing of hypermedia documents. In *Proceedings of the ACM Symposium on Document Engineering*. 165–175.

COSTA, R. M. R., MORENO, M. F., AND SOARES L. F. G. 2008. Intermedia synchronization management in DTV systems. In *Proceedings of the ACM Symposium on Document Engineering*.

DUCHENEAUT, N., MOORE, R. J., OEHLBERG, L., THORNTON, J. D., AND NICKELL, E. 2008. SocialTV: Designing for distributed, social television viewing. *Int. J. Hum.-Comput. Interac. 24*, 2, 136–154.

GILL, P., ARLITT, M., LI, Z., AND MAHANTI, A. 2007. Youtube traffic characterization: A view from the edge. In *Proceedings of the SIGCOMM Conference on Internet Measurement*. 15–28.

HARBOE, G., MASSEY, N., METCALF, C., WHEATLEY, D., AND ROMANO, G. 2008. The uses of social television. *Comput. Entertain. 6*, 1.

KIRK, D., SELLEN, A., HARPER, R., AND WOOD, K. 2007. Understanding videowork. In *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems*, 61–70.

LUGMAYR, A., NIIRANEN, S., AND KALLI, S. 2004. *Digital Interactive TV and Metadata Future Broadcast Multimedia*. Springer-Verlag, Berlin, Germany.

MORRIS, S. AND SMITH-CHAIGNEAU, A. 2005. *Interactive TV Standards: A Guide to MHP, OCAP and JavaTV*. Focal Press.

REIMERS, U. 2005. *DVB: The Family of International Standards for Digital Video Broadcasting*. Springer-Verlag, Berlin, Germany.

SHAMMA, D. A., BASTEA-FORTE, M., JOUBERT, N., AND LIU, Y. 2008. Enhancing online personal connections through the synchronized sharing of online video. In *Extended Abstracts on Human Factors in Computing Systems*, 2931–2936.

SHAW, R., AND SCHMITZ, P. 2006. Community annotation and remix: A research platform and pilot deployment. In *Proceedings of the International Workshop on Human Centered Multimedia*. 89–98.

SMYTH, B. AND COTTER, P. 2000. A personalized television listings service. *Comm. ACM. 43*, 8, 107–111.

SOARES, L. F. G., AND RODRIGUES, R. F. 2006. Nested context language 3.0: Part 8 - NCL (Nested Context Language) digital TV profiles. *Tech. rep.* No. 35/06. Informatics Department, PUC-Rio.

SOARES, L. F. G., RODRIGUES, R. F., AND MORENO, M. F. 2007. Ginga-NCL: The declarative environment of the Brazilian digital TV system. *J. Brazilian Comput. Soc. 13*, 37–46.