Foundations and Trends® in
Human–Computer Interaction
Vol. 2, No. 4 (2008) 279–373

DOI: 10.1561/1100000008

now
the essence of knowledge

The Evolution of TV Systems, Content, and Users Toward Interactivity

By Pablo Cesar and Konstantinos Chorianopoulos

Contents

Foundations and Trends® in
Human–Computer Interaction
Vol. 2, No. 4 (2008) 279–373

DOI: 10.1561/1100000008

now
the essence of knowledge

The Evolution of TV Systems, Content, and Users Toward Interactivity

Pablo Cesar[1] and Konstantinos Chorianopoulos[2]

[1] *CWI, The Netherlands, P.S.Cesar@cwi.nl*
[2] *Department of Informatics, Ionian University, Corfu, Greece, choko@ionio.gr*

Abstract

Interactive TV research spans across a rather diverse body of scientific subfields. Research articles have appeared in several venues, such as multimedia, HCI, CSCW, UIST, user modeling, media and communication sciences. In this study, we explore the state-of-the-art and consider two basic issues: What is interactive TV research? Can it help us reinvent the practices of authoring, delivering, and watching TV? For this purpose, we have reviewed the research literature, as well as the industrial developments and identified three concepts that provide a high-level taxonomy of interactive TV research: (1) content editing, (2) content sharing, and (3) content control. We propose this simple taxonomy (edit–share–control) as an evolutionary step over the established hierarchical produce–deliver–consume paradigm. Moreover, we demonstrate how each disciplinary effort has contributed to and why the full potential of interactive TV has not yet been fulfilled. Finally,

we describe how interdisciplinary approaches could provide solutions to some notable contemporary research issues.

> *'Interactive Television is an oxymoron. On the other hand, television provides the most common ground in our culture for ordinary conversation, which is arguably the most enjoyable interaction a person has. We should try to leverage the power of television while creating some channel back from the audience to provide content, control or just a little conversation.'*[*]

[*] Although we have tried to summarize previous research as much as possible, we still find that the quote by Dan O'Sullivan (Interactive Telecommunication program, New York University, Tisch School of the Arts) has been the most comprehensive definition of interactive television, so far. Retrieved from: http://itp.nyu.edu/∼dbo3/proj/#tele (July 2008).

1

What is Interactive Television

The user activities that surround television creation, distribution, and viewing have been interactive long before the digitization of television systems. For example, viewers compete mentally against quiz show participants. Moreover, viewers react emotionally to TV content, they record and share TV content with friends and discuss about shows either in real-time, or afterward. Currently, the digitization of TV systems and TV content has only increased the opportunities for interactivity. A major question that should be answered before we describe the details of this research area is: "what is interactive TV (iTV)?" Despite the widespread use in industry and academia, the term "iTV" is still quite ambiguous.

For a long time, the answer to the question "what is iTV" has been dependable on the discipline or the industry concerned, which might have been one source of ambiguity when the respective disciplines had to coordinate:

(1) i**TV as infrastructure:** A telecom engineer assumes digital broadcast, return channel, or broadband Internet infrastructure (e.g., IPTV);

(2) i**TV as user terminal:** A multimedia designer refers to interactive graphics and dynamic editing on the user terminal;
(3) i**TV as media format:** A media manager describes new content formats such as betting, interactive storytelling, and play-along quiz games; and
(4) i**TV as social actor:** A sociologist's definition focuses on the interaction between people about TV shows.

While none of the above definitions seems to agree with each other, each one stands for an approach followed by iTV researchers so far. In particular, each one makes some assumption about one or more of the following elements: (1) infrastructure, (2) user terminal, (3) content, and (4) social behavior, respectively. Therefore, in order to setup a unifying definition of iTV we need to abstract from the particularities of disciplinary approaches and their implicit assumptions. We have found that there are at least two high-level approaches for defining iTV. The first one considers iTV as an artifact or experience. The second approach considers iTV as an area of academic study.

In terms of user experience, we consider interactive TV (iTV) to hold the following properties: (1) mash-ups of fixed (pre-edited) video-clips, which have linear narrative (2) low-to-mild levels of user input, and (3) dynamic graphics that are employed mostly for video-overlays. Nevertheless, the borderline between other media formats (e.g., videogames) and iTV is sometimes vague. For example, there are song-contest videogames that follow the format of the respective TV-shows. At the same time, there are iTV formats that have been modeled after adventure videogames. For the sake of consistency within this study, we do not treat borderline applications, but we provide a few references to developments from the industry and mainly focus on the academic treatments of iTV.

In terms of academic discipline, iTV research studies the interaction among users and video-clip-based content, which is presented on networked multimedia computers. Therefore, iTV research builds and extends upon established disciplines such as Human–Computer

Interaction, Multimedia, and Communication Science. Again, there might be borderline cases, in which research methods in iTV have been transferred from other disciplines. Nevertheless, iTV research focuses on those interdisciplinary cases that have guided researchers to leverage existing disciplinary methods, in order to address the development and use of iTV systems.

The goal of this work is to provide a common framework for future iTV research by surveying the most relevant publications and the most innovative industry developments. In order to abstract from the different disciplines and views, we structure the framework on three basic television concepts that we believe capture the basics of all previous approaches: (1) content editing, (2) content sharing, and (3) content control. In the following section, we provide further details regarding the scope and assumptions that we made in the course of this work.

1.1 Framework and Delimitation of Scope

This section provides a detailed description and rationale of the framework we utilize to position the different initiatives around iTV research. It delimits the scope of the study and highlights key assumptions.

Firstly, the intention of this study is not to enumerate the most significant technological achievements in terms of television delivery. Although several iTV developments (e.g., Web-based TV, IPTV, and broadcast TV) have followed parallel or even competing paths, we prefer to elaborate on the common themes from the viewpoint of the human, as a creator, distributor, and viewer of content. For example, broadcast developments have been in competition with video streaming approaches, and the TV as device has been in conflict with the PC. Nevertheless, the convergence of network and rendering platforms has made such distinctions somewhat superficial. Even though there are still significant differences between the networking and rendering platforms, those differences regard mostly to the context and the preferences of the user, rather than to the capabilities of the technology.

As introduced by Pine and Gilmore [143] in The Experience Economy, we are living a shift from a service economy to an experience economy. In other words, if the first technological challenge was to

provide efficient delivery mechanism, now the challenge is to provide enhanced experiences [13]. While during the 1990s iTV research concentrated in the provision of digital television and on how efficiently broadcast digitalized television, the challenge now is to provide interactive television experiences as represented by the efforts of personalization, social television, interactive narratives, and ambient technology.

With the goal of being as inclusive as possible, this work takes a pragmatic view and considers research coming from the industry and the academia. Notably, many significant iTV developments have been realized by industrial players (content producers, network operators, and device manufacturers), who have very different strategies and interests.[1,2] For this reason, in addition to academic literature we have also examined iTV developments published in the popular press. Nevertheless, it is outside the scope of the present work to provide an overview of all commercial trials and products, which are described elsewhere [89, 88, 141].[3,4,5]

The goal of this study is to provide a comprehensive overview of iTV research around a unifying concept: television as a set of activities that include content edition, content sharing, and content control. In the rest of this article we organize previous research and development efforts along the three major categories, which have an immediate impact on the way people interact and participate in the TV lifecycle.

Content editing, apart from professional content edition, considers the casual viewer as an active node in the content creation value chain. Contemporary viewers have the expectation of producing digital content by employing easy-to-use applications. Although the current shift has important implications in the television value chain, we do

[1] Frank Rose, The Televisionspace Race, Wired 6.04, 1998 http://www.wired.com/wired/archive/6.04/mstv.html

[2] Frank Rose, TV or not TV, Wired Issue 8.03, 2000, http://www.wired.com/wired/archive/8.03/bskyb.html

[3] Sean Dodson, A short history of interactive TV, guardian.co.uk, Thursday 5 April 2001, http://www.guardian.co.uk/technology/2001/apr/05/onlinesupplement5

[4] Robert X. Cringely, Digital TV: A Cringely Crash Course, PBS, http://www.pbs.org/opb/crashcourse/

[5] Tracy Swedlow, Interactive Enhanced Television: A Historical and Critical Perspective, http://www.itvt.com/etvwhitepaper.html

not expect that professional content production will disappear in the future. Indeed, high-quality production values and massively attended events function as a reference point and as social glue for society [100]. At the same time, the popularity of services like YouTube and MySpace demonstrates that there is an increasing demand for user-generated content. In conclusion, there is a need to accommodate both approaches by providing lightweight authoring tools for end-users.

Content sharing corresponds to a meta-content activity, "have you seen that goal?" or "you should definitely watch this clip!" When a viewer calls a friend to chat about a current program, he is following a communication process. This process can be synchronous (while viewing) or asynchronous (after viewing). Research on communication process includes, among others, providing chat-enabled television channels, real-time voice communication, or synchronous avatars that indicate the current status of a viewer.

Content control corresponds to the selection process, "what to watch?" and to the consumption process "Where to watch it?" For example, after scanning the program guide, when the viewer changes to another channel he is controlling the television content. Research on content control can be divided into a number of subtopics such as the input devices to be utilized, automation and personalization, and the available rendering devices.

1.2 Methodological Considerations

Researchers have employed several methodologies in the study of TV viewing and they have established a rich body of knowledge, which has been expanded by the design, development, and study of novel iTV content and applications. In the following, we highlight relevant methods from selected research in the iTV field.

Although researchers have identified the differences between the TV, the personal computer and the Web,[6,7] the majority of the research and many commercial products have been created in the face of usability

[6] Jakob Nielsen, WebTV Usability Review, February 1997: http://www.useit.com/alertbox/9702a.html

[7] Jakob Nielsen, TV Meets the web, February 1997: http://www.useit.com/alertbox/9702b.html

Fig. 1.1 In addition to the contrast between lean-back versus lean-forward user posture, the TV environment considers a shared dispaly and social activities in a relaxed domestic setting.

measured as efficiency. Several aspects of video search and navigation could be modeled after the traditional HCI tasks and goals. For example, the usability of the Electronic Program Guide (EPG) is very similar to the usability of productivity software, because it involves more information processing than enjoyment of iTV content. Still, there are some aspects of the EPG design and many other types of iTV applications that would benefit by a consideration of the affective dimension [45, 46]. The focus on the affective dimension of iTV applications was motivated by the realization that users' subjective satisfaction is at odds with the established notion of efficiency.

A usability test of a video skipping user interface (UI) revealed that user satisfaction was higher for the UI that required more time, more clicks, and had the highest error rate. In other words, the most usable UI was not the most favored one [58]. This result is opposite to the assumptions of the efficient usability paradigm, which conceives the efficient as more usable and thus preferable. One could not blame the designers of those efficient UIs (the widely acclaimed TiVo and ReplayTV), which have been designed according to the established UI principles (e.g., "provide shortcuts"). Nevertheless, the satisfaction questionnaires exposed that users preferred the most relaxing UI over the most efficient one [58]. Therefore, UI in ITV applications should be tested in the face of affective goals, in addition to the traditional efficient usability conceptualizations. In other words, upcoming user experience evaluation methodologies should be applied in the iTV domain.

In addition to the evaluation conceptualization, there are methodological differences with regard to the techniques and processes employed during the development of new iTV products and services. Monk [123] argued that there is a need to adapt the traditional UI design and evaluation methods to the home environment (Figure 1.1). Since iTV applications serve entertainment goals and domestic leisure activities for a diverse user population [101], there is a need to re-examine the traditional usability engineering concepts and evaluation methods, under the light of existing results from the field of media studies. Indeed, the intersection between the human-computer interaction (HCI) and the mass communication disciplines has been highlighted as a significant area for further research [112].

Chorianopoulos and Spinellis [47] have integrated the research from affective HCI with media studies, in order to devise a conceptualization for UI evaluation that facilitates the universal access to iTV applications. Mass communication has explored the effects of broadcast electronic media messages to the TV audience. It has developed several important concepts, such as the "uses and gratifications" theory [159], which describes the motivations for watching TV. The uses and gratifications theory does not assume an attentive user like the traditional usability engineering methods do, but measures explicitly a continuum of viewer involvement with a TV program [142]. Moreover, the "selective exposure" paradigm [190] regards the viewer as an active receiver of the media messages, who changes TV channels and actively selects TV content to be exposed to. The selective exposure concept contrasts with the traditional usability conception of a specific task to be performed by a user.

An important element in the process of usability evaluation is the notion of the user task. A user task consists of a finite number of steps and it has an exact ending. Accordingly, a usability evaluation session includes a few tasks that should be performed by a user. Tasks might not be suitable in the context of many iTV applications. Indeed, Maguire [113] raised the research question of whether tasks should be fixed, or users should be allowed to use the service as freely as they wish. It has been argued that the users should be allowed to use the service for a predefined, but flexible duration of time (e.g., 15–30 minutes),

without any particular task to complete [47]. Because viewers select TV channels and watch TV programs in order to regulate their mood, the evaluation of an iTV UI should facilitate free exploration and enjoyment of the iTV application. The emphasis on an affective methodology for iTV applications does not entail a complete abandonment of the efficient usability paradigm. For example, an iTV news application used in the morning before leaving home for work should afford efficient information retrieval and navigation. The same application, used in the evening after returning home from a long day at work, should be more automated and encourage relaxed use.

Shrimpton-Smith et al. [168] provide an empirical comparative evaluation study of two versions of the traditional think-aloud method. In particular, they suggest that since TV is a social medium it must be tested in a social context as well. For this purpose, they employed real life couples in think-aloud usability testing. The same usability test was also performed with single users. It was found that couples detected more usability issues than single test users. Furthermore, the test session was considered to require less effort in the couple condition. Besides collocated groups, there is also a need for evaluation methods in the context of distance communication among multiple TV viewers. Duchenaut et al. [59] performed an elaborate analysis of the voice communication between two remote groups of TV viewers. The evaluation was based on video-taping and detailed transcripts (both spoken and non-verbal) of the interpersonal communication, within the same room and between the two remote rooms.

In continuation to the past qualitative analysis of traditional TV audience [110], ethnographic studies in the living room are popular evaluation methods [130]. More recently, Obrist et al. [132] performed an extensive ethnographic study of interactive TV use. They employed diaries and cultural probes, and evaluated a broad range of iTV applications. They found that the preferences of different user groups (e.g., couples, singles, flatsharing, and seniors) could only be fulfilled with an equally diverse set of iTV applications, and they put special emphasis on social communication. Elderly users have been involved in the design of navigation interfaces [155]. In complement to qualitative studies, Sperring and Stradvall [174] employed multiple usability and media

evaluation methods including eye-tracking, questionnaires, and physiological measurements. They report that the viewers' behavior during the show and involvement in the game varied depending on whether they participated together with friends or alone.

While usability tests are suitable during the development process, ethnographic methods are more useful for requirements collection and for investigating the long-term effects of iTV applications. Bernhaupt et al. [22] developed two variations of cultural probes by introducing creative cultural probing cards and extending it toward playful cultural probing. Creative cultural probing material is based on the idea that creative stimuli will motivate participants in their self-observation to provide more insightful information on daily routines and technology usage. For the playful probing approach, traditional games are adopted for the research needs to enhance participants' involvement. For example, they extended "card games" by including research-related question cards. These questions were answered by participants while playing the game. Furthermore, they experimented with modeling clay as a means for answering design oriented questions. Overall, they found that the playful approach motivates participants to reflect on the research topic more thoroughly.

In summary, the contemporary usability techniques are necessary for the evaluation of iTV applications, but it seems that they are not sufficient. In particular, the TV audience has been accustomed to expect much more than ease of use. In particular, the TV audience receives information and expects to be entertained, in a lay-back posture and through an emotionally loaded visual language. In this way, having satisfied the basic usability requirement, everybody should be receiving a reasonable level of entertainment.

1.3 Timelines and Basic Concepts

After many decades of development, iTV has remained one of the most elusive consumer technologies [99]. Several reasons have been cited, such as pervasiveness of basic TV infrastructure (Figure 1.2),

[8] The sources for all timelines are wikipedia, http://www.fcc.gov, http://www.digitaltelevision.gov.uk

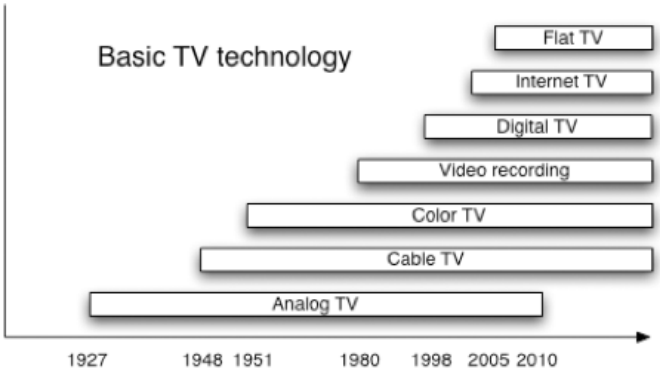


Fig. 1.2 Summary of basic technical advancements.[8]

unrealistic expectations, slow evolution of iTV technologies, and conflicting viewpoints of the stakeholders.[9,10,11]

There are several ways to look into the development of TV technology, applications, content, and social practices over time. In this section, we discuss multimedia content flow in TV and we study historical development for each building block of the TV value chain. In the timelines, we have selected the most significant technological advancements in terms of the impact they had in the way users (producers, distributors, and viewers) employed TV. It is worth noting that the proposed framework to study iTV research (edit–share–control) stands as an evolutionary step over the traditional model of authoring–delivery–consumption.

The flow of multimedia content is started when the media is captured. The raw material might be captured using digital means or can be, later, converted into digital format. Then, the content is encoded and might be authored by aggregating various media elements into one presentation, by determining the layout characteristics of each media element, and by introducing handlers for user interaction. Finally, the

[9] Kevin Kelly, Becoming Screen Literate, NY Times, November 23, 2008, http://www.nytimes.com/2008/11/23/magazine/23wwln-future-t.html

[10] Bill Rosenblatt, 500 channels and nothing's on, http://www.cnn.com/TECH/computing/9812/04/500channels.idg/index.html

[11] Bill McConnell, The Shape of Things To Come, Broadcasting & Cable, 1/5/2004, http://www.broadcastingcable.com/article/CA372624.html

Fig. 1.3 Simplified view of the hierarchical content flow.

content is delivered to the end-user's device for consumption. Figure 1.3 shows a simplified version of the established content flow [31]. The hierarchical content flow is useful in order to define a benchmark against which we are going to measure the progress toward alternative or complementary paradigms, such as the participatory model edit–share–control (ESC), which we propose in this article.

Based on the hierarchical flow of multimedia content, we can distinguish major research topics: content production and authoring, content delivery, and content consumption.

Regarding content authoring, television content has been traditionally produced in expensive studio settings using digital means. As a matter of fact, previous research has emphasized large video libraries and professional settings with desktop computers [184], instead of living room arrangements. Major research in this area included the provision of efficient video encoding mechanisms for effective video stream rendering and retrieval. The most popular solutions include MPEG-2[12] and MPEG-4[13] video formats. Even though encoded video is an efficient manner for rendering, it provides very limited interactive capabilities. Contemporary technical developments (Figure 1.4) have introduced lightweight content authoring tools for viewers as well [96].

Apart from video encoding, higher level or integration tools allow the composition of multimedia presentations by integrating and synchronizing different media elements. Some examples include Synchronized Multimedia Integration Language (SMIL) [33], Flash,[14] MHEG [57], and MPEG-4 [140]. Integration tools permit to generate multimedia presentations by defining the spatial and temporal relationships of the media elements. In addition, interactivity can be achieved

[12] http://www.chiariglione.org/mpeg/standards/mpeg-2/mpeg-2.htm
[13] http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm
[14] http://www.adobe.com/products/flash/

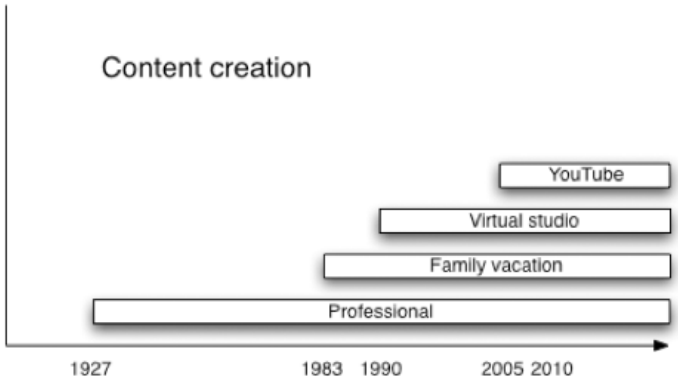


Fig. 1.4 The home-made family video is anything but new, but the popularity of YouTube has been the tipping point for a democratization of the multimedia authoring and editing processes.

by the inclusion of internal and external links. These solutions have been mainly investigated by the research community, but have not been widely deployed by the industry. Instead, in the later 1990s and the beginning of the 2000s the industry concentrated on standardizing an open middleware for iTV set-top boxes, which led to a set of Java-based standards such as Multimedia Home Platform (MHP) [124, 42] in Europe, OpenCable Platform (OCAP) [124] and Advanced Common Application Platform (ACAP) [42] in the USA, Broadcast Markup Language (BML) [42] in Japan, and Ginga [171, 172] in Brazil. Unfortunately, their acceptance and popularity have never met the initial expectations.

Regarding content delivery (Figure 1.5), the first most important challenge for the broadcast community was to actually distribute television content in an efficient manner, so research focused on the transmission mechanisms. This body of research was influenced by the unexpected success of the DVD technology and reused a number of underlying concepts (e.g., using MPEG-2 streams to deliver the content). This wave of research concluded with the deployment of digital television systems [121, 69, 150, 157, 161, 151] and three major regional standards were defined. Advanced Television Systems Committee (ATSC) in North America, Integrated Services Digital Broadcasting (ISDB) in Japan, and Digital Video Broadcasting (DVB)

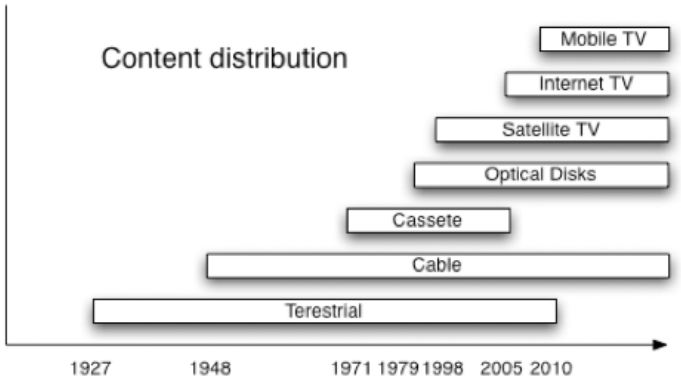


Fig. 1.5 Although the distribution of content has been an hierarchical one-way process, the development of broadband Internet and mobile infrastructures has released content from monolithic distribution mechanisms.

in Europe, [42]. The Japanese solution has been selected in other countries such as Brazil [171]. In addition to broadcast to home, interesting advances have occurred for the delivery of mobile television. Mobile transmission of television content can now be achieved using a number of standards like DVB-H,[15] Digital Multimedia Broadcasting (DMB),[16] and Mobile Broadcast Services Enabler Suite (BCAST).[17]

The previous paragraphs quickly summarize the story behind broadcast television and its content flow. Since this study is not restricted to broadcast transmission, the following paragraphs will discuss about Web-based TV and IPTV solutions.

Web-based TV (or Internet TV) and online video sharing have become a primary activity in the World Wide Web. Some relevant examples include services like YouTube, Yahoo! video, and MySpace. The common characteristics of these systems are that they provide easy-to-use interfaces for uploading, searching, viewing, rating, and most notably for sharing videos. They are intended for personal computer usage and mostly focus on user-generated material. At the same time, a number of Web-based TV solutions are targeted for consuming

[15] http://www.dvb-h.org/
[16] http://eng.t-dmb.org/
[17] http://www.openmobilealliance.org/Technical/release_program/bcast_v1_0.aspx

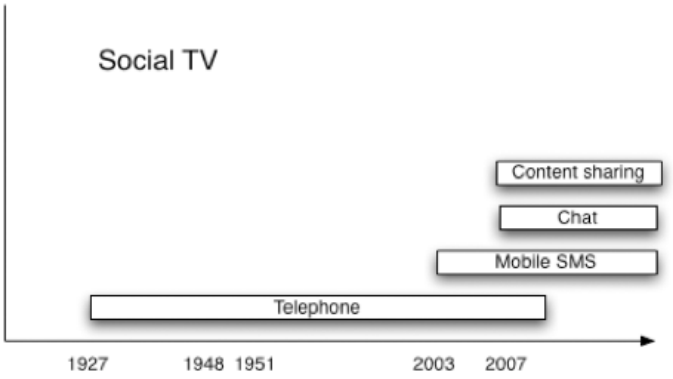


Fig. 1.6 Content sharing and social communication about TV content have taken place over out-of-band channels (e.g., telephone, mobile SMS) but contemporary services have introduced integrated services (e.g., Joost).

professionally produced videos. Some examples include Joost and Lycos Cinema.

Finally, IPTV systems [4, 9, 43] reuse the Internet infrastructure for delivering television content. Over the past years, IPTV systems have been steadily evolving and now they have become a key technology for future television. In many cases, IPTV systems are as well upgrading their infrastructure in order to provide social communications (Figure 1.6). We refer to these solutions as social interactive television. For example, CollaboraTV [127], from AT&T, permits to record the viewer's comments while watching a television program. Then, such comments are replayed when a friend is watching the video using avatars to identify who has said what. At the same time, synchronous communication features have been introduced by Motorola's SocialTV [82, 119] and Alcatel's AmigoTV [50].

The development story of closed captioning might provide further ground for understanding the shortcomings as well as the potential of iTV. In the beginning, closed captioning was conceived as a service for people with hearing disabilities (Figure 1.7). It was implemented by exploiting an invisible part of the television signal, known as the Vertical Blanking Interval (VBI). Closed captioning was initially available to viewers through special caption decoder boxes that were attachable to televisions. Lately, closed caption technology has been used for

Fig. 1.7 TV has been an inclusive technology from very early. Closed captioning was conceived as a way to communicate voice to hard-of-hearing-people.

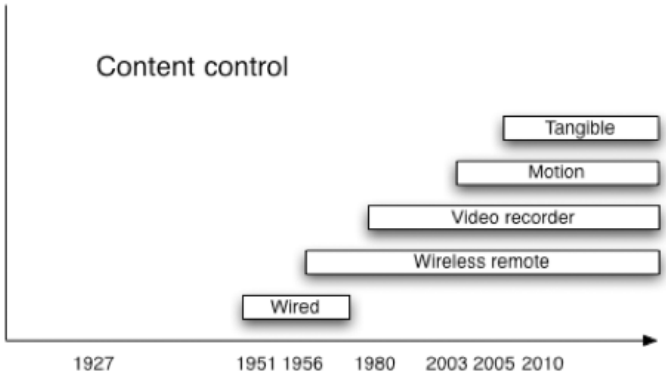


Fig. 1.8 The wireless remote control has been a pervasive input device in user terminal configurations, but novel paradigms have emerged.

a number of different services including BT's talk TV video editing tool [24]. There are a number of lessons to be learned by the story of closed captioning: (1) the VBI technology has been later on exploited to introduce the TeleText service, very popular in Europe, as a first solution toward accessing the Web from the television set (popular services sometimes are not the one the designers had in mind); and (2) the integration of novel technologies into TV sets is necessary for wide adoption by viewers and broadcasters (Figures 1.8 and 1.9).

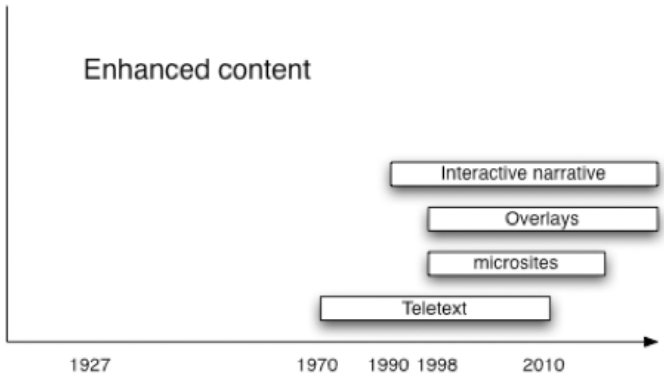


Fig. 1.9 Starting with Teletext systems the audiovisual experience in TV has been extended with additional material, which is rendered at the viewer's terminal.

Although there are many technological, creative, and behavioral changes in the way TV is authored, distributed, and consumed, we do not expect that the established paradigm (author–deliver–watch) will be replaced by the emerging paradigm (edit–share–control). Indeed, television is an information and entertainment medium that has occupied the largest share of domestic leisure time [189] and has become a rather pervasive activity. Therefore, we expect that the emerging paradigm will either build upon or complement existing practices.

Table 1.1 summarizes the traditional view on content flow and compares it with the current view on how television content will be produced, delivered, and consumed. The proposed developments are not meant to replace the traditional practices, rather to complement and enhance them. Our assumption is that traditional television watching will be enhanced with current trends on Web-based television systems, it will incorporate user-generated content and will allow for social communication between viewers. In summary, this work argues that television consumption is composed of three basic components: content control, content sharing, and content editing. Hence, we argue that research topics aimed to improve any of these categories will make a difference in the interactive television landscape.[18,19]

[18] http://www.fcc.gov/
[19] http://www.digitaltelevision.gov.uk/

Table 1.1 Comparison between the traditional view on the content flow and the emerging paradigm.

Topic	Subtopic	Established paradigm	Emerging paradigm
Editing content	Metatada	Professionally produced Professional metadata	User-generated User tags (folksnomy)
	Middleware	Proprietary frameworks	Web-based frameworks IPTV solutions
	Mash-ups	Studio enhancements	User enhancements
Sharing content	Topology	Terrestrial, cable, and satellite transmission Static user terminals Broadcaster or client-server architecture	Mobile distribution of content User Broadcast of content P2P technologies
	Content rights management	Closed system	Content sharing between users
Controlling content	Content selection	Recommender systems	Contextual-based searches
	Content navigation	EPG-like functionality VCR-like functionality	Group-based searchers Semantic navigation of content

1.4 Reaching Its Full Potential

The story of television, as the story of many other technologies, is a constant trial of new ideas and innovations. This section discusses a number of promising technologies and system that did not achieve their full potential. Such exercise will help us to understand how to better provide services and technologies in the future. The first generation of iTV applications has been influenced by the traditional computer paradigms such as the desktop and hypertext. Application developers put most of the efforts on issues that were familiar to them, sometimes forgetting the unique characteristics of the television experience. For example, iTV applications are deployed in a domestic environment and users have entertainment goals when compared to desktop applications, which are deployed in a work environment for productivity goals (Figure 1.10). Other issues that have been sometime forgotten include: (1) television watching is a social and shared experience, (2) contextual information is essential for content rendering and selection, and (3) nowadays in a house there are more rendering components than the television set and more interactive devices than the remote control.

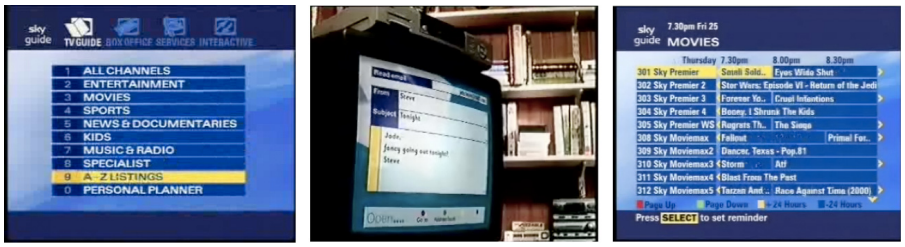


Fig. 1.10 Many interactive TV applications have been designed with the "look and feel" of personal computer applications.

Table 1.2 Topics that have not fulfilled expectations and reason why they did not.

Topic	Major reasons
Video on demand	High demands on the network. After ten years, desktop-based Video on Demand is becoming a reality.
Return channel	There has not been a wide deployment of solutions. Nowadays, the use of SMS messages is the most popular return channel.
Interactive content	Interactivity as video overlays might be disruptive to the entertainment experience due to the intrusiveness of the content
Usability (based on productivity)	Consistency and task efficiency might not be adequate for pleasurable interactive television experiences
Electronic Program guide	Lack of contextual searching (television as a planned activity, different layout, and options depending on the people in the room or the time of the day)
User modeling and personalization	Limited research on group modeling and social communication support (apart from collaborative filters)

Table 1.2 summarizes the topics that we consider did not meet the high expectations generated, when proposed. We must admit, nevertheless, that they provided valuable results. The following sections will inspect each of the topics in detail.

1.4.1 Multimedia Technology and System Architectures

Interactive television in Europe is normally associated with the provision of a return path from the user to the broadcaster. Since terrestrial television is the predominant technology, setting up an efficient interaction channel to the content provider was a research topic of its own at the turning of the century. According to [87], currently discounted solutions such as SMS voting are most widely used and accepted. This was due to the fact that previous implementations of the return channel did not allow for much more than control signals and short messages.

A symmetrical return channel would allow bidirectional distribution of audiovisual content. Today, on the other hand, IPTV standards and cable TV would provide a full working return channel.

Video-on-demand was the central element in the early vision of iTV services [105]. Correspondingly, the academic community put effort into server-side architectures, broadband delivery, and thin network clients [29, 70]. In terms of the commercial success, a retrospective evaluation of the respective research might lead to the conclusion that video-on-demand was not worthwhile pursuing. Nevertheless, a more careful examination may reveal that there were also numerous benefits from that approach, such as the broadband Internet, current IPTV standards, and Web-based TV systems (e.g., Joost,[20] Miro,[21] YouTube,[22] Amazon video on Demand[23]) which are becoming very popular services ten years after.

If we consider graphics capabilities, iTV set-top boxes have only provided the lowest common denominator. The main reason has been that the graphics are controlled at a high level in the middleware, resulting in slow execution and in a complex application composition model. Such inefficiency clearly contrasts with latest game consoles or even with DVDs, where the video-graphics are fundamental to the product architecture. Their architectures are optimized for sequential video presentation with graphics and mainstream DVD titles such as Minority Report include elaborate forms of interactivity linked with good visual effects, which are part of the user interface.

Finally, in terms of content gathering, there were high expectations for the combination of dynamic information coming from the Web with broadcast data. Still, a seamless integration of the different networks bringing video content at home has not been achieved. Basic broadband Internet access and advanced peer-to-peer systems (e.g., BitTorrent) have enabled efficient distribution of content on the PC. While wireless broadcast distribution is becoming suitable for the delivery of high-demand, high-bit rate items, which have a real-time appeal

[20] http://www.joost.com/
[21] http://www.getmiro.com/
[22] http://www.youtube.com/
[23] http://www.amazon.com/Video-On-Demand/b/ref=sv_d_7?ie=UTF8&node=16261631

(e.g., popular sport events, news, and movies). Nevertheless, we can foresee that an EPG could be employed for re-scheduling the favorite show of a family into a more convenient time and day that fits that family's particular schedule, independently of the delivery infrastructure. The fact that some of TV viewing is considered to be 'ritualistic' [160] does not preclude the exploitation of out-of-band techniques for collecting the content at user's premises.

1.4.2 Content Navigation and Personalization

During the 1990s there had been a lot of speculation about the 500 channels[24] to be provided by the future iTV [99]. As a matter of fact, new technologies such as video recorders, cable television and the Web have increased the channel repertoire of TV viewers [63, 90]. This increased availability of TV channels and content has become one of the main drivers for the development of technologies that assist content selection and navigation, such as the EPG and content personalization.

The EPG technology has been mostly associated with the products and services of the Gemstar company. Gemstar began to operate in Europe in 1991, when it launched the patented ShowView VCR recording technology, which simplified the process of taping television programs through the use of unique barcodes associated with each TV show (Figure 1.11). Although there have been some popular consumer products (e.g., TiVo), currently there is no standard navigation method neither for the input, nor for the output human interface [49].

Communication scientists reported that viewers could recall fewer than a dozen of TV channels [64]. Moreover, it has been estimated that one needs at least 15 minutes to browse through 500 channels, assuming a less-than-a-second channel switch delay and assuming an approximately one second glance before pressing the next-channel button. These two issues often have not been adequately addressed by research on EPGs. At the same time, studies have revealed that in some cases TV watching is a planned activity, which is a finding that contrasts with the monolithic focus on the EPG as a method to select a program

[24] 500 Channels and Nothing to Watch, Time, Dec. 14, 1992 http://www.time.com/time/magazine/article/0,9171,977204,00.html

Fig. 1.11 Gemstar has patented several technologies and services related to EPG and print-based input of recording data.

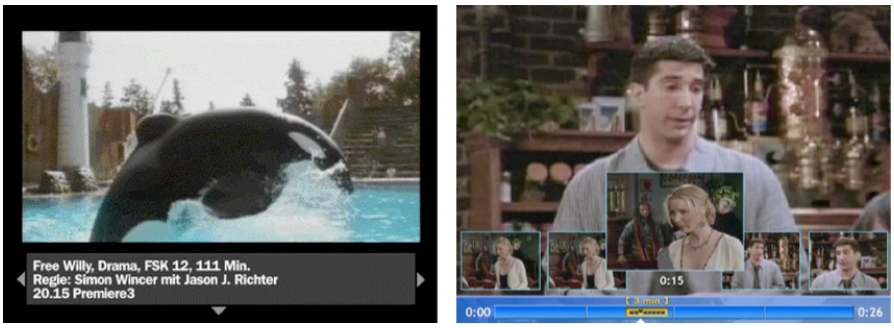


Fig. 1.12 An EPG does not have to take over the whole screen and it could also allow relaxed navigation through the information on available channels without changing the current one [34]. In addition, an EPG could employ additional modalities, such as summarization [58, 94].

to watch each time a user switches-on the TV. On the other hand, there is a fraction of the viewers that impulsively selects a program to watch, especially among the younger demographic [71].

The majority of previous research about iTV applications has addressed the EPG (Figures 1.12 and 1.13) and has proposed a few design guidelines for it [16, 34, 27, 180]. Unfortunately, the EPG as a file explorer-like UI is not appropriate for long TV listings, since it contains less information per screen than a printed TV magazine. Moreover, both methods for navigating TV content are based on a

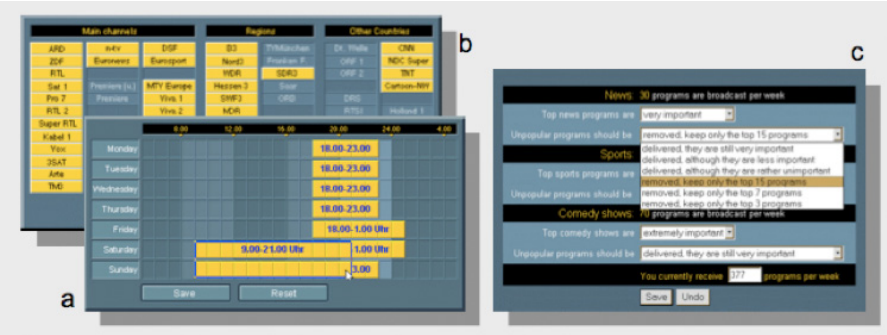


Fig. 1.13 The Electronic Program Guide has been a popular theme in HCI and user modelling research [16].

simple visual mapping of the underlying data structures, without much consideration for the established TV channel selection behavior.

The envisioned 500 channels future was turned upside-down by the user modeling research community [10], as well as from industry, who put forward the vision of a single personalized channel. Nevertheless, it is acknowledged that TV content is a conversation starter [109] and, thus, personalization reduces the chances that any two might have watched the same program.

TV personalization has been one of the most important research directions applying and extending recommendation methods from other interactive media (e.g., Web). Adomavicius and Tuzhilin [6] referred to a number improvement to current recommender systems such as a better understanding of the users and items, inclusion of contextual information, and a provision of less intrusive types of recommendations. He wrote "However, in many situations, the utility of a certain product to a user might depend significantly on time. It may also depend on the person(s) with whom the product will be consumed or shared and under which circumstances." Even though mainstream research focused on imported models from the Internet [170], there are a number of systems that actually followed the main four categories indicated by Adomavicius. Some examples included Masthoff [117], Masthoff et al. [118], and Goren-Bar [76] who considered television watching as a shared experience, and other researchers [11, 17] who considered contextual information for television program recommendations.

Finally, we believe that personalized TV should take into account the social value of the shared TV experience as well. If content recommendation algorithms are indeed tuned and successful to discover new content all the times or content that satisfies the particular tastes of each viewer, then there will be less opportunity to watch and to discuss about the familiar content. Therefore, personalization researchers should also consider the sociability dimension of content recommendation and tune their algorithms accordingly [117]. In summary, EPG research should consider television as a planned activity or television as a shared experience before reaching its full potential.

1.4.3 Designing Interactive Content

The "red-button" of the 1990s[25] from BBC interactive television system, in which the user had to press the red button to launch interactive applications,[26] was an interesting trial about interactive content. According to Baker [13], there are many reasons why the red button has not fulfilled the expectations; the most relevant for our discussion are the following: intrusiveness of the extra content in the main screen, poor resolution of the standards, and slowness of the solutions. In addition, we can argue that such standards did not take into account the social nature of television consumption. Moreover, most of the services provided to users, such as online banking (Figure 1.14), did not fit the television paradigm and were services directly imported services from the Internet.

The introduction and wide adoption of the Web has been promoted by and attributed to the interactive nature of the new medium. It often goes without much thought, that if something is interactive then it is also better and it will be preferable [182]. Interactivity with the user might seem as the major benefit of iTV, but this does not necessarily need to be true and designers should further evaluate it in the context of entertainment applications [92]. Most notably, there is evidence that in some cases interactivity may be disruptive to the entertainment experience. Vorderer et al. [183] found that there are some categories

[25] BBC Red Button, http://en.wikipedia.org/wiki/BBC_Red_Button
[26] Using the red button, http://www.bbc.co.uk/digital/tv/tv_interactive.shtml

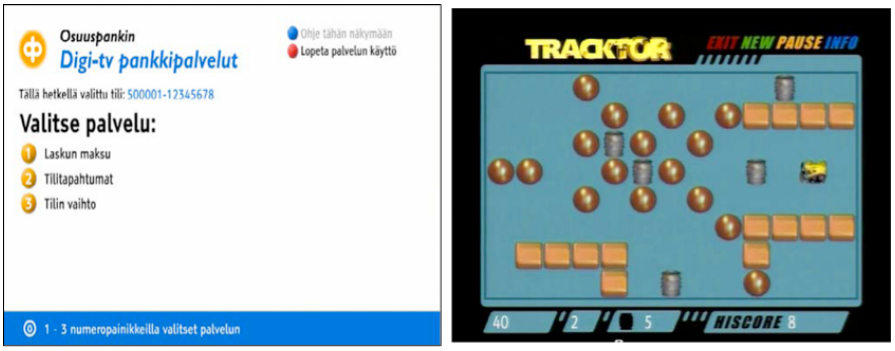


Fig. 1.14 Television banking application and game. These are examples of interactive applications not connected to the television program.

Fig. 1.15 The show 'Pyramid Challenge' by BBCi encourages the viewer to get in the place of the main hero, to make choices and to follow alternative paths along an interactive storyline.

of users who do not like to have the option to change the flow of a TV story (Figure 1.15); they just prefer to watch passively. Nevertheless, there are also situations that users appreciate some extra interactivity, such as sports, where users have enjoyed the control of camera angles.

Indeed, the passive uses and emotional needs gratified by the broadcast media are desirable, [159]. Unfortunately, many iTV applications support the presentation of generic information on the screen, instead of considering the augmentation of the entertainment experience. Although TV offers a wide variety of content that spans from pure entertainment to pure information, the content is usually presented in a captivating way, regardless of the type (e.g., documentary, news).

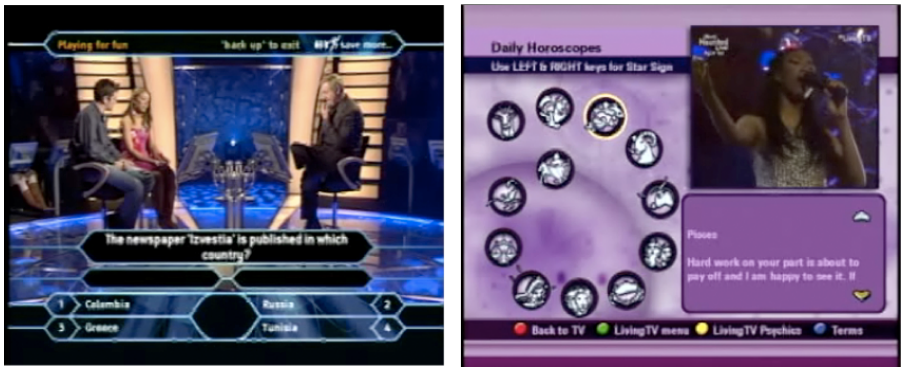


Fig. 1.16 Quiz games such as the "Who wants to be a millionnaire" have been a straightforward domain for adding interactivity with the audience. The Living TV channel on Sky offers on-demand horoscopes, which matches the gossip, celebrity, and paranormal programming of the channel.

Therefore, it is suggested to employ informational elements, in order to augment the entertainment content [106]. For example, a music video channel could insert interactive information related to the video clips, such as trivia, discography, or motivate direct sales and downloads of music (Figure 1.16). Furthermore, a quiz game might introduce an iTV application that allows viewers to play-along the contestants in the studio, to compete in the home or over distance. As a principle, designers should provide interactive entertainment elements or on-demand information elements that match the main TV content.

Another popular research stream has considered iTV for educational programs. Aarreniemi-Jokipelto [1] provides a historical description of educational content offered through TV in Finland. The background information about Finnish educational TV is complementary and runs almost in parallel with that of USA, as reported by Revelle [153]. Both of the above efforts started with the motivation to use traditional TV to educate children in the home and in the classroom. The main rationale cited for the adoption of TV as a learning medium is its pervasiveness. Television watching is a familiar and reliable consumer device with more than 90% penetration in developed countries. Although computers and the Web have been very popular in some developed countries too, they have not reached the pervasiveness of TV [15]. Nevertheless,

iTV systems hold many opportunities for enhancing distant education, such as messaging between the students and the eventual formation of online learning communities, which are interlocked with TV content.

A common belief is that TV viewers are always concentrated on the TV content, but there is ample evidence that TV usage takes many forms, as far as the levels of attention of the viewer are concerned. Jenkins [86] opposes to the popular view that iTV will support only the needs of the channel surfers by making an analogy: "With the rise of printing, intensive reading was theoretically displaced by extensive reading: readers read more books and spent less time on each. But intensive reading never totally vanished." Indeed, an iTV study has empirically confirmed the existence of readers and skimmers, as two distinct groups of TV viewers [35]. Therefore, the creators of content should consider the full continuum of viewer roles between skimmer and reader.

1.5 Lessons Learned and Open Research Issues

Notably, the most successful use of interactivity in TV has been achieved by external means, such as the VCR, the DVD, and game consoles. In fact, interactivity on the DVD players was in creative terms much better than any concept devised by the broadcast industry. It was so good that at one point it seemed as if the DVD middleware would become the default standard for all TV platforms. Another successful story has been the one of TiVo (Figure 1.17). It offers a UI for stored programs and has been popular in the USA for sometime already.

Due to the diversity of scientific subfields, 20 years of research on interactive TV has not produced a unified set of results. Interactive TV research as a whole is a loosely interwoven body of findings, broadly divided into a collection of separate research fields (e.g., content distribution system, graphics architectures, user interface development, user modeling, etc.) and commercial products. Each scientific field brought its expertise to bear on a separate facet of interactive TV, generating important results but not assembling them into common threads that could define how the main issues relate to one another or ideally how each finding builds upon each other. Moreover, most of the innovations

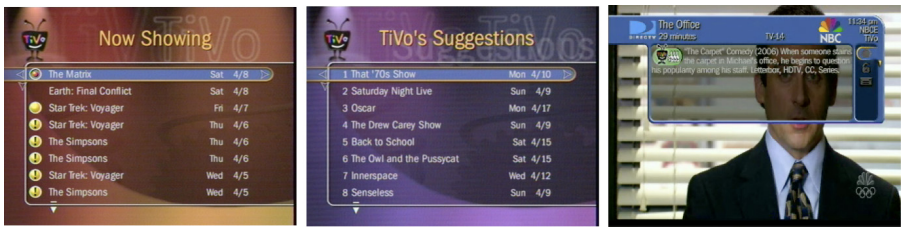


Fig. 1.17 The popular TiVo system (a combination of set-top box and service) always provides a choice of pre-recorded content and suggestions about what to watch, based on collaborative filtering. Information related to the running program is placed in a semitransparent box that does not distract viewing.

have been introduced by the industry in commercial products. As a matter of fact, the design process for those developments has remained very much undocumented. In contrast to the broadcast TV area, the networked TV one has been initiated on pragmatic expectations, feasible infrastructures, and most importantly lower barriers of entry for researchers and users.

In the following sections, we organize iTV research into three concepts, which stand as an evolutionary step over the traditional model of production–distribution–consumption: (1) editing, (2) sharing, and (3) controlling content. In particular we consider the end-user having an active role in each one of these activities, instead of being just a "viewer." In the rest of this article, we organize existing literature into these three distinct concepts.

2

Editing Content

This section focuses on the first proposed paradigm: interactive television as a content creation activity. Research on this topic concentrates on: (1) the provision of authoring tools that are suitable and build upon the established television visual language, (2) the development of adequate content and metadata modeling formats for enriched media experiences, (3) the employment of 3D graphics in television content, and (4) the empowerment of viewers as authors of media content and services.

This section is structured around four main topics: authoring tools, content and metadata modeling, virtual and augmented reality systems, and user-generated content. First, we discuss the state-of-the-art of authoring tools and indicate future directions that better fit the interactive digital television paradigm. Then, we introduce different content and metadata modeling technologies used for interactive digital television services. Next, we introduce virtual and augmented reality systems and their potential impact on the future of television. After that, we consider how user-generated content will reshape the media landscape. Finally, this section concludes with a summary of the different topics studied.

2.1 Authoring Tools Principles

Similarly to any new medium, researchers tried to shape the future of interactive digital television by transferring and using traditional development techniques and tools from the PC and the Web. As a matter of fact, the respective authoring tools provided limited support for television-specific issues that could facilitate the production of innovative television content and services. Next, we describe the above two issues in more detail.

First, TV audiences have become familiar with a visual grammar that is common to many television programs. As well, the presentation style needs to be dynamic and surprising [120], which is in sharp contrast with the traditional usability principle of consistency [129]. As a result, designers should enhance the core and familiar TV notions (e.g., characters and stories) with programmable behaviors (e.g., objects, actions). Regarding the iTV UI, it should not look like a button or a dialog box (Figure 2.1). Instead, it could be an animated character, which features multimodal behaviors (e.g., text, motion, and speech). Furthermore, user selections that activate scene changes should be performed in accordance with the established TV visual grammar (e.g., dissolves, transitions, fade-outs).

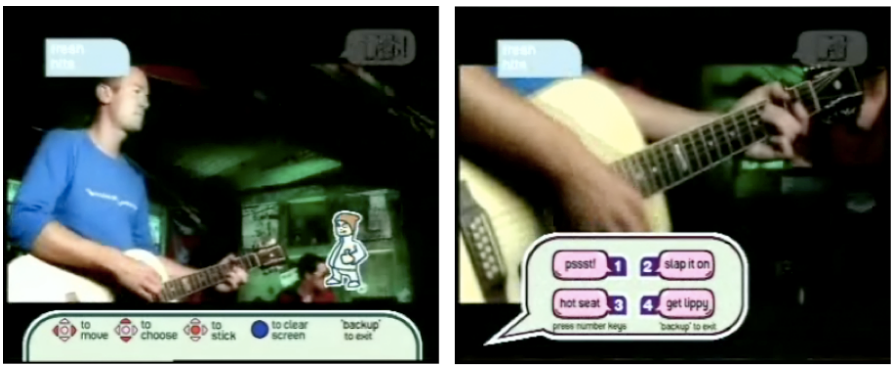


Fig. 2.1 Electronic stickers that can be selected and placed over music video clips are hardly any useful, but they are very suitable for the audience of MTV. The main menu of that application does not follow any particular dialog design guidelines, but it might follow the aesthetics of the MTV audience.

A common problem in UI design for iTV applications is the employment of UI widgets that have been derived from the PC and the Web programming toolkits, such as form buttons, icons, and links [45]. In contrast, it has been established that iTV producers prefer a TV-based design and story-driven content [85].

Any authoring tool can be categorized based on: (1) its target audience, (2) its functionality, and (3) its underlying paradigm. For example, the target audience of Adobe Premiere Pro CS3[1] is professionals and advanced end-users, the intended functionality is to integrate media presentations, and it is based on the timeline paradigm for video creation. On the other hand, Eclipse[2] is intended for professional developers, the functionality is to develop complex software programs, and it is based on the object-oriented paradigm.

In the past, the target audience of interactive digital television authoring tools has been professionals working in broadcast companies. The functionality of these tools includes the aggregation of different media objects, such as subtitles overlaying the video content, and some event handling mechanism for user interaction. The two most prominent authoring paradigms include scene-based [42] and timeline paradigms [32]. The scene-based paradigm deploys the application as a set of scenes and it is normally used for interactive productions, the timeline paradigm on the other hand is intended for more linear productions with limited interactivity.

Typical video authoring tools such as Movie Maker,[3] iMovie,[4] Adobe Premiere Pro CS3, and Adobe Flash Pro CS3,[5] use the timeline paradigm. This paradigm is intended for linear productions offering script-like interactivity. The major drawback of such tools is that they assume a passive viewer, thus limiting the potential interactivity of video content.

[1] http://www.adobe.com/products/premiere/
[2] http://www.eclipse.org/
[3] http://www.microsoft.com/windowsxp/downloads/updates/moviemaker2.mspx
[4] http://www.apple.com/ilife/imovie/
[5] http://www.adobe.com/products/flash/

On the other hand, scene-based authoring tools such as Osmosys,[6] Sofia Digital,[7], Cardinal Systems,[8] and Alticast[9] allow the author to construct interactive productions. The author describes a set of scenes, linked to the video content, that are shown depending on the user actions. The major drawback of such tools is that they are intended for professional authors, it is expected that the authors know the underlying programming environment, and that the user interaction is restricted to one remote control.

Based on the previous discussion, we can conclude that the foreseen research topics on authoring tools for iTV are:

- To take into account the television grammar and aesthetics rules.
- To follow a television authoring paradigm where the television is an invisible appliance, the end-user might use a number of special input devices for interaction including gesture and pen-based technologies, and where the services follow concepts adapted from hyperlinked multimedia. Thus, time management becomes a primary concern, instead of the predominant spatial organization of the WIMP style.

2.2 Content and Metadata Modeling

Content and services authoring involves three major stages: semantic modeling, presentation integration, and inclusion of interactive capabilities. Figure 2.2 shows these stages. Semantic modeling refers to the description of the content and media elements using existing metadata standards, presentation integration consists in the description of the layout and synchronization of the media elements composing the presentation. Finally, authors might provide interactive capabilities related to the media elements.

Semantic modeling is needed for content description, thus it is essential for content selectivity and searching. The result of the semantic

[6] http://www.osmosys.tv/
[7] http://www.sofiadigital.com/
[8] http://www.cardinal.fi/
[9] http://alticast.com/

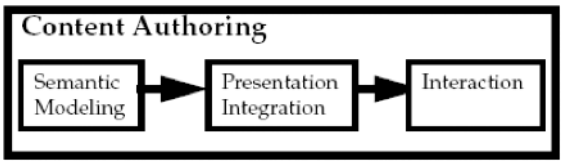


Fig. 2.2 Content authoring flow diagram.

Fig. 2.3 Joost user interface includes semi-transparent interactive widgets.

modeling stage is the association of metadata to the content. Presentation integration, on the other hand, corresponds to the composition of enriched content from different media assets. For example, presentation integration is the inclusion of subtitles or of an audio track accompanying the video. Finally, interactive capabilities provide support for user interaction. As a result of these stages, the author might provide a rating system to be used while watching videos as illustrated in Figure 2.3.

The technology employed for creating multimedia presentations has major consequences on the way end-users consume the media. So far, television viewers have been considered as media consumers with little impact on the content. Thus, research has focused on efficient rendering mechanism, user selection of content, and adaptation of that content to various devices. But, the research community has not dedicated enough

resources to study what the end-user can do with the content, after it has been rendered.

In relation to standards for semantic modeling, significant work has been performed by the TV-Anytime Forum (based on MPEG-7) [125, 126]. Interesting research in this area includes the UP-TV project [93, 91]. This project presented a program guide that can be controlled and managed using personal devices (e.g., handheld devices). Unfortunately, apart from some exceptions, current digital television broadcast only uses low-level descriptions such as the ones included in the MPEG-2 tables [108, 124].

In terms of specific-purpose semantic modeling, learning has been a popular topic for iTV research. Rey-Lopez et al. [154] proposed that iTV content can be classified in an ontology that consists of learning objects. A learning object is defined as "any digital resource that can be reused to support learning." The main idea of learning objects is to break educational content into self-contained items that can be reused in various learning environments. Learning objects are tagged with descriptive information, known as metadata that allows them to be searched for easily. The description can be searched and provides the means for finding learning objects of interest, including those that may be non-textual such as a video clip.

While semantic modeling is used for describing the content and thus for easy selectivity and recommendations, there are tools and standards utilized for presentation integration. One solution for modeling television content is by using a declarative television standard such as Digital Video Broadcasting Hypertext Markup Language (DVB-HTML) or Broadcast Markup Language (BML). Even though they permit Document Object Model (DOM) modifications, they lack the temporal relationship inherent to television content. Moreover, these solutions are text-centric solutions to a media-centric problem [42]. Other solution is to use Moving Pictures Expert Group (MPEG)-4 for modeling the content and MPEG-7 for describing the content [51, 52, 78, 14]. But it does not seem that the industry is interested in its adoption.

At the same time, the W3C SYMM working group has been working on a television set-top box profile [30, 39] based on the Synchronized Multimedia Integration Language (SMIL). SMIL provides the means

for integrating different media objects into one presentation. While there are a number of systems using SMIL as an intermediary format for mixing and adapting media content prior to being encoded for broadcast, there are not so many end-to-end solutions available in the market. Nevertheless, there are successful stories such as the European Project NM2.[10] This project has developed an authoring tool for content producers intended for the creation of interactive narratives. The authoring tool is based on an internal semantic metadata language, called NSL, which produces an SMIL file. Using the SMIL language, dynamic modifications of the broadcast content are possible. The resulting television program has been termed as "ShapeShifting" productions, since they are interactive programs in which the narrative can be shifted in real-time based on audience feedback [179, 178].

After a number of European, American, and Japanese proposals intended for the set-top box middleware [42], in the past years Brazil has taken a decision about their own iTV standard. The solution is called Ginga-NCL [171, 172] and it is a multimedia-based declarative language — closely related to SMIL. Thus, it provides a media-centric solution for modeling television content and applications.

As a summary, by content description we refer to data that defines content, and thus it is used for selecting content. On the other hand, presentation description refers to the actual composition of interactive applications including its synchronization, its layout, and interaction. There are a number of content description standards including TV-Anytime, MPEG-2, and MPEG-7. While MPEG-7 is mostly a researchers-oriented solution, TV-Anytime is widely accepted and promoted by the industry. Finally, MPEG-2 content stream description is already used in digital television deployment to, for example, populate the electronic program guide. In terms of presentation description, there are a number of standards such as MHP/OCAP for Europe and North America, BML for Japan, and Ginga-NCL for Brazil. Nevertheless, it seems that the deployment of any of these solutions has not met the initial expectations. Other academia-oriented solutions such as MPEG-4 and SMIL provide enhanced functionalities.

[10] http://www.ist-nm2.org/

2.3 3D Interfaces

Traditional Virtual Reality (VR) tools have been applied to construct a complete new world that is shown to the user or in which the viewers can participate [145]. This section summarizes the state-of-the-art research in this topic.

One popular example is the service called *News at Seven*[11] (Figure 2.4), developed at Northwestern University's Intelligent Information Laboratory (US). This service provides automatically generated news, gathered from a variety of sources, which are presented by avatars. In addition, they use Text to Speech technologies for presenting the news.

In addition, Sony has developed Blendo,[12] a declarative markup language derived from Virtual Reality Markup Language (VRML) and intended to move the control of the broadcast enhancements from the studio to the living room [116]. Figure 2.5 shows some examples of using Blendo. One implemented prototype is an interactive sports application

Fig. 2.4 In News at Seven, an avatar provides personalized news.

[11] http://newsatseven.com
[12] http://www.plasm.com/rob/portfolio/Blendo/

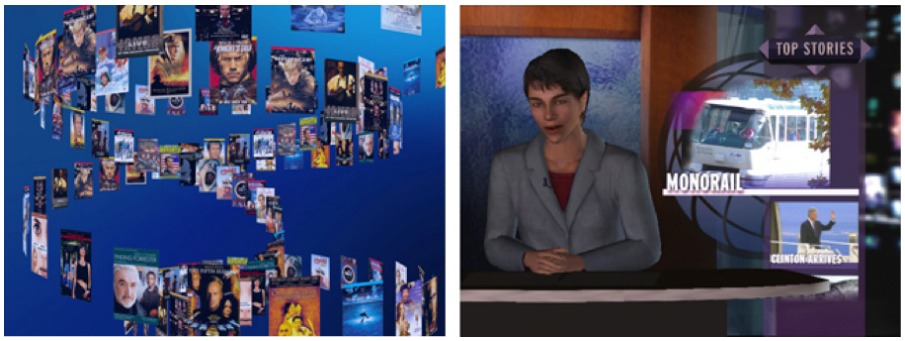


Fig. 2.5 Examples of using Blendo language.

Fig. 2.6 3D and time-based navigation [144].

for car racing. It includes polling questions as well as overlay information (e.g., real-time telemetry data) [147]. Other approaches for incorporating 3D graphics in the television environment include an innovative solution for the EPG [144] (Figure 2.6) and extensions to the current middleware standard MHP [42, 177] (Figure 2.7).

3D-based television has been studied by the Personalized, Immersive Sports TV Experience (PISTE) project [114] and the 3DTV Network of Excellence [137]. PISTE studied the possibilities of MPEG-4 for integrating animated 3D content and television sports broadcasts.

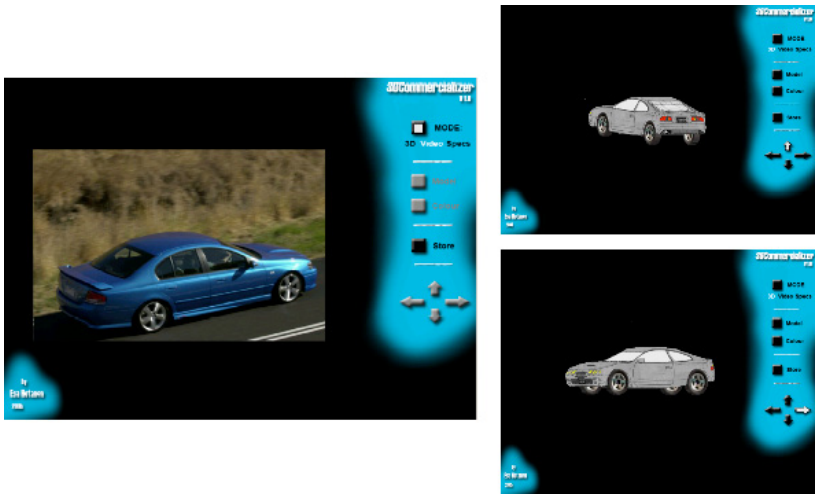


Fig. 2.7 3D graphics as an extension to current middleware standards. These pictures show an enhanced commercial for television.

On the other hand, the 3DTV project focuses on the whole content chain including capture, scene presentation, coding by specifying the exchange format of the data, transmission, and display of the 3DTV signal.

Finally, the popularity and potential of online 3D worlds such as Second Life[13] has motivated many corporations for actual integration of 3D technologies in the television domain.

2.4 Cooperative Editing

Content editing is usually considered and modeled after a single person activity, but the process of creating and enriching media content is deeply collaborative. In the past, television fans have setup forums to discuss about popular programs and exchange alternative narratives and point of views. More recently, the act of uploading a video on the Web might produce a stream of video responses that build up on the original narrative. In addition, the availability of online video editing tools has enabled viewers to easily edit and create mash-ups of readily

[13] http://secondlife.com/

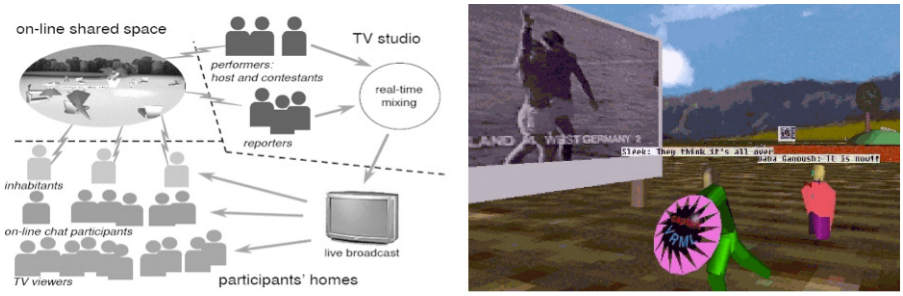


Fig. 2.8 Examples of inhabited Television [18].

available content uploaded by others. In this section, we investigate cooperative spaces that enable viewers to collaboratively edit content. In particular, we have identified two directions of research and practice, one that considers online virtual worlds and one that considers online cooperative tools for video editing.

The concept of a more immersive television experience has been a theme of previous research. The project YORB[14] introduced a virtual environment (3D world), where viewers could contribute in the production process. In particular, viewers within the world encountered pictures, sounds and video that had been sent in by other viewers. Inhabited TV [18] involves the development of collaborative virtual environments within which viewers collaboratively produce and become part of the final content that could be broadcasted. This extends the TV experience by enabling social interaction among participants and by offering them new forms of control over narrative structure and greater interaction with content. In an inhabited TV application (Figure 2.8), the television becomes part of a group interaction within the virtual online world as well as in the living room. In this situation, "the television becomes not only a social actor, but also a place to be" [5].

Besides the virtual 3D worlds, others have investigated the collaborative production of content within real or online spaces. The notion of "TV as space" has been transferred from the metaphorical to

[14] Project YORB, Dan O'Sullivan, NYU, http://itp.nyu.edu/~dbo3/proj/yorb.htm, Visited March 2009

the real domain in the artistic project "Park Bench TV."[15] In "Park Bench TV," commuters are invited to use their wireless computers and to log-in a local community TV station that broadcasts content uploaded and edited by citizens themselves. Community TV stations have been also created in Italy in response to increased media control consolidation.[16] More recently, the European Project Citizen Media has actively researched the use of IPTV and mobile phones production and sharing of media in local communities [131]. At the same time, Hamasaki et al. [81] have described the network effects of collaborative video editing on online Web site that supports song-writing. Moreover, other researchers have described positive effects [62], when local communities produce TV content by themselves.

In summary, researchers should consider extending the functionality for cooperative user generated content. For example, the ability to upload personal music, photos, and videos might be used to achieve communication through content. In particular, the automated production of personal TV channels that keep track of individual life streams (e.g., music, photos, and personal videos) could be multiplexed with broadcast TV watching behavior. Indeed, Kubey and Csikszentmihalyi [100] have found that everyday life experience is correlated with TV watching behavior.

2.5 Summary: Viewer As a Content Editor

The establishment of Web technologies that support social networking had major implications on content creation paradigms and methods: from the traditional professional creation of content to a more fresh and immediate user-generated content. User-generated content might regard the content itself, as well as content enhancements, or the metadata about the content.

Traditionally, the content follows a hierarchical flow from the professional studio to the broadcaster to the user. In that case, the end-user had limited interactivity. In the 1990s, the red button era came along

[15] Park Bench TV, Pete Gomes, Architectureal Association, http://www.parkbenchtv.org/project.html

[16] Telestreet, Italian pirate short-range TV stations, http://en.wikipedia.org/wiki/Telestreet

which offered browsing of information pages. Contemporary research regards that the viewer is not the end of the content value chain. Instead, the viewer is considered to be an active node in the production–distribution–consumption chain. In this way, the user can play additional roles, such as distributor or even producer of content.

The success of Web sites that distribute users' content (e.g., YouTube) has demonstrated that users are ready to overcome traditional conceptions on how they consume iTV. Researchers have been developing novel iTV systems that allow users to create their own media and share it (Figure 2.9). For example, Cattelan et al. [36] present an MPEG-4 system, which allows the end-user to enrich television content. Cesar et al. [38, 37] have developed a similar system by employing SMIL as the content modeling format. Therefore, designers should involve the user in lightweight content editing, such as annotations and virtual edits.

While the previous examples provide a solution for incidental authoring of television content, there are a number of systems in the Web that provide an interface to remix [166] and repurpose [138] multimedia content. The main idea behind these systems is to reuse existing material on the Web, or on the television channels repositories, in order to create new media content ready for consumption. According to Shamma et al. [165] there is a need to shift from semantics to pragmatics in multimedia content authoring systems.

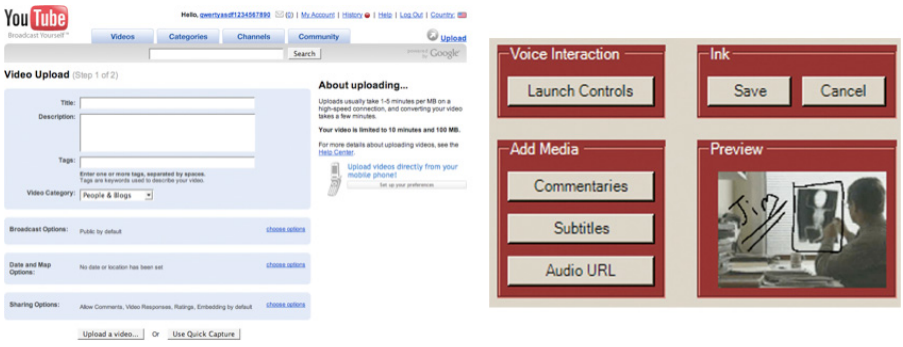


Fig. 2.9 Hierarchical top-down content distribution has been disrupted with end-user content uploading (YouTube) and TV content enrichment [38].

Table 2.1 State-of-the-art regarding content creation.

Topic	State-of-the-art	Research agenda
Authoring tools	Authoring tools based on the scene paradigm and on the timeline paradigm	Development of authoring tools based on the specific television paradigms Authoring tools targeted to end-users
Content and metadata modeling	A number of standards for describing media and to model television content	Adoption of standardized solutions for describing the media (e.g., TV-Anytime) Adoption of time-based standardized solutions to model television content that supports user interactivity
User-generated content	Web-based systems; not enough research in the iTV domain	Development of television-oriented tools for user-generated and user-enriched content

The main conclusion here is that we are moving from a monolithic conception of television content to a more open alternative, in which segments of the content can be remixed and shared with other people. Authoring tools need to provide capabilities for supporting such shift.

This section has focused on one aspect of interactive television: content creation. The support for user-contributed content raises the potential to revolutionize the hierarchical TV production process, by introducing the viewer as part of content delivery chain. Table 2.1 summarizes the state-of-the-art regarding content creation.

3

Sharing Content

In this section we describe content sharing as a technical and social system. Besides hierarchical networks, TV content can be efficiently distributed over peer-to-peer (P2P) networks. In this way, the variety of video delivery paths has been increased with the support of new Internet technologies, which allow new ways of distributing video (e.g., broadband connected TV boxes). Thus, iTV applications are neither limited to the traditional TV device and broadcast delivery, nor to the typical channels of satellite, cable, and digital terrestrial networks. Alternative and complementary devices and distribution methods have been considered, such as mobile phones (mobile DTV).

In addition to distribution technology, content sharing also refers to the social practice of enjoying together TV. There has been a significant body of computer supported co-operative work (CSCW) research on supporting interaction among geographically distributed co-workers, but there is limited investigation in the context of leisure activities, and in particular distributed use of audiovisual content, such as TV [80]. As a matter of fact, there is not much knowledge on designing applications for leisure or informal TV sociability. Social TV applications have a wide appeal as audiovisual content becomes more closely integrated with the social structure of Web video services, such as YouTube.

Even though television, since its inception, has been considered a social link between people, actual social television systems have been scarce in the literature. This section is structured around the following topics: breaking the mass communication hierarchy, sharing content between users and devices, talking about content, and sharing experiences. Firstly, we examine how the traditional content flow from the broadcast station to the end-user is breaking down, and new paradigms for interactive television consumption should be proposed. Then, we review previous studies on media consumption and how the habits of the people are changing toward digital media. Next, we explore social communications and presence awareness when interacting with television content. Finally, this section concludes with a summary, in which we argue that television viewers are becoming an active node that adds value in the distribution of content and not only mere spectators.

3.1 Beyond Hierarchical Content Distribution

In the past, TV content in the living room has been provided either by broadcast, or optical discs. A basic iTV system includes a set-top box that decodes the signal and provides processing and storage capabilities that enable interactive applications. Nevertheless, the disagreement on a common open middleware platform has been an obstacle for the development of sophisticated interactive applications that are independent from the set-top box hardware. On the other hand, there is agreement over the specifications for the digital video broadcasting (DVB-S/C/T/H specifications satellite, cable, terrestrial, and mobile). Furthermore, TV content can be efficiently distributed P2P networks.

We can differentiate a number of new challenges:

- Mobile television: where the content is transmitted using the mobile network.
- P2P television: where content can be downloaded from other peers.
- Television sharing: where users are becoming broadcasters of content.

Digital mobile TV systems have been designed to complement mobile networks with broadcast and multicast capabilities for spectrum-efficient delivery of multimedia services on mobile devices in both outdoor and indoor environments, without introducing constraints on the user terminal or the consumer itself. The available technologies provide broadcasts that send content to all the mobile terminals within the footprint of a base-station. The presence of multiple situated base-stations is one of the competitive advantages of digital broadcasting over traditional analog video broadcasting, because the content could be personalized to fit the terminal and physical location preferences.

On the technological side, digital broadcasting technologies and standards allow cell phone and personal digital assistant (PDA) users to watch terrestrial digital television on their portable communications devices. In particular, the DVB-H standard is based on the widely deployed series of DVB standards (DVB-S/C/T) and includes enhancements for mobile terminals, such as reduced power consumption and reception while on the move. Although the technical standards are suitable for mobile TV reception, it is clear that mobile TV prospects should be examined as a complement to traditional living room TV. This is because the perceived quality of TV on a mobile phone and the solitary experience are not favored mode of watching TV. Interesting research in this topic is provided by Knoche [98, 97] and by O'Hara [133].

Moreover, relevant research is occurring in P2P-TV coming both from the academia [185, 68, 84] and the industry (Joost, BBC). For example, the European Union project P2P-Next[1] intends to build the next generation P2P content delivery platform, while the European Broadcast Union (EBU) has started a group on the topic.[2] The major benefits of P2P delivery are the efficiency of transmission, the diminution of traffic for a specific node (the server), and the social nature of P2P that can be utilized for enhancing the process of watching television.

[1] http://www.p2p-next.org/
[2] http://tech.ebu.ch/groups/dp2p

In summary, iTV builds upon the convergence between different technological infrastructures, such as broadcasting, telecommunication, and the Internet. The convergence has been realized in different forms. On the one hand, Internet content may be accessed through television Web browsers, or linked to iTV programs (e.g., interactive advertisements). On the other hand, TV content has become available in other platforms, such as online video sites and mobile TV.

3.2 Sharing Content Between Users and Devices

Apart from mobile television and P2P TV, significant research has been performed on content sharing between users and devices. This strand of research does not consider viewers as passive agents, but as another active node in the distribution of media content. In the context of iTV, networked television systems (e.g., mobile TV, P2P TV, etc.) have been the necessary technological infrastructure for advancing the state-of-the-art in human-centered iTV research. In addition to the flow within a dedicated distribution network, the flow of content is also realized between multiple devices owned by viewers. For example, viewers might record broadcast TV content, transfer to the Web, and finally synchronize to a mobile device. In this way, the traditional hierarchical distribution of content has become just a sub-case of content sharing between users and devices.

On the technical side of content sharing, we have distinguished two types of systems: (1) distribution of content between network nodes, and (2) delegation of content rendering between user terminal devices. On the user side of content sharing, we have identified two activities: (1) content sharing between users, and (2) content sharing between the terminals of a user. In both cases it is essential to take into consideration the digital rights management (DRM) for all the content owners (including the end-users as content creators). Nevertheless, many approaches to DRM have been considered harmful to the usability of TV-related activities.[3]

[3] Simson Garfinkel, Losing Control of Your TV, MIT Technology Review, March 3, 2004, http://www.technologyreview.com/computing/13512/

In previous research, Bentley et al. [19] provided interesting results on music and photo sharing and Taylor and Harper [175] reported on a study on gift-giving and reciprocity. Complementary results for television content have been provided by Cesar et al. [37], while Schatz [162] extended mobile television with P2P interaction, and Miller [122] has argued for MMS television. Other research focused on specific television services such as games or learning. For example, Ressin and Haffner [152] presented how television games can be combined with SMS. Moreover, Fallahkhair et al. [61], Pemberton et al. [139], and Jokipelto [1] combined television and mobile phones to enhance the learning experience of viewers.

Sharing content does not only regard how a particular content item moves from one device or user to another one. In this context, we need to study sharing of fragments of television programs. A potential direction for further research might consider how content should be rendered when multiple users and multiple devices are present in the living room. The availability of small broadband multimedia devices has facilitated the development of multimodal systems that split the user interface over multiple screens [156]. Additional results have been provided by Cesar et al. [40] on sharing fragments of television content by employing secondary screens. The most challenging part in those works is how the user interface and the content is delegated, instead of mirrored between the complementary devices.

In summary, gift-giving and sharing of fragments of television content are potentially strong business models, although topics such as copyright control, versioning control might be obstacles in the way. Although content sharing usually makes most copyright owners very uncomfortable, it also makes up for an efficient and intelligent adaptation to user needs, which is the ultimate goal of most decent media business.

3.3 Interfaces Between People

During the last decade, there have been many iTV systems that facilitate social communication between viewers. Indeed, communication applications such as messaging, chatting, or voting during certain

programs (quizzes, contests, etc.) strengthen viewer's loyalty to the specific program. Social TV systems consist of technological solutions for integrated interpersonal communication and content distribution. Those systems provide support for buddy lists, talking about content, as well as sharing personal photos and home videos. Interpersonal communication is based on voice, text, and video formats, as well as animated avatars.

The study of social interactive television is not new. Back in 1979 Wellens [188] wrote: "interactive television represents means of linking individuals together by providing each with an electronically mediated representation of the other's voice and visual presence." In contrast to video mediated communication, Social TV does not use video for social communication, but as point of reference. We define a "Social TV" application to be part of a content distribution system and to allow distant viewers to communicate with each other using several interpersonal communication modalities, such as open audio channel, instant messaging, emoticons, etc. Similarly, research on interpersonal communication in the human–computer interaction (HCI) field has regarded video-mediated communication at work [181].

One of the first approaches to Social TV was the "Inhabited TV" research effort [54], which developed a collaborative virtual environment, where viewers could interact with other viewers or virtual objects. In this case, viewers were watching TV within the virtual environment and not within physical space. Thus, the TV experience was extended by enabling social interaction among participants and increased interaction with content. In an Inhabited TV application, the television becomes an actor and a part of a group interaction within a virtual online world.

There is a wealth of research on Instant Messaging (IM), but limited application to iTV. Moreover, there are many tools for online discussion, but computer-based IM and chat about a TV program do not provide a bridge to the program itself. For example, a drama series could provide facilities for online community building along the storyline of the broadcast [3]. The Media Centre Buddies system integrated TV technology into an instant messaging application [149]. The main aim was to allow multiple users to log into an instant messaging client that

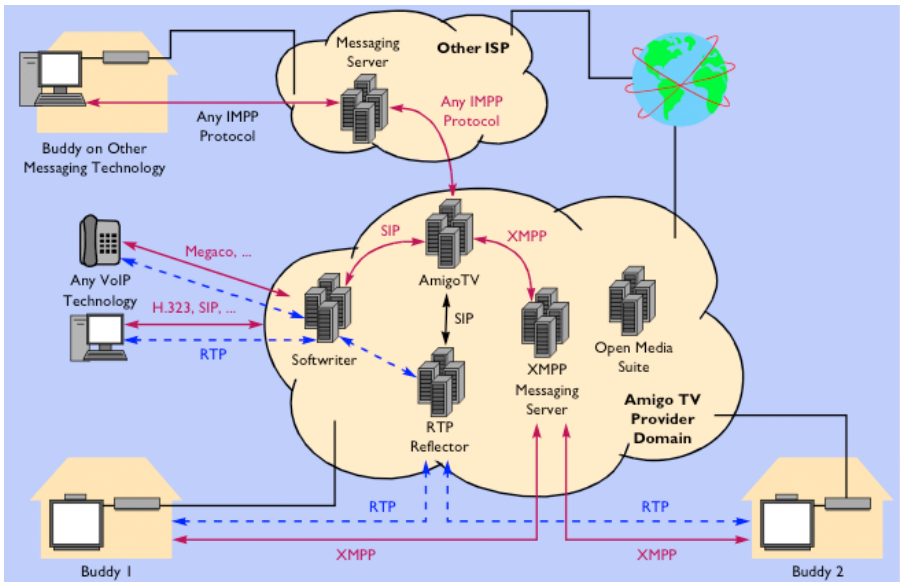


Fig. 3.1 Overview of a Social TV architecture [50].

was running next to a TV channel. The 'Amigo TV' system provides a technological platform (Figure 3.1) for integrating content delivery, communities, and interpersonal communication [50]. In addition, the broadcasted content can be personalized by sharing personal photos and home videos. Amigo TV supports online user meetings and buddy lists. Interpersonal communication is based on voice, text, and video formats, as well as animated avatars.

More recently, several researchers have revisited the topic and they are reshaping the field [7, 28, 59, 44, 74]. Agamanolis [7] presented a number of projects that enhanced the human connectedness using the television setting. Brown and Barkhuus [28] formulated essential basic questions such as "how new media technologies are affecting family structures?" in relation to the new media landscape. Ducheneaut et al. [59] provided a number of design guidelines and innovative ideas for implementing social television systems including presence awareness. Chorianopoulos and Lekakos [44] defined a taxonomy based on two dimensions of the social aspects of television. The first dimension

Fig. 3.2 Joost provides instant messages, chat over distance and user groups organized around TV content. AmigoTV by Alcatel employs avatars or photos to represent distant viewers.

concerns the presence (co-located viewers or non-co-located viewing) and the second dimension identifies the type of communication (synchronous or asynchronous communication). Finally, Geerts and De Groff [74] provided a set of heuristics for the evaluation of Social TV systems.

Social interactive television systems (Figure 3.2) have been popular in the industry, with systems such as Telebuddies [111], ConnectTV [26], Amigo TV [50], CollaboraTV [127], and SocialTV [82, 119]. All these systems (with the exception of CollaboraTV) focus on synchronous communication mechanism between television viewers. Fortunately, not only systems have been developed, but user studies also have been published. Apart from the studies mentioned above, ConnectTV, CollaboraTV and Social TV, Geerts [72] compared voice and text chat and highlighted the difficulty of text entry for a television environment, Weisz et al. [187] and Weisz and Kiesler [186] indicated the connectedness factor, when providing chat capabilities for television viewers, and Geerts et al. [73] studied how the television genre affects the social behavior of the viewers. At the same time, Shamma et al. [164] studied synchronous sharing of media content by using Zync, Yahoo!'s instant messenger system extended with media sharing functionality.

Although most of the work has considered synchronous and distant viewing, there are also opportunities for research in collocated

situations. For example, an iTV quiz game might provide opportunities for competition between family members.

3.4 Sharing Experiences

Within media studies, television has received significant attention, although it has remained a controversial electronic medium. Some researchers have blamed television for a fall in civic engagement [146]. In contrast, there are researchers who argue that TV creates a shared and common experience that bonds together members in an extended society [169]. Indeed, people lead widely diverse lives and activities, but TV and other mass media (radio and newspaper) provide a common point or reference or a kind of "social glue" that bonds both strangers and acquaintances together.

Despite the many criticisms on the quality of TV content and on the passive nature of the watching activity, the social uses of TV have been documented [100]. In particular, the use of audiovisual content as a point of reference for starting and sustaining relationships (e.g., "watercooler" discussions about yesterday's football match, or a popular TV series) is an everyday experience for the majority of TV users. Nevertheless, the pressures of daily life and the increase in the number of diasporic households make joint television viewing increasingly difficult.

In comparison to technological support for chatting over distance, broader support for sharing experiences through iTV has received little attention by researchers. Previous research in Social TV has addressed the verbal aspects of social communication to a great extent, as described in the previous section. Indeed, verbal communication is the most obvious way of social communication, but previous research in sociology has highlighted that much of the essence of social communication lies within the non-verbal realm [167].

As an active node, the television viewer might want to communicate (verbally or non-verbally) with others while watching [44, 59], to leave notes and comments for friends at specific moments of a television show [127], and to share enriched fragments of multimedia content with others [37]. For example, the Social Television project [82, 119], by

Motorola, provides an unobtrusive awareness system based on ambient devices. The final goal of these approaches is to provide enriched communication between separate parties, when watching television content.

Finally, in the case of iTV educational programming, it is likely that Social TV encourages collective efficacy and enhances learning. Accordingly, Aarreniemi-Jokipelto [1] describes the design and the evaluation of an instant messaging service that supports educational TV programs. The objective of the Instant Messaging service is to form a reciprocal community of users and to support this community, as well as to utilize the interaction of the community. Indeed, the above instant messaging application showed potential in engaging children with the traditional educational TV program.

3.5 Summary: Viewer As a Node in the Network

The first conclusion of this chapter is that the traditional hierarchical content flow is only one of the many distribution options for interactive television. Recently, viewers are enjoying television content on computers and on the move. But more importantly, viewers are becoming an active node that might add value and distribute media content. Figure 3.3 shows a comparison between the typical client/server architecture (Broadcast station/set-top box) and the current one: a hybrid approach. This hybrid approach highlights that clients might become more active nodes, and technical achievements on P2P networks and mobile/TV convergence support this paradigm shift.

The second conclusion is that television is not a passive and solitary watching activity and thus there is a need for further developed social interactive systems. These systems should focus on synchronous/asynchronous communications, as well, as on providing non-intrusive means to indicate the presence of the viewer's peers. There are some situations that could benefit from Social TV systems:

- Synchronous viewing over distance: for example, distant viewers should be able to watch together popular Social TV content, such as sports, quiz shows, series, or reality shows. A good starting point is to consider ways to disclose presence and status of viewers, to continue with support

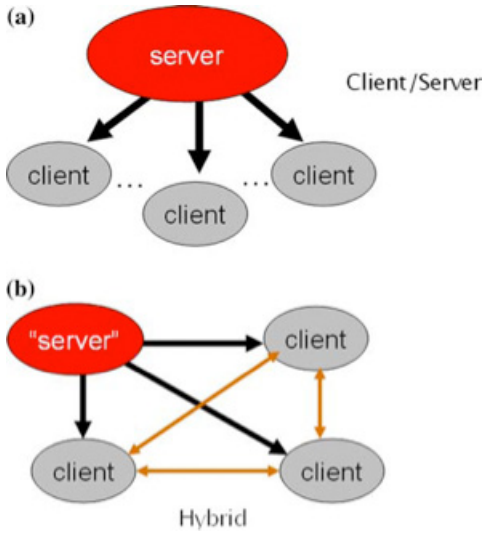


Fig. 3.3 Comparison between the traditional client/server architecture and the current hybrid architecture [31].

for multiple interpersonal communication modalities (non-verbal most notably), and to summarize the social experience with automated highlight production, which could motivate further discussion and social bonding between the distant viewers.

- Asynchronous viewing over distance: This a feasible scenario if we consider that distance viewers might have different time-schedules, patterns of daily life activities, or even live in distant time zones. Then, the probability of synchronous co-viewing is rather limited. In this case, a Social TV system could record and share shows and viewing habits with the members of the social circle. In addition, a Social TV system should allow annotation of content and recording of interactions, such as pausing, skipping, replaying, and content browsing. In this way, each time a particular TV program is accessed, a trace is kept, which is exploited at the next access, in order to personalize the content and most notably to provide a motivation for asynchronous communication.

This could be rather subtle, such as visual annotation of the content highlights, or could be more explicit such as audio and text comments.

- Asynchronous viewing at the same place: The main motivation for the development of Social TV systems is based on the need to bridge the distance between social circles of people, but there is also the case that co-located groups of people do not manage to meet as often as they wish for a Social TV night. A subset of the functionality that was described in the previous case might be the most appropriate here (Figure 3.4).

In summary, there are two dimensions of the social aspects of TV. The first dimension concerns the *presence* of the viewers: (1) co-located viewing in groups, and (2) distance viewing. The second dimension concerns the type of *social communication* between viewers: (1) synchronous communication that happens in real-time, and (2) asynchronous communication that happens with a time-lag. This

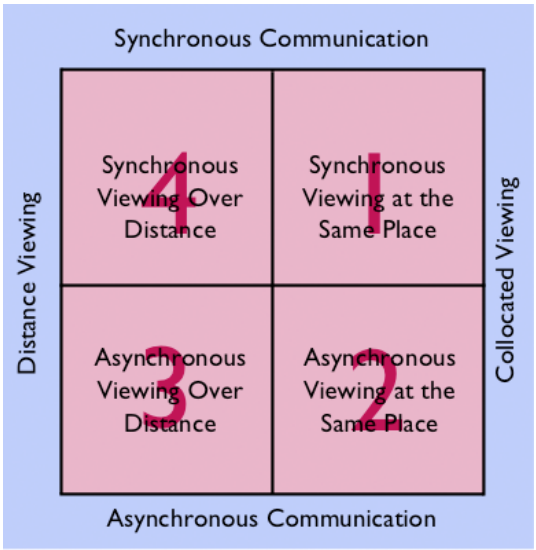


Fig. 3.4 Taxonomy of TV sociability along two dimensions (*presence, social communication*).

Table 3.1 State-of-the-art in content sharing.

Topic	State-of-the-art	Research agenda
Content distribution	Mobile TV field trials are available	The user as an active node in the distribution chain should be further investigated
	P2P Television is starting	
Social communications	There are many system that provide social capabilities	Non-verbal communication
		Collocated Social TV
		Inhabited television
		Sharing of fragments of video as gift-giving
Presence	Buddy-list like presence awareness systems	Privacy, scalability
		Non-obtrusiveness
		Social interactive television etiquette

matrix might be helpful in categorizing available and emerging content enriched communication services with iTV.

Traditionally, the content flows from the professional studio to the broadcasters to the viewers. Then, the viewers consume the media and might talk about it the following day with colleagues and acquaintances. The end-user, thus, had limited interactivity: to watch or not to watch. Contemporary research assumes that the viewer is not the end of the chain (or the sea where the river goes). In contrast, the viewer becomes just another node in the production–distribution–consumption chain. That is, other node that can play different roles: distributor or even producer of content. Table 3.1 summarizes the state-of-the-art regarding social communications.

4

Controlling Content

Content control corresponds to the activity, human-driven or automatic, of controlling the television content and services. Because of the amount of available digital content, personalization is applied to help the viewer in selecting the actual content he wants to consume. Content selection is by itself one of the major research topics in interactive television in the form of the Electronic Program Guide (EPG) and personalization of television content and advertisements initiatives.

After the content is selected, the user can navigate within it. Current personal video recording systems such as TiVo provide DVD-like capabilities: play, pause, forward and rewind. Interesting research in this area is to provide logical links within the content to parts of interest for the user. That is, the user might not be interested to jump to Section 4 of the content, but probably to a specific scene (e.g., "when he holds her while on the ship").

Finally, the content is rendered using the end-device's middleware. For simple encoded content this only requires an efficient decoding processor connected to the video card. For more complex video compositions, a software rendering stack might be needed. Even though

devices with a powerful middleware have not been particularly popular, government subside in countries like Italy has resulted in a high adoption of more powerful terminals.

This section is structured around the following topics: flexibility of TV schedules, personalization of content, input and output devices, and cooperative control of content. The first part deals with personalization of television content and advertisements. Then, we study extensions to the traditional remote control and television set setting. We conclude with a proposed metaphor that considers the end-user assuming several of the roles that have been centralized at the broadcasting studio.

4.1 Beyond Fixed TV Channels

In order to provide useful iTV applications and services there is a need for enhanced content. Examples are countless from the scores of a football match to the latest news. Current interactive television standards (e.g., MHP/OCAP) provide a Java runtime environment and an XHTML-based presentation environment for supporting interactivity linked to the content (Figure 4.1). Some recent results about this topic were presented by Costa et al. [53]. Their system supports real-time provision of enhanced services by the broadcaster. Nevertheless, as mentioned elsewhere the slow adoption of these standards has generated serious doubts about their usefulness in the broadcast community.

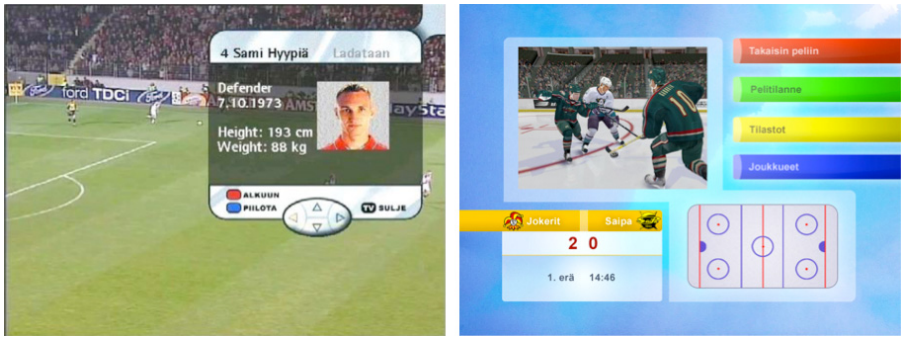


Fig. 4.1 Typical interactive enhanced content. (Left) An application that provides extra information about the football players. (Right) An application that provides extra information about an ice hockey match. In both cases, the extra information is obtrusive.

Apart from enhanced content, currently the most important iTV service is the EPG, since it helps the user to decide what to watch. Besides the commercial implementations of an EPG, there is a large body of academic research in the topic. Previous research for the EPG has addressed issues such as conceptual model, navigation, Web-based video retrieval, zooming interface, and natural language interaction. In the above cases, the research approach has followed the IT perspective by assuming that viewers need to select a channel or a program each time they start the TV. However, long established research in the media psychology discipline has identified that TV viewers settle down with a small number of channels and that they adhere to a ritualistic process in watching the same programs everyday or every week. In this respect Chorianopoulos and Spinellis [46] proposed the virtual channel metaphor for television viewers.

Another popular research stream within iTV has been the study of interactive narrative. Interactive narrative research addresses the production and creation of story-driven TV content. From that respect, it is probably the most innovative proposition toward iTV content, since story-driven content is the most popular type of TV content. For example, Ursu et al. [178] proposed a complete solution and grammar for the production of interactive programs, in which end-users are capable of modifying the outcome of a television program. They presented "Accidental Lovers" a production broadcasted by YLE (Finnish public broadcast company). Interactive narrative is a rather broad research topic, usually oriented for video games; nevertheless, researchers such as Agamanolis and Bove [8], Bocconi et al. [25], and Direu [56] have provided interesting results for television programs.

4.2 Automation and Personalization

Research on personalized television has brought together researchers from the communities of user modeling and adaptive hypermedia to discuss the applicability of previous theories in the field of iTV. Within that paradigm, there are a few discreet sub-streams of research: (1) personalized EPG, (2) personalized TV content, and (3) personalized advertising. A significant part of the personalized TV research focuses

in the design of a personalized EPG. Researchers have long realized that the EPG is merely a dataset to be linked with the user preferences and the watching behavior, in order to provide an input for recommendation engines or for automatically recording TV content on video servers. There is always a need for a visual user interface that can display program recommendations to the users. The majority of the respective research has neglected the issues of presenting the recommendations or automatically assembling a TV program. Moreover, there are a few works about personalized TV news. Last but not least, the Personalized TV stream of research has treated the special case of TV advertising content [103].

The typical recommendation process [170], called content-based filtering, takes as input ratings/preferences on previously observed content items (or item features) and/or user data, which are used to develop the user model (or profile). The input data are loaded on a recommendation engine that produces predictions concerning the interestingness of a user for unobserved items. The presentation of the results to the user depends on the type of the application typically including a ranked list or a limited list of the top-n interesting items. On the other hand, collaborative filtering is based on the likings of similar people. Thus, based on the user profile, it recommends those items that similar people have liked or selected previously. There are solutions such as AVATAR [23] that combine both approaches for providing most accurate results. These solutions have been applied for some time in the Web (e.g., Amazon) and have a number of problems [6] such as lack of contextual information and intrusive recommendations.

In the past, the recommendation process has been extensively applied in personalized Web-based applications. Most recently the recommendation process has been adjusted for the iTV domain by taking into account the particular characteristics of the medium, such as the low-involvement environment, input devices, viewing distance, group viewing [10, 76, 117, 118], and contextual information [11]. Fokker et al. [68] introduces Tribbler, a personalized P2P television system that connects social networks and can predict more accurately the likes of the users. Fink et al. [65] exploits the fact that the television content consists of audio data as well. Their approach offers an unobtrusive way

of collecting viewing and contextual information, based on a real-time ambient audio identification, and can be utilized in several applications. Finally, Lopez et al. [107] proposes a solution for personalized non-intrusive commercials.

Complementary research in user modeling has investigated how adaptation works for groups of people, such as a family. Researchers have argued that for a given group of people the recommended TV content might be better liked when the system considers the profiles of the respective group. For example, a study of an iTV adaptive instructional program confirmed that people tend to choose the TV content that would fit the preferences of a certain group of viewers [117].

Apart from recommendation systems, online video sharing interfaces include a rating system for collecting the user's preferences and, more interestingly, they allow for direct recommendation of content to other users. In that respect, Cesar et al. [37] have proposed micro-personal recommendation messages (Figure 4.2), direct recommendations of fragments of television content with optional media overlays. One important feature of such messages is the finer level of granularity, fragments of content, in comparison to traditional recommendations based on full programs.

Research in multimedia information retrieval has been trying to solve complex problems. For example, Ekin [60] developed techniques that allow automatic summarization of a sports game. In addition to patterns within the content he has also exploited knowledge about the cinematic structure of a game, in order to allow a computer program

Fig. 4.2 Generation of micro personal recommendation messages [37].

understand which are the highlights. However, significant (indeed, many of those algorithms and techniques have been patented to protect the hard effort), these alternative approaches to multimedia information retrieval in a networked TV infrastructure.

There are TV activities, such as aggregate replays, pauses, and recordings of content that hold potential for Social TV and personalization. In this context, Shamma et al. [165] states that there is a need to shift from semantics to pragmatics in multimedia information retrieval systems. The shift from semantics to pragmatics holds a great promise as a shortcut solution to some hard research issues in multimedia information retrieval and it is now becoming feasible due to Networked TV systems that facilitate the uninterrupted flow of both content and user activity between peers.

In summary, the social aspect of TV viewing might also point toward new fruitful directions for personalization, which are based on the behavior of small social circles of affiliated people. This type of "social" personalization in combination with social communication tools could enhance the sociability of TV well beyond a single household. Therefore, the research community should refocus the mainstream research on content personalization for interactive television, by taking into account that: television is a shared experience, contextual information is essential, and data gathering should be non-intrusive processes.

4.3 Interfaces Between Devices

The remote control has been the focus of research because of its inherent limitations. For example, text input for content searching is not an easy task to perform, since the number of keys is limited to the arrows, channel numbers, the OK button, and the color buttons. In this respect, Bernhaupt et al. [22] reported on the remote control complexity and their limitations for a shared experience.

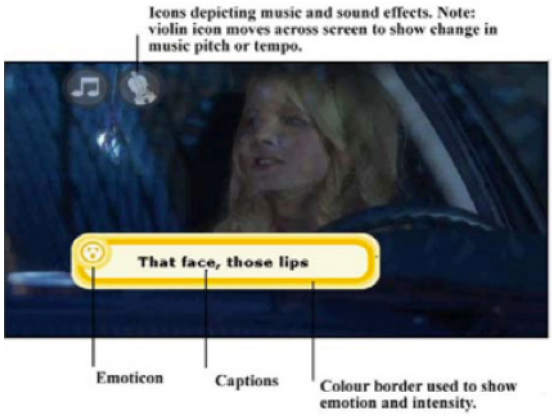
Lee and Lee [101] suggested that there is a wide diversity of attention levels to the television set — from background noise to full concentration. A viewer may sit down and watch a TV program attentively or leave the TV open as a radio and only watch when something interesting comes-up [48]. These findings contrast "to the image of the

highly interactive viewer intently engaged with the television set that is often summoned up in talking about new possibilities" [101]. Instead of assuming a user, who is eager to navigate through persistent dialog boxes, designers should consider that users might have varying levels of attention to the main display device, or to the complementary ones.

Contemporary iTV development tools for software user interface assume that the final product will be some variation of the WIMP paradigm. Nevertheless, previous research [79] has identified four types of non-WIMP user interfaces: (1) virtual reality for special input and output devices, such as gloves and VR helmets, (2) embedded, whereas the computer is invisible in an appliance, (3) notebook for mobile use with pen and handwriting recognition, and (4) hypermedia for hyperlinked multimedia. WIMP interfaces do not fit the interactive digital television paradigm, since it is a combination of the (2) embedded and (4) hypermedia styles. Furthermore, products such as the Wii[1] remote control and Philips research on ambient intelligence[2] are becoming popular examples of non-WIMP interaction with the television set (Figure 4.3).

The proportion of elderly people is increasing and will most likely continue to do so. They are the fastest growing proportion of society and they also possess significant material wealth and free time. Nevertheless, aging is associated with several mental or physical health problems. In this direction, Lee et al. [102] have devised a user interface that represents the emotional tone of the content through the use of special icons, in order to assist users that are deaf or hard of hearing. Indeed, some of the audio information such as music, sound effects, and speech prosody are not generally provided for in captioning. To include some of this information in closed captions, researchers have proposed graphical representations of the emotive information that is normally represented with non-dialog sound [102]. The graphical representation of emotive captions consists of color, icons, text, and emoticons. In addition to affective captioning, researchers have been devising novel interfaces that increase the accessibility of TV. O'Modhrain and Oakley [135]

[1] http://wii.com/
[2] http://www.research.philips.com/technologies/syst_softw/ami/



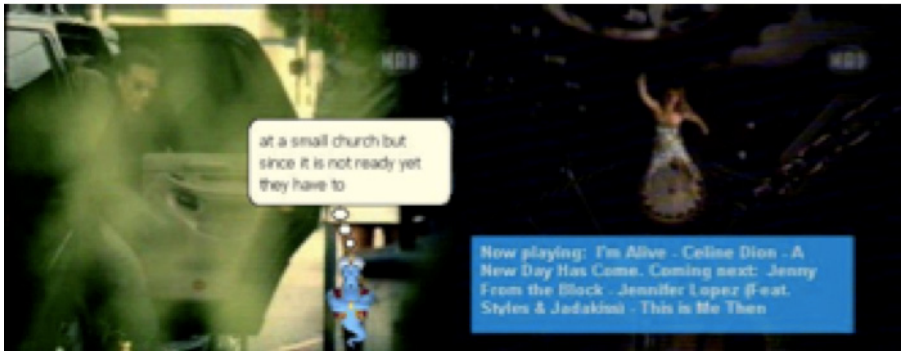


Fig. 4.3 Emotive captions is a proposal for enhancing captions [102]. Animated characters had an unfortunate reaction by desktop users, but anthropomorphic interfaces might be more suitable for ITV applications [45].

has presented a palpable remote control that vibrates depending on the TV content, in order to assist users who blind or partially sighted.

Contemporary research has extended the remote control capabilities. Some examples include the usage of everyday objects such as pillows [12, 163], paper-based devices [20], gestures recognizers [95], voice recognizers [21], and digital devices at home such as mobile phones and PDAs [40, 93, 104]. We can divide the research on input devices into three major directions: (1) extension of traditional remote controls, (2) re-utilization of everyday objects such as pillows or paper, and (3) adaptation of other personal devices such as mobile phones.

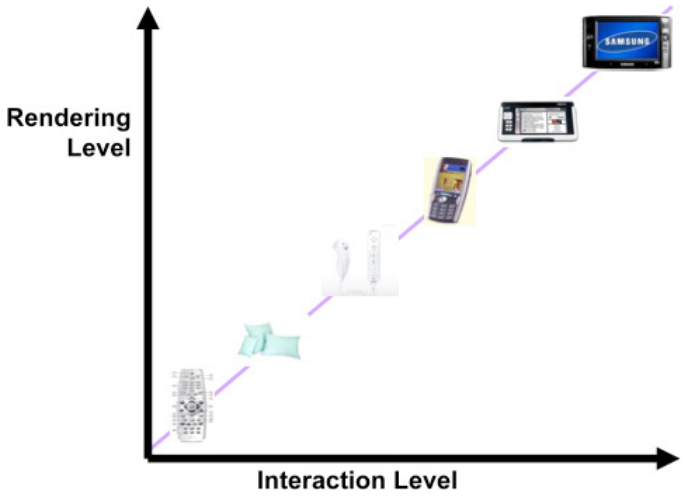


Fig. 4.4 Categorization of interactive devices based on their rendering capabilities. Apart from these devices, voice, gestures, and ink can be used for interaction with television content.

Figure 4.4 shows a categorization of interactive devices based on their rendering capabilities.

The first research technique is to extend current models for new usages; results following this approach include Berglund's research on extending remote controls with multi-modal capabilities [21] or with gesture capabilities [95] (Figure 4.5).

In addition to extending traditional remote controls, interesting findings have resulted by rethinking the possibilities of everyday objects such as paper and pillows. Berglund et al. [20] have presented an extensive user study about using digital paper and ink for selecting television programs. In addition, an active pillow with a host of user and situational sensors, such as the ambient technology solution proposed by Philips (Figure 4.6), can provide a mechanism for indirect content control [12, 163]. Finally, other everyday objects such as tables have been considering for manipulating media content, some examples include Microsoft Surface product[3] and the work performed in the ITEA Passepartout[4] project (Figure 4.7).

[3] http://www.microsoft.com/surface/
[4] http://www.passepartout-project.org/

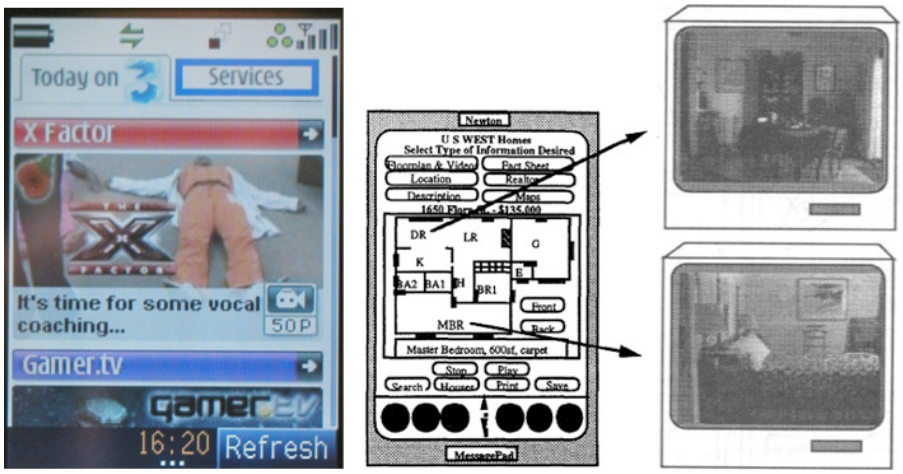


Fig. 4.5 The ITV experience might take place on the move [98], or on complementary devices [156].

Fig. 4.6 Ambient technology as proposed by Philips. For example, the lighting and fans in the room get activated depending on the current television content.

Digital devices such as pen-enabled tablet PCs have been used for annotating a manipulating audiovisual content. For example, Cesar et al. [38, 37] presented the "authoring from the couch" paradigm, which allows viewers to enrich television content by using a tablet PC or a mobile phone from the sofa. Similar results were presented by Goularte et al. [77], Ramos and Balakrishnan [148], and Abowd et al. [2]. These systems are based on a pre-defined pen-gesture language that permits extensive interaction and on ink scribbling support (Figure 4.8).

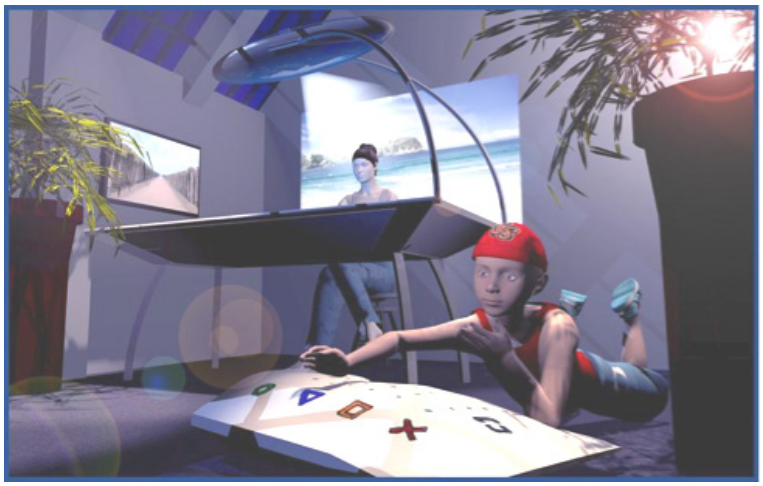


Fig. 4.7 Home setting investigated in the Passepartout project. The home includes a number of rendering and interactive devices such as screens, a table, and an interactive pillow.

Fig. 4.8 Using ink for enriching multimedia content. (Left) "Authoring from the couch" by including overlay media (ink, in this case) over media content. (Right) Authoring from the desktop for home movies.

Apart from annotation and manipulation, mobile devices have been used as extended remote controls for, for example, controlling the personal video recorder at home. Mobile devices can be used as a secondary and personal display at home that can be used for controlling interactive television (Figure 4.9), as presented elsewhere [40, 91, 93, 156, 176], and the convergence of mobile networks and interactive television [104, 158]. In this respect, the work on PC/mobile phone convergence

Fig. 4.9 Navigation of content in a non-intrusive manner. (Left) Navigation of scenes within a television program in a handheld device. (Right) EPG displayed in a handheld device.

and on the development of the universal remote control done during the Plebbes[5] project [128] is very relevant.

Regarding output devices, the industry has largely invested in two topics: high-definition television and mobile television. In particular, the availability of small broadband multimedia devices (Figure 4.4) has motivated the study of content for mobile video formats [98, 97]. There has also been a growing research interest on the usages of mobile television. Results indicate that mobile television is an immediate and short activity, thus television programs have to be tailored specifically for this medium; not only in terms of resolution but as well in terms of length. [55, 134, 158, 173].

An interesting research line is to consider the handheld device screen as a secondary screen [66, 67, 176] intended for personal information [40]. In this case, the distinction between shared devices (TV screen) and private screens at home (handheld device) can determine which information, depending on the nature of the information and the contextual situation of the user, is rendered where. This research considers the convergence of devices not as a substitute approach, but as an additive approach.

4.4 Cooperative Viewing

While TV is considered a shared screen, it is also associated with an input device that encourages solitary usage patterns. Indeed, TV

[5] http://www.pebbles.hcii.cmu.edu/

sets are usually equipped with one remote control, which excludes the possibility for interactivity to all, but the one who keeps the remote control. Despite this shortcoming, TV usage has been always considered a social activity [71, 110] and it might provide a better experience when watched with family members [100]. Early iTV settings considered the remote control as the single entry point, which imposes a number of restrictions on how groups of people can interact with content at home.

If TV watching takes place over distance or at different times, then the main requirement is to facilitate the communication of basic information that discloses status, preference, and activity of distant viewers. Indeed, an important functionality of an iTV system would be to create the impression of watching TV alongside a group of friends. For example, a Social TV system could offer a real-time indicator, or a history trace of TV content that the rest of the viewers in a social circle have been watching. In this way, Social TV provides a shared social context for conversations about the media that they have enjoyed, although not at the same time or place. These types of communications could be considered as the non-verbal part of the social communication between iTV users (Figure 4.10).

Social TV systems offer many possibilities and scenarios for remote social bonding. In the case of synchronous watching, users could remotely sense presence of other viewers that watch the same or a

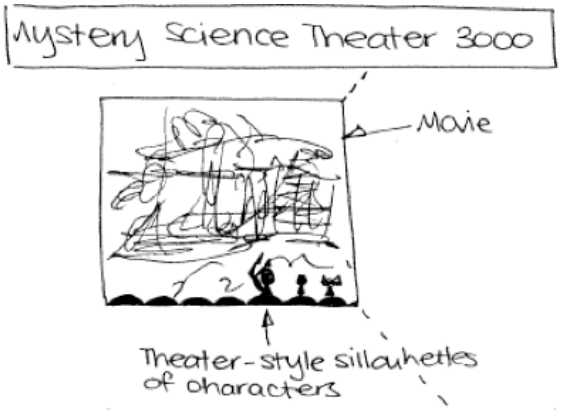


Fig. 4.10 Ducheneaut et al. [59] have proposed a conceptual design that facilitates seamless awareness of the presence of remote viewers.

different TV program. Then, the buddy list would be the first stage of an interface that would allow one to see what one's social circle is watching in real-time — whether they are watching live television or something stored on their local storage. In addition, they should have the option to stream a particular program directly from the STB of a friend, which would be very appealing when living in different countries. In this way, each STB becomes a virtual TV channel [46] that broadcasts a particular TV program to a social circle of viewers. This concept is also referred to as the virtual living room or the virtual couch [75].

When adding new functionality to TV programs, there will be a respective visual interface added onto one screen. Then, there is a risk of having on-screen interactions interrupt the enjoyment of TV content. This is a major issue in the case of TV, because viewers have become familiar with an established set of audiovisual techniques that keep the video area clean of other visual distractions. Most notably, the use of avatars and emoticons promotes a seamless and non-verbal communication among distant viewers. On the other hand, if the users wish to switch from a relaxed content enriched communication session to an intense audio or text chat, or even a video-conference session, a Social TV system should be able to support it.

Harboe et al. [83] performed an ethnographic study of a traditional TV set-up, which is enhanced with novel communication devices that support lightweight remote awareness (Figure 4.11). They reported that the ambient communication devices have enhanced the shared experience of TV watching over distance. The main contribution of that study is that it considers an alternative channel for the interpersonal communication aspect of Social TV. Nevertheless, current users of media and social communication technologies might not be willing to combine distinct communication channels if they could employ a single computer for that purpose [136].

TV watching in groups is governed by a set of cultural practices and interaction rules, which have evolved in a way so that co-located viewers can enjoy each other's company. These rules should be reflected in the case of mediated sociability [59]. Therefore, Social TV should facilitate distributed, sociable television viewing by processing each user's activities and ensuring that they fit within the established

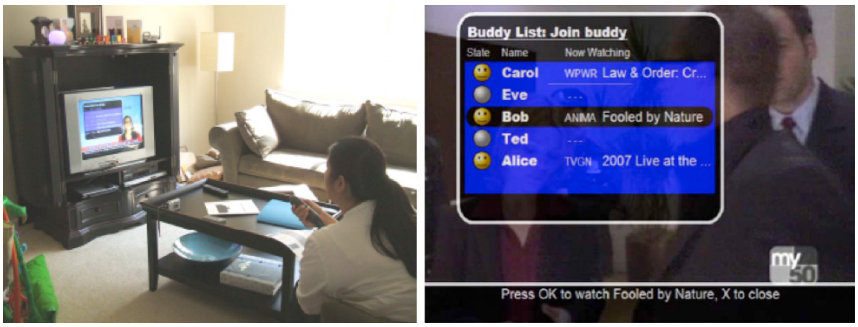


Fig. 4.11 The ambient orb is visible in the upper left (purple ball), above the television(left), and on-screen buddy-list (right) by Harboe et al. [83].

interaction rules that exist when watching TV together in the same room. For example, there should be an option to express high attention to the running TV program and "mute" audio or other modalities of communication by peer distant viewers. In the same way, there should be an option to express availability for intense social interaction. These types of status could be informed by the design of instant messaging software. In addition to the above, Social TV designers should consider the traditional TV watching scenario, where a group of viewers gather in the same place to enjoy a favorite TV program. Although this is a case that content enriched communication is least needed, there might worthwhile benefits in employing a Social TV system.

In summary, cooperative viewing could start with a screen displaying media use of each party during the past few days or hours. Current challenges in terms of presence and awareness, apart from the related privacy issues, include non-obtrusive mechanisms for indicating the presence of buddies and the development of social etiquettes. For example, new social protocols are needed to for example decide what to watch by a group of distant viewers. Finally, potential problems of scalability in terms of how many people fit in a person's virtual living room should be taken into account.

4.5 Summary: Viewer As a TV Director

Overall, there are many opportunities to enhance the TV content, beyond the bandwidth efficient broadcast delivery. Designers should

justify the use of persistent local storage and broadband Internet connections, which are standard to many iTV products (e.g., media center boxes). Digital local storage technology takes viewer control one step further — from simple channel selection with the remote — by offering convenient time-shifting through content. As a principle, designers should try to release the content from the fixed broadcast source and augment it with out-of-band content delivery. Therefore, an appropriate UI for content control should allow the user to customize the preferred sources of additional information and video content.

There is a need to support familiarity with a small set of visible or short-term memory choices [45]. As a consequence, designers should assume that most TV viewing starts with familiar content, but it might continue with browsing of relevant items. As a principle, instead of information seeking, iTV applications should facilitate relaxed exploration. Therefore, content navigation support should be subtle and not be enforced to the users. For example, there could be a hardwired remote control button for changing the flow of the running program (e.g., channel-up or down), in which the channel browsing brings the user to thematically adjacent content, instead of switching to a preprogrammed channel position.

There might be cases such as video games, in which the addition of interactive elements enhances the entertainment experience [115]. Therefore, the inclusion of overlay content should be topical (e.g., closely related to the underlying content). A starting point would be to make interactive versions for most of the information that is embedded in the video signal at the TV studio. For example, users could control the display of sports statistics, vote, play-along the players of quiz games, or customize news and stock-market tickers (Figure 4.12). Therefore, we suggest the empowerment of the viewer with features borrowed from a TV production studio.

Many aspects of the TV experience are gradually mediated through a computer program, thus making possible any kind of manipulation or control of the TV content. Being able to control remotely the TV experience at consumers' households has considerable ethical implications. Computer programs in digital STBs may store and analyze a wide variety of interactions for every household. Then, the interactions may

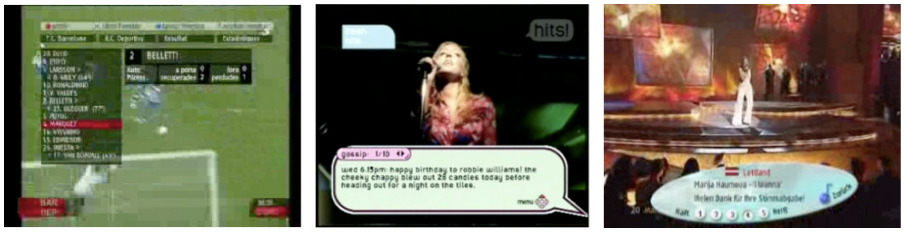


Fig. 4.12 On-demand sport statistics (TVC), music video clip trivia (MTV), and voting (NDR) bring the control of popular and familiar features from the TV studio directly into viewers' homes.

Table 4.1 Research agenda for content control.

Topic	State-of-the-art	Research agenda
Interactivity in content	Enhanced television and (some) interactive narratives	Aggregation of content based on Web 2.0 paradigms Productions based on interactive narratives
Personalization	Content filtering Collaborative filtering	Recommendations based on user context Recommendations of fragments of content Non-intrusive gathering of data Privacy Buddy-list recommendations
Input and output devices	Remote control	Extension of traditional remote controls (e.g., voice, gestures) Re-utilization of everyday objects (e.g., paper and ink) Adaptation of other devices (e.g., mobile phones) Non-monolithic rendering (different parts of the content making use of the most suitable rendering terminal devices)

be connected to personality characteristics, which is a major privacy issue. On the other hand, the availability of detailed user models makes possible a personalized approach for each individual, thus improving a medium, which has been characterized as mass and passive. Therefore, the balance between privacy and personalization features will be a major issue as iTV becomes more widespread.

5

Directions for Further Research

This article has identified three concepts, which are inherent in interactive television research: (1) editing, (2) sharing, and (3) controlling content. Based on these three concepts, we have surveyed the literature. Therefore we summarize opportunities for further research along three main topics:

(1) **Editing content:** The role of the viewer might overlap with that of the creator of TV content. We do not foresee that professional content will disappear, because high-quality and mass-appeal content has an important role in any society. On the other hand, user-generated content could co-exist and be interoperable with professionally produced content. Thus, methods, techniques, and tools for modeling and employing user-generated content and annotations are a major research area. In addition, tools for "authoring-from-the-sofa" and enriching television content would be beneficial for social participation and universal access to the information society.
(2) **Sharing content:** Although hierarchical distribution of content might have a place as an efficient and large-scale shared

experience (e.g., real-time sports), it might have to co-exist with network infrastructures that enable users as distributors of content, as well. The first solutions of iTV treated the users as consumers, but forgot that television watchers have established several social practices. Television users talk about programs with other people in the sofa, and even take phone calls while watching. Moreover, they recommend programs to friends and even make compilation of content for them to watch.

(3) **Controlling content:** In order to control the content at home a single remote control is not enough. Contextual information should be gathered in a non-intrusive manner. In addition, other devices such as mobile phones will be used as remote controls for, for example, preview enhanced content.

Fortunately, the iTV research community has been working toward such goals. The next paragraphs indicate the directions in these three main areas.

In terms of authoring there is a necessity to provide tools for accommodating user-generated content in the television distribution chain. Moreover, appropriate tools for remixing and repurposing multimedia content are needed. One example is the provision of tools to share fragments of multimedia content among peers. For example, to allow a user to share with her social network a part of a video material that might interest her. Such tools will provide users truly interactive capabilities; if we define interaction as the potential impact of the user on the content being consumed. From the content production viewpoint, next-generation authoring tools should provide functionality to create innovative television programs. For example, interactive narratives are an important direction to be further explored. We can conclude that tools that consider multimedia authoring as an incidental activity will play a key role in the future of multimedia creation.

In terms of control, there is a need for a better interoperability of the diverse multimedia content spread in different domains (e.g., Web, mobile, and TV domains). Normally, metadata standards are intended

for being used in one particular domain limiting the access to interesting video material. The final goal, thus, is that Web-based television converges with broadcast TV. At the same time, it is essential to take into account the contextual situation of the viewer. From the user perspective it is not the same to watch television alone than together with other people, it is not the same to watch it at home than in the bus; such contextual situation determines not only what the viewer wants to watch, but as well which rendering device needs to be used. Finally, in terms of user interaction, enhanced remote controls should be provided, probably with rendering capabilities. For example, the mobile phone can be used as an extended remote control in a number of situations. We can conclude that metadata integration together with contextual information will reshape the way viewers consume and interact with television content.

Finally, in terms of communication and sharing there is a need for a clearer integration of the viewing experience with the viewer's social network. There is a broad set of research and results into social networks and how they will define the way people communicate with each other. Television is not an exception, so a better integration of the social network in the living room is required. In this case both synchronous and asynchronous communication methods are required. Nevertheless, it is essential to take into account that such communications might disturb the television watching experience, as a noisy neighbor might destroy the cinema experience. Thus, more user-oriented research that indicates how, when social communications should be started and ended is needed. At the same time presence awareness will help to solve the obtrusive nature of communication with another peer while watching television. Nevertheless, research in this direction has to be cautious and to take into consideration all the privacy concerns related to presence.

In summary, most researchers have reached consensus that television use is not a passive and solitary watching activity, and thus there is a need for further development of social interactive television systems. These systems should focus on both synchronous and asynchronous social communications, which might be verbal as well as non-verbal. Moreover, iTV systems should support non-intrusive means to indi-

cate the presence of the viewer's peers and build upon the activity of like-minded (distant or close-by) viewers. Moreover, researchers have realized that the viewer is not the end of the chain. In this way, the viewer becomes another node that can play different roles: distributor or even producer of content. Or in other words, we need to move toward human-centered television [41].

6

Conclusion

In this study, we have examined the state-of-the-art in interactive television user developments. We have described how iTV, as a product offering (device and/or service), has been an unfulfilled promise for a long time. Moreover, we have realized that academic researchers have studied iTV behavior or have been developing iTV systems, all within their disciplinary boundaries. As a matter of fact, iTV has been either pushed to users as a product, or studied as behavior toward traditional TV systems, or developed as revolutionary digital system. In contrast, we suggested that emerging TV practices might be rather evolutionary rather than revolutionary. Moreover, we have found that the most effective approach toward iTV is a multidisciplinary one that concerns technology, user behavior, and media studies.

We proposed three basic concepts, namely, content editing, content sharing, and content control. Content editing corresponds to the activity of developing or organizing multimedia material, which can be done by professionals but could include user-generated content. Content sharing refers to all kinds of social activities that might occur around the television watching, such as chatting about television content and sharing content. Finally, content control corresponds to the activity of

deciding what to watch and how to watch it. Most notably, the proposed classification (edit–share–control) is a superset of the traditional content lifecycle (produce–distribute–view).

The proposed taxonomy (edit–share–control) is an evolutionary step over the established hierarchical produce–distribute–consume value chain. In addition, the taxonomy is not meant for categorizing future research issues or commercial products into one of the three concepts. In contrast, we find that many current efforts might be described by a combination of the three basic concepts. For example, an online video sharing system, such as YouTube, is mainly intended for content sharing, but also provides several features for sociability (e.g., comments), control of content (e.g., favorites, channels, etc.), as well as content creation. In this way, most existing research and commercial products could be described using this simple taxonomy.

Overall, interactive television research should leverage the simple and established viewer practices, such as recording, browsing (e.g., pause, repeat, and skip), sharing with others, and talking about content to become significant determinants in the value chain of content distribution on any TV network. In practice, both users and content providers are looking forward for improved content navigation systems. For this purpose, the user modeling research stream has to focus on a solution that enhances sociability. If TV content is such an important placeholder for discussion, as argued by many researchers before, then the traditional approaches to personalization reduce the chances that two persons might have watched the same program. On the other hand, personalization could become an effort that enhances social bonding, in addition to providing recommendations about novel content.

In addition to content control and content sharing, content creation is an important ingredient for the future of television. We have identified that user-generated content has become an important genre in three different forms: custom play lists, user-generated short clips, and most importantly user-augmented content, such as content annotations and mash-ups. User-augmented material can take different shapes as well, for example, as micro-personal recommendations of fragments of television content. Finally, we do not foresee the end of professionally produced content anytime soon. Although user-generated content has

been growing, viewers will also appreciate mass appeal, professionally made story-telling, and high-quality productions for their distracting value, their elaborate aesthetics, and in particular for creating a shared experience.

In conclusion, while counter-intuitive to many, watching television could be a very fulfilling, sociable, and creative activity. Therefore, the main objective of iTV research should be to develop technological support for the personal, social, and creative practices that surround the iTV content lifecycle. At the same time, we should retain the centrality of TV as a leisure pursuit and enhance the opportunities for mediated shared experiences that have become the social glue of modern information societies.

Acknowledgments

This study reports and builds upon the collaborative work of numerous individuals and teams, who have been active in the interactive television research community during the past decade. Moreover, we are grateful to the community of the European Interactive Television Conference (EuroITV), which has been gathering together so many diverse and interesting individuals. Finally, we are grateful to Keith Baker, Ben Bederson, David Geerts, Gunnar Harboe, and Ian Kegel for their constructive comments on early drafts of this article.

Pablo Cesar would like to thank the funding from the NWO BRICKS PDC3 project, the ITEA project Passepartout, the EU FP6 IST project SPICE, the EU FP7 ICT project Ta2, and the EU FP7 ICT project iNEM4U.

Konstantinos Chorianopoulos was partially supported by the following European Commission Marie Curie Fellowships: MEDIACITY (MTKD-CT-2004-517121) and CULT (MC-ERG-2008-230894).

Bios

Pablo Cesar is a researcher at CWI (The National Research Institute for Mathematics and Computer Science in the Netherlands) in Amsterdam. He has (co)authored over 40 articles (conference papers and journal articles) about multimedia systems and infrastructures, media sharing, interactive media, multimedia content modeling, and user interaction. He has given tutorials about Interactive Digital Television in prestigious conferences such as ACM Multimedia and WWW Conferences. http://homepages.cwi.nl/∼garcia

Konstantinos Chorianopoulos is a lecturer in the Department of Informatics, at the Ionian University (Greece). In 2002, he founded UITV.INFO (http://uitv.info), which is a newsletter and Web portal for interactive television research resources. He has been researching interactive TV since early 2000. http://www.ionio.gr/∼choko

> *"Television? The word is half Greek and half Latin. No good will come of it." C.P. Scott*

References

[1] P. Aarreniemi-Jokipelto, "Instant messaging in informal learning via interactive television: Online communities among children in a "get along" program," *Computers in Entertainment*, vol. 5, no. 2, article number 9, 2007.

[2] G. D. Abowd, M. Gauger, and A. Lachenmann, "The family video archive: An annotation and browsing environment for home movies," in *Proceedings of the ACM International Workshop on Multimedia Information Retrieval (MIR)*, pp. 1–8, 2003.

[3] J. F. Abreu, P. Almeida, and V. Branco, "2BeOn: Interactive television supporting interpersonal communication," in *Proceedings of EuroGraphics*, pp. 199–208, 2001.

[4] J. F. Abreu, P. Almeida, R. Pinto, and V. Nobre, "Implementation of social features over regular IPTV STB," in *Proceedings of the European Conference on European Interactive Television Conference (EuroITV)*, pp. 29–32, 2009.

[5] P. C. Adams, "Television as gathering place," *Annals of the Association of American Geographers*, vol. 82, no. 1, pp. 117–135, 1992.

[6] G. Adomavicius and A. Tuzhilin, "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*, vol. 17, no. 6, pp. 734–749, 2005.

[7] S. Agamanolis, "At the intersection of broadband and broadcasting: How interactive TV technologies can support human connectedness," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 121–135, 2008.

[8] S. Agamanolis and V. M. Bove, "Viper: A framework for responsive television," *IEEE Multimedia*, vol. 10, no. 3, pp. 88–98, 2003.

[9] A. Al-Hezmi, O. Friedrich, S. Arbanowski, and T. Magedanz, "Interactive multimedia services over open NGN testbed," in *Proceedings of the International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities*, article number 39, 2008.

[10] L. Ardissono, A. Kobsa, and M. T. Maybury, eds., *Personalized Digital Television Targeting Programs to Individual Viewers*. Kluwer Academic Publishers, 2004.

[11] L. Aroyo, P. Bellekens, M. Bjorkman, G. J. Houben, P. Akkermans, and A. Kaptein, "SenSee framework for personalized access to TV content," in *Proceedings of the European Conference on Interactive Television*, pp. 156–165, 2007.

[12] L. Aroyo, F. Nack, T. Schiphorst, H. Schut, and M. KauwATjoe, "Personalized ambient media experience: Move.me case study," in *Proceedings of the International Conference on Intelligent User Interfaces*, pp. 298–301, 2007.

[13] K. Baker, "Intrusive interactivity is not an ambient experience," *IEEE Multimedia*, vol. 13, no. 2, pp. 4–7, 2006.

[14] K. Baker, "Jules verne project: Realization of reactive media using MPEG4 on MHP," in *Proceedings of the IEEE International Conference on Consumer Electronics*, pp. 439–440, 2006.

[15] P. J. Bates T-learning Consultation Paper. pjb Associates http://www.pjb.co.uk/t-learning.htm, November 2002.

[16] P. Baudisch and L. Brueckner, "TV scout: Guiding users from printed TV program guides to personalized TV recommendation," in *Proceedings of the Workshop on Personalization in Future TV*, 2002.

[17] P. Bellekens, G.-J. Houben, L. Aroyo, K. Schaap, and A. Kaptein, "User model elicitation and enrichment for context-sensitive personalization in a multiplatform TV environment," in *Proceedings of the European Conference on Interactive Television (EuroITV)*, pp. 119–128, 2009.

[18] S. Benford, C. Greenhalgh, M. Craven, G. Walker, T. Regan, J. Morphett, and J. Wyver, "Inhabited television: Broadcasting interaction from within collaborative virtual environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 4, pp. 510–547, 2000.

[19] F. Bentley, C. Metcalf, and G. Harboe, "Personal versus commercial content: The similarities between consumer use of photos and music," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 667–676, 2006.

[20] A. Berglund, E. Berglund, A. Larsson, and M. Bang, "Paper remote: An augmented television guide and remote control," *Universal Access Information Society*, vol. 4, no. 4, pp. 300–327, 2006.

[21] A. Berglund and P. Johansson, "Using speech and dialogue for interactive TV navigation," *Universal Access in the Information Society*, vol. 3, no. 3–4, pp. 224–238, 2004.

[22] R. Bernhaupt, M. Obrist, A. Weiss, E. Beck, and M. Tscheligi, "Trends in the living room and beyond: Results from ethnographic studies using creative and playful probing," *ACM Computers in Entertainment*, vol. 6, no. 1, article 5, 2008.

[23] Y. Blanco, J. J. Pazos, A. Gil, M. Ramos, A. Fernandez, R. P. Diaz, M. Lopez, and B. Barragans, "AVATAR: An approach based on semantic reasoning to recommend personalized TV programs," in *Proceedings of the World Wide Web Conference*, pp. 1078–1079, 2005.

[24] E. Blankinship, B. Smith, H. Holtzman, and W. Bender, "Closed caption, open source," *BT Technology Journal*, vol. 22, no. 4, pp. 151–159, 2004.

[25] S. Bocconi, F. Nack, and L. Hardman, "Supporting the generation of argument structure within video sequences," in *Proceedings of the ACM Conference on Hypertext and Hypermedia*, pp. 75–84, 2005.

[26] E. Boertjes, "ConnecTV: Share the experience," in *Adjunct Proceedings of the European Conference on Interactive Television*, pp. 139–140, 2007.

[27] S. Bonnici, "Which channel is that on? A design model for electronic programme guides," in *Proceedings of the European Conference on Interactive Television*, pp. 49–57, 2003.

[28] B. Brown and L. Barkhuus, "The television will be revolutionized: Effects of PVRs and filesharing on television watching," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 663–666, 2006.

[29] H. Bryhni, H. Lovett, E. Aaartmann-Moe, D. Solvoll, and T. Sorensen, "On-demand regional television over the internet," in *Proceedings of the ACM International Conference on Multimedia*, pp. 99–107, 1996.

[30] D. C. A. Bulterman, "A rationale for creating an interactive-TV profile for SMIL," in *Adjunct Proceedings of the European Conference on Interactive Television*, pp. 593–597, 2006.

[31] D. C. A. Bulterman, "User-centered control within multimedia presentations," *Multimedia Systems Journal*, vol. 12, no. 4–5, pp. 423–438, 2007.

[32] D. C. A. Bulterman and L. Hardman, "Structured multimedia authoring," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 1, no. 1, pp. 89–109, 2005.

[33] D. C. A. Bulterman and L. Rutledge, *SMIL 3.0: Interactive Multimedia for Web, Mobile Devices, and Daisy Talking Books*. Heidelberg: Springer-Verlag, 2008.

[34] M. Burrnester, F. Koller, and T. Konig, "A user interface for interactive pay TV: PREMIERE goes online," in *Proceedings of ITV conference*, 1996.

[35] A. Cappelletti, M. Nardon, F. Pianesi, and M. Zancanaro, "Readers and skimmers watching iTV: Evidences from interactive video-on-demand," in *Proceedings of the Workshop on Personalization in Future TV*, 2003.

[36] R. G. Cattelan, C. Teixeira, R. Goularte, and M. D. Pimentel, "Watch-and-comment as a paradigm toward ubiquitous interactive video editing," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 4, no. 4, article 28, 2008.

[37] P. Cesar, D. C. A. Bulterman, D. Geerts, J. Jansen, H. Knoche, and W. Seager, "Enhancing social sharing of videos: Fragment, annotate, enrich, and share," in *Proceedings of the ACM Conference on Multimedia*, pp. 11–20, 2008.

[38] P. Cesar, D. C. A. Bulterman, and J. Jansen, "An architecture for viewer-side enrichment of TV content," in *Proceedings of the ACM International Conference on Multimedia*, pp. 651–654, 2006.

[39] P. Cesar, D. C. A. Bulterman, and J. Jansen, "Benefits of structured multimedia documents in iDTV: The end-user enrichment system," in *Proceedings of the ACM Symposium on Document Engineering*, pp. 176–178, 2006.

[40] P. Cesar, D. C. A. Bulterman, and J. Jansen, "Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content," in *Proceedings of the European Conference on Interactive Television*, pp. 168–177, 2008.

[41] P. Cesar, D. C. A. Bulterman, and L. F. G. Soares, "Human-centered television: Directions in interactive digital television research," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 4, no. 4, article 24, 2008.

[42] P. Cesar, P. Vuorimaa, and J. Vierinen, "A graphics architecture for high-end interactive television terminals," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 2, no. 4, pp. 343–357, 2006.

[43] M. Cha, P. Rodriguez, J. Crowcroft, S. Moon, and X. Amatriain, "Watching television over an IP network," in *Proceedings of the ACM SIGCOMM Conference on Internet Measurement*, pp. 71–84, 2008.

[44] K. Chorianopoulos and G. Lekakos, "Introduction to social TV: Enhancing the shared experience with interactive TV," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 113–120, 2008.

[45] K. Chorianopoulos and D. Spinellis, "Affective usability evaluation for an interactive music television channel," *ACM Computers in Entertainment*, vol. 2, no. 3, article number 14, 2004.

[46] K. Chorianopoulos and D. Spinellis, "User interface development for interactive television: Extending a commercial DTV platform to the virtual channel API," *Computers and Graphics*, vol. 28, no. 2, pp. 57–166, 2004.

[47] K. Chorianopoulos and D. Spinellis, "User interface evaluation of interactive TV: A media studies perspective," *Universal Access in the Information Society*, vol. 5, no. 2, pp. 209–218, 2006.

[48] M. Clancey, "The television audience examined," *Journal of Advertising Research*, vol. 34, no. 4, pp. 2–11, 1994.

[49] W. Cooper, "The interactive television user experience so far," in *Proceeding of the International Conference on Designing Interactive User Experiences for TV and Video (UXTV)*, pp. 133–142, 2008.

[50] T. Coppens, F. Vanparijs, and K. Handekyn, "AmigoTV: A social TV experience through triple-play convergence," *Alcatel, White Paper*, 2005.

[51] J. Cosmas et al., "CustomTV with MPEG-4 and MPEG-7," in *IEEE Colloquium on Interactive Television*, vol. 9, pp. 1–6, 1999.

[52] J. Cosmas et al., "Customized television: Standards compliant advanced digital television," *IEEE Transactions on Broadcasting*, vol. 48, no. 2, pp. 151–158, 2002.

[53] R. M. R. Costa, M. F. Moreno, R. F. Rodrigues, and L. F. G. Soares, "Live editing of hypermedia documents," in *Proceedings of the ACM Symposium on Document Engineering*, pp. 165–175, 2006.

[54] M. Craven, S. Benford, C. Greenhalgh, J. Wyver, C. Brazier, A. Oldroyd, and T. Regan, "Ages of avatar: Community building for inhabited television," in

Proceedings of the International Conference on Collaborative Virtual Environments, pp. 189–194, 2000.

[55] Y. Cui, J. Chipchase, and Y. Jung, "Personal TV: A qualitative study of mobile TV users," in *Proceedings of the European Conference on Interactive Television*, pp. 195–204, 2007.

[56] M. Deriu, "A viable approach to deliver television events, having multiple and selectable plots: MultiChannelStory," in *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pp. 358–360, 2005.

[57] Digital TV Group (DTG), *Digital Terrestrial Television MHEG-5 Specification*. v1.06, 2003.

[58] S. M. Drucker, A. Glatzer, S. D. Mar, and C. Wong, "Smartskip: Consumer level browsing and skipping of digital video content," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 219–226, 2002.

[59] N. Ducheneaut, R. J. Moore, L. Oehlberg, J. D. Thornton, and E. Nickell, "SocialTV: Designing for distributed, social television viewing," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 136–154, 2008.

[60] A. Ekin, "Sports video processing for description, summarization, and search," Unpublished PhD thesis, Rochester Institute of Technology, 2003.

[61] S. Fallahkhair, L. Pemberton, and J. Masthoff, "A dual device scenario for informal language learning: Interactive television meets the mobile phone," in *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, pp. 16–20, 2004.

[62] B. J. Felton, M. L. Moss, and R. J. Sepulveda, "Two-way television: An experiment in interactive programming for the elderly," *Experimental Aging Research*, vol. 6, no. 1, pp. 29–45, 1980.

[63] D. A. Ferguson, "Channel repertoire in the presence of remote control devices, VCRs, and cable television," *Journal of Broadcasting and Electronic Media*, vol. 36, no. 1, pp. 83–91, 1992.

[64] D. A. Ferguson and E. M. Perse, "Media and audience influences on channel repertoire," *Journal of Broadcasting and Electronic Media*, vol. 37, no. 1, pp. 31–47, 1993.

[65] M. Fink, M. Covell, and S. Baluja, "Social– and interactive–television: Applications based on real-time ambient-audio identification," in *Proceedings of the European Conference on Interactive Television*, pp. 138–146, 2006.

[66] M. Finke and D. Balfanz, "Interaction with content-augmented video via off-screen hyperlinks for direct information retrieval," in *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pp. 187–194, 2002.

[67] M. Finke and D. Balfanz, "A reference architecture supporting hypervideo content for ITV and the Internet domain," *Computers & Graphics*, vol. 28, no. 2, pp. 179–191, 2004.

[68] J. Fokker, H. Ridder, P. Westendorp, and J. Pouwelse, "Psychological backgrounds for inducing cooperation in peer-to-peer television," in *Proceedings of the European Conference on Interactive Television*, pp. 136–145, 2007.

[69] B. Fox, "Digital TV comes down to earth," *IEEE Spectrum*, vol. 35, no. 10, pp. 23–29, 1998.

[70] B. Furht, "Interactive television systems," in *Proceedings of the ACM Symposium on Applied Computing*, pp. 7–11, 1996.

[71] D. Gauntlett and A. Hill, *TV Living: Television, Culture and Everyday Life*. Routledge, 1999.

[72] D. Geerts, "Comparing voice chat and text chat in a communication tool for interactive television," in *Proceedings of the Nordic Conference on Human-computer Interaction*, pp. 461–464, 2006.

[73] D. Geerts, P. Cesar, and D. Bulterman, "The implications of program genres for the design of social television systems," in *Proceeding of the International Conference on Designing Interactive User Experiences for TV and Video (UXTV)*, pp. 71–80, 2008.

[74] D. Geerts and D. De Grooff, "Supporting the social uses of television: Sociability heuristics for social TV," in *Proceedings of the International Conference on Human Factors in Computing Systems*, pp. 595–604, 2009.

[75] S. Goldenberg, "Digital video recorders and micro-social networking: Recreating the shared watching experience of television," *Adjunct Proceedings of the European Conference on Interactive Television*, 2007.

[76] D. Goren-Bar and O. Glinansky, "Family stereotyping — A model to filter TV programs for multiple viewers," in *Proceedings of the Workshop on Personalization in Future TV*, 2002.

[77] R. Goularte, R. G. Cattelan, J. A. Camacho-Guerrero, V. R. Inacio, and M. da Graca C. Pimentel, "Interactive multimedia annotations: Enriching and extending content," in *Proceedings of the ACM Symposium on Document Engineering*, pp. 84–86, 2004.

[78] R. Goularte, E. S. Moreira, and M. G. C. Pimentel, "Structuring interactive TV documents," in *Proceedings of the ACM Symposium on Document Engineering*, pp. 42–51, 2003.

[79] M. Green and R. Jacob, "Siggraph '90 workshop report: Software architectures and metaphors for non-wimp user interfaces," *SIGGRAPH Computer Graphics*, vol. 25, no. 3, pp. 229–235, 1991.

[80] T. Gross, M. Fetter, and T. Paul-Stueve, "Toward advanced social TV in a cooperative media space," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 155–173, 2008.

[81] M. Hamasaki, H. Takeda, and T. Nishimura, "Network analysis of massively collaborative creation of multimedia contents: Case study of hatsune miku videos on nico nico douga," in *Proceeding of the International Conference on Designing Interactive User Experiences For TV and Video*, pp. 165–168, 2008.

[82] G. Harboe, N. Massey, C. Metcalf, D. Wheatley, and G. Romano, "The uses of social television," *ACM Computers in Entertainment*, vol. 6, no. 1, article 8, 2008.

[83] G. Harboe, C. J. Metcalf, F. Bentley, J. Tullio, N. Massey, and G. Romano, "Ambient social TV: Drawing people into a shared experience," in *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2008.

[84] P. Hulsen, E. Vlemmix, and K. Baker, "Peer-to-Peer Internet television," in *Proceedings of the First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution*, pp. 267–270, 2005.

[85] K. Jaaskelainen, "Strategic questions in the development of interactive television programs," PhD thesis, University of Art and Design Helsinki, 2001.

[86] H. Jenkins, *TV Tomorrow*. MIT Technology Review, 2001.

[87] J. F. Jensen, "Interactive television: New genres, new format, new content," in *Proceedings of the Australasian Conference on Interactive Entertainment*, pp. 89–96, 2005.

[88] J. F. Jensen, "The concept of interactivity — Revisited: Four new typologies for a new media landscape," in *Proceeding of the International Conference on Designing Interactive User Experiences for TV and Video*, pp. 129–132, 2008.

[89] J. F. Jensen, "Interactive television — A brief media history," in *Proceedings of the European Conference on Interactive Television*, pp. 1–10, 2008.

[90] M. H. Kang, "Interactivity in television: Use and impact of an interactive program guide," *Journal of Broadcasting and Electronic Media*, vol. 46, no. 3, pp. 330–345, 2002.

[91] A. Karanastasi, F. G. Kazasis, and S. Christodoulakis, "A natural language model for managing TV-Anytime information in mobile environments," *Personal and Ubiquitous Computing*, vol. 9, no. 5, pp. 262–272, 2005.

[92] C. Karat, J. Karat, J. Vergo, C. Pinhanez, D. Riecken, and T. Cofino, "That's entertainment! designing streaming, multimedia web experiences," *International Journal of Human-Computer Interaction*, vol. 14, no. 3 & 4, pp. 369–384, 2002.

[93] F. G. Kazasis, N. Moumoutzis, N. Pappas, A. Karanastasi, and S. Christodoulakis, "Designing ubiquitous personalized TV-Anytime services," in *Proceedings of the International Workshop on Ubiquitous Mobile Information and Collaboration Systems*, 2003.

[94] J. Kim, H. Kim, and K. Park, "Towards optimal navigation through video content on interactive TV," *Interacting with Computers*, vol. 18, no. 4, pp. 723–746, 2006.

[95] S. H. Kim, J. Ok, H. J. Kang, M. C. Kim, and M. Kim, "An interaction and product design of gesture based TV remote control," in *Extended Abstracts in Human Factors in Computing Systems*, pp. 1548–1548, 2004.

[96] D. Kirk, A. Sellen, R. Harper, and K. Wood, "Understanding videowork," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 61–70, 2007.

[97] H. Knoche, J. McCarthy, and M. Sasse, "Can small be beautiful? Assessing image size requirements for mobile TV," in *Proceedings of ACM Multimedia*, pp. 829–838, 2005.

[98] H. Knoche and J. D. McCarthy, "Design requirements for mobile TV," in *Proceedings of MobileHCI*, pp. 69–76, 2005.

[99] R. Kubey, "TV and the Internet: Pitfalls in forecasting the future," *Knowledge, Technology, & Policy*, vol. 13, no. 2, pp. 63–85, 2000.

[100] R. Kubey and M. Csikszentmihalyi, *Television and the Quality of Life: How Viewing Shapes Everyday Experiences*. Lawrence Erlbaum, 1990.

[101] B. Lee and R. S. Lee, "How and why people watch TV: Implications for the future of interactive television," *Journal of Advertising Research*, vol. 35, no. 6, pp. 9–18, 1995.

[102] D. G. Lee, D. I. Fels, and J. P. Udo, "Emotive captioning," *Computers in Entertainment*, vol. 5, no. 2, article number 11, 2007.

[103] G. Lekakos and G. M. Giaglis, "A lifestyle-based approach for delivering personalized advertisements in digital interactive television," *Journal of Computer-Mediated Communication*, vol. 9, no. 2, 2004.

[104] C. C. Lin and M. S. Chen, "On controlling digital TV set-top-box by mobile devices via IP network," in *Proceedings of the IEEE International Symposium on Multimedia*, pp. 52–59, 2005.

[105] T. D. Little and D. Venkatesh, "Prospects for interactive video-on-demand," *IEEE MultiMedia*, vol. 1, no. 3, pp. 14–24, 1994.

[106] J. Livaditi, K. Vassilopoulou, C. Lougos, and K. Chorianopoulos, "Needs and gratifications for interactive TV applications: Implications for designers," in *Proceedings of the Annual Hawaii International Conference on System Science*, 2003.

[107] M. Lopez, J. J. Pazos, J. Garcia, Y. Blanco, and A. Gil, "Non-invasive and personalized advertising through MPEG-4 processing and semantic reasoning," in *Proceedings of the IEEE International Conference on Consumer Electronics*, 2007.

[108] A. Lugmayr, S. Niiranen, and S. Kalli, *Digital Interactive TV and Metadata Future Broadcast Multimedia*. Springer, 2004.

[109] J. Lull, "Family communication patterns and the social uses of television," *Communication Research*, vol. 7, no. 3, pp. 319–333, SAGE Publications, 1980.

[110] J. Lull, *Inside Family Viewing*. Routledge, 1990.

[111] K. Luyten, K. Thys, S. Huypens, and K. Connix, "Telebuddies: Social stitching with interactive television," in *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1049–1054, 2006.

[112] N. Macdonald, "Can HCI shape the future of mass communications?," *Interactions*, vol. 11, no. 2, pp. 44–47, 2004.

[113] M. Maguire, "Applying evaluation methods to future digital TV services," in *Pleasure with Products Beyond Usability*, (W. Green and P. Jordan, eds.), pp. 353–366, Taylor and Francis, 2002.

[114] C. Malerczyk, K. Klein, and T. Weibeisek, "3D reconstruction of sports events for digital TV," in *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization, and Computer Vision*, 2003.

[115] T. W. Malone, "Heuristics for designing enjoyable user interfaces: Lessons from computer games," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 63–68, 1982.

[116] C. Marrin, R. Myers, J. Kent, and P. Broadwell, "Steerable media: Interactive television via video synthesis," in *Proceedings of the International Conference on 3D Web Technology: Virtual Reality Modelling Language Symposium*, pp. 7–14, 2001.

[117] J. Masthoff, "Group modeling: Selecting a sequence of television items to suit a group of viewers," *User Modeling and User Adapted Interaction*, vol. 14, no. 1, pp. 37–85, 2004.

[118] J. Masthoff, W. W. Vasconcelos, C. Aitken, and F. S. Correa da Silva, "Agent-based group modeling for ambient intelligence," in *The AISB Symposium on Affective Smart Environments*, 2007.

[119] C. Metcalf, G. Harboe, J. Tullio, N. Massey, G. Romano, E. M. Huang, and F. Bentley, "Examining presence and lightweight messaging in a social television experience," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 4, no. 4, article 27, 2008.

[120] P. Meuleman, A. Heister, H. Kohar, and D. Tedd, "Double agent — presentation and filtering agents for a digital television *recording system*," in *CHI 98 Conference Summary on Human Factors in Computing Systems*, pp. 3–4, 1998.

[121] M. Milenkovic, "Delivering interactive services via a digital TV infrastructure," *IEEE Multimedia*, vol. 5, no. 4, pp. 34–43, 1998.

[122] S. Miller, "Interactive TV: Taking on the masses with mobile messaging TV," *ACM Computers in Entertainment*, vol. 3, no. 2, article number 6, 2005.

[123] A. Monk, "User-centred design: The home use challenge," in *Home Informatics and Telematics: Information Technology and Society*, (A. Sloane and F. van Rijn, eds.), pp. 181–190, Kluwer, 2000.

[124] S. Morris and A. Smith-Chaigneau, *Interactive TV Standards: A Guide to MHP, OCAP and JavaTV*. Focal Press, 2005.

[125] F. Nack and A. T. Lindsay, "Everything you wanted to know about MPEG-7 (part 1)," *IEEE Multimedia*, vol. 6, no. 3, pp. 65–77, 1999.

[126] F. Nack and A. T. Lindsay, "Everything you wanted to know about MPEG-7 (part 2)," *IEEE Multimedia*, vol. 6, no. 4, pp. 64–73, 1999.

[127] M. Nathan, C. Harrison, S. Yarosh, L. Terveen, L. Stead, and B. Amento, "CollaboraTV: Making television viewing social again," in *Proceeding of the International Conference on Designing Interactive User Experiences for TV and Video*, pp. 85–94, 2008.

[128] J. Nichols, B. Myers, M. Higgins, J. Hughes, T. Harris, R. Rosenfeld, and M. Pignol, "Generating remote control interfaces for complex appliances," in *Proceedings of the ACM Annual Symposium on User Interface Software and Technology*, pp. 161–170, 2002.

[129] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1994.

[130] J. O'Brien, T. Rodden, M. Rouncefield, and J. Hughes, "At home with the technology: An ethnographic study of a set-top-box trial," *ACM Transactions on Computer-Human Interaction*, vol. 6, no. 3, pp. 282–308, 1999.

[131] M. Obrist, E. Beck, S. Kepplinger, R. Bernhaupt, and M. Tscheligi, "Local communities: Back to fife (live) through IPTV," in *Proceedings of the European Conference on Interactive Television*, pp. 148–157, 2008.

[132] M. Obrist, R. Bernhaupt, and M. Tscheligi, "Interactive television for the home: An ethnographic study on users' requirements and experiences," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 174–196, 2008.

[133] K. O'Hara, A. S. Mitchell, and A. Vorbau, "Consuming video on mobile devices," in *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pp. 857–866, 2007.

[134] V. Oksman, V. Ollikainen, E. Noppari, C. Herrero, and A. Tammela, "'Podracing': Experimenting with mobile TV content consumption and delivery methods," *Multimedia Systems Journal*, vol. 14, no. 2, pp. 105–114, 2008.

[135] S. O'Modhrain and I. Oakley, "Touch TV: Adding feeling to broadcast media," in *Proceedings of the European Conference on Interactive Television: From Viewers to Actors?*, 2003.

[136] M. Oumard, D. Mirza, J. Kroy, and K. Chorianopoulos, "A cultural probes study on video sharing and social communication on the Internet," in *Proceedings of the International Conference on Digital Interactive Media in Entertainment and Arts*, pp. 142–148, 2008.

[137] H. M. Ozaktas and L. Onural, eds., *Three-Dimensional Television: Capture, Transmission, and Display*. Springer Verlag, 2007.

[138] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert, "The diver project: Interactive digital video repurposing," *IEEE MultiMedia*, vol. 11, no. 1, pp. 54–61, 2004.

[139] L. Pemberton, S. Fallahkhair, and J. Masthoff, "Learner centred development of cross platform language learning support system," *Journal of Educational Technology and Society*, vol. 8, no. 4, pp. 52–63, 2005.

[140] F. Pereira and T. Ebrahimi, *The MPEG-4 Book*. Prentice Hall, 2002.

[141] T. S. Perry, "The trials and travails of interactive TV," *IEEE Spectrum*, vol. 33, no. 4, pp. 22–28, 1996.

[142] E. M. Perse, "Media involvement and local news effects," *Journal of Broadcasting and Electronic Media*, vol. 34, no. 1, pp. 17–36, 1990.

[143] B. J. Pine and J. H. Gilmore, *The Experience Economy: Work is Theatre and Every Business a Stage*. Harvard Business School Press, 1999.

[144] F. Pittarello, "Time-Pillars: A 3D cooperative paradigm for the TV domain," in *Proceedings of the Workshop on Personalization in Future TV*, 2002.

[145] R. Pose, "Steerable interactive television: Virtual reality technology changes user interfaces of viewers and of program producers," in *Proceedings of the Australasian User Interface Conference*, pp. 77–84, 2001.

[146] R. Putnam, *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, 2001.

[147] R. Rafey et al., "Enabling custom enhancements in digital sports broadcasts," in *Proceedings of the International Conference on 3D Web Technology: Virtual Reality Modeling Language Symposium*, pp. 101–107, 2001.

[148] G. Ramos and R. Balakrishnan, "Fluid interaction techniques for the control and annotation of digital video," in *Proceedings of the Symposium on User Interface Software and Technology*, pp. 105–114, 2003.

[149] T. Regan and I. Todd, "Media center buddies: Instant messaging around a media center," in *Proceedings of the Nordic Conference on Human-Computer Interaction*, pp. 141–144, 2004.

[150] U. Reimers, "Digital television broadcasting," *IEEE Communications Magazine*, vol. 36, no. 1, pp. 104–110, 1998.

[151] U. Reimers, *DVB: The Family of International Standards for Digital Video Broadcasting*. Springer-Verlag, 2005.

[152] M. Ressin and C. Haffner, "Human-Centered design of interactive TV games with SMS backchannel," in *Proceedings of the European Conference on Interactive Television*, pp. 37–46, 2007.

[153] G. L. Revelle, "Educating via entertainment media: The sesame workshop approach," *Computers in Entertainment*, vol. 1, no. 1, article number 7, 2003.

[154] M. Rey-López, R. P. D. Redondo, A. F. Vilas, and J. J. P. Arias, "Entercation: Engaging viewers in education through TV," *ACM Computers in Entertainment*, vol. 5, no. 2, article number 7, 2007.

[155] M. Rice and N. Alm, "Designing new interfaces for digital interactive television usable by older adults," *ACM Computers in Entertainment*, vol. 6, no. 1, article number 6, 2008.

[156] S. Robertson, C. Wharton, C. Ashworth, and M. Franzke, "Dual device user interface design: PDAs and interactive television," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 79–86, 1996.

[157] M. Robin and M. Poulin, *Digital Television Fundamentals: Design and Installation of Video and Audio Systems*. McGraw-Hill, 1998.

[158] A. C. Roibás and S. Johnson, "Unfolding the user experience in new scenarios of pervasive interactive TV," in *Extended Abstracts on Human Factors in Computing Systems*, pp. 1259–1264, 2006.

[159] A. M. Rubin, "Television uses and gratifications: The interaction of viewing patterns and motivations," *Journal of Broadcasting and Electronic Media*, vol. 27, no. 1, pp. 37–51, 1983.

[160] A. M. Rubin, "Ritualized and instrumental television viewing," *Journal of Communication*, vol. 34, no. 3, pp. 67–77, 1984.

[161] C. P. Sandbank, "Digital television in the convergent environment," *IEEE Computer Graphics and Applications*, vol. 21, no. 1, pp. 32–36, 2001.

[162] R. Schatz, S. Wagner, and N. Jordan, "Mobile social TV: Extending DVB-H services with P2P-interaction," in *Proceedings of the International Conference on Digital Telecommunications*, pp. 14–19, 2007.

[163] T. Schiphorst, F. Nack, M. KauwATjoe, S. de Bakker, L. Stock Aroyo, A. P. Rosillio, H. Schut, and N. Jaffe, "PillowTalk: Can we afford intimacy?," in *Proceedings of the International Conference on Tangible and Embedded Interaction*, pp. 23–30, 2007.

[164] D. A. Shamma, M. Bastea-Forte, N. Joubert, and Y. Liu, "Enhancing online personal connections through the synchronized sharing of online video," in *Extended Abstracts on Human Factors in Computing Systems*, pp. 2931–2936, 2008.

[165] D. A. Shamma, R. Shaw, P. L. Shafton, and Y. Liu, "Watch what I watch: Using community activity to understand content," in *Proceedings of the International Workshop on Multimedia Information Retrieval*, pp. 275–284, 2007.

[166] R. Shaw and P. Schmitz, "Community annotation and remix: A research platform and pilot deployment," in *Proceedings of the ACM International Workshop on Human Centered Multimedia*, pp. 89–98, 2006.

[167] J. A. Short, E. Williams, and B. Christie, *The Social Psychology of Telecommunications*. New York: John Wiley & Sons, 1976.

[168] T. Shrimpton-Smith, B. Zaman, and D. Geerts, "Coupling the users: The benefits of paired user testing for iDTV," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 197–213, 2008.
[169] R. Silverstone, *Television and Everyday Life*. Routledge, 1994.
[170] B. Smyth and P. Cotter, "A personalized television listings service," *Communications of the ACM*, vol. 43, no. 8, pp. 107–111, 2000.
[171] L. F. G. Soares and G. L. S. Filho, "Interactive television in Brazil: System Software and the digital divide," in *Adjunct Proceedings of the European Conference on Interactive Television*, pp. 41–43, 2007.
[172] L. F. G. Soares, R. F. Rodrigues, and M. F. Moreno, "Ginga-NCL: The declarative environment of the Brazilian digital TV system," *Journal of the Brazilian Computer Society*, vol. 13, pp. 37–46, 2007.
[173] C. Södergard, ed., Mobile television — technology and user experiences. Report on the Mobile-TV project. *VTT Information technology*. 2003.
[174] S. Sperring and T. Strandvall, "Viewers' experiences of a TV quiz show with integrated interactivity," *International Journal of Human-Computer Interaction*, vol. 24, no. 2, pp. 214–235, 2008.
[175] A. Taylor and R. Harper, "Age-Old practices in the 'New World': A study of gift-giving between teenage mobile phone users," in *Proceedings of the Conference on Human Factors and Computing Systems*, pp. 439–446, 2002.
[176] E. Tsekleves, R. Whitham, K. Kondo, and A. Hill, "Bringing the television experience to other media in the home: An ethnographic study," in *Proceedings of the European Conference on European interactive Television*, pp. 201–210, 2009.
[177] A. Ugarte, I. García, A. Ortiz, and D. Oyarzun, "User interfaces based on 3D avatars for interactive television," in *Proceedings of the European Conference on Interactive Television*, pp. 107–116, 2007.
[178] M. F. Ursu, I. A. Kegel, D. Williams, M. Thomas, H. Mayer, V. Zsombori, M. L. Tuomola, H. Larsson, and J. Wyver, "ShapeShifting TV: Interactive screen media narratives," *Multimedia Systems Journal*, vol. 14, no. 2, pp. 115–132, 2008.
[179] M. F. Ursu, M. Thomas, I. Kegel, D. Williams, M. Tuomola, I. Lindstedt, T. Wright, A. Leurdijk, V. Zsombori, J. Sussner, U. Maystream, and N. Hall, "Interactive TV narratives: Opportunities, progress and challenges," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 4, no. 4, article 25, 2008.
[180] J. van Barneveld and M. van Setten, "Involving users in the design of user interfaces for TV recommender systems," in *Proceedings of the Workshop on Personalization in Future TV*, 2003.
[181] E. S. Veinott, J. Olson, G. M. Olson, and X. Fu, "Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 302–309, 1999.
[182] P. Vorderer, "Interactive entertainment and beyond," in *Media Entertainment: The Psychology of Its Appeal*, (D. Zillmann and P. Vorderer, eds.), pp. 21–36, Lawrence Erlbaum Associates, 2000.

[183] P. Vorderer, S. Knobloch, and H. Schramm, "Does entertainment suffer from interactivity? The impact of watching an interactive TV movie on viewers' experience of entertainment," *Media Psychology*, vol. 3, no. 4, pp. 343–363, 2001.

[184] H. D. Wactlar, M. G. Christel, Y. Gong, and A. G. Hauptmann, "Lessons learned from building a terabyte digital video library," *Computer*, vol. 32, no. 2, pp. 66–73, 1999.

[185] J. Wang, J. Pouwelse, J. Fokker, and M. J. T. Reinders, "Personalization of a peer-to-peer television system," in *Proceedings of the European Conference on Interactive Television*, pp. 147–155, 2006.

[186] J. D. Weisz and S. Kiesler, "How text and audio chat change the online video experience," in *Proceeding of the International Conference on Designing Interactive User Experiences for TV and Video*, pp. 9–18, 2008.

[187] J. D. Weisz, S. Kiesler, H. Zhang, Y. Ren, R. E. Kraut, and J. A. Konstan, "Watching together: Integrating text chat with video," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 877–886, 2007.

[188] A. R. Wellens, "An interactive television laboratory for the study of social interaction," *Journal of Nonverbal Behavior*, vol. 4, no. 2, pp. 119–122, 1979.

[189] D. Zillmann, "The coming of media entertainment," in *Media Entertainment: The Psychology of Its Appeal*, (D. Zillmann and P. Vorderer, eds.), pp. 1–20, Lawrence Erlbaum Associates, 2000.

[190] D. Zillmann and J. Bryant, *Selective Exposure to Communication*. Lawrence Erlbaum Associates Hillsdale, 1985.