Leveraging the user impact

Cesar, P.; Bulterman, D.C.A.; Jansen, J.

published in
Multimedia Systems
2009

DOI (link to publisher)
10.1007/s00530-009-0159-z

document version
Publisher's PDF, also known as Version of record

Multimedia Systems (2009) 15:127–142
DOI 10.1007/s00530-009-0159-z

REGULAR PAPER

Leveraging user impact: an architecture for secondary screens usage in interactive television

Pablo Cesar · Dick C. A. Bulterman · Jack Jansen

Published online: 17 June 2009

Abstract This paper reports on an architecture, and a working implementation, for using secondary screens in the interactive television environment. While there are specific genres and programs that immerse the viewer into the television experience, there are situations in which people perform as well a secondary task, whilst watching. In the living room, people surf the web, use email, and chat using one or many secondary screens. Instead of focusing on unrelated activities to television watching, the architecture presented in this paper aims at related activities, i.e., to leverage the user impact on the content being watched. After a comprehensive literature review and working systems analysis, the requirements for the secondary screen architecture are identified and modelled in the form of a taxonomy. The taxonomy is divided into three high-level categories: control, enrich, and share content. By control we refer to the decision what to consume and where to render it. In addition, the viewer can use the secondary screen for enriching media content and for sharing the enriched material. The architecture is validated based on the taxonomy and by an inspection of the available services. The final intention of our work is to leverage the viewers' control over the consumed content in our multi-person, multi-device living rooms.

P. Cesar (✉) · D. C. A. Bulterman · J. Jansen
CWI: Centrum voor Wiskunde en Informatica,
Science Park 123, 1098 XG Amsterdam, The Netherlands
e-mail: p.s.cesar@cwi.nl

D. C. A. Bulterman
e-mail: dick.bulterman@cwi.nl

J. Jansen
e-mail: jack.jansen@cwi.nl

1 Introduction

The living room is a shared space, in which families carry out their daily routines. Families gather around the television for entertainment, for information, as a conversation starter, or just to be together. The proliferation of personal devices, such as handhelds or laptops, has modified traditional family behaviours towards the television.

In previous research it has been reported that viewers "frequently use their laptop to surf the web, use email, or shop online whilst watching television" [23]. We can classify this kind of activities as incidental, seldom related to the content being watched. However, our research goal is on activities related to the content that serve to leverage end user control over the content being watched. This article focuses on how the viewer can use a variety of digital devices to enrich the viewing experience, either alone or together with others. Digital devices acting as secondary screens can provide the viewer with extra information about the current show or an actor. They can allow the user to browse recorded content for later playback, without disturbing the television experience. And they can empower the viewer with lightweight authoring functionalities.

This article introduces a taxonomy that captures the way in which viewers can interact with television content. The taxonomy includes three super-classes: content control, content authoring, and content sharing. Moreover, we provide a detailed architecture, for television content consumption in a multi-user, multi-device environment. The major innovations of the proposed architecture reside in (1) the possibility of using secondary screens for affecting the television content while watching and (2) the utilization of high-level description languages for empowering the viewer with finer-grained control over such content.

Fig. 1 Mark and Maria at Home [in this work we propose to use a secondary screen as shown in the *right* side instead of (or in addition to) the traditional remote control shown in the *left* side]

This article will elaborate on these topics and will report on our experiences implementing the underlying architecture. Moreover, the initial results of a business analysis and a summary of a user evaluation are provided. Our paper is structured as follows. Section 2 provides a real world scenario that exemplifies the architecture and implementation reported in this article. Section 3 provides a detailed description of the taxonomy for secondary screen usages in an interactive television environment. Section 4 reviews the related work and validates the taxonomy introduced in the previous section. Section 5 introduces the contribution of this article. Section 6 presents the architecture and the design goals. Section 7 presents the initial results in terms of implemented services, while Sect. 8 introduces the initial business analysis and user evaluation results.

2 Scenario: potential user behaviour

In order to introduce the ideas presented in this article, this section presents an example of secondary screens usage together with an analysis of the benefits over traditional interactive digital television systems.

Mark, a USA national, his wife Maria, and the kids are watching television at their home in the Netherlands. Mark has a personal device (e.g., a mobile phone) that he uses as an extended remote control. As it has been shown in previous research, in specific situations people tend to multitask whilst watching television by using laptops, for example, to check emails or to obtain extra information about television programs. Mark can navigate through personal media content based on his preferences or can access to extra information about the current television program that might not be of interest to the rest of the family.

The personal device can inform him when extra (personalized) clips have been detected. The personal content might include instant translation of sentences he might not yet understand in Dutch, personalized commercials, or extra features extracted from web services. In this case, the extra material and information is rendered in the personal device, instead of as overlays in the main screen.

Finally, the system provides functionality for connecting Mark, and the rest of the family, to friends in different locations. For example, at some moment during the show Mark can quickly annotate and enrich the television show, which then can be shared with his brother living in the USA using a P2P network. Such enrichments are done by using the personal device. Figure 1 illustrates the scenario.

One of the differentiating characteristics of the work presented in this article is that secondary screens are used to actively affect the television content *while* watching.

We do not claim that by using a secondary screen the family shared experience is enriched—Mark could as well ask Maria about the meaning of a sentence he has not fully understood and thus start a conversation. We only recognize the diversity of attitudes and level of engagement of the viewers in front of the television. And we argue that secondary screens provide a less obtrusive mechanism for affecting television content than traditional solutions in the form of television overlays. We agree with O'Hara et al. [20] that shared experience can take place in different forms, sometimes it is about sharing an activity, while other times it is about sharing a common space—in our case the living room.

At the same time, we are aware that currently households have more than one television display, which in some cases moves the watching experience from the living room to the bedroom or to the kitchen. Nevertheless, we can argue that still the living room remains the focal social point of the household.[1]

3 Taxonomy: control, enrich, and share

The scenario sketched above determines a number of functionalities viewers can perform, when consuming television content. We can structure them into a formal framework [5]: control, enrich, and share. Table 1 introduces the taxonomy that models the user intentions in front of the television. In addition, the table includes the most relevant subclasses of each category.

[1] As demonstrated in the Ta2 project: http://www.ta2-project.eu/.

Table 1 Taxonomy of the different usages of the secondary screen in the television environment

	Definition	Subclasses
Content control	To decide what and how to consume television content	Select content
		Browse content
		Manage content
		Select path
		View extra material
		Video controls
		Session transfer
Content enrich	To actively manipulate the television content	Create content
		Annotate content
		Fragment content
		Enrich content
Content sharing	To socially communicate with others	Share personal content
		Share media fragments

We assume that content may be provided via different paths. This paper will not look into detail the actual protocol or transmission technology used, but on the potential impact the user has over the content. There are a number of end user terminals surrounding the user with a diverse set of rendering and interaction capabilities. Some of the devices can act as primary screens, while others can act as secondary screens. On a secondary screen the end user, apart of viewing the content, can affect the content in a number of ways. As shown in Table 1, we can differentiate three types of activities.

Content control empowers the viewer to decide what to watch and how/where to render it. Some subclasses of this category include the provision of traditional player control functionality (e.g., play/pause), the possibility of transferring a running session from one device to another (e.g., from the television to a mobile phone). In addition, content selection, management, and browsing can be considered as instances of controlling content. Moreover, form filling for selecting a navigation path (e.g., to fill an exam for a Television learning, T-Learning, service) is another way of content control. Finally, the software media player might be capable of differentiating between shared data and private data and thus providing multimedia content to the television screen and private extra information (or tailored commercials) to the user's handheld device.

Content enrichment empowers the viewer to manipulate multimedia content. There are a number of sub-classes within this category such as fragmenting, or clipping, television content, generating new content, and enriching or annotating existing material. By content fragmentation we refer to the capability of identifying a specific portion of a television program. Enrichment, on the other hand, is the process of aggregating extra material, such as an audio commentary, to the content. Finally, the user might want to annotate the television content, by for example adding extra information, or metadata, about a program being watched (e.g., actors, director, rating).

Content sharing empowers the viewer to share television content. The viewer might want to share a whole program, a fragment of it, additional annotations, or enrichments over television program with his social network.

The final goal of our work is to empower the user with a varied array of interactive capabilities over the consumed content, while watching. The major benefit of the proposed taxonomy is that it provides us a number of high-level functional requirements for architecting a solution. If we can demonstrate that such requirements cover the major usages of secondary screens, we can then justify that the architected solution is useful and can be used for different particular services. The following section provides a formal validation of this framework in the form of a literature review.

4 Related work

The intention of this section is to validate the taxonomy (control, enrich, share) identified in the previous section. We will first describe a number of systems and proposals for secondary screen usages in the interactive television environment. Then, we will classify each of the approaches using our categorization and demonstrate that all the use cases can be placed within one super-class.

First of all, we should highlight that lately several studies [3,9,23] suggest that more capable remote controls will help the viewers become more active in the interactive television environment than the traditional *lean back* view. Firstly, Bernhaupt et al. [3] found that remote controls are often considered as unusable; moreover the results indicate that

the rendering capabilities of the remote control could be exploited. Cruickshank et al. [9] concluded that "a more sophisticated form of input and control needs to be introduced for iTV to reach its full potential". While Seager et al. [23] say that "there was a preference for accessing different services on different display panels rather than overloading one shared display channel". Based on these results, even though we do not think that the "television + beer" viewing experience will ever disappear, appropriate tools for active viewer involvement with the watched content should be provided.

4.1 Secondary screens usages: services

We cannot claim that the idea of using a secondary screen in the television environment is new. Back in 1996 Roberston et al. [22] presented a system where handheld devices are used for interacting with the television. Such system was used in an application for house hunting. A floor map of the house was provided in the handheld device. Then, the user could select a room and activate a movie of the room in the television set. More recently, as interactive television systems and handheld devices are becoming more popular a body of research is emerging about the usage of secondary screens.

Two specific areas of research that can take advantage of the secondary screen are T-Learning and content selection. As presented by Fallahkhair et al. [12], non-desktop technologies fit learning activities. The authors use the secondary screen for a number of scenarios such as to provide help for difficult cultural or language items, to provide extra information about specific concepts, and to manage the personal learning sphere.

Other active research field for the secondary screen is the electronic program guide (EPG). Park et al. [21] provided rough guidelines about a number of foreseen services, while Cruickshank et al. [9] presented a detailed study on system design and implementation, together with a comprehensive user study. They reported on a working system in which a PDA is used for displaying a personalized EPG. Moreover, the PDA provided functionality such as volume controls and channel navigation. In terms of content management, Karanastasi et al. [14] provided a solution for a ubiquitous personal video recording interface in handheld devices. Their system was capable of recording, deleting, and summarizing recorded television content. Other specific services explored for secondary screens include advertisement and commerce [10], and viewer participation. Davis and Yung et al. [10] reported on systems in which the viewer can use his mobile phone for sending messages that will late appear as an overlay of the television content as a broadcast chat and Miller [19] discussed how the end-user can become a participant in the show.

Apart from enhanced information rendering, content management and communication, handheld devices can be used as a primary screen for content viewing. One relevant scenario is session transfer. The viewer is a mobile being, thus when leaving the place in which the television screen is located, he might want the media content to follow him. Previous work in this subject is mainly focused on the enabling technology. While most of the research around mobile television studies technical factors or the perceived quality (e.g., [17]), there are also studies that focus on the current usages of mobile devices. For example, O'Hara et al. [20] provided an extensive study on how and why video material is consumed in mobile devices. They concluded that even though consuming video in mobile devices is a privatizing technology, it might facilitate togetherness in the home as people can watch "their own content while being in proximity to family".

Finally, people use mobile phones for capturing and sharing images. There have been several studies on the motives of people to share media material. Taylor and Harper [25] observed the social norms around gift-giving and the demand of reciprocity with mobile phones. Another study categorized the uses of photo sharing in two dimensions: social/individual and functional/affective [16]. House et al. [13] provided a similar categorization: creating and maintaining social relationships, personal and group memory, self expression, self presentation, and functional. In most of the cases, the usage was mainly incidental and not a planned activity.

4.2 Taxonomy validation

The next step, after an exhaustive description of related work and systems is to validate the taxonomy introduced in Table 1. Table 2 associates the functionality described above to the high-level category it belongs to (control, enrich, share).

Within the control behaviour we can position content selection services, such as the ones proposed by Cruickshank et al. [9] and Karanastasi et al. [14]. While the first one focused on the implementation of an EPG for handheld devices, the second one reported on a mobile-based content management system. Robertson et al. provided a control service, in which extra material that helps the user finding a house was shown in the television screen. Finally, Mate et al. [18] concentrated on session transfer, where the secondary screen becomes the primary screen, when the user moves away from the television screen.

In terms of content enrichment, we can identify the work of Fallahkhair et al. [12] and Davis and Yung [10]. In the first case, apart from using the mobile phone for rendering extra information that helps the user learning a language, the authors provided a system to create their own vocabulary while watching television. Davis and Yung considerd scenarios such as the broadcasting of viewers' text messages.

Table 2 Mapping of high-level usage categories and related work

	Service	Main usage
Cruickshank et al. [9]	Content selection	Control
Karanastasi et al. [14]	Content manager	Control
Robertson et al. [22]	Extra material	Control
Mate et al. [18]	Session transfer	Control
Fallahkhair et al. [12]	Extra material	Control
	Content manger	
Fallahkhair et al. [12]	Content annotation	Author
Davis and Yung [10]	User participation	Author
Kindberg et al. [16]	Content capture	Author
		Share
House et al. [13]	Content capture	Author
		Share
O'Hara et al. [20]	Content viewing	Control
		Share

Finally, in terms of content sharing we find the work from Kindberg et al. [16], House et al. [13], and O'Hara et al. [20]. In the first two cases, mobile phones are used to capture a specific moment of the user life. Such captured content can then be shared with friends and family by using the phone connectivity (e.g., as an MMS). Capture and sharing in both cases are done as an unintended activity. Even though their research does not consider television content, there is a clear analogy to capture television content as a non-planned activity [6]. Finally, the exhaustive research work from O'Hara et al. indicated the importance of the mobile phone for content sharing.

The intention of this section was to validate, through an exhaustive survey of previous research and systems, the proposed taxonomy for secondary screen usage. Thus, we can conclude that the taxonomy can be used as key functional requirements for the design and implementation of an usable architecture. Such architecture should be capable of supporting controlling, enriching, and sharing functionalities in order to be useful for different specific scenarios.

5 Contribution

The contribution of this article is the provision of a detailed architecture for secondary screen usage in an interactive television environment. The special characteristics of such a connected home are twofold. Firstly, it leverages the viewer impact on the television content—thus we do not focus on content unrelated activities as other previous works. Secondly, the viewer is capable of affecting the content while watching it—thus the system is not limited to content management activities or simple EPG functionalities.

The functional requirements of the architecture have been gathered based on a high-level taxonomy of viewer intentions toward television content. In order to ensure the validity of the requirements an exhaustive analysis of related work and systems is provided. The architecture is used for constructing a working system that then is validated. The validation of the system is done using three different mechanisms. First of all, we prove that the system complies with the viewers' requirements as identified in the taxonomy. Secondly, a market analysis on the proposed solution was conducted in order to understand the business opportunities. Finally, we provide a summary of the results of a controlled user study of our system. This article constitutes a step forward in empowering the viewer with interactive tools that provide a higher-degree of control over consumed television content.

6 Architecture

Our research studies new paradigms for interacting with television content that is available to social groups. Our home architecture consists of a home media server which might be implemented in a set-top box, a home networking gateway or a separate server device that stores content that is provided via a number of delivery channels. Such channels include standard broadcast, peer-to-peer content sharing networks, and high-density optical disks such as DVDs or Blu-Ray HD. The first assumption is that the television content will be fetched on the viewers' behalf using an intelligent recommender system [1,2], and may be post-processed in the home to allow differentiated viewing based on the individual interests of family members. The contribution of this paper focuses on the consumption stage, and not so much on the different group recommendation algorithms used for content selectivity.

The distinguishing characteristic of this architecture is that multiple remote control devices are presented to the home users. The devices, connected and synchronized with the home media server, range from a conventional remote control, through low-powered handheld devices like telephones and minimal pen-based devices such as the Nokia 770, up to full-featured (but reduced size) tablets such as the ultra-mobile tablet PC Samsung Q1. In the home environment the devices are connected using a wireless network, while when transferring a session to a mobile phone the mobile network is used to retrieve the media content. The personal remote devices provide the viewer with differentiated content delivery and differentiated personal recommendation delivery and generation. A schematic diagram of the home environment is provided in Fig. 2.

As shown in the figure, there are a number of innovative features provided by the architecture. Firstly, we provide a distributed architecture in which different devices can be

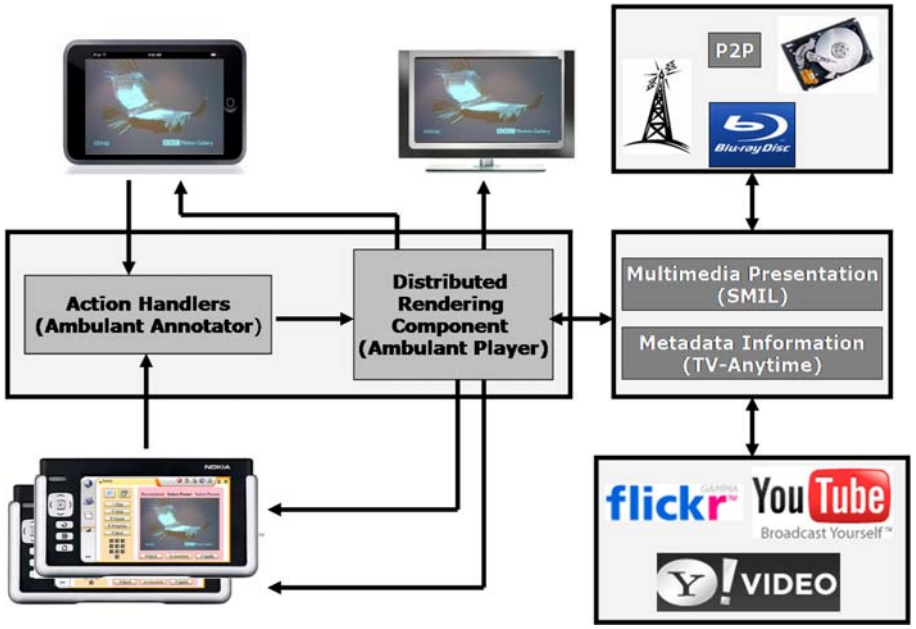

Fig. 2 High-level architecture of the home media server

used for rendering and/or controlling content. Secondly, we provide a rich content modelling format, in our case Multimedia Integration Language (SMIL)[2] and TV-Anytime,[3] that can semantically encode the content, parts of the content, extra information related to the content, and more complex data such as the nature of the content (e.g., if the content is sharable for all the viewers or it is intended for one specific viewer). Finally, the architecture supports to gathering content from different sources and to share it using different channels.

At a high-level, the home media server can be divided into two differentiated components: (1) the Rendering Component and (2) Action Handlers. The Rendering Component is in charge of displaying the content, while the Action Handlers expose media-based interaction functionality to other devices (e.g., handheld device). Such separation of viewing versus interacting is essential for providing secondary screen functionality.

Our concern was not to study the commercial models for home media storage, but to study a model in which multiple control clients could be managed in a home environment. For this reason, we made the pragmatic decision to use a small size personal computer (in our case, a Mac-Mini) upon which our server infrastructure could be implemented.

This section elaborates on how the proposed architecture was transformed into a working system and how it meets the requirements identified by the proposed taxonomy. Firstly, some implementation details of the home server architecture are provided. Secondly, we report on the different secondary screens and their interfaces. Finally, the content modelling choices and its benefits are discussed.

6.1 Implementation details

The home server architecture is an extension of traditional television receivers, which allows the viewer with richer interactive functionality in the form of controlling, enriching, and sharing recorded television content. It is implemented as an extended Personal Digital Recorder (PDR+). As indicated earlier the PDR+ has access to the conventional broadcast content, to high-density disks, and to television content located elsewhere in the web. In addition, a number of personal devices at home can be connected to the PDR+ via the home network. These devices include a conventional television set, the remote control, and personal devices. The conventional television set acts as a shared rendering component.

The functional requirements identified for such architecture are to provide tools, or interfaces, for controlling, enriching, and sharing content. This subsection focuses on the home media server, while more information on personal devices can be found later in the article.

Figure 3 introduces the low-level architectural details of the home media server. We can identify a number of components: Presentation, User Interaction, Control, Lightweight Authoring, and Sharing module.

The Presentation Module is in charge of rendering the content. One important differentiating factor is that the input to this module is a high-level description document that describes the content to be displayed. Thus, sub-components

[2] http://www.w3.org/AudioVideo/.

[3] http://www.tv-anytime.org/.

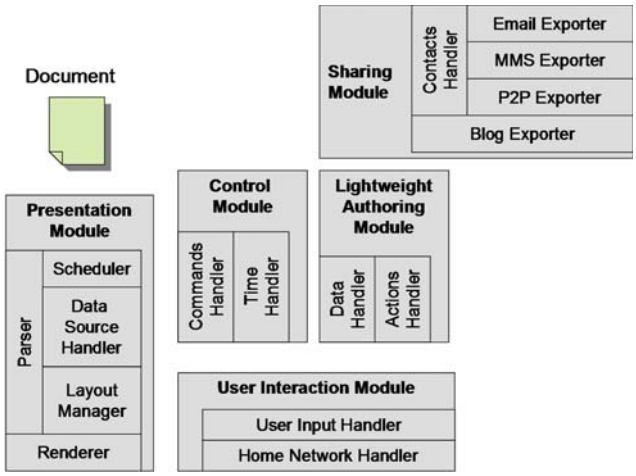

Fig. 3 Low-level architecture of the home media server

such as a scheduler, a parser, a layout manager, and data source managers (to handle different content gathered from different sources) are included. In our working system this component is the open source Ambulant Player. It is implemented in C++ and provides complete support for Synchronized SMIL [4] language. It is a multi-platform player that runs on Linux, Macintosh, Windows, and various handhelds. The Presentation Module is connected to the television screen for rendering purposes, but can handle rich multimedia presentation formats for an enriched user experience. Even though a software media player is used to manage the presentation of enriched content, the user is presented with a conventional television-centric view of the content. While watching television, the viewer can affect the television content in two differentiated forms: content control and content enrichment.

The Control Module takes care of content control. Content control is a basic viewing functionality that allows the user to navigate through a set of content objects (and where appropriate, within content objects) to find and activate particular content sequences of interest—including extra material. The rendering and control is provided through the usage of secondary screens, instead of using television overlays that might disturb the rest of the viewers in front of the television display. Apart from content selection, management, navigation, and activation, the Control Module provides session transfer capabilities. That is, functionality for allowing the movable viewer to continue watching television content in his personal device. In this case, the current state of the content stream, as kept by the Presentation Module, should be stored, the presentation should be streamed to his personal device, and the presentation should be restored there. This process should be as seamless and dynamic as possible. It is important to notice that all the functionality provided by the home media server is time-aware.

The Lightweight Authoring Module handles content enrichment, i.e., viewers' manipulations over the consumed content. Such manipulations can consist of content annotation, new content generation, content fragmentation, or actual enriching of existing material. Thus, the viewer can explicitly fragment the television stream and save the fragment separately from the base content. He can as well enrich the television fragment with user-generated overlays. We name the modules as Lightweight Authoring, since the user actions are incidental in nature taking place as the content is watched. This module is connected to the Sharing Module, which is used for sharing the enrichments with others. The viewer can use an array of channels for sharing content such as the Internet for uploading content into a blog or for sending email messages. In addition, it includes a SMS generator, specifically Wireless Application Protocol (WAP) Push functionality, for recommending enriched television content to mobile phones for people on the move.

The modules described above (Control, Lightweight Authoring, and Sharing) are the Action Handlers shown in Fig. 2. Together their implementation forms the Ambulant Annotator, which is implemented in Python. This component is an extension to the Ambulant Player that acts as an extended DOM interface to and from the player. The Ambulant Annotator can handle the actions requested by the interactive devices and affect the player's behaviour, while it is rendering content.

The functionality supported by the Ambulant Annotator is exposed to the external world using well-define interfaces. For implementation convenience such interface is provided as XML-Remote Procedure Call (XML-RPC) and as HTTP-Requests. This paper focuses on the actual actions provided to the user while viewing multimedia content, and not so much on the underlying home network infrastructure. In Fig. 3, the interfaces to the external world are provided by the User Interaction Module. The User Interaction Handler provides the actual interfaces, while the Network Handler provides the connection to the different extended remote controls in the living room.

As can be seen in Figs. 2 and 3, we provide a clear separation between the action handling, incorporated in the home media server, and the user interfaces. The major advantage of such separation is that any client can implement the user interface in the most convenient manner, that is, natively. As a result, our architecture provides a universal remote plug-in infrastructure.

6.2 Component integration and interfaces

A key requirement for our system is to dynamically distribute media content rendering and media control capabilities to the most suitable device(s) surrounding the user. At some times,

Table 3 Categorization of the client devices

	TV	Remote control	Apple iPod	Nokia N770	Samsung Q1	PC
						
Social experience	Shared	Private	Private	Private	Private	Private
Viewing/feedback experience	HD screen	None (on-screen)	Small screen	Small screen	Medium screen	Big screen
Interaction role	Lean back	Control	Control enrich share	Control enrich share	Control enrich share	Multi purpose
Role	Primary screen	None	Secondary screen	Secondary screen	Secondary screen	Primary screen

the user might want to watch video material on his television screen. At other moments, he might want to use a handheld device to watch extra information of a show being rendered in the television display. But at other moments, for example when on the move, the user might want to render video clips in his/her mobile device. Similar requirements can be found in current research literature. It is clear that consumers are becoming mobile [18] and that there is a need to better integrate the devices surrounding the user [11]. Our architecture solves all the mentioned requirements.

As indicated before, our working implementation is provided based on a combination of the Ambulant Player (rendering component), the Ambulant Annotator (action handler), and a well defined interface that provides user interaction, enrichment, and sharing functionality. The different clients are connected to the home media server and request an action using the appropriate exposed features. Thus, each client allows the personalized user experience in the living room, while the shared content is being displayed on the television. The functionality provided by each client device will vary with its complexity and richness. Each of the devices allows for content selection and most non-trivial devices allow for direct navigation and recommendation management.

A more general classification on devices that can be used in the interactive television environment was documented in a previous article [7]. Table 3 illustrates the characteristics of the digital devices used in our systems. The table differentiates between shared and private devices and elaborates on the different potential usages of each of the devices. For example, the television is a shared device with high rendering capabilities, intended for entertainment and normally used as a primary screen. The iPod is a private device with a small screen that can be used as a secondary screen while watching television. The Nokia 770 is a tablet-based device similar to the iPod, in which 2D gestures can be used for interaction, while the Samsung Q1 is an ultra-mobile PC with bigger screen and higher computing power.

While, in the past research on adaptive multimedia concentrated on the provision of the same service to different devices, we concentrate on the usage of secondary devices that increase the viewer capabilities over the content. Thus, the problem is not about content adaptation per se, but task to device(s) mapping. In addition, the "Interaction Role" includes how the increasing capabilities look like. In terms of viewing experience, it is clear that the combination of remote control and television force the service developer to utilize video overlays for providing extra functionality—like for example in the DVDs. Our solution, on the other hand, provides rendering capabilities on the client side, thus minimizing the obtrusiveness of the interfaces. Finally, based on the nature of the content, as encoded in the high-level television description language, our system is capable of differentiating between shared and private content—as well represented in the table.

Based on the taxonomy presented earlier, the client devices implement a user interface for the following end-user actions:

Content control

Content interrogation extra information about the current displayed program (or fragment), which can be rendered on the secondary screen. It allows for personal interrogation without bothering others.

Personal browsing EPG-like functionality that can be rendered on the local display, allowing personal browsing without endangering family relationships.

Player control main screen content can be selected and controlled (e.g., pause/play) via a conventional multi-button interface.

Session transfer seamless transfer of the current content to a secondary screen, thus becoming a personal primary screen.

Content management control over the content stored in the home media server, with functionality such as remove, rate, or activate specific content or parts of it.

Fig. 4 Client interface (Nokia 770)

Content enrichment

Fragmentation this is the ability to fragment a television program, while watching, by using the client interface.

Enrichment to enrich an existing television program, or fragments of the program.

Annotation to add personal extra information, or metadata, to a television program, while watching.

Content sharing

Content sharing to share television programs or the result of the enrichment functionality with others.

The look and feel for the user interface will vary between client devices. The implementation done for the Nokia 770 device, for example, is depicted in Fig. 4. The left side of the interface contains the player control area that includes traditional video player controls (e.g., play/pause). On the left-top corner the user can use two buttons. The left one is used for current content interrogation, while the right one is used for fragmenting the current viewed content. The right side of the interface contains the content rendering region. It can be used for personal browsing and viewing of content, and for advanced enriching and sharing of the television content (e.g., ink overlays). Currently, it shows how the viewer can personally navigate within content stored in the PDR+.

6.3 Content modelling

While the previous subsection presented the detailed architecture and implementation of our system, this subsection will investigate how content is modelled as a rich description format. In our research we propose modelling the content using a combination of SMIL and TV-Anytime metadata descriptions. This provides innovative features such as the possibility of providing semantic information over the content—and most importantly, fragments of it—, temporal control over the content being watched, manipulation of the content without modifying the base content, and the possibility for quite and clean layout transformation for session transfer. Source Code 1 shows a snapshot of the description of recorded television program.

```
… <!-- Part 1 -->
 <metadata xml:id="meta-rdf">
  <rdf:RDF
   xmlns:rdf =http://www.w3.org/1999/02/22-rdf-syntax-ns#
   xmlns:rdfs = http://www.w3.org/2000/01/rdf-schema#
   xmlns:tva="urn:tva:metadata:2005"
   xmlns:tva2="urn:tva:metadata:extended:2005">
  <!-- Metadata about the video -->
  <rdf:Description about="crid://.../video.mp4" />
  …
  </rdf:Description>
 </rdf:RDF>
</metadata>

… <!-- Part 2 -->
<layout>
 <region id="v" width="1024" height="800"/>
</layout>

… <!-- Part 3 -->
 <seq>
<!-- Fragment 1-->
<video region="v" src=".../video.mp4" clipEnd=”30s”>
 <metadata xml:id="meta-rdf">
    <!-- Metadata Fragment 1 -->
    <tva2:Item … >
    …
    </tva2:Item>
  </metadata>
</video>
<!-- Fragment 2 -->
<video region="v" src=".../video.mp4" clipBegin=”30s”
clipEnd=”60s”>
 …
</video>
…
  </seq>
```

Source Code 1: Snapshot of a recorded television program

The XML fragment presented above describes a television program fetched by the home media server. The first *metadata* section (Part 1) provides general information about the video program using the TV-Anytime standard. In addition, the TV-Anytime universal identifier, or CRID, is included for video location purposes. Such identifier can be mapped to different versions of the video residing in different locations (e.g., BluRay disk or content provider's video server). Secondly, a description of the *layout* of the service is provided. The layout information (Part 2) indicates regions on the screen for displaying the media content, in our case a

television screen with a resolution of 1,024 × 800. Finally, the synchronization information (Part 3) describes the video content. Unlike previous approaches, the actual television program is fragmented in different segments of 30 s each. The video part of the program is intended for the television screen, thus the region *v* is associated to it. Information about the duration of each segment is provided using the *clipBegin* and *clipEnd* attributes. Moreover, the source code includes *metadata* information for each fragment. Such metadata is composed of a poster, a relevant image summarizing the contents of the fragment, and extra information related to it. The posters, and extra information, can be defined by the content provider prior to delivery, by an intermediary recommender system or an automatic segmentation tool, or directly by the user.

There are a number of benefits of modelling television content as explained above. Firstly, the television content is not composed of a monolithic video stream to be rendered by the television set; instead the content is fragmented into different parts, thus providing fined grained selectivity and activation features. Secondly, we use the new version of the SMIL standard (3.0), in which metadata information can be included not only in the *head* of the document, but in any other location as well. In our system, extra metadata, or private information such as a poster, is associated to a fragment, thus providing richer semantics to the content. As we shall see later, by using a high-level description language, manipulation functionality becomes non-destructive and session transfer becomes easier to implement.

The description format described in this subsection provides the necessary infrastructure for enriched viewer interaction while watching recorded television programs. For example, the system can identify the public/private nature of specific content by including personal identity associated with a specific fragment's information (in the metadata section) for rendering in the secondary screen, such as a commercial tailored to an individual user. In addition, because we use a high-level description language, instead of content streams, it is open to external web services such as BabelFish[4] or Wikipedia,[5] which can provide extra material related to a fragment of the content (by using linking functionality in the metadata part). Finally, the benefits of this approach is that SMIL code is small, it is easily verifiable, it allows content associations to be defined, it provides a separation between logical and physical content, and it provides as base for license-free implementation on a wide range of platforms.

[4] http://babelfish.yahoo.com/.

[5] http://www.wikipedia.org/.

6.4 Summary

This section elaborated on the architecture presented in this article. We have proposed a high-level architecture, in which a variety of client devices are interconnected with the home media server. The devices can be used for rendering and interacting with the television content, while watching. Then, we introduced the detailed architecture, and its implementation, in order to validate that the functional requirements identified in the proposed taxonomy are met. After that, we proposed a device modelling categorization based on issues such as interaction capabilities, role of the client devices, and provided social experience. The categorization helps us to identify the functionality provided by each client device; moreover, it shows that our solution is targeted to the usage client devices in conjunction with the television screen. Finally, a rich television content description model was proposed, which provides richer semantic over the content and higher degree of manipulation functionality over previous solutions. We believe, based on the identified functional requirements, that our architecture provides a useful, and innovative, solution for consuming television content in our multi-device and multi-person living rooms. The next section provides a first validation of the implementation by proposing a walkthrough of its functionality.

7 Control, enrich, and share

The first set of results on our system is reported in this section as a walkthrough of a number of representative implemented services. Previous sections (cf. Sects. 3, 4) provided our taxonomy, in the form of key functionalities that should be provided, for secondary screen usage in the interactive television environment. The identified activities include content control, content enrich, and content sharing. This section elaborates on the specific functionalities associated to these activities, supported by our system. The intention is to validate the system, with respect to the key features identified in the taxonomy.

7.1 Control

Even though the efforts by the community to bring multimedia to other devices have been impressive, the results have been disappointing. Current systems only focus on the efficiently display of a single content stream, coming from a single source, into one device. They do not take advantage of the contextual situation of the user, the increased number of devices he is surrounded by, and the increasing number of sources that might be used to aggregate enriched content.

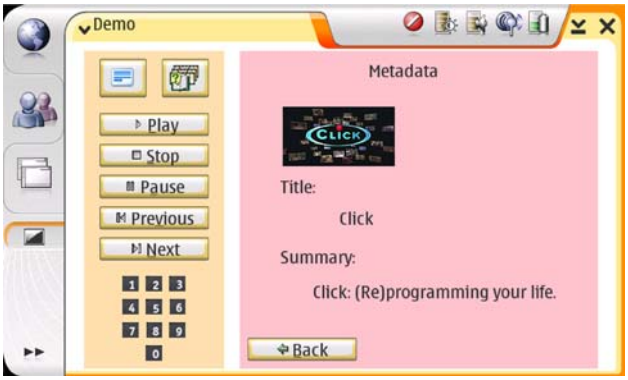

Fig. 5 Screenshot of the extra material interface (Nokia 770)

The control activity corresponds to the selection of content, the rendering of enhanced information, and the viewing of content on a specific device. In order to achieve it, there is a need to (1) identify the devices surrounding the user in a specific moment and (2) to differentiate the shared/personal nature of the media content. For example, the video content is shared media to be enjoyed in the living room, while enhanced material is private information to be rendered by a private device. In previous sections we have described the different ingredients of our system. SMIL and TV-Anytime provides a rich description format for recorded television material, the different possible client devices can be categorized based on their characteristics as illustrated in Table 3, and those devices are connected to the home media server using a wireless home network.

Two relevant services that can be classified as control activities are to request extra information about an active television fragment and to navigate within a television program. For example, due to its private nature the enhanced material might not be of interest to the rest of the viewers on the couch. Thus, it should be displayed on the secondary screen of the requesting user. Figure 5 shows an example of such interface. As shown in the example, the home media server provides to the secondary screen extra material of a current fragment of the television program. In this case, the title, the summary, and a representative poster are rendered. In fact, the secondary screen receives a subset of the code included in Source Code 1. Specifically, the *metadata* related to the active video fragment as shown in Source Code 2. Two clear advantages should be highlighted. Firstly, the handheld device only receives a fragment of XML code, which includes the extra material. This is possible thanks to the rich description format used in our implementation. Moreover, linking functionality to external services can be provided within the XML fragment. Secondly, the information contained in the file is associated to only the current fragment being watched, or current scene, and not to the whole television program.

```
…
<metadata xml:id="meta-rdf">
    <!-- Metadata Fragment 1 -->
    <tva2:Item … >
    …
    </tva2:Item>
</metadata>
…
```

Source Code 2: Snapshot of the current television program's metadata

A second activity that belongs to the control paradigm is content selection. In this case, the user can browse and select a video program recorded by the home media server, but even more interestingly the user can inspect fragments within the video. Once again, such browsing functionality is provided in the secondary screen. This functionality is similar to the EPG in interactive television domain or the DVD menus, but with the particularity of not using overlays over the main content. Figure 6 illustrates the user interface provided to the user.

In this case the Ambulant Annotator, after a request from the client device, responds with the *metadata* sections of all the fragments of the video. The user interface of the client, then, organizes and displays the content information in the allocated secondary screen space. The viewer might as well act on each of the fragments as indicated previously in Fig. 5.

In addition to enhanced information displaying, content browsing and selection, the secondary screen can be used for presentation continuity. In this case, the user decides that he wants to bring along the presentation shown in the television set. Thus, the secondary screen will become the primary screen after this action. In this case a dynamic evaluation of the context of the user in terms of available devices is needed. Figure 7 shows a screenshot of transferring the session from a television screen to a mobile device.

When the user decides to transfer the session to the secondary screen, a decision algorithm based on the presentation description and on the characteristics of the devices surround-

Fig. 6 Screenshot of the content browsing and selection interface (Nokia 770)

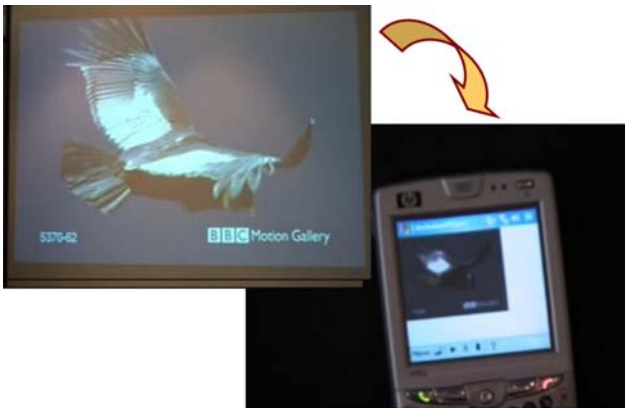

Fig. 7 Session transfer from a TV display to a Windows Mobile 5 device

ing the user is applied. Such decision follows a number of steps: (1) to store the current state of the running video, (2) to adapt the media description to the targeted platform, (3) to transfer the modified presentation to the new device, and (4) to restore the content in the new device at the same time offset as when the *transfer_session* button was pressed. Unlike previous approaches, session transfer is done by sending an XML file to the secondary display, thus minimizing the transferring time. Since modifications on the layout, timing, and source media content are done in the XML document, its implementation is more powerful and easier to implement.

Source Code 3 shows the modified media description when transferring the session after 20 s of watching the video. Firstly, the *layout* information is updated to fit the characteristics of the new device, in this case a low-resolution mobile phone. Secondly, the current viewing time is maintained by fragmenting the video at the transferring time. In our case, an extra video fragment is generated starting at *clipBegin=20s* and ending at *clipEnd=30s*. Moreover, the source code indicates that the first video fragment has been seen already, hence when restoring the presentation, that fragment can be skipped. We achieve so by adding a new *customTest* attribute. Finally, the actual media assets have to be adapted to the new rendering device, and the *src* attribute in the source code has to be updated. Such transformation of the content can be performed through transcoding or by locating (thanks to the CRID identifier) another version of the media content, more adequate in terms of video resolution and encoding for the new device.

After all the needed transformations have been applied, the new media description is delivered to the selected device. This device then can restart the description file in order to continue enjoying the multimedia content. The deployment behind this functionality is incorporated in the Ambulant Annotator; further details on the algorithms and on the device discovery and description systems can be found in a previous publication [8].

```
… <!-- Part 1 -->
<!-- CustomTests for saving the state -->
<customAttributes>
<customTest defaultState="false" override="hidden"
id="show"/>
</customAttributes>

… <!-- Part 2 -->
<layout>
 <region id="v" width="176" height="220"/>
</layout>

… <!-- Part 3 -->
 <seq>
<video region="v" src=".../video.3gp" clipEnd=”20s”
customTest="show">
…
</video>
<video region="v" src=".../video.3gp" clipBegin=”20s”
clipEnd=”30”>
…
</video>
<video region="v" src=".../video.3gp" clipBegin=”30s”
clipEnd=”60s”>
…
</video>
…
 </seq>
```

Source Code 3: Snapshot of the resulting description file, when transferring the session

7.2 Enrich and share

Multimedia content sharing systems have slowly grown in popularity over the past decade. Initially, the shared content consisted mainly of photographs or short video clips that contained re-purposed studio content, such as music, music videos and individual news items. The last few years have seen an unprecedented growth of video material available in the World Wide Web. Some examples of this growth include websites such as YouTube[6] and MySpace,[7] and innovative products such as Joost.[8] Social services, such as posting comments about videos, are among the most attractive features of these sites. Another popular service is sharing video material with others by sending a message that includes the Uniform Resource Locator (URL) of the video. Nevertheless, in spite of their success, the majority of the current systems contain a number of serious restrictions. First, the user is unable to share a bounded fragment of the video. Second, the user cannot customize the recommended video by including, for example, a voice commentary or strategically placed line art

[6] http://www.youtube.com/.

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

[8] http://www.joost.com/.

Fig. 8 Screenshot of the shared television display

Fig. 9 Screenshot of the enrichment interface (Nokia 770)

overlays. For a full analysis of current video sharing system, the interested reader can consult a previous publication [6].

Apart from content controlling, the other two high-level requirements identified in this article are to enrich television content and to share such enriched material with others. Let's first take a look at the enrichment interface, as illustrated in Figs. 8 and 9.

Figure 8 shows the main television screen displaying a recorded television program, in this case a documentary about South America. The scenario, as in all previous cases, is that Mark, Maria, and the kids are together in the living room in front of the television set. Mark wants to enhance the video by adding a set of overlays to send it to a friend later on. This activity does not have to disturb his wife, since he can use his secondary display in order to do so. Figure 9 shows the implemented interface for content enrichment. The interface allows the end user to fragment the recorded television program and to add a number of enrichments in the form of ink overlays. In addition, the viewer can add personal annotations on the content such as extra information, his own summary, or a rating of the television program—or fragment.

The feature of content enrichment while watching provide a number of benefits over traditional solutions. Firstly, the user can select parts of the video and does not have to share the full-length material. Secondly, the user can personalize the fragments by adding extra media assets. Finally, all the process is done using a secondary screen and thus not disturbing the other family members in the living room. We have termed this feature as micro-personal recommendations [6].

In terms of implementation, the enrichment process benefits from the rich multimedia description model we use. Source Code 4 shows the resulting document after a number of enrichments have been performed over the original material. As can be seen in the source code, the user has added a new *audio*, possibly an audio commentary that introduces the video, and a new *image*. In this case, the audio material has been inserted before the video content, and then a $\langle par \rangle$ element has been included for the first fragment of the video. This element indicates that the first fragment should be played at the same time as the added image. Most probably the image is an ink overlay that enhances the video being shown. After that, the video material continues playing as before.

It is important to note that the bits of the base content are not modified, but only the description file of the media content.

```
… <!-- Part 2 -->
<layout>
 <region id="v" width="1024" height="800"/>
 <region id="a" width="1024" height="800"/>
</layout>

… <!-- Part 3 -->
 <seq>
<audio src="…/audio_annotation001.mp3" … />
<!-- Fragment 1-->
<par>
 <video region="v" src=".../video.mp4" clipEnd="30s">
     …
   </video>
   <img src="…/ink_annotation.svg" region="a"/>
</par>
<!-- Fragment 2 -->
   <video region="v" src=".../video.mp4" clipBegin="30s"
clipEnd="60s">
…
</video>
…
  </seq>
```

Source Code 4: Snapshot of the resulting description file, after enrichment

After the enrichments are performed, the functional feature identified in this article is television sharing. Our system allows the user to share the personalized fragments with others. The user can select a member, or members, of his community and deliver the resulting enriched document. It is important to note that the user is not delivering the original television content, but the enrichments description, such as

Fig. 10 Screenshot of the sharing interface (Nokia 770)

the one shown in Source Code 4. The recipient of the enrichments needs to access to the original material, for which we use the TV-Anytime metadata. Thus, our system takes into account copyright and DRM constraints. In order to share the enriched media fragments, the viewer uses his secondary screen that is connected to the home media server. By initiating the *share* action, shown in Fig. 10, and selecting the desired recipients, the media server can use a number of different channels for delivering the resulting enrichments description file. Messaging technologies supported by our system include email, blog posting, and MMS for users on the move.

8 Evaluation

This section reports on the results of two independent tests of the services and architecture introduced in this article. It is important to note that the market analysis, technology development and user testing were performed by independent groups, across three countries.

8.1 Business analysis

In order to analyse the business opportunities the company Gradient/LUTIN, located in France, organized a panel with six professionals involved in interactive video and service providers. Specifically, the panel consisted of representatives from a European quadruple-offer player, a European equipment provider, an international advertising company, and a major international mobile video service provider. The panel was organized in Paris, during the Passepartout[9] project, and the developers of the system acted as mere spectators, so they could not "defend" their system. The panellists were not part of the project, but independent evaluators. The goal of the focus group was not to jointly design a new interface, but to analyse the commercial prospects of a distributed home control paradigm from a non-technical perspective. The panel was presented with a stylized presentation of the capabilities presented in this paper. The panel was told that the intention of the project was to define a value-added service that could be offered as an enhancement to a home PVR offering.

The participants were enthusiastic about the services presented in this article. The majority of them think that this sort of application should be deployed as soon as possible because some of its main features would be covered by major market players in the short term future. Nevertheless, the application should be deployed progressively.

It was felt that the content enrichment and sharing should not be restricted to a family or neighbourhood circle, but should focus on networked communities. Users should have a choice about being included in a local or global community of recommenders. All participants agreed that a distributed control application should be offered to the end user by a service operator as a part of a larger package. Its functionalities would not only benefit the end user, but also help to expand the base of the operator and to position the operator as a value-added supplier. Finally, one business case highlighted by the panel was the possibility to use the secondary screen for displaying targeted and personalized advertisements.

8.2 Initial user testing

In addition to the initial business analysis, we modelled a representative user community of up to three people watching television together. The viewing environment consisted of three handheld control devices, a small library of recorded content, a high-definition television set, and the prototype server. Each of the participants, sometimes three of them at the same time, were given a handheld control device, which was a personal device they could carry around as a mobile phone. The goal was to get feedback on the services, so we encouraged them to explore the different capabilities, to play around, and to complete a number of predefined tasks (e.g., to share a fragment of a video with some friends).

This article only includes a brief summary of the obtained results in order to show the usefulness of our architecture. The interested reader can consult a previous report, in which the full set of results has been reported [6]; it includes as well more detailed information on how the tests were conducted and analysed.

From the results we can highlight that users were attracted by the possibility of having a personal display that allowed for browsing, personalizing, and enriching content. While sharing content with other people outside home was seen as a value-added service, the end-users did not find it appealing to share the content within the home.

From the obtained results we can conclude that in order of relevance, the secondary screen was most valued for previewing and viewing content, as well as for accessing enriched information. The second preferred activity was to share frag-

[9] http://www.passepartout-project.org/.

ments of television content. Finally, most users liked the idea of enriching television content with personal overlays.

9 Conclusion

This paper presented an architecture, together with a working implementation, that leverages the viewer impact on the watched content. The presented architecture supports the usage of handheld devices that are used in conjunction with other electronic appliances for consuming and manipulating television content. The distinct characteristics over previous approaches are that the viewer can affect television content *while* watching and that the content is modelled using a high-level multimedia description format, and thus richer semantics are available.

The working assumption is the viewer's potential impact on content can be classified in three high-level categories: control, enrich, and share content. The taxonomy was validated after an exhaustive literature review and an analysis of existing solutions. Such taxonomy, the high-level categories together with the associated sub-classed, provided us a number of useful functional requirements for our architecture. Moreover, we can argue that the architecture could be reused, with minimal modifications, for a variety of services and viewers' situations in the living room.

The implemented working system and the architecture were validated, first, by analysing if the functional requirements were met. Moreover, the results of an initial business analysis indicate that secondary screens are considered as a value-added service for the home Personal Digital Recorder offerings. At the same time, users that tested the prototype system were attracted by the possibility of using secondary screens integrated into the home environment.

The final goal of our research is to empower the viewer with appropriate tools that leverage his impact on the content he is consuming. The work as presented in this article goes beyond a detailed architecture and implemented scenarios. We placed this work in a spectrum of activities that included an initial market assessment by professionals in the areas of media creation and distribution, and we subjected our prototype implementation to be tested by a dozen groups of users in a social setting. So far, the results are encouraging, even though more implementation work is needed and further testing remains to be done.

Acknowledgments Dick Bulterman and Jack Jansen are also affiliated with the Vrije Universiteit in Amsterdam. This work has been funded by the NWO BRICKS PDC3 project, the ITEA Project Passepartout, the FP6 IST project SPICE and the FP7 IP project TA2. Development of the implementation of this work within the open source Ambulant Player has been funded by the NLnet foundation. We are grateful for this support.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005). doi:10.1109/TKDE.2005.99
2. Ardissono, L., Kobsa, A., Maybury, M.T. (eds.): Personalized Digital Television Targeting Programs to Individual Viewers. Kluwer, Dordrecht (2004)
3. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tscheligi, M.: Trends in the living room and beyond. In: Proceedings of the European Conference on Interactive Television, pp. 146–155 (2007)
4. Bulterman, D.C.A., Rultedge, L.: SMIL 2.0: Interactive Multimedia for the Web and Mobile Devices. Springer, Berlin (2004)
5. Cesar, P., Chorianopoulos, K.: Interactivity and user participation in the television lifecycle: creating, sharing, and controlling content. In: Proceedings of the International Conference on Designing Interactive User Experiences for TV and Video, pp. 125–128 (2008)
6. Cesar, P., Bulterman, D.C.A., Geerts, D., Jansen, A.J., Knoche, H., Seager, W.: Enhancing social sharing of videos: fragment, annotate, enrich, and share. In: Proceedings of the ACM International Conference on Multimedia, pp. 11–20 (2008)
7. Cesar, P., Chorianopoulos, K., Jensen, J.F.: Social television and user interaction. ACM Comput. Entertain. **6**(1) (2008)
8. Cesar, P., Vaishnavi, I., Kernchen, R., Meissner, S., Hesselman, C., Boussard, M., Spedalieri, A., Bulterman, D.C.A., Gao, B.: Multimedia adaptation in ubiquitous environments: benefits of structured multimedia documents. In: Proceeding of the ACM Symposium on Document Engineering, pp. 275–284 (2008)
9. Cruickshank, L., Tsekleves, E., Whitham, R., Hill, A., Kondo, K.: Making interactive TV easier to use: interface design for a second screen approach. Des. J. **10**(3), 41–52 (2007)
10. Davis, R., Yung, D.: Understanding the interactivity between television and mobile commerce. Commun. ACM 48(7), 103–105 (2005). doi:10.1145/1070838.1070869
11. Dearman, D., Pierce, J.S.: It's on my other computer!: computing with multiple devices. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, pp. 767–776 (2008)
12. Fallahkhair, S., Pemberton, L., Griffiths, R.: Dual device user interface design for ubiquitous language learning: mobile phone and interactive television (iTV). In: Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education, pp. 85–92 (2005)
13. House, N.V., Davis, M., Ames, M., Finn, M., Viswanathan, V.: The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing. In: Extended Abstracts on Human Factors in Computing Systems, pp. 1853–1856 (2005)
14. Karanastasi, A., Kazasis, F.G., Christodoulakis, S.: A natural language model for managing TV-anytime information in mobile environments. Pers. Ubiquitous Comput. **9**, 262–272 (2005). doi:10.1007/s00779-004-0330-7
15. Keukelaere, F.D., Sutter, R.D., de Walle, R.V.: MPEG-21 session mobility on mobile devices. In: Proceedings of the International Conference on Internet Computing, pp. 287–293 (2005)
16. Kindberg, T., Spasojevic, M., Fleck, R., Sellen, A.: I saw this and thought of you: some social uses of camera phones. In: Extended Abstracts in Human Factors in Computing Systems, pp. 1545–1548 (2005)
17. Knoche, H., McCarthy, J.D., Sasse, M.A.: Can small be beautiful? assessing image resolution requirements for mobile TV. In: Proceedings of the ACM International Conference on Multimedia, pp. 829–838 (2005)
18. Mate, S., Chandra, U., Curcio, I.D.D.: Movable-multimedia: session mobility in ubiquitous computing ecosystem. In: Proceedings

of the International Conference on Mobile and Ubiquitous Multimedia (2006)
19. Miller, S.: Taking on the masses with mobile messaging TV. Comput. Entertain. **3**(2) (2005)
20. O'Hara, K., Mitchell, A.S., Vorbau, A.: Consuming video on mobile devices. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, pp. 857–866 (2007)
21. Park, J., Blythe, M., Monk, A., Grayson, D.: Sharable digital TV: relating ethnography to design through un-useless product suggestions. In: Extended Abstracts on Human factors in computing systems, pp. 1199–1204 (2006)
22. Robertson, S., Wharton, C., Ashworth, C., Franzke, M.: 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)
23. Seager, W., Knoche, H., Sasse, M.: TV-centricity—requirements gathering for triple play services. In: Adjunct Proceedings of the European Conference on Interactive Television, pp. 274–278 (2007)
24. Shacham, R., Schulzrinne, H., Thakolsri, S., Kellerer, W.: Ubiquitous device personalization and use: the next generation of IP multimedia communications. ACM Trans. Multimed. Comput. Commun. Appl. **3**(2) (2007)
25. Taylor, A.S., Harper, R.: Age-old practices in the 'New World': a study of gift-giving between teenage mobile phone users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 439–446 (2002)