Journal of Virtual Reality and Broadcasting, Volume 3(2006), no. 9
JVRB EuroITV 2006 Special Issue, no. 2

An Architecture for End-User TV Content Enrichment

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

CWI: Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ Amsterdam
+31 20 592 43 00
www: {P.S.Cesar, Dick.Bulterman, Jack.Jansen}@cwi.nl

Abstract

This paper proposes an extension to the television-watching paradigm that permits an end-user to enrich broadcast content. Examples of this enriched content are: virtual edits that allow the order of presentation within the content to be changed or that allow the content to be subsetted; conditional text, graphic or video objects that can be placed to appear within content and triggered by viewer interaction; additional navigation links that can be added to structure how other users view the base content object. The enriched content can be viewed directly within the context of the TV viewing experience. It may also be shared with other users within a distributed peer group. Our architecture is based on a model that allows the original content to remain unaltered, and which respects DRM restrictions on content reuse. The fundamental approach we use is to define an intermediate content enhancement layer that is based on the W3C's SMIL language. Using a pen-based enhancement interface, end-users can manipulate content that is saved in a home PDR setting. This paper describes our architecture and it provides several examples of how our system handles content enhancement. We also describe a reference implementation for creating and viewing enhancements.

First presented at the 4th European Interactive TV Conference EuroITV 2006, extended and revised for JVRB

Keywords: Interactive Digital TV, User-generated content, End-user enrichments, SMIL, Smart environments

1 Introduction

Digital television promises several improvements over analogue television, including advanced services such as different subtitle tracks and graphics-based teletext, remote interactivity using a return path, and personalisation of content. While all these issues have captured the attention of standardisation consortia and research groups, they provide only incremental improvements for end-users. While selectivity across content streams is likely to increase, very little research has focused on advanced interactive capabilities for television users to manipulate actual content sequences.

This paper investigates advanced interactive capabilities for viewers by proposing a software architecture that permits end-users to enrich broadcast content that is stored locally on a home Personal Digital Recorder (PDR). We initially assume that such content is captured either from an incoming digital content stream or that it is stored on a high-density optical device. The content itself is considered to be atomic and unstructured. Our goal is to provide users with an ability to personalize the content by allowing one or more layers of virtual edits to be added to any particular content objects. The combined base material and enhancements can then be viewed as a whole, or selectively (based on the intended audience or the available user interface).

This paper is structured as follows. First, Section 2 provides a set of basic use scenarios that clarify what we mean by content enrichment. This discussion will motivate the technology described elsewhere in this article. Section 3 identifies the requirements an interactive television enrichment environment should meet and Section 4 discusses the related work. Based on the requirements, Section 5 describes our proposed architecture. Section 6 reports on our initial experiences in developing the system. Finally, Section 7 concludes the paper and shapes the future work we will carry out in this area.

2 Basic Use Scenarios

This section provides three scenarios of our view of end-user content enrichment. These scenarios are neither exhaustive nor limiting; they are presented to provide a framework for understanding our architecture.

In all three cases, our assumption is that a home PDR is able to receive and store broadcast content. The content is subsequently accessed and enriched for later use. The enriched content could be intended for personal use, for use by other members of a household in a family setting, or for sharing within a (distributed) peer group.

Scenario 1: Content Control via Virtual Edits
Mark is the father of two children, a boy of 13 and a girl of 9. Three years ago, father and daughter agreed on buying a little bird after she has finished the first cycle at school. Mark is concerned with the impact the news about the bird-flu can have on his daughter. He wants his daughter to be informed; to know that there is a bird-flu epidemic and that it is dangerous. But, he does not necessarily want his daughter to see how families across Europe kill their own birds. He uses an enrichment interface to gain control over what his daughter watches. His goal is to eliminate certain scenes from the news. He does this by adding a series of virtual edits to the underlying content; when his daughter accesses the program, she does not see parts that he considers objectionable. However, he is still able to see the complete version with his wife because the base content has not been modified.

Scenario 2: Content Enhancement
Katrina notices that interactive television can improve the learning process of her two children by allowing them to access additional content on selected topics that she would like to insert in a base presentation. She is willing to spend some time on inserting this content, but she is not a professional video editor. She would like to easily be able to insert links to additional content that the kids can activate while viewing; during this additional content, the base program pauses. The enrichments that she wants to create consist of some text slide, a few pictures from the family scrapbook and a video sequence she shot during the last vacation together.

Scenario 3: Repackaging and Sharing Content
Leonard, the teenager son of Katrina and Mark, loves music. He always shares information on his favourite bands with his two friends, Philip and Sonny. Recently, a concert that the three attended was released on a high-density disk. Leonard is willing to spend extra time to include in-line text annotations and photographs from the concert to be shown on top of the video footage on the disk. In addition, he would like to draw some figures such as a circle with the text ”we were sitting in here, remember?”. Later, he wants to distribute his enrichments to Philip and Sonny, so that they can view them when they play their copy of the program disk.

These three simple scenarios exemplify the advanced user interaction paradigm we are proposing in this paper. Today, two solutions exist to support content enrichment: use of a content editing system on a PC or use of an interactive TV development environment. The PC desktop option uses software editing tools (such as Flash[1], Director[2], iMovie[3], GRiNS[4]). Unfortunately, such tools do not fit the television paradigm and often require extensive end-user training. The second alternative is to use interactive TV authoring tools (e.g., Cardinal[5], Sofia Arena[6], and Alticast[7]), but such tools require a level of programming and architecture knowledge that is beyond the scope of nearly all home viewers.

The approach that we describe in this paper assumes that a simple editing interface is available for use on the couch. This interface (which can range from a simple remote control to a TabletPC), communicates with the PDR; the PDR acts as a home server and allows an intermediate file to be created that is used to store content enrichments. As we will see, this intermediate file is based on the World Wide Web Consortium (W3C)'s Synchronized Multimedia Integration Language (SMIL).

[1] http://www.macromedia.de
[2] http://www.macromedia.de
[3] http://www.apple.com
[4] http://www.oratrix/GRiNS/
[5] http://www.cardinal.fi
[6] http://www.sofiadigital.com
[7] http://www.alticast.com

In the case of the first scenario, the content producer or the broadcaster may have included metadata about the video content. Then, the TV equipment filters those scenes that are objectionable. But the scenario presented in this paper would require from the filter engine a high degree of personal knowledge about the family and an immediate learning/response to particular events (three years ago they agreed on buying a bird, there is a bird-flu epidemic, and people are killing their birds).

3 Requirements Discussion

In this section, a number of requirements for an enrichment system in the interactive television environment are identified. The requirements are divided into the following categories: source of the base content, nature and timing of the annotations, user interface functionality, and infrastructure.

3.1 Source of the Base Content

Within a television context, we assume that the basic source of program content will be via conventional program sources. This is what we call the *base content*. In this paper, we describe a model in which the base content is stored on a hard disk in a PDR, or on a high-density optical disk (e.g., DVDs and Blue-Ray Discs). This base content may consist of a single audio/video file or a structured set of assets, such as audio + multi-channel audio + teletext.

The base content can be augmented with other content. Television viewers can access multimedia content from a number of sources: World Wide Web (WWW), stored material, Peer-to-Peer (P2P) networks, and broadcast. In addition, equipment at home is an important source of augmented digital material, such as home movies, images or audio fragments. It is expected that all of these sources could eventually provide content that a given viewer may want to integrate into a presentation. (We are not alone in this assumption: even broadcasters such as the BBC consider that P2P networks provide a mechanism for distribution of content.[8])

In order to delimit the scope of this paper, we assume that a user has access to all these sources from his television receiver. We are aware, though, that some of the topics such as WWW access, P2P networks, and Blue-Ray technology are current research topics. But we believe, based on current results in this area, that in the near future these technologies will be integrated in television receivers. Some interesting results include, for example, Cesar's work on WWW access from the television receiver [Ces05], University of Vigo research about the integration of a P2P client in the MHP architecture [LNEEBF$^+$04], and the I-Share project that is investigating the use of BitTorrent in the interactive television environment [PGES05].

The main question in our work is not so much: *from whence comes the content?* but: *who can integrate which content when, and how?*

3.2 Nature of the Annotations

There are different ways of understanding annotations. For some researchers, annotations are metadata that can be used to locate or classify media items within a presentation. Other researchers consider annotations as end-user created external media items that are used to augment object content. [Bul04] differentiates them as hierarchical and peer-level annotations, respectively. Table 1 shows the differences between them in the television environment.

The first type, hierarchical, provides information about specific media items (e.g., name of a movie) that can be used for searching or analysis. This type of annotation is essential, in interactive television, for content personalisation. Recommender systems [SC00, BPG$^+$05] base their decisions by comparing hierarchical annotations with user profiles. These annotations can be generated manually or automatically, are encoded as text strings, and lie in the responsibility of the content provider.

The second type, peer-level annotation, is created to enrich specific media items. For example, in the beginning of a movie while the name of the actors is in the screen, an end-user can highlight with a yellow pen, those actors he likes most. Because these annotations are enhanced information created by end-users, they can be any multimedia type such as text, graphics, images, animations, audio, and videos. Finally, this kind of annotations does not follow a restrictive vocabulary, but provides freedom of expression to the user.

This paper proposes advanced interactive capabilities over TV content by peer-level enrichments. The major requirement for such enriched material is that its inclusion must not alter the base content. This requirement protects base content copyrights and allows the reuse of original sources many times in other con-

[8] http://www.bbc.co.uk/imp/

	Responsible	Type	Example
Hierarchical	Broadcaster	Text String	Movie name
Peer-Level	End-user	Any kind	Highlight actors' names

Table 1: Comparison of Hierarchical and Peer-Level Annotations in the Television Environment.

texts.

In addition, the enriched material may be later distributed using mechanisms such as pal-sharing [PGES05] or family-sharing (e.g., family members living outside home). Digital Rights Management (DRM) issues of the base content are taken into consideration by restricting the end-user to only distribute the enriched content.

3.3 Timing of the Annotation

The creation of enriched content by the end-user can take place in two separate moments: previewing time or in real-time. Previewing time corresponds to the case, in which one member of the family has time to preview and enrich the base content that is going to be watched the following day by the rest of the family. On the other hand, real-time enrichment corresponds to the case, in which the enrichment is performed while watching the base content. Because of the complexity of the user interface that imposes real-time annotation, together with the temporal restrictions (e.g., you might want to include a link, but the scene is over), this kind of annotations will be investigated in the future. Still, the requirement for our architecture is that it has to be extensible enough for supporting real-time enrichment.

3.4 User Interface Functionality

[Tan98] considers user interaction as the potential impact the user can have on the application. Other researchers, such as [Lau93], [Ale98], and [Bol01], have proposed more concrete measures of the level of interaction. For example, [Ale98] defines four possible levels of interaction: passive (i.e., user has no control), reactive (i.e., control remains in the user interface), proactive (i.e., user gains control and can for example navigate a path), and directive (i.e., user authoring).

Based on the previous definitions, in the television environment we can differentiate the following categories:

- *Content consumption* (passive level): traditional role, in which the user only watches/hears the broadcast content.
- *Setting selection* (reactive level): traditional role, in which the user can change the channel or the volume of the television.
- *Navigation/Selection* (proactive level): DVD paradigm, in which the user can navigate (e.g., skip content) and select, for example, special features or specific subtitling. Within this category belong, as well, broadcast Multimedia Home Platform (MHP) [ETS05] applications such as showing the current score on a match game when by pressing a colour button.
- *Authoring* (directive level): highly interactive mechanism, in which the user authors content. A simple example is to fill forms. But this paper contemplates more advanced use cases such as the inclusion of enriched material.

The enrichment system should provide the following functionality: inclusion, virtual edits, and creation of modifications. By inclusion we refer to the creation of new media objects (i.e., Katrina and Leonard scenarios). By virtual edits we refer to alterations on the base content (i.e., scenario 1). Finally, by modification we refer to the creation of additional navigation links.

In the case of inclusion, the media objects can be in the form of audio (e.g., commentaries to a football match), text (e.g., textual notes or subtitles), handmade drawings using a pen-based device, animations, and pictures or videos stored in a digital camera. We can differentiate two kinds of inclusions: sequential and parallel. In the first case, the base content is stopped while the included object is being presented. In the second case the enriched material is synchronized with the base content.

Virtual edits include two tasks: alteration of the presentation timeline and exclusion of media objects. In this case, as discussed in the *nature of the annotations* requirement (section 3.2), the user must not be allowed to actually modify the base content. For example, in the case of exclusion of media objects, the user/author should be able to impose restrictions in specific media items, but never delete those media items. That is even though he might not want his kids to watch it, he can still enjoy it with his wife. This requirement imposes the necessity of a container format, such as SMIL [BR04], that permits to have media items tagged as optional/conditional.

Finally, the creation of alternative navigation paths can enhance the television viewing process. Viewers can select different paths within the presentation depending on their interests or mood. Once more, we do not want the user/author to play around and destroy the original material, but he should be allowed to, for example, include different branches in the presentation that the viewers can follow.

In conclusion, the functionality of our enrichment system include:

- Inclusion of media objects
  - Sequential to base content
  - Synchronized with base content
- Virtual edits
  - Alteration of the presentation timeline
  - Exclusion of media objects
- Creation of alternative navigation paths

3.5 Infrastructure

The basic requirement in this case is that our system has to be compatible with end-user equipment. Thus, this subsection deals with some practical issues such as enrichment system integration, interaction devices, and relation to current standards, and distribution.

Because the intention is to enrich the broadcast TV material, the enrichment system should be integrated in existing interactive television receivers such as PDRs. Examples of such receivers are the popular Tivo[9] solution and Blue-Ray players as researched within the European Passepartout project[10]. In addition, even though not an essential requirement, the PDR should be able to access the WWW, P2P networks, and other devices at home (e.g., digital video camera).

In order to enrich the base content, the user can utilise a number of interactive devices. Some examples include:

- *Television devices such as remote control*: in this case the user has a limited number of options: the arrow keys, the numerical pad (used to change channels), and the OK buttons. [Ber04] proposes interesting results in this topic by extending remote control capabilities.

- *Other devices*: the number of digital devices at home is numerous. These devices, such as a Personal Digital Assistant (PDA), a mobile phone, and a tablet augmenter (e.g., Samsung Q1) can be used for content enrichment. This solution is less limited than using the remote control and exploits the fact that the user already has those devices.

- *Voice Recogniser*: voice can be used as a command language (e.g., delete). [BJ04] has performed relevant research about this topic.

- *Gesture Recogniser*: based on a pre-defined language, gestures permit extensive interaction. Devices such as a tablet augmenter can simplify the problem because the gestures are performed over a flat surface.

- *Combination of interfaces*: A combination of the above mentioned option is, as well, possible. For example, gestures can be used for specifying a position and voice commands for actions.

In relation to current standards, there are a number of solutions that can be used. Digital television transmission follows, in Europe, the DVB standard, thus the receiver should be DVB-enabled. In addition, current systems, and Blue-Ray players, integrate a platform-independent middleware called Globally Executable MHP (GEM) based on Java technology. For interoperability issues, the enrichment system can be developed using that standard. Still, the creation of the enriched material has to be performed in a declarative manner. Some of the benefits of such solution include modality independence. For example, the final enriched content can be embedded in a SMIL presentation that can be run by advanced digital television receivers [LCHV03, Ces05].

3.6 Summary

The requirements identified in the previous section are the following:

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

[10] http://wwww.hitech-projects.com/euprojects/passepartout/

- The user should be able to access a diversity of content sources (e.g., broadcast, WWW, and stored).

- The enriched content is created as an intermediate content enhancement layer. Thus, the base content is not altered.

- The enrichment system should provide inclusion of media objects, virtual edits, and creation of alternative navigation paths capabilities.

- The process of enrichment will take place while the user previews the content.

- The enrichment system is integrated into a PDR.

- The interactive devices the user can use include a PDA, a mobile phone, and a tablet augmenter.

- Issues such as real-time enrichment and post-distribution of the enriched content are left as future work

4 Related Work

In our opinion, advanced interactivity in the digital television environment has not yet been deeply explored. For example, Jensen defines three basic types of interactive television [Jen05]: enhanced (e.g., teletext) personalized (e.g., pause/play content stream using a PDR), and complete interactive (i.e., return channel). MHP's Interactive profile is defined as the provision of a return channel for the television receiver. Finally, commercial digital television services, such the Super Teletext, include user interfaces, where navigation/selection are the only options left to the user. While, more advanced interactive applications can include, for example, filling forms.

This article defines interactive television as the potential impact the user has on the content, instead of as the provision by the receiver of a return path. Thus, we align with researchers that intend to define new metaphors for television and proposals that extend television capabilities.

First, Chorianopoulos, in his doctoral dissertation, argues that traditional metaphors cannot be applied to digital television [Cho04]. He proposes a metaphor called Virtual Channel, ”dynamic synthesis of discrete video, graphics, and data control at the consumer's digital set-top box”. [Aga01] presented the Viper system that allows the user to skip content (i.e., enhanced interactivity), while watching it. Nevertheless, the enhancements over the base content take place at the broadcast station (in this paper, they take place at home). Finally, Berglund has studied extensibility usage of the remote control and audio input in television [Ber04, BJ04].

We extend Jensen's categorization with a new television paradigm: *viewer-side content enrichment*. In this paradigm, the viewer is transformed into an active agent, exercising more direct control over content consumption, creation and sharing. A key element of our paradigm is that, unlike with the PC, the television viewer remains essentially a content consumer who participates in an *ambient* process of incremental content editing. Our motto is ”authoring and pal-sharing from the sofa, not the desk”. This paradigm includes three activities: intra-program selection (selection of the content to be enriched), enrichments authoring (the content enrichment process), and pal-sharing (post-enrichment distribution).

4.1 Content Selection

Much of the research on content selection within a digital television environment has focused on the macro-level concerns of selecting an entire program among a wide range of content available to a user. This is often done by some form of recommender system [SC00, BPG$^+$05]. While we agree that recommender systems will play an important role in the future, they provide little or no assistance in navigating through content once it arrives in the home.

A micro-level of content selection is often required if content-based navigation through a program is to be supported. Finding a program that fits the profile of a user is useful, but being able to navigate or select stories within than newscast is equally essential. A micro-level recommender system is required that processes the base content and then selects fragments of interest. (This is not an exclusive selection, but could be used in a customized navigation control interface).

Nevertheless, our system can be used together with a recommender system. First, the recommender engine is in charge of selecting content matching user preferences and stores the content in the PDR. Then, the user can enrich the recommended content by providing more personal enhancements such as audio commentaries, drawings, and images.

In relation to macro-level selection, the most relevant standard today is the TV-Anytime Forum[11]. Interesting research in this area includes the UP-TV[12] project. The UP-TV project has presented [KKC05] a program guide that can be controlled and managed (e.g., delete programs) from personal devices (e.g., handheld devices).

In relation to micro-level selection, we have focused on a content tagging and selection interface that allows an end-user to effectively scan a large content set and then provide a personalized version of that content for members of his/her peer-group. We expect to focus on the semi-automatic generation of micro-level recommenders in follow-on work.

4.2 Incremental Content Enrichments

Both the research community and the consumer electronics industry have put a considerable effort into designing and developing authoring tools for the masses. Most of this work has, however, focused on the editing of media content within a PC setting. Our work shifts the focus to providing intuitive editing support ”from the sofa” by non-specialist television viewers.

Similarly to our work, [FB02, FB04] proposes the use of hyper-video in the television domain in order to integrate the Internet and the broadcast. Their system is divided into a client and a server side. The server side provides three different views: video/annotation, navigation, and information/communication views. The annotation view allows the user to edit hyper-video links, while the communication view manages the news group forum. The major difference with our work is that Finke focuses on TV-based Web content, while we focus on direct-view IDTV content. Because of the importance of timing in the broadcast environment, the use of a temporal-based language such as SMIL is more effective than the use of the static solution proposed by Finke in the form of HTML.

4.3 Content Sharing

We see peer-group content sharing as a post-enhancement distribution activity. There are two post-enhancement distribution scenarios of particular interest. First, the enriched content can be shared with your pals or family members living outside home. The most relevant technology in this case is the use of P2P networks, which are becoming a research topic in the television domain. Some examples include the University of Vigo research about the integration of a P2P client in the MHP architecture [LNEEBF$^+$04] and the I-Share project that is investigating the use of BitTorrent in the interactive television environment [PGES05]. In the current version of our home server, the use of the I-Share P2P environment has been architected into the system, but not implemented in the current demonstrator. Second, the annotated content can be distributed to other devices such as an iPod or a mobile phone.

The most important challenge in this case is how to adapt the content to other devices. The problem of adaptability of content is out of the scope of this paper, but we will dedicate future work to it.

5 Proposed Architecture

Based on the requirements identified in the previous section, the first decision was to integrate the enrichment system into a digital television receiver such as an extended PDR (i.e., PDR+). The PDR+ has access to the conventional broadcast content, to high-density disks, and includes a P2P client. In addition, a number of devices at home can be connected to our PDR+ via fixed or wireless networks. These devices include the conventional TV set and the remote control, a PDA/Mobile device, and an interactive augmenter device. Figure 1 shows the proposed architecture and its components. The conventional TV acts as a viewer. It has a broad presentation path (i.e., to show the final enriched content), but a restricted interaction path (i.e., remote control). Other devices, such as the tablet augmenter, have a broader interaction path, so they can provide further interactivity.

The enrichment system is divided into two components: server and client side. The server side is located in the home-server PDR+. It implements the functionality described in the requirements. The client side, on the other hand, resides in the interactive device (e.g., PDA). It corresponds to the user interface of the enrichment system and it exposes to the user the functionality of the server side. Server and client side communicate using a well-defined interface, allowing multiple client implementations. The following subsections elaborate on the different parts of our enrichment system.

5.1 Server Side

The server side is located in the PDR+. It implements all the actions that permit the user to enrich content:

[11] http://tv-anytime.org

[12] http://music.tuc.gr/Research/UPTV.htm

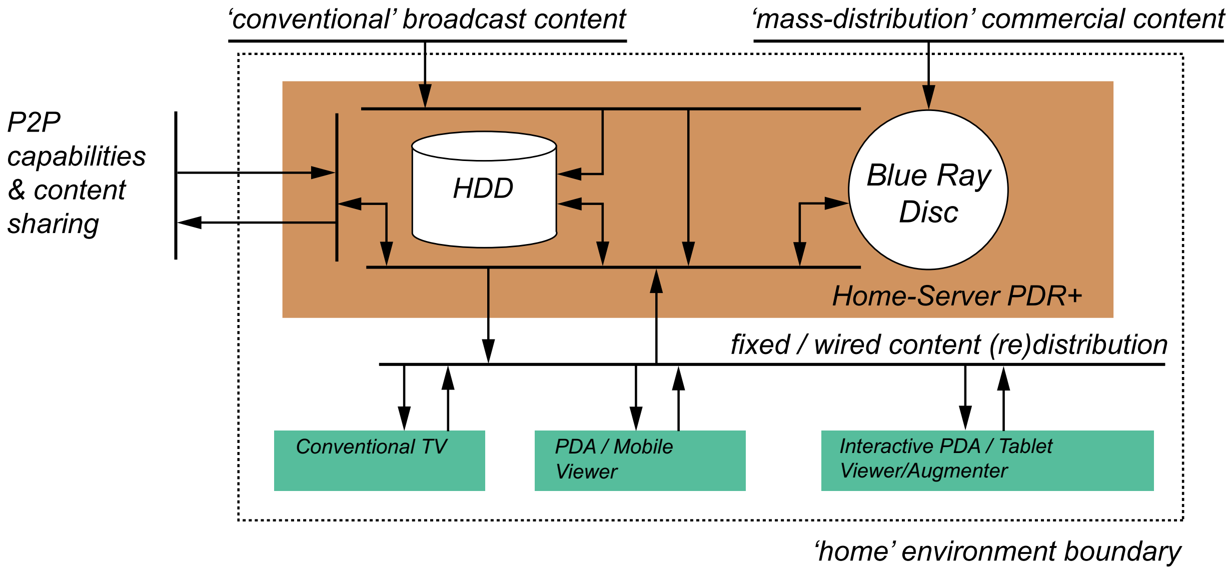
'conventional' broadcast content
'mass-distribution' commercial content
P2P capabilities & content sharing
HDD
Blue Ray Disc
Home-Server PDR+
fixed / wired content (re)distribution
Conventional TV
PDA / Mobile Viewer
Interactive PDA / Tablet Viewer/Augmenter
'home' environment boundary

Figure 1: Proposed Architecture of the Enriched Environment

creation of the navigational markup structure, navigation of the content, and selection of the action to be performed.

By creation of the navigational markup, we refer to the capability of imposing a navigational structure to the base content. As mentioned before, one of our assumptions was that the base content was unstructured. Hence, the user needs a mechanism to determine when the enrichments will take place (i.e., a navigational markup structure). In order to impose the navigational structure of the base content, the user can select points (moment in time) and ranges (lapse of time or story) in the base content. A set of roles are associated with points/ranges; these roles identify either a system user (such as the name of a child or a group of friends) or a semantic component (such as a story ID) for use during editing.

By navigation of the content we refer to the process of scanning the base content and deciding when the enriched content should be included. Basic functionality such as pause/restart is needed. Further functionality include, for example, ”go to the next point”.

By selection of the action to be performed, we refer to the process of including the enriched content. As described in the requirements there are three types of actions: virtual editing, inclusion of objects, and creation of alternative navigation paths.

The deletion of scenes is an easy task that can be performed by selecting a range (story) and by marking it as deleted or conditional (e.g., inappropriate for my kids). Restructuring the timeline of the base content is a more complicated task that requires reordering ranges from the base content.

As discussed in the requirements, the inclusion of objects can be performed in sequential or parallel mode. In the sequential mode conditional objects are shown to the user, so the user can pause the base content and trigger them. This action is simple to implement because it only requires indicating the timing of the object and the content of the object (i.e., point selection). The second case, parallel, requires as well indicating information such as when the object should stop being visible and the placement of the object within the screen.

Finally, the creation of alternative navigation paths are created by establishing virtual links within the base content.

In all the cases, the final enriched content is created as a SMIL 2.1 document that can contain, for example, Scalable Vector Graphics (SVG) elements (e.g., drawings). The enriched content is, then, uploaded in the server side. Later, the server side can distribute it through the P2P network or show it in the television set.

5.2 Client Side

The client side of the system resides in the interactive device used for enriching the content. The client side is in charge of implementing the user interface of the enrichment system, while the server side actually implement the functionality.

First, the client side needs to use an interface to communicate to the server side. This interface describes the semantics of the functionality (e.g., add + what + where + where). Second, the client side needs to share with the server side the notion of time (e.g., pause/restart). Finally, the client side, depending on its native capabilities (some of them might be less resource constrained than others), exposes the full set of server side actions or only a subset of them.

The main assumption for our enrichment system is that the user does not have previous knowledge about MHP authoring tools (i.e., programming languages) or video editing. Hence, the user interface should be intuitive, simple, and restricted to the actual actions the enrichment system can perform. Then, the user interface should be independent on the modality (e.g., gestures, voice commands, or Keyboard-like input) used for interaction. Hence, the decision of separating the implementation of the functionality (i.e., server side) from its user interface (i.e., client side) supports that requirement.

6 Implementation and Results

This section discusses why SMIL was selected as the multimedia description language, it describes the implemented system, and explains the workflow of enriching content using an example.

6.1 Multimedia Description Language

The SMIL language was selected as the final format of the content because of many reasons: it is a container format, thus the base content is not modified; it provides a rich set of media timing and activation primitives; it provides a simple and flexible layout model; it provides both system- and user-test attributes for content control; and it can be used together with other languages such as SVG. Even though the MHP standard does not yet include support for them, extensions to MHP have been proposed and demonstrated [Ces05].

By using SMIL, all the required functionality for the enrichments is provided:

- Inclusion of media objects: as SMIL elements (e.g., image or video) or as SVG images (e.g., drawing).
  - Sequential to the base content: using the *seq* element.
  - Parallel: using the *par* element.
- Virtual Edits: content control.
- Creation of alternative navigations paths: anchors (linking).

6.2 Implemented System

The system should provide the four levels of interactivity described in the requirements: passive, reactive, proactive, and directive.

- *Passive*: the user only watches/hears the base content
- *Reactive*: the user can pause/restart the content.
- *Proactive*: the user can select whether or not to see the enriched content (e.g., a picture or a drawing).
- *Directive*: the user can enrich the content.

For the three first categories (passive, reactive, and proactive), we have integrated the Ambulant player[13] [BJK$^+$04] in a Linux-based television receiver. The open-source Ambulant player provides a complete implementation of SMIL 2.1. It is a multi-platform player that can run on Linux, Macintosh, Windows, and Zaurus Linux-PDA. Note that while a software media player is used to manage the presentation of enriched content, the user is presented a conventional television-centric view of the content.

The Ambulant Player has been used in desktop and mobile phone/PDA environments. The only limitation was that its user interface did not fit the television paradigm (e.g., remote control interaction). Thus, we reused the core functionality and accommodated for the current requirements.

For the directive category, the enrichment system should be capable of three user activities: imposing a navigation structure for the base content, navigating the content, and editing the enriched material. In order to support these activities, we have integrated in our system the Ambulant Annotator [BR04].

Figure 2 shows the functional elements of the implemented system. The user interacts with the enrichment

[13] http://ambulantplayer.org

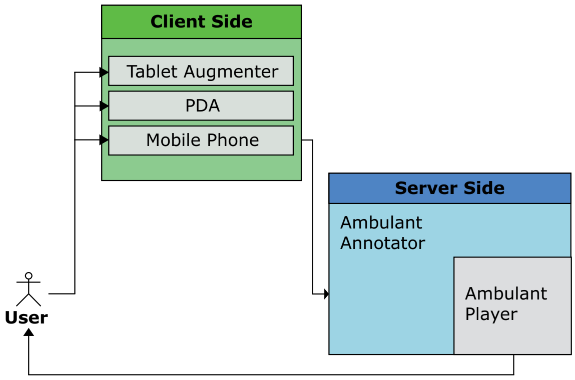

Figure 2: Functional Elements of the Enrichment System.

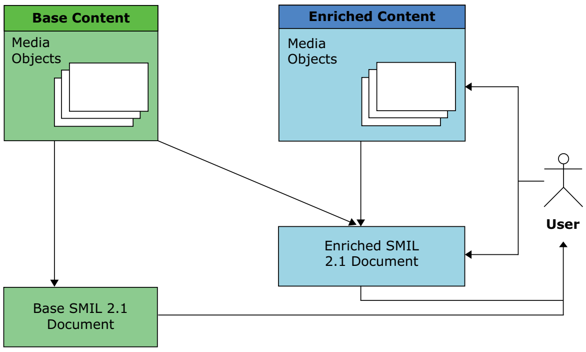

Figure 3: Workflow of the Enrichment System.

system through the client side. The user interface is located in an interactive device (e.g., tablet augmenter) and it uses the functionality implemented by the Ambulant Annotator (server side) using the XML-Remote Procedure Call (XML-RPC) protocol (semantics). The output of the Ambulant Annotator is a SMIL 2.1 document that later can be played by the Ambulant Player.

In the scope of this paper, we consider the client side to be a pen-based tablet augmenter. This device is capable of handwriting recognition, gesture recognition, and free-hand pen-strokes (e.g., free text or drawings). We are currently working on a number of issues. First, we are implementing other clients such as a PDA. Second, we are investigating other modalities such as voice commands. Finally, we are developing the creation of the markup structure in the Ambulant Annotator (currently, we assume that the content has been previously structured as a set of stories).

6.3 Workflow of the Enrichment System

In all the scenarios proposed in Section 2, the first assumption is that the content is already downloaded in the PDR. Obviously, in this case the original base content is viewed as a stored program rather than a real-time broadcast.

The workflow of our system is shown in Figure 3. After the base content is stored in the PDR, the user has two options: the base content can passively viewed or it can be enriched. In the first case, the user acts as a consumer and she can watch the base content as a SMIL document (base SMIL document) using the Ambulant Player. In the second case, the user acts as a producer and she can enhance the content. These enhancements, in the form of text, graphics, drawings, or videos, are created while previewing the base content. When the enrichment is finished, the integration describing both the base content and the enrichments is described in the form of a SMIL 2.1 file. (The base content is left unchanged; this ensures that there are no rights conflicts associated with the virtual edits.) Finally, the description of the base content and the enrichments (i.e., enriched SMIL document) can be viewed locally as a compound object. It can also be post-distributed using, for example, a P2P. Note that in the context of P2P distribution, it may be necessary for each remote user to have a local copy of the base content if its redistribution is restricted.

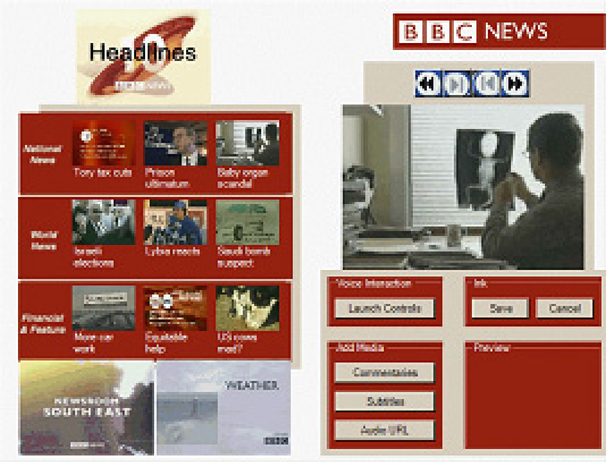

Figure 4: User Interface of the Client Side Implemented in a TabletPC.

Figure 4 shows the user interface of the Ambulant Annotator implemented in a tabletPC. The screen is divided into four main areas. On the left side the markup navigation region is presented as a set of stories. On the right side three regions can be differentiated: media controls (pause/restart) on the top, story edition area on the centre, and a button-based interface on the bottom.

Figure 5 zooms the markup navigation region. In this region user interaction is based on gesture recognition. Table 2 describes some of the gestures we have implemented for our system. Figure 5 (right) shows the markup region after the user has deleted a story ("/" gesture over "Lybia Reacts" story) and move to its right another story ("→" gesture over "Tory Tax Cut" story).

By double tapping on one story, the user brings it to the story edition area as shown in Figure 4 (the user has double tapped the "Baby Organ Scandal" story). There, the user can play the content. At any moment, he can pause the content, and enrich the material. For example, Figure 6 shows how the user has drawn an image and scribbled some text over the content. Finally, the user can utilize the buttons shown in Figure 7 to generate the XML-RPC call that would be sent to the server-side. The server-side, then, updates the presentation with the user-generated image as a transparent overlay

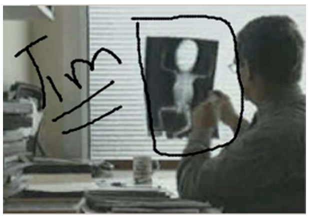

Figure 6: Story Edition Area, after the User has Scribbled.

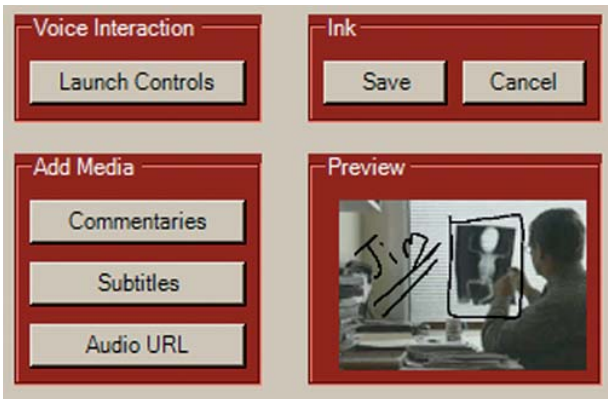

Figure 7: Button-Based Area.

7 Conclusion and Future Work

In conclusion, this paper constitutes the first milestone in our intention to research advanced interactive techniques for TV viewers and post-distribution capabilities. This work is related to a number of current research topics such as recommender systems, standardisation procedures, and definition of new TV paradigms. In relation to recommender systems, our enrichments can be made on top of the recommendations, thus giving more control over the content the viewer is watching or generating a new model of family experience. In relation to standardisation procedures, we argue that digital television receivers should include support for a declarative environment based on SMIL 2.1 language. Finally, with this paper we try to define a new TV paradigm in which the user can control the content he watches and can enrich it.

The current status of the system permits an end-user to author enrichments on top of the broadcast content stored in a PDR device. It permits a user to include, exclude, and modify content without altering the original content. In addition, its user interface simplifies the task in comparison to current authoring tools (e.g., iMovie or Flash) while providing advanced user interaction.

Future work includes the study of voice commands. Our current system allows the inclusion of voice commentaries to the content, but in the future user voice commands will be supported. The proposed architecture is modality independent, hence we can implement the user interface as a set of voice commands and use the functionality already provided by the server side. Finally, issues such as post-distribution of the enriched content in, for example, a P2P network and real-time authoring are work under development.

We are starting collaborations with other research groups that are specialized in related fields. First, content distribution and sharing using Peer-to-Peer networks will be integrated in our system by collaborating with the I-Share project. Second, we will perform usability tests and study business opportunities within the Passepartout project. Finally, the actual legal implications of distributing user-generated content (even if the underlying content was not modified) will be further investigated within the SPICE project.

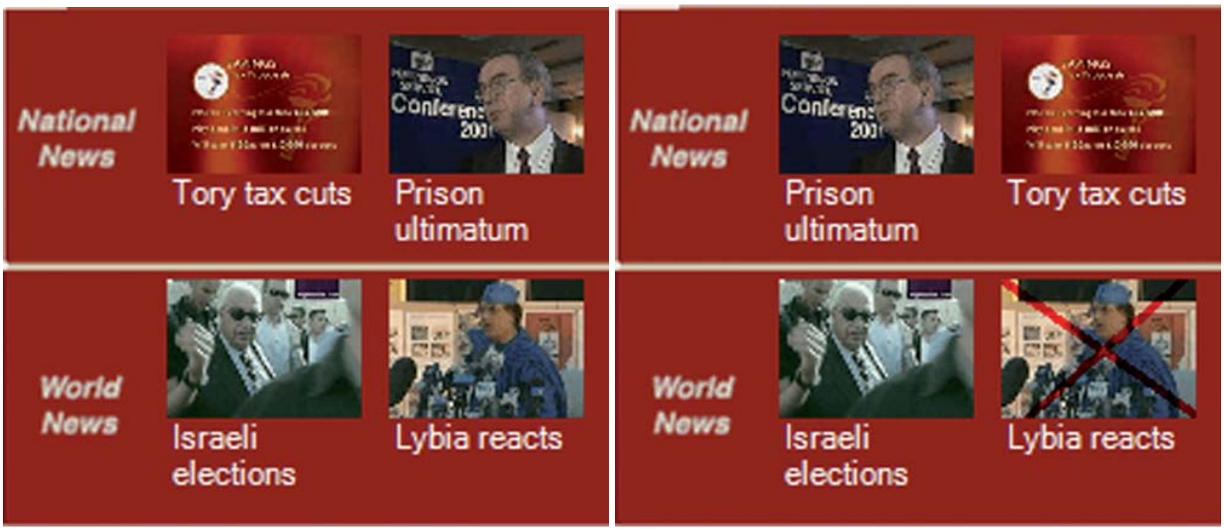

Figure 5: Markup Navigation Region. (left) before user interaction. (right) after user interaction.

Gesture	Double tap	/	←	→
Meaning	Start editing	Remove Scene	Move story left	Move story right

Table 2: Sample of the implemented gestures.

8 Acknowlegdements

This work was supported by the ITEA project Passepartout, by the NWO project BRICKS. The development of the Ambulant Player is supported by the NLnet foundation.

References

[Aga01] Stefan Panayiotis Agamanolis, *Isis, Cabbage, and Viper: New tools and strategies for designing responsive media*, Ph.D. thesis, Massachusetts Institute of Technology, USA, 2001.

[Ale98] T.A. Aleem, *A Taxonomy of Multimedia Interactivity*, Ph.D. thesis, The Union Institute, USA, 1998.

[Ber04] Aseel Berglund, *Augmenting the Remote Control: Studies in Complex information navigation for digital TV*, Ph.D. thesis, Linköping University, Sweden, 2004, ISBN 91-7373-940-5.

[BJ04] Aseel Berglund and Pontus Johansson, *Using Speech and Dialogue for Interactive TV Navigation*, Universal Access in the Information Society **3** (2004), no. 3-4, 224–238, Springer, ISSN 1615-5289.

[BJK^{+}04] Dick C.A. Bulterman, Jack Jansen, Kleanthis Kleanthous, Kees Blom, and Daniel Benden, *Ambulant: A Fast, Multi-Platform Open Source SMIL Player*, Proceedings of the 12th ACM International Conference on Multimedia, October 10-16, 2004, New York, NY, USA, 2004, ISBN 1-58113-893-8, pp. 492–495.

[Bol01] Susanne Boll, *ZYX: Towards Flexible Multimedia Document Models for Reuse and Adaptation*, Ph.D. thesis, University of Vienna, Austria, 2001.

[BPG^{+}05] Yolanda Blanco, José J. Pazos, Alberto Gil, Manuel Ramos, Ana Fernández, Rebeca P. Díaz, Martín López, and Belén Barragáns, *AVATAR: an approach based on semantic reasoning to recommend personalized TV programs*, Special interest tracks and posters of the 14th international conference on World Wide Web, 2005, ISBN 1-59593-051-5, pp. 1078–1079.

[BR04] Dick C.A. Bulterman and Lloyd Rutledge, *SMIL 2.0: Interactive Multimedia for Web and Mobile Devices*, X.media.publishing, Springer-Verlag, Heidelberg, Germany, 2004, ISBN 3-540-20234-X.

[Bul04] Dick C.A. Bulterman, *Animating Peer-Level Annotations Within Web-Based Multimedia*, 7th Eurographics Workshop on Multimedia, 2004, ISBN 3-905673-17-7, pp. 49–57.

[Ces05] Pablo Cesar, *A Graphics Software Architecture for High-End Interactive TV Terminals*, Ph.D. thesis, Helsinki University of Technology, Finland, 2005.

[Cho04] Konstantinos Chorianopoulos, *Virtual Television Channels: Conceptual Model, User Interface Design and Affective Usability Evaluation*, Ph.D. thesis, Athens University of Economic and Business, 2004.

[ETS05] ETSI: European Telecommunications Standards Institute, *Digital Video Broadcasting (DVB) - Multimedia Home Platform (MHP) Specification 1.1*, 2005, TS 101812 v1.

[FB02] Matthias Finke and Dirk Balfanz, *Interaction with Content-Augmented Video via Off-Screen Hyperlinks for Direct Information Retrieval*, The 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002, WSCG 2002, 2002, ISSN 1213-6964, pp. 187–194.

[FB04] Matthias Finke and Dirk Balfanz, *A reference architecture supporting hypervideo content for ITV and the internet domain*, Computers and Graphics **28** (2004), 179–191, ISSN 0097-8493.

[Jen05] Jens F. Jensen, *Interactive Television: New Genres, New Format, New Content*, Second Australasian Conference on Interactive Entertainment (Sydney, Australia), ACM International Conference Proceeding Series; Vol. 123, 2005, ISBN 0-9751533-2-3, pp. 89–96.

[KKC05] Anastasia Karanastasi, Fotis G. Kazasis, and Stavros Christodoulakis, *A Natural Language Model for Managing TV-Anytime Information in Mobile Environments*, Personal and Ubiquitous Computing **9** (2005), 262–272, ISSN 1617-4917.

[Lau93] Brenda Laurel, *Computers as Theatre*, Addison-Wesley, Reading, MA, 1993, ISBN 0-201-55060-1.

[LCHV03] J. Luc Lamadon, Pablo Cesar, Carlos Herrero, and Petri Vuorimaa, *Usages of a SMIL Player in Digital Television*, 7th IASTED International Conference on Internet and Multimedia Systems and Application, 2003, pp. 579–584.

[LNEEBF$^+$04] Martín López-Nores, Andres Elexpuru-Eguia, Yolanda Blanco-Fernández, José J. Pazos-Arias, Alberto Gil-Solla, Jorge García-Duque, Belén Barragáns-Martínez, and Manuel Ramos-Cabrer, *A Technological Framework for TV-Supported Collaborative Learning*, Sixth IEEE International Symposium on Multimedia Software Engineering (ISMSE'04), 2004, ISBN 0-7695-2217-3, pp. 72–79.

[PGES05] Johan A. Pouwelse, Pawel Garbacki, Dick H. J. Epema, and Henk J. Sips, *The Bittorrent P2P File-sharing System: Measurements and Analysis*, 4th International Workshop on Peer-to-Peer Systems (IPTPS 2005), Lecture Notes in Computer Science, vol. 3640, 2005, ISBN 3-540-29068-0, pp. 205–216.

[SC00] Barry Smyth and Paul Cotter, *A Personalized Television Listings Service*, Communications of the ACM **43** (2000), no. 8, 107–111, ISSN 0001-0782.

[Tan98] Robert Sher Tannenbaum, *Theoretical Foundations of Multimedia*, 1st edition ed., Computer Science Press, New York (NY), 1998, ISBN 0-7167-8321-5.

Citation

P. Cesar, D.C.A Bulterman and A.J Jansen, *An Architecture for End-User TV Content Enrichment*, Journal of Virtual Reality and Broadcasting, 3(2006), no. 9, Special Issue on EuroITV 2006, June 2006, urn:nbn:de:0009-6-7594, ISSN 1860-2037.