Enhancing Social Sharing of Videos: Fragment, Annotate, Enrich, and Share

Pablo Cesar,[A] Dick C.A. Bulterman,[A] David Geerts,[B] Jack Jansen,[A]
Hendrik Knoche[C] and William Seager[C]

[A]CWI: Centrum voor Wiskunde en Informatica
Kruislaan 413
Amsterdam 1098 SJ, The Netherlands
{p.s.cesar, dick.bulterman, jack.jansen}@cwi.nl

[B]Centre for Usability Research
Parkstraat 45 Bus 3605
3000 Leuven, Belgium
david.geerts@soc.kuleuven.be

[C]University College London
Gower S
WC1E 6 BT London, UK
{h.knoche, w.seager}@cs.ucl.ac.uk

ABSTRACT

Media consumption is an inherently social activity, serving to communicate ideas and emotions across both small- and large-scale communities. The migration of the media experience to personal computers retains social viewing, but typically only via a non-social, strictly personal interface. This paper presents an architecture and implementation for media content selection, content (re)organization, and content sharing within a user community that is heterogeneous in terms of both participants and devices. In addition, our application allows the user to enrich the content as a differentiated personalization activity targeted to his/her peer-group. We describe the goals, architecture and implementation of our system in this paper. In order to validate our results, we also present results from two user studies involving disjoint sets of test participants.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Communications Applications - *Information browsers*. H.5.1 [**Information Interfaces and Presentations**]: Multimedia Information Systems - *Audio, Video.* I.7.2 [**Document and Text Processing**]: Document Preparation - *Format and notation, hypertext/hypermedia, Languages and Systems, Multi/mixed media.*

General Terms

Design, Documentation, Experimentation, Languages

Keywords

Asynchronous Media Sharing, Content Enrichment

1. INTRODUCTION

Multimedia content sharing systems have slowly grown in popularity over the past decade. Initially, shared content consisted of photographs or short video clips that contained re-purposed studio content, such as music, music videos and individual news items that were intended to reach a wide, anonymous audience. During the past two years, there has been a growing awareness of the power of personal media productions that are geared to a more limited audience. These include holiday videos, focused product demonstrations and intra-project demonstration videos. Interfaces such as YouTube[1] and MySpace[2] have captured the imagination of a broad community of users and investors: contributing personal media is 'hot'.

While the production model for digital media has undergone a fundamental shift in the form of user-generated content, the user model for viewing media has evolved much more slowly. The dominant UI model for content sharing and social interaction is still centered around a single user who manipulates an interface containing a display, mouse and keyboard. At a time when users have a multitude of media experience interfaces at their disposal — from personal media players, telephones and other mobile devices, through a wealth of PCs and upto high resolution HDTV displays — most sharing systems provide only the limited support for manipulating, customizing and enhancing all of this media being presented. Put another way, users are expected to consume media, not socially interact with it.

This paper describes an architecture and implementation of an inherently more social approach to viewing and sharing media. Building on top of popular on-line video interfaces such as YouTube, our system enables copyright-safe personal recommendation and forwarding of fragments of content in a family/friends network, as well as the further personalization of content via an enhancement interface. The emphasis of the work reported in this paper is on asynchronous sharing and enhancing, when groups of users interact with media over a broad time interval. Our interest extends beyond basic on-line sharing, by providing a useful framework for evaluating extended end-user functionality that may be expected in future media systems. A major difference from previous approaches is the distribution and previewing of content over a collection of personal control devices, in which media viewing is dynamically differentiated between personal and social (group) media viewing.

This paper is structured as follows. Section 2 provides a motivational user scenario for our work. Section 3 surveys existing systems, highlighting features that we feel are ill supported. Section 4 focuses on the architecture and implementation effort of the system. Section 5 reports on two qualitative studies that we have used to evaluate both the feature set supported by our work

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-303-7/08/10...$5.00.

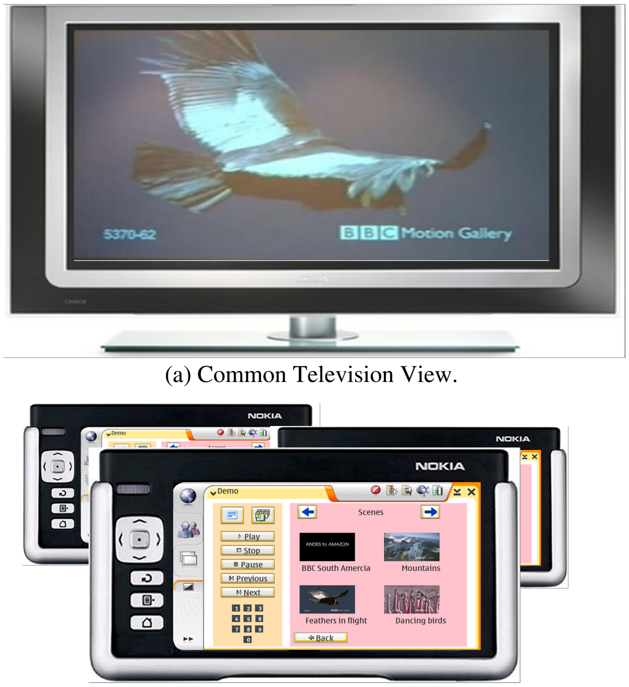
[1] http://www.youtube.com/
[2] http://www.myspace.com/

and the underlying interaction architecture. Section 6 discusses the results and contributions of this paper, highlighting the architectural implications for next-generation video sharing systems. Finally, Section 7 concludes the paper with a discussion of future directions.

2. MOTIVATION

Mark and Katrina are a couple with two children. During a business trip to Vancouver, Katrina views an interesting 60-minute documentary about South America, where the family is going this summer. She decides to share relevant parts of this presentation with her family back home in Amsterdam. This sharing consists of identifying the video, overlaying a personal navigation structure highlighting portions of interest, and adding a number of voice-over annotations to some of the fragments — such as 'I want to go there' or 'we have to visit this spot'. She then sends the family a message with a pointer to "her" version of the video. Note that such enrichments do not generate a new version of the video, they create a (set of) content wrapper(s) containing a link to the base video, along with a navigation map and a set of — possible personalized — annotations. Katrina uses asynchronous sharing: the 9-hour time difference, plus the differentiated content, make synchronous sharing (such as via a chat-based system) impossible.

Back home, Mark and the kids receive the recommendations sent by Katrina. While they have multiple viewing options, they choose to watch the message as a family on their high-definition television set. Mark uses his *Nokia 770* as a secondary screen during the presentation: this screen shows the navigation structure Katrina designed. As illustrated in Figure 1, the content in the shared display is unobstructed for the kids while Mark is navigating from his personal device.



(a) Common Television View.

(b) Personal Secondary Screens View

(Nokia N770 interface).

Figure 1. Personal Selection Interface.

Neither Katrina nor Mark are in the video editing or post-production business. When Katrina made the original recommendation, she used a portable device without keyboard or mouse — only a touch interface. The interface is similar to that shown in Figure 2: a poster image is selected for each navigation point, which is added to the collection of posters for that program based on its temporal positioning in the content. Katrina may add optional text captions, voice-over descriptions and even line-art overlays if she so desires (either for everyone, or for a particular user's secondary screen), but this isn't required. She may even time-limit some annotations, knowing (in this case) that they won't be important after she returns.

This scenario highlights the major contributions of our work:

1. An interface that supports the direct recommendations of content to others in a social network, using light-weight [14,15] or full feature editing systems. The recommendations can reference all or part of a base piece of media by introduce non-destructive fragmenting of content.
2. The development of a personal remote control model that allows users to manipulate, view metainformation and preview content in a mixed social setting, providing a private space in a socially-crowded living-room.

3. BACKGROUND

Modern on-line video services provide social features such as posting comments about a specific video, rating them, and sharing video material with others by embedding a fragment of HTML code that includes the video's location. In spite of their success, there are a number of serious restrictions in such interfaces. First, videos are addressed as atomic objects, without any partitioning in time or space. (We call such partitions *fragments.*) Second, the lack of content-based fragmentation brings with it a lack of intra-object navigation. Third, the lack of user-defined fragmentation results in an inability for users — rather than producers — to share bounded portions of an object among sub-groups of viewers. For shorter videos this might not be needed, but for longer videos it is often useful to define a short fragment of the base video that can be used to illustrate a particular point. Finally, the user cannot customize the recommended video by including, for example, a voice commentary or strategically placed line art overlays.

In order to categorize basic and innovative features provided by current video sharing systems we have selected four representative examples: YouTube, Asterpix[3], Yahoo! Videos[4], and Lycos Cinema[5]. Together these systems provide video description and

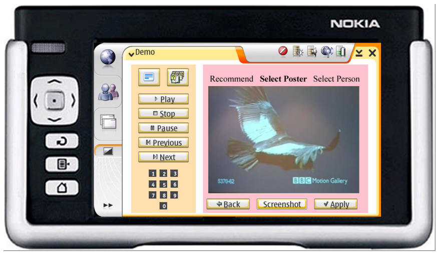


Figure 2. Creating a navigation / recommendation poster.

[3] http://www.asterpix.com/
[4] http://video.yahoo.com/
[5] http://cinema.lycos.com/

manipulation functionality both in a synchronous and asynchronous manner [4]. The intention of this paper is to focus on asynchronous video manipulation capabilities for sharing; we find that this provides the most realistic operational use of a content enrichment facility.

3.1 Asynchronous Manipulation Features

There are several activities performed by the participants in a media sharing system. We differentiate the functionality required by content owners and content users.

Content owners are defined as the initial parties to share a piece of media. They require the following functionality:

- *Upload:* a facility to add media to the content server.
- *Describe:* a facility to describe the entire media object; to be used for searching and for display during viewing
- *Tag:* a facility to add keywords about the media content.

Content viewers are defined as parties that reference the owner's content; they may send others pointers to the content. The basic functionality required by viewers are:

- *Share:* a facility to send recommendations to others. This can be done via an e-mail with a link to the media or as an embedded HTML fragment on a social website.
- *Comment:* a facility to post comments about an object, in whole or part.
- *Rate:* a facility to indicate the popularity of a video. Users might rate it, make it a favorite, or the site might use non-intrusive metrics such as the number of views.

Advanced user features are those extensions to conventional sharing behavior that allow personalized, focused sharing and enhancement of content by non-owners (without compromising the rights that owners have)*:*

- *Fragment:* a facility that allows a user to define one or more ranges of clips within a base media object. These fragments can be explicitly or implicit exposed to parties with whom the user shares content.
- *Annotate:* a facility to add user-generated notes or comments to a particular media fragment. The annotations may be audio, text, image or line art in nature. The annotations may be exposed to all parties sharing the media, or only a user-defined subset.
- *Enrich:* a facility to add new temporal links, subtitles, captions, remixing [19,20], repurposing [16], overlaid media or voice intro to a baseline object. These enrichments may be layered — that is, a particular media object may expose a history of enhancements by various parties.

The purpose of our work has been to design and implement a prototype environment for studying advanced user features, and to evaluate these features in two independent user trials.

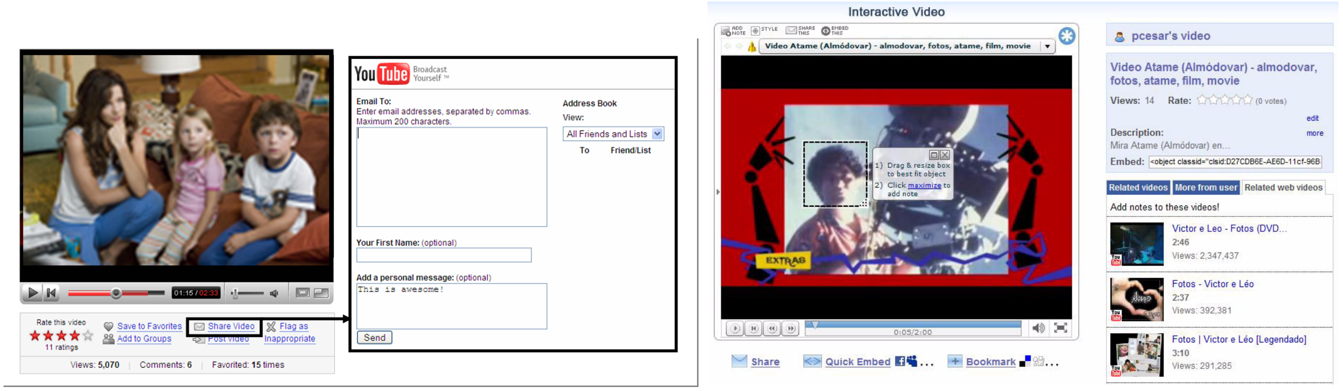
3.2 Selected Examples

Before describing our sharing architecture, we consider the features available in current-generation sharing systems. Of these, YouTube is probably the most popular service for hosting user-contributed content. As shown in Figure 3(a), it includes functionality such as title-based search, simple annotation of a full clip, popularity tracking, content rating, third-party commenting and an external referencing interface for embedding source material in (other) social websites such as FaceBook[6] or MySpace. Other features include flagging content, the possibility of responding a video by uploading a new video, the ability to create personal playlists, and community features for inviting friends or forming groups.

Asterpix is a web service that provides access to media content from different sources such as DailyMotion[7]. It includes functionality similar to YouTube, but adds an important feature: the viewer is capable of enriching the video by adding temporal links such as commentaries or related videos. As shown in Figure 3(b), the temporal links are rectangular shaped and when surrounding an object the system uses a tracking system to automatically follow such object. After the enrichment process, Asterpix uploads a new version of the video under a unique URI.

Yahoo! videos provides a similar interface as the one of YouTube, but includes a 'tag it' feature, for facilitating the indexing and searching of videos. In addition, personalization functionality such as to select a different thumbnail for a video when embedding a video in a social website, is provided. While watching a clip using Yahoo! videos the user can start Jumpcut[8], a web-based video editor, in order to change from a viewing mode to an authoring mode.

[6] http://www.facebook.com/
[7] http://www.dailymotion.com
[8] http://jumpcut.com/



Email To:
Enter email addresses, separated by commas. Maximum 200 characters.

Your First Name: (optional)

Add a personal message: (optional)
This is awesome!

Video Atame (Almódovar) - almodovar, fotos, atame, film, movie

1) Drag & resize box to best fit object
2) Click maximize to add note

Video Atame (Almódovar) - almodovar, fotos, atame, film, movie
Views: 14 Rate: ☆☆☆☆☆ (0 votes)
Description:
Mira Atame (Almódovar) en...
Embed: <object classid="clsid:D27CDB6E-AE6D-11cf-96B

Add notes to these videos!

Victor e Leo - Fotos (DVD...
2:46
Views: 2,347,437

Fotos - Victor e Léo
2:37
Views: 392,381

Fotos | Victor e Léo [Legendado]
3:10
Views: 291,285

(a) YouTube media sharing interface

(b) Asterpix media ehnacement interface

Figure 3. Two media interaction interfaces.

In contrast to the other examples, Lycos Cinema service provides a synchronized experience. Lycos Cinema can be classified as a virtual cinema theater, in which the user can invite people to join for watching a movie together. It offers synchronous features such as chat, presence awareness, and join invitations, similarly to other social interactive television systems such as Joost[9] and Motorola's Social TV/TV2 [10]. Lycos Cinema includes as well an asynchronous feature, in which the user can clip a video. Moreover, similar to Zync[10] it allows synchronized watching of online video. Another application along the same lines is CollaboraTV [11], which allows a user to add temporal comments that will be shown during the selected video fragment when the video is watched by the peers of the user.

3.3 Under-Supported Features

Many sharing sites provide institutionalized support for community building around their media. The effectiveness of such support is questionable[9]. Latest data suggests that '...users are directed to YouTube by friends sending them specific videos' [7] and that 'the aggregate views of these linked videos account to 90% of the total views' [3]. This confirms our belief that people are directly guided by other people in the media selection process. Our work illustrates two major directions in which sharing facilities of media can be improved: media manipulation and the ability to support shared but differentiated social viewing.

Table 1 shows a comparison among current systems for a selected set of functionality. The table assumes that all the systems already provide the basic functionality of media description support for facilitating searching. Only a minority of the systems allow fragmentation, annotation, and enrichment of media. Moreover, such systems do not provide support for differentiated shared viewing.

TABLE 1: Comparison of Selected Functionality Across Representative Systems.

	Share	Annotate	Fragment	Enrich	Multi-Layered Enrichments	Differentiated Social Sharing
YouTube	+	-/+	-	-/+	-	-
DailyMotion	+	-	-	-	-	-
Asterpix	+	+	-	+	-	-
Yahoo!	+	-	-	-	-	-
Lycos	+	-	+	-	-	-
CollaboraTV	-	+	-	+	-	-
Zync	+	-	-	-	-	-
Joost	+	-	-	-	-	-

Of the characteristics in Table 1, the facilities for sharing, annotation and fragmentation have been cover earlier in this article. By *multi-layered enrichment*, we define a facility that allows multiple enhancements that are logically layered to be managed from a single user interface. The opposite of this functionality is that each enrichment gets published as a separate, unrelated object. Such multi-publishing is not uncommon: most videos on YouTube already have one to four aliases [3]. Since these are all published as independent objects, no content management user interface is available to aid in fragment searching.

The final characteristic in Table 1 is labelled differentiated social sharing. This facility allows one viewer to obtain — either by direct request or as a property of the recommendation — extra information that may not be visible to others, even when all viewers are watching a single common media stream on a shared display such as a TV. By using a secondary display device, users can obtain additional information on the displayed content without disturbing others, they may be able to manipulate a separate control interface (for example, Mark's use of a secondary display for navigation in Section 2), or it may allowed targeted content to one of the shared viewers, such as a personal hint in an otherwise shared-experience game. Current research indicates that 'our devices should collaborate to support a notion of user-centric activities that span multiple devices' [6].

4. ARCHITECTURE AND INFRASTRUCTURE

The goal of our research has been to evaluate the usefulness and the feasibility of providing media manipulation functionality as a spontaneous activity in a social environment, like the living room. In order to support this evaluation, we implemented a working prototype that allowed users to view third-party media, to construct content fragments and to add annotations, and to share these annotated fragments within a small-scale social network. In this network, one size did not need to fit all: individual users could have tailored messages, and all users had the option of using a secondary screen for supporting navigation.

This section elaborates on the architecture of our prototype, with special focus on content modeling, the system software, and the social network aspects. The main contribution of this prototype is the development of an open-ended testbed system in which the creation of media fragments are easy to perform, in which such fragments can be enriched and targeted to specific viewers, in which video manipulation does not result in a new encoded version of the media, and in which and all such activities can be performed in the context of differentiated social sharing.

The results of the UI aspects of this work are summarized in Section 5. There are also more generic results to report. One of this is that, by making media annotation a non-destructive, layered task, the number of video servers required to store media content can be significantly reduced. At the same time, real personalization of video content, as a postcard send by a friend, is provided. We find these aspects to be significant.

4.1 Design Goals

The primary focus of the work reported in this paper is on the development of a scalable interface model for capturing control sequences to support intra-program selection/navigation, content enhancements and peer group sharing.

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

[10] http://timetags.research.yahoo.com/zync/

The design goals that guided this work were:

- *Investigate distributed, concurrent control*: a home media environment is a complex combination of people, devices and content. The collection of people who consume content will vary from a single person to a local collection of family members to a distributed collection of remote viewers. Rather than assume a single point of control (with a single hand-held device in one room), we wanted to study a broader base of content and control interaction.
- *Separate rendering of control information from (shared) content*: given the diverse user, content and rendering environments, we wanted to explicitly separate out the viewing and interacting with control information from the viewing and interacting with media content.
- *Focus on individual users instead of shared devices*: most digital television systems are centered around a set-top box: this box is connected to an external content stream, it puts content on a TV display and it interacts with the user's remote control. Architecturally, the user is an appendage to the system. In many households, there are multiple users who each have their own content needs. We wanted the user to be central in our system.
- *Enable a framework for micro-recommendations*: content recommendations are typically managed both at a device level - that is, figuring out which content the set-top box should store - and at the full-program level. We wanted to develop an environment where individual programs could be partitioned (by the user or system) into a collection of fragments of interest, and in which collections of programs could be grouped in to packages. This framework would be the basis for future study on automated micro-level recommendations.
- *Enable a framework for sharing of recommendations*: it is becoming common for recommender systems to gather content for a user, and for a user to rate content. We also wanted to investigate ways of having individual users send personal recommendations within their family or social network.
- *Interface with practical user environment*: we wanted our system to build on existing use models for broadcast content. Rather than assuming that conventional broadcast outlets would disappear, we wanted to study ways of working within a common home consumer electronics framework.

The last design goal had the consequence that we decided to focus on relatively passive viewing of content in a mixed personal/social setting. The mixed setting provides a wealth of interesting interaction problems that are not encountered when one assumes that a user is sitting behind a personal computer. We feel that results from the family couch can scale to the PC world, but that PC-based solutions can not scale to the passive couch-top environment. A non-technical goal was to frame the technical progress in our research in a context that would likely have broad impact on the way that real people consume real media. This is the motivation for the integration of several external user testing.

4.2 Content Modeling

Currently, online video systems widely support the capability of sharing video material with others either sending an e-mail or by embedding it into a social network. As indicated before, the major form of accessing video on the web is via this sharing capability [7]. Moreover, the generation of new videos has other negative implications: to saturate the already overloaded video servers [3] and the impossibility to manage video editions performed by different parties. The saturation of video servers do not have only an economic implication, but as well an authorship implication. That is, the owner of a video does not have any idea of who and how other people is manipulating his/her production. The latter implies that a viewer cannot selectively view enrichments. Moreover, essential information (such as who has manipulated what when), very useful for archival and indexing purposes, is lost when a new encoded version of the video is provided.

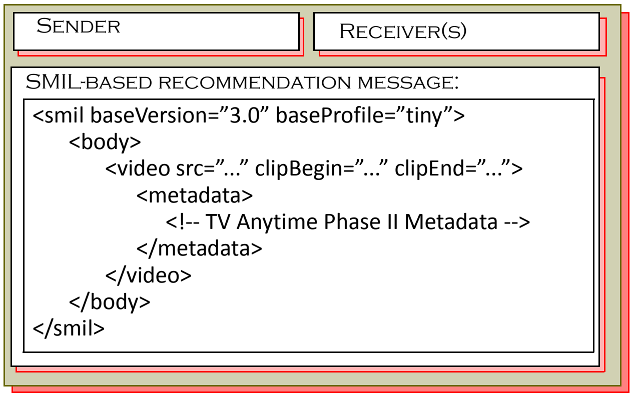
Implementing video manipulation activities such as fragmenting, annotating, and enriching as a non-destructive operations provides a solution to these problems. Current research provides two major directions: temporal URIs [17,18] and structured media documents [5,8]. The first approach permits sharing video fragments by replacing a base URI with an annotated URI that indicates the starting and ending point of the fragment. Such solution can be easily implemented and enhances the video sharing functionality. Nevertheless, because of the amount of information one can fit into a URI is limited, such solution is not powerful enough for more complex video manipulation capabilities. Furthermore, essential metadata such as who has created the fragment, when, why, and to whom might be very difficult to include. Hence, our system relies on structured documents for providing enhanced media sharing features in the form of SMIL 3.0 and TV-Anytime Phase II.

In our environment, the starting point is an unstructured, or raw, video. We then define a structured shell in the form of a SMIL description by using authoring templates. If an automatic scene selection tool is available, the video can be further structured into sections of interest. All navigation points are encoded as a series of content events that are placed in a dynamically-created SMIL file. Each of these function, illustrated in Figure 2, is captured via a *poster interface*. Each poster contains the following information:

1. *the temporal moment of the navigation point*: this is the time at which the navigation point appears in the content.
2. *enrichments*: Any text, ink, audio, link enhancements that have been added by the user. Note that similar content is generated for studio-produced navigation points if they are distributed as TV-Anytime markup.

Our content hierarchy allows navigation/recommendation points to be described at three levels: the package level, where collections of programs are stored and grouped by package name; the program level within packages, where individual programs are identified; and the fragment level within programs, where individual navigation points are identified. It is a longer-term interest to have the partitioning of content between packages, programs, and fragments occur automatically. However, at every level, a user should be able to exert a personal influence over content scheduling.

When video manipulation takes place, the SMIL file is used to control the interactive display at the client. The SMIL file is transformed into a TV-Anytime description at the point that one or more recommendations are transformed into a recommendation message as illustrated in Figure 4. (Details are available in [2]**.)** At no time is actual content integrated, manipulated or shared among users. All communication is done via an abstracted SMIL file or as a set of TV-Anytime markup. Hence this is a non-destructive



```
<smil baseVersion="3.0" baseProfile="tiny">
    <body>
        <video src="..." clipBegin="..." clipEnd="...">
            <metadata>
                <!-- TV Anytime Phase II Metadata -->
            </metadata>
        </video>
    </body>
</smil>
```

Figure 4. Micro-Personal Recommendation Modeling.

solution for the video manipulation challenge. One very important advantage is that viewers might select to see the manipulations performed by a specific user or turn-off all enrichments at once.

Figure 4 shows the structure of the SMIL-based recommendation message resulting from our lightweight manipulation of video content. As shown in the figure, SMIL 3.0 is used as the host language; the end-user manipulations on the video are encoded in the SMIL file, using SMIL constructs. They do not change the base media objects in any way. For example, the end-user can fragment the video by using clipBegin and clipEnd attributes. Moreover, the end-user can insert audio overlays by using the audio element or any other kind of multimedia overlays. The actual process of SMIL creation/manipulation is hidden from the end-user, who simply sees an interactive UI on the secondary device.

An important feature included in SMIL 3.0 is the expansion of the metainformation facilities. In previous versions, meta-information was restricted to the head element, meaning all metainformation referred to the whole document. SMIL 3.0 now allows placing meta-information on any element within the document body. This makes it possible to provide information on semantic intent within the presentation by making the binding of that information with relevant nodes more local. In Figure 4, the video fragment indicated in the video element is annotated with the universal TV-Anytime content identifier, CRID, in order to help the localization process. Each fragment of video could be further annotated by using, for example, Friend of a Friend (FOAF) for indicating the creator of the fragment or the person intended as recipient.

4.3 Home Network

Another challenge addressed by this article is the dynamic distribution of media content and control to the most suitable device at home. The goal is to dynamically monitor the end-user environment for available rendering/interactive devices. Moreover, such devices should provide an accurate description of their physical, rendering, and interactive capabilities. Based on those descriptions, the context of the user, and the nature of the media to be watched, we can take a decision on the actual distribution mechanisms to be utilized. It is out of the scope of this paper to describe the device description mechanisms and decision algorithms, the interested reader can refer to [12]. Device discovery is done by an exchange of invitations to join the network using Bluetooth, Wireless LAN or IP Multimedia Subsystem (IMS). After devices have been discovered, they provide the descriptions of their physical, rendering, and interactive capabilities. In order to support a wider set of devices than mobile handsets, we have extended UAProf[11] for describing devices' capabilities. Applications are described as a Software Oriented Architecture (SOA), in which the input and the output model are described using XML. Based on the device and service descriptions, a matching algorithm together with basic recommendation facilities are used in order to take a decision to where to render the media and how to provide user interaction in the most suitable way depending on the context of use.

Within the home, the central content storage/management and service publishing component is a home media server. This server can ultimately be implemented in many different forms (as a PC Media Center, as a conventional set-top box, as a network controller hidden in a utility closet). The result is a hybrid architecture in which devices can directly be connected to each other, in a Peer-to-Peer (P2P) fashion, within the home. Our main concern was not to study the commercial models for home media servers, but (1) to study a model in which multiple control clients could be managed in a home environment and (2) dynamically distribute the media content to the most suitable rendering device(s). 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.

Apart from device discovery, dynamic content distribution, and service publishing, the server performs the following functions:

1. *It connects to an external content pool*: This pool consists of a connection to (digital) broadcast content, to a peer-to-peer content infrastructure for sharing non-professional content, and to a collection of physical optical devices such as DVD and Blu-Ray disk players.
2. *It caches content that is differentiated per user*: Each of the users of the environment is managed separately. Each maintains their own content preferences and their own user group. For each user, non-protected content is cached using PVR-like functionality.
3. *It implements a content recommender system*: the server manages the recommender environment for the home. This includes communication with external recommender systems, forwarding and receiving recommendation messages, and enabling users to add and share personal recommendations within program content.
4. *It provides a home management interface*: Not all content recommendations generated by external systems will actually be suitable for all members of a family or social network. In addition to simply storing recommendations, our server environment provides a management system in which a hierarchy of control allows privileged users to manage (override, augment) the recommendations provided for others.
5. *It communicates with the client devices within the home*: this includes communicating with the distributed remote control devices and actually sending information to clients.

[11] http://www.openmobilealliance.org/

Our architecture has been designed to be aware of DRM issues in the home. All operations on actual media content are abstracted from the actual media encoding into a higher-layer structure. A portion of this structure uses the TV-Anytime specification for program and package descriptions. The local user operations are implemented by dynamically generating SMIL presentations that describe the transient structure of content modifications and annotations within a program. The work reported in this article does not focus on recommender strategies (other than gathering and forward personal micro-recommendations). Our focus is on the communication aspects of distributing control in a home environment. This is a function of the communication with the collection of client devices.

Content navigation is performed based on the set of posters that have been defined for a particular program. These posters may be defined by the content owner (in this case, the BBC), they may be automatically induced or they may be defined by a viewer. When the user fragments a video, a screenshot poster is taken from the primary screen. The user may enrich the poster, or the video fragment, with additional information (annotations), line art or an audio voice-over.

4.4 Social Network

Current online video sharing systems are based on two models: person-to-person and person-to-world. The first one makes use of an active e-mail address in which a link to the video is included, while the second one provides an HTML fragment to be posted in a webpage in which a link to the video is embedded. In this paper, we argue that a broader social network architecture is needed. Our architecture is based on social network capabilities in which individuals can be gathered as solitary users (in front of their personal devices) or as a group of people such as families that interact with a common social medium such as the TV. We use XMPP[12]and the Google extensions[13] to retrieve contact information about the peers.

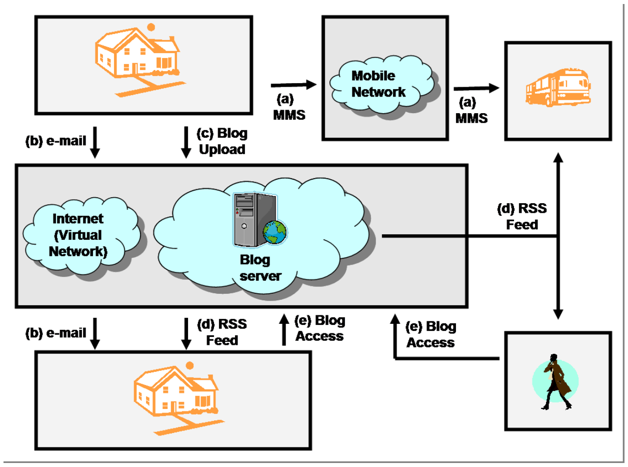


Figure 5. Micro-Personal Recommendation Sharing.

Figure 5 divides the social communication model into three categories: immediate communication/mobile, immediate communication/static, and world communication. In Figure 5(a), the home server is connected to a message gateway[14]that generates a Push WAP message in the form of a SMS, so the recipient can display the video fragment using his/her mobile phone. In Figure 5(b), the message can be sent in the form of an e-mail to the recipient's home server, which then informs the user using his/her active connection device (TV/PC). Finally, the sender can post the message in his/her Blogger[15] account — as illustrated in Figure 5c — in case he/she wants to share the enriched fragment of video with the world [2].

4.5 Current Status and Limitations

In order to gather useful information from the user studies, a full working implementation of the system was developed. The implementation is composed of a number of elements as introduced previously. The software base for the implementation of the fragmenting/enriching tool is the Ambulant Player[16], an open-source media playback engine. Functional extensions have been made to support the functionality presented in this paper. Individual clients (e.g., Nokia 770) make use of local applications that perform enriching operations. The range of operating systems and media libraries make the development of custom clients unavoidable, but all of the control operations have been harmonized in a layer of interface specifications.

Currently, the system status is a working prototype, suitable for use in guided field trials. The initial encouraging results from the user testing described in this article has motivated an extended investment in the demonstrator to improve the stability and performance of the system, as well as to extend its functionality. One essential intention is to provide the infrastructure to enable end-user enrichment and sharing from popular Internet video feeds, such as YouTube and the creative commons content provided by the BBC. Another extension is to migrate the current stand-alone Ambulant engine as a portable plug-in player for browser integration. This work is anticipated within the coming year. Once completed, a full public implementation of the Ambulant Annotator is expected. Developments of the Ambulant Engine are updated on the Ambulant website.

5. EVALUATION RESULTS

In order to evaluate the functionality presented by our architecture, we submitted our system to user testing, focusing on a qualitative analysis, in two different countries: the UK and Belgium. The tests were temporally separated, allowing the systems to be improved and more functionality included based on the results from the first tests. For example, during the test in UK functionality such a sharing was not fully available for the participants.

5.1 UK Study

We modeled a representative user community by constructing twelve groups of friends with up to three people to participate as paid subjects in the study. In total 27 participants took part in the

[12] http://www.xmpp.org/

[13] http://code.google.com/apis/talk/jep_extensions/extensions.html

[14] We use Clickatell (http://www.clickatell.com/)

[15] We use the Blogger API (http://code.google.com/apis/blogger/)

[16] http://ambulantPlayer.org/

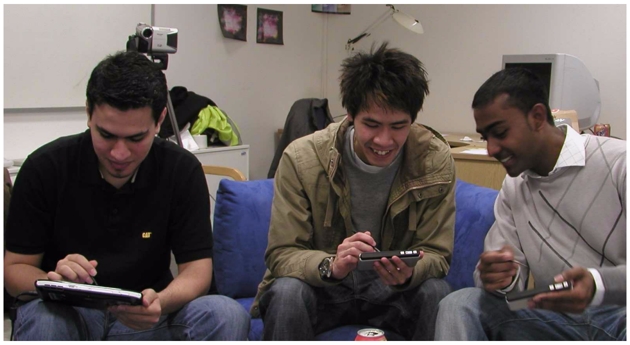


Figure 6. Recommending the Content.

test (average age: 28 old; meridian age: 23 years old). As for gender, 18 participants were male and 9 participants were female. One test session lasted for 90 minutes and were audio and video recorded. Finally we administered questionnaires which collected preferences about the individual preferences of the features our system. A viewing environment was developed consisting of the media server, and three hand-held control devices. The shared display was a 50" 16:9 screen. We encouraged them to explore the system during a a co-located test situation and to comment on the features or any problems they ran into at any time. A photograph of one of the evaluation groups is shown in Figure 6.

The co-located test situations were followed up by a group discussion phase, in which we asked questions such as What did you like about the system? What other things would you like to be able to do? What didn't you like about the system? What were the most annoying things that should be changed? Would you be interested in using such a system at home or elsewhere? What was your experience in a group with multiple remote controls like? If you could annotate content how would you like to do it? When did you make your decision on what to watch next? The user groups provided a rich and detailed set of comments on many aspects of the system. The results on the participants' preferences for the available and potential features is shown in Figure 7.

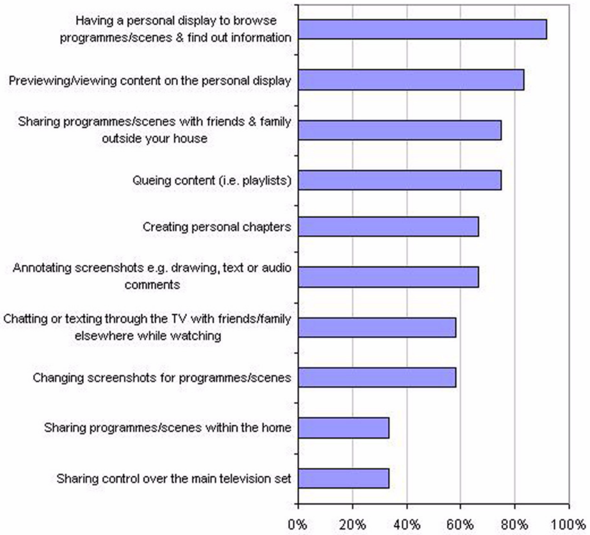


Figure 7. Expected Funcionality From the First User Testing (in percentage).

The following conclusions came out of the user study:

- *Benefits of a secondary personal display*: For many participants, a key advantage of the system was the fact that they were able to browse content and choose programmes without interrupting what was playing on the main screen. Thus, in a group viewing situation, everyone could browse for themselves without interrupting others. However, this aspect did not appeal to everyone. A number of participants did not want to browse while something was already showing.
- *Fragment and Enrich Videos*: Most users liked the idea of having such functionality, either for creating personal bookmarks or for defining personal scene selection points. Some users wanted to define not only a single point, but also a range of content that could be recommended as a whole sub-item. In other words, they wanted functionality that enabled them to gain quick access to certain points in the video.
- *Sending and receiving fragments of videos*: Participants were generally very enthusiastic about the idea of sharing content. For the most part, they viewed this as something they would like to do with friends or family who did not live with them although they could also see the value of sending content to people living in the same household if they were in another room or else not currently present.

Having a secondary display that allowed for browsing and annotating content, as well as sending and receiving fragments of videos, appealed to the majority of the participants. Sharing within the home did not have a large appeal to the groups of friends but this might be different for families and in shared households. We can see the desire to reciprocate recommendations as a potentially strong driver for the adoption of impromptu recommendations from the couch. As observed in other studies the social norms around gift-giving include the demand for reciprocity [21].

5.2 Belgian Study

A second study was carried out in Belgium approximately 10 months later, with a system containing extra functionality (identified during the UK study), such as improved delivery across heterogeneous devices in an IMS environment.

Twelve groups of users were recruited for this test. One test session involved one single group, lasted for two hours, and was audio and video recorded. Each group consisted of two to five people that knew each other well, either as friends or as family members, and sometimes a mix of both. In total 36 participants took part in the test, with ages ranging from 14 to 72. As for gender, 13 participants were male and 23 participants were female. The test sessions took place in a simulated living room and consisted of four main parts: an explanation of the system, a co-located test situation, a remote test situation, and a group interview. After the second and third part, a couple of questionnaires were filled in by each participant.

The first part of the test was used for an extensive explanation of the system, to make sure all participants could use the system well (the purpose was not to improve the usability, but to examine the way users would experience the system). During the second part of the test, all members of the group stayed in the same room, and were asked to browse through the available content, select these items they wanted to share with someone they knew, clip and annotate these clips if they wanted to, and finally send them to someone specific. In the third part of the test, the group was split into two sub-groups. One sub-group (sometimes a single person) stayed in the simulated living room, whereas the other sub-group (or single person) was led to a separate room. For this part of the test, the participants in the living room were asked to edit and annotate clips to send to the participants in the other location. This way, the participants knew their edited clips were really received by someone. Finally, the fourth part of the test consisted of a group interview lasting about twenty minutes, which covered several aspects such as reasons for sending and annotating clips, the use of separate screens, or useful extra functionality.

When analyzing the results, the video observations and participant's answers were coded and clustered. Only those concepts that were observed in or mentioned by at least two groups were taken into account. Some issues were repeated by almost all groups, and were treated as more important when interpreting the results. In the discussion below, quotes from single persons are used to illustrate a greater concept mentioned by several participants.

The following conclusions, relevant to the main topics of this paper, came out of this user study:

- *Separate screens and differentiated sociability:* The fact that a separate screen was used for creating, annotating and sharing clips was considered beneficial for letting other viewers continue watching the program. Two participants referred to how changing the settings of current video or DVD players disturbs the viewing experience, because a menu comes up on the screen, overlaying or even replacing the television program. During the observations, it was also clear that the private screen stimulated interaction between the participants. When one person was creating clips or annotating, the person sitting next to her would look over her shoulder to see what she was doing, comment on what she should do next or even point at the screen to indicate certain actions to be taken. The private device was often passed on to or claimed by the other person, so he could take over control.
- *Annotating clips:* The possibility of making annotations was positively perceived by most participants. Several people explicitly said they liked the personal aspect of making annotations. They said it allows you to give your own opinion, a commentary or a suggestion. One person said it was the same as talking to each other while watching, but then to people at another location. Another participant would put a signature on the clip, so the recipient would be sure who the sender was, and that it was not spam. Some participants said they would write the reason for sending something, as sending the clip alone would be unclear. Figure 8 shows the results of the answers to questions about how good or bad the participants rated annotating clips and having a secondary screen. The results show that most participants preferred annotating clips as well as having a secondary screen.

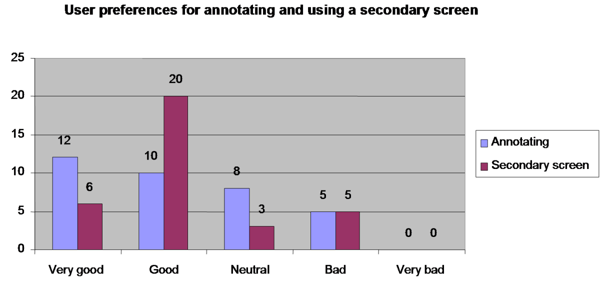


Figure 8. User preferences for annotating and using a secondary screen.

6. CONCLUSIONS

The personal content management application described in this paper is one example of how relative passive users can be allowed to interface more directly with content they watch. In this paper we do not report on (yet another) authoring tool [13,19] or an easy-to-use media browsing tool [1]. Instead, we proposes extending popular video sharing systems with video manipulation features such as video fragmentation, annotation, and enrichment as a spontaneous activity [14,15]. This section analyzes different issues that we have learned from the implementation and evaluation of the system.

Interestingly, as shown in Table 2, both studies provide similar results on the user expectations towards media sharing systems. We can identify a number of commonalities between the results of the tests.

Video fragmentation and enrichment: users miss the capability to fragment and enrich content in current commercial systems. The qualitative tests indicate that people want such functionality and that, when available, it will be widely used. Users like the capability of creating clips from videos, either for better navigation or to be able to send specific parts of a program to someone. Moreover, they enjoyed being able to enrich such clips, and sending them to other people. This is much in line with the current social practices of television watching, e.g. talking during a program but also discussing it afterward.

Secondary screen: It seems clear that in both tests, the users liked the idea of a secondary screen for other usages than controlling media. In both tests, participants said that a secondary screen would allow them to edit and send content without interrupting co-viewers who are watching television. As watching television in a social setting with multiple people (either friends or family) is still a dominant paradigm, this is an important conclusion that supports the design choices for our system.

The user testing provides us useful data on the architectural extensions needed for enriching the sharing of media experience. Such extensions include: inclusion of social features, better integration with current home networks, and modifications on the sharing interfaces.

TABLE 2: User Studies Results.

	UK Test	**Belgian Test**
Secondary Screen	A key advantage of the system was the fact that they were able to browse content and choose programmes without interrupting what was playing on the main screen	It was considered beneficial for letting other viewers continue watching the program
Clip Creation	Most users liked the idea of fragmenting and enriching the video.	People really wanted to refer to specific parts of a video

From an architectural perspective, we feel that one of the fundamental limitations of current commercial systems is the lack of use of a structured container in which media enhancements can be encoded. By viewing media as a closed container, systems limit the flexibility that users (rather than owners or producers) have in encoding personal extensions.

At the same time, current delivery systems do not integrate well with the home environment, meaning that it is difficult to synchronize media consumption across devices at home. This limitation restricts the quality of experience, since in most cases the users might want to use the television screen to watch video material. Our system attempts a seamless integration in the home environment providing the possibility of using primary and secondary screens for video consumption and video manipulation activities.

We expect to continue our efforts to gain a deeper understanding of the types of content management and manipulation that can be support in a dynamic, social community. This includes new systems for managing recommendations and migrating content. We also expect an impact on media standards such as SMIL and TV-Anytime for encoding runtime behavior and persistent sharing of recommendations and annotations.

7. 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.

8. REFERENCES

[1] G. D. Abowd, M. Gauger, A. Lachenmann, *The Family Video Archive: an annotation and browsing environment for home movies*, Proc. 5th ACM SIGMM Int. Workshop on Multimedia Information Retrieval, 2003, pp 1-8.

[2] P. Cesar, D.C.A. Bulterman, and A.J. Jansen, *Social Sharing of Television Content: An Architecture*, Proc. IEEE Symposium on Multimedia (ISM Workshops 2007), TaiChung, Taiwan, December 10-12, 2007, pp. 145-150

[3] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, *I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system*, IMC '07: Proc. 7th ACM SIGCOMM Conf. on Internet Measurement, 2007, pp. 1–14.

[4] K. Chorianopoulos, *Content-enriched communication supporting the social uses of TV*, Journal of the Communications Network **6** (2007), no. 1, 23–30.

[5] R. M. Resende Costa, M. Ferreira Moreno, R. Ferreira Rodrigues, and L. F. Gomes Soares, *Live editing of hypermedia documents*, Proceedings of ACM DocEng, 2006, pp. 165–175.

[6] D. Dearman and J. S. Pierce, *It's on my other computer! : computing with multiple devices*, CHI '08: Proc. 26th SIGCHI Conf. on Human Factors in Computing Systems, 2008, pp. 767–776.

[7] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, *Youtube traffic characterization: a view from the edge*, IMC '07: Proc. 7th ACM SIGCOMM Conf. on Internet Measurement, 2007, pp. 15–28.

[8] R. Goularte, E.S. Moreira, and M.G.C. Pimentel, *Structuring interactive TV documents*, Proc, of ACM DocEng, 2003, pp. 42-51.

[9] M. J. Halvey and M. T. Keane, *Exploring social dynamics in online media sharing*, WWW '07: Proc. 16th Int. Conf. on the World Wide Web, 2007, pp. 1273–1274.

[10] G. Harboe, C. J. Metcalf, F. Bentley, J. Tullio, N. Massey, and G. Romano, *Ambient social TV: drawing people into a shared experience*, CHI '08: Proc. 26th SIGCHI Conf. on Human Factors in Computing Systems, 2008, pp. 1–10.

[11] C. Harrison and B. Amento, *CollaboraTV: Using asynchronous communication to make TV social again*, EuroITV '07: Adjunct Proceedings of the European Conference on Interactive Television, 2007, pp. 218–222.

[12] C. Hesselman, P. Cesar, I. Vaishnavi, M. Boussard, R. Kernchen, S. Meissner, A. Spedalieri, A. Sinfreu, and C. Raeck, *Delivering Interactive Multimedia Services in Dynamic Pervasive Computing Environments*, Proc. Int Conf on Ambient Media Systems, 2008.

[13] X.-S. Hua and S. Li, *Interactive video authoring and sharing based on two-layer templates*, HCM '06: Proc. 1st ACM Int. Workshop on Human-Centered MM 2006, pp. 65–74.

[14] T. Kindberg, M. Spasojevic, R. Fleck, and A. Sellen, *I saw this and thought of you: some social uses of camera phones*, CHI '05: CHI '05 Extended Abstracts on Human Factors in Computing Systems, 2005, pp. 1545–1548.

[15] D. Kirk, A. Sellen, R. Harper, and K. Wood, *Understanding videowork*, CHI '07: Proc. SIGCHI Conf. on Human Factors in Computing Systems, 2007, pp. 61–70.

[16] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert, *The DIVER project: interactive digital video repurposing*, IEEE Multimedia, 11(1), 54-61, 2004.

[17] S. Pfeiffer, C. Parker, and C. Schremmer, *Annodex: a simple architecture to enable hyperlinking, search & retrieval of time–continuous data on the web*, MIR '03: Proc. 5th ACM SIGMM Int. Workshop on Multimedia Inf. Retrieval, 2003, pp. 87–93.

[18] L. Rutledge and P. Schmitz, *Improving media fragment integration in emerging web formats*, Proc. ACM Multimedia Modeling Conference, 2001, pp. 147–166.

[19] R. Shaw and P. Schmitz, *Community annotation and remix: a research platform and pilot deployment*, HCM '06: Proc. 1st ACM Int. Workshop on Human-Centered MM 2006, pp. 89–98.

[20] D.A. Shamma, R. Shaw, P.L. Shafton, and Y. Liu, *Watch what I watch: using community activity to understand content*, MIR '07: Proc. of the International Workshop on Multimedia Information Retrieval, 2007, pp. 275-284.

[21] A. S. Taylor and R. Harper, *Age-old practices in the 'new world': a study of gift-giving between teenage mobile phone users*, CHI '02: Proc. SIGCHI Conf. on Human Factors in Computing Systems, 2002, pp. 439–446.