An Architecture for Non-intrusive User Interfaces for Interactive Digital Television

Pablo Cesar[1], Dick C.A. Bulterman[1], Zeljko Obrenovic[1], Julien Ducret[2], and Samuel Cruz-Lara[2]

[1] CWI: Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
p.s.cesar@cwi.nl, dick.bulterman@cwi.nl, zeljko.obrenovic@cwi.nl
[2] LORIA / INRIA Lorraine
Campus Scientifique - BP 239, 54506 Vandoeuvre-lès-Nancy, France
Samuel.Cruz-Lara@loria.fr, Julien.Ducret@loria.fr

Abstract. This paper presents an architecture for non-intrusive user interfaces in the interactive digital TV domain. The architecture is based on two concepts. First, the deployment of non-monolithic rendering for content consumption, which allows micro-level personalization of content delivery by utilizing different rendering components (e.g., sending video to the TV screen and extra information to a handheld device). Second, the definition of actions descriptions for user interaction, so that high-level user interaction intentions can be partitioned across a personalized collection of control components (e.g., handheld device). This paper introduces an over-all architecture to support micro-personalization and describes an implementation scenario developed to validate the architecture.

1 Introduction

Watching television is usually a shared experience: a family (or group of friends) watch a shared output device (the screen) and interact with that device using a single shared control object (the remote control). A social protocol exists that determines the content displayed on the output device. This protocol is required because there is no differentiation between common content for the group and optional information that may be of interest to only a sub-set of the group. In this paper, we propose a model in which personal devices and sensory enhanced everyday objects can be used to render and interact with television content. We refer to this model as a non-intrusive user interface because the selection and interaction with personal content streams do not disturb the television experience of other viewers.

Our research considers the last stage of the media distribution chain, when the user is actually consuming and interacting with TV content. In order to limit the scope of this paper, we constrain our interest to the interactions with an active content stream that is stored on a local home media server such as a Personal Digital Recorder (PDR).

P. Cesar et al. (Eds.): EuroITV 2007, LNCS 4471, pp. 11–20, 2007.

2 Related Work

The main motivation of our research is to provide a user, or group of users, with advanced control over the content they are viewing. We share the view presented by Baker [1] that current intrusive interfaces are not the solution. We feel that a value-added experience must rely on non-intrusive user interfaces for content selection, navigation, rendering, and interaction.

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 [2]. 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. To add personal value to the viewing experience, we feel that micro-level content personalization is also required.

Macro-level content selection is supported by the TV-Anytime Forum[1]. Interesting research in this area includes the UP-TV project. The UP-TV project [3] presents a program guide that can be controlled and managed (e.g., delete programs) from personal handheld devices. Our work also studies navigation using personal devices, but focuses on a finer level of granularity: how fragments within a program can be managed and personalized, and then controlled using a variety of light-weight end-user devices.

In order for micro-level personalization to be effective, a structured content model is useful. Several approaches to content structuring have been proposed world-wide: Digital Video Broadcasting - HyperText Markup Language (DVB-HTML)[2] (Europe); Advanced Common Application Platform - X (ACAP-X)[3] USA; and Broadcast Markup Language (BML)[4] (Japan). These solutions are based on a number of World Wide Web Consortium (W3C) technologies, such as eXtensible Markup Language (XML), Cascading Style Sheets (CSS) and the Document Object Model (DOM).

Unfortunately, they also rely on a non-declarative framework for modeling the temporal relationship between media elements in the document, such as ECMAScript or Java. Because declarative data is easily converted to other formats, is more likely to be device-independent and tends to live longer than programs [4]; we use a complete declarative solution, SMIL.

Regarding content rendering and interaction, Jensen [5] defines three basic types of interactive television: enhanced (e.g., teletext) personalized (e.g., pause/play content stream using a PDR), and complete interactive (i.e., return channel). In this paper, 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

[1] http://www.tv-anytime.org
[2] http://www.mhp.org/
[3] http://www.atsc.org/
[4] http://www.arib.or.jp/english/index.html

consumption, creation and sharing. A key element of our paradigm is that the television viewer remains essentially a content consumer who participates in an ambient process of incremental content editing. Similar results, but intended to broadcasters, has been presented by Costa [6].

Finally, Chorianopoulos argues that traditional metaphors cannot be applied to digital television [7]. He proposes a metaphor called the Virtual Channel: dynamic synthesis of discrete video, graphics, and data control at the consumer's digital set-top box. In this paper we extend that notion by providing a system that can retrieve enriched content from external web services.

3 Contribution

The main contribution of this paper is a model and an architecture that support an enhanced experience of the user in comparition to traditional interactive television services such as the red button or SMS voting solutions. We can divide this contribution into three different categories: content modeling, content consumption, and user interaction.

First, we propose modeling the content using rich-description standards such as SMIL in combination with TV-Anytime metadata descriptions [8]. The major benefit of this approach is that the content can be enriched by different parties at different times. For example, content creators can include enriched material at the creation stage, while individuals might further enhance the content at viewing time. Moreover, at viewing time, content enrichment can be obtained from different freely available resources such as Wikipedia.

Second, we study the differences between the private and the shared space. For example, the television in the living room is a shared space between family members, while a handheld device is a private space. This paper proposes as an innovation the development of a non-monolithic multimedia player that is capable of rendering parts of content into different output devices depending on the share/private nature of the content.

Finally, we propose a model for user interaction that focuses on more abstract concepts of actions instead on particular interfaces [9,10]. We define three types of components: actions, handlers, and activators. The action is the description of the user intention (e.g., pause content or add media), the handler is the implementation of the action, and the activator is the user interface for the action (e.g., play button, speech recognition engine, or gesture). The activators can be implemented in a variety of ways (e.g., gestures, voice, or sensory enhanced everyday objects). The major benefit of this solution is that the user is not limited to the remote control interaction, but can use his personal device or even enhanced everyday objects to interact with the content.

4 Architecture

This section introduces our architecture for providing non-intrusive user interfaces in the home environment. Figure 1 shows the architecture of our system, that includes the following components:

– an intelligent and flexible middleware component, called AMICO
– a non-monolithic SMIL rendering component, the Ambulant Player
– the actions handler called Ambulant Annotator.

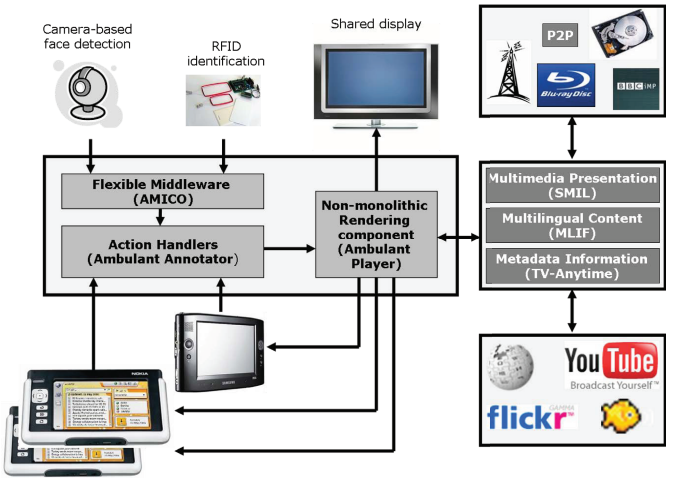


Fig. 1. System Software of the Proposed Architecture

4.1 Content Modeling

The television experience, we propose, uses an enriched description of the multimedia content that includes SMIL files linked to TV-Anytime metadata description and to Multi Lingual Information Framework (MLIF) [11] textual content. Our architecture is also open to external services such as BabelFish, Flickr, YouTube, and Wikipedia that might provide additional content. SMIL code is small, it is easily verifiable, it allows content associations to be defined, it provides a separation between logical and physical content, and it provides as base for license-free implementation on a wide range of platforms. MLIF provides a unified conceptual representation of multilingual content and its related segmentation (i.e. linguistic granularity). The advantage of MLIF is in its ability to deal with different hierarchies of textual segments: linguistic granularity (i.e. sentences, words, syllables), document structure (i.e. title, paragraph, section), or any other personalized textual segmentation which may allow, for example, to associate time and format to any specific segment.

4.2 The Brokering Infrastructure: AMICO

Supporting novel interaction modalities with TV requires usage of many heterogeneous software modules, such as sensors, reasoning tools, and web services. The desired functionality is often available in a form of open-source and free software, such as libraries for vision-based interaction modalities, lexical tools,

and speech input and output for many languages. The main problem when using these components is that they are developed for other purposes, in diverse implementation environments, following standards and conventions often incompatible with multimedia and TV standards.

We have developed Adaptable Multi-Interface Communicator (AMICO)[5], an infrastructure that facilitates efficient reuse and integration of heterogeneous software components and services. The main contribution of AMICO is in enabling the syntactic and semantic interoperability between a variety of integration mechanisms used by heterogeneous components.

Our brokering infrastructure is based on the publish-subscribe design pattern. It is well suited for integration of loosely-coupled parties, often used in context-aware and collaborative computing. AMICO provides a unified view on different communication interfaces, based on a common space to interconnect them. It supports several widely used standard communication protocols. AMICO is extensible, and it is possible to add new communication interfaces.

4.3 Non-monolithic Rendering of Content and Actions Handlers: Ambulant

In previous work, we have described the first prototype implementations of the Ambulant Player [12] and Annotator [13]. The player is a multimedia rendering environment that supports SMIL 2.1, while the annotator is an extension of the player that is a bidirectional DOM-like interface to the player implemented in Python. Together, player and annotator, provides viewer-side enrichment of multimedia content functionality at viewing time.

In addition to those capabilities, this paper introduces two extensions to Ambulant:

- end-user actions handler: the Ambulant Annotator handles the user actions. These actions can come from personal activators (e.g., Nokia770) or from AMICO middleware. Some simple actions the annotator understands are *play/pause*; more complex actions include, for example, *provide me extra information in French about the movie I am watching now.*
- non-monolithic rendering: the Ambulant Player is responsible of targeting different parts of the presentation and content to different rendering devices. For example, the Ambulant Player can render extra information or commercials in my personal device.

4.4 Component Integration and Interfaces

In order to enable integration of components, our architecture support a variety of interfaces. Figure 2 shows the interfaces that the AMICO middleware uses in our studies. Through these interfaces, AMICO integrates a number of services that provide the following functionality:

[5] http://amico.sourceforge.net/

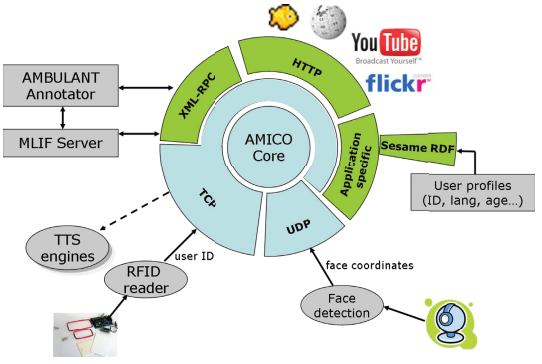


Fig. 2. AMICO Interfaces

- Non-intrusive activators: catches, handles, and interprets the input from non-intrusive activators (e.g., RFID reader).
- User Profile: retrieves and utilises the different user profiles encoded in Resource Description Framework (RDF) files.
- External Services: allows content integration from external services.
- Output Transformation: provides output transformation features.

Figure 3 shows the interfaces of the Ambulant Player and Annotator with the rest of the components of the architecture. It provides the following functionality

- Action Handling: the Ambulant Annotator handles user input coming from different activators.
- Content Retrieval: the Ambulant Player accesses different content resources, including broadcast, optical disk, P2P network, and local storage. The Ambulant Player can also request content from external services via AMICO.
- Non-monolithic Rendering: the Ambulant Player can divide and target the multimedia presentation to different rendering components. These components include the television set and other personal devices.

5 Implemented Example

In order to validate the ideas presented in this paper, this section presents an implemented example and an analysis of the benefits of our solution over traditional interactive digital television systems.

Non-Monolithic rendering: Dick (a US national) and his (Dutch) wife are watching TV. Dick uses a personal device (e.g., Nokia 770) as an extended remote control to navigate through media content based on his personal preferences, as shown in Figures 4(a) and 4(b). The personal devices can inform him when extra (personalized) fragments have been detected by the non-monolithic player. In both cases the content is rendered in the personal device and, thus, do not

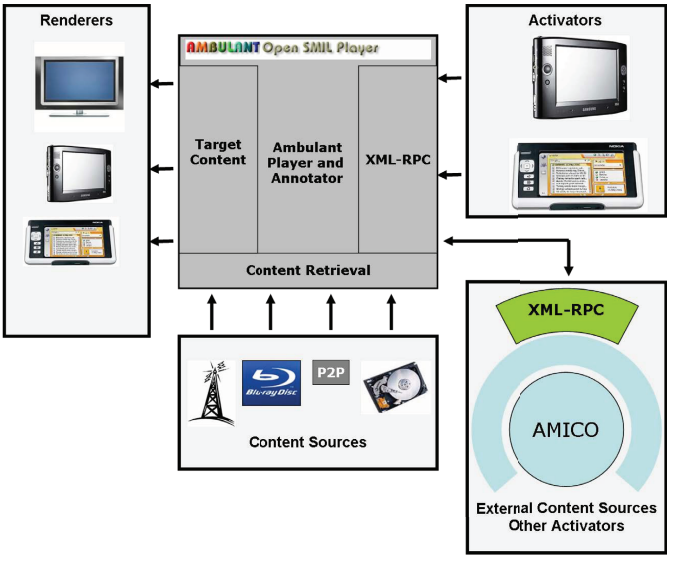
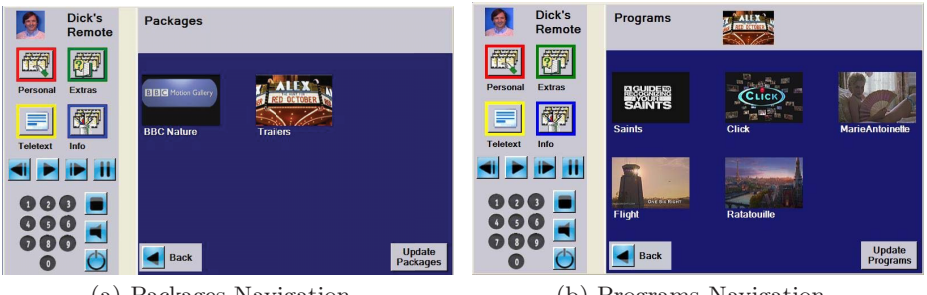


Fig. 3. Ambulant Player and Annotator Interfaces

disturb the television viewing experience of the rest of the family. The personal content might include, for example, instant translation of sentences he might not yet understand in Dutch, personalised commercials, or extra features extracted from web services.

Non-Intrusive input: We apply non-intrusive activators such as RFID readers and camera-based face detectors to assist in interaction with the television content. For example, the sensors detect the identity of Dick and his wife and their distance to the television set, and identify the language settings for subtitle language selection.

User Interaction Capabilities: our system provides as well a shared experience for connected people. For example, Figure 5(a) shows the content Dick



(a) Packages Navigation. (b) Programs Navigation.

Fig. 4. Screenshots of Micro-Level Personalization in a Handheld Device

is watching on the television screen. At some moment, he uses the Ambulant Annotator to enrich the content. Figure 5(b) shows the interface in his handheld device. This enriched content is then shared with, for example, his brother living in the USA by using a P2P network.

This example shows the two main contributions of this paper: non-monolithic rendering of content and non-intrusive user input. Based on a rich television content model, we believe they are the cornerstone of valued group experiences. Clear advantages of our system over current solutions include the capacity of targeting the personal content to where it belongs: to personal devices; the possibility of linking media content in packages or experiences, and the support for a variety of input mechanisms due to the action descriptions. For example, in addition to the ones mentioned earlier, AMICO supports voice input and, even more interestingly, an intelligent pillow interface for controlling the media playback. The intelligent pillow was developed by the company V2_ [14].

(a) Content in the TV Screen.

(b) Enrichment Interface in a Handheld Device.

Fig. 5. Non-Intrusive Rendering

6 Conclusion and Future Work

This paper has presented an architecture for non-intrusive user interfaces in the home environment. This architecture takes into account the differences between the share space (e.g., television set) and the private space (e.g., handheld device) at home. The main contribution of this paper is the proposal of an architecture based on non-monolithic rendering of content and description of user actions. In the first case, the architecture provides the mechanisms to target specific parts of the digital content at home to different rendering components (e.g., high-definition content to the television set and personal material to handheld devices). This way, the personal experience of the user is enriched, while the television viewing experience is not disturbed. In the second case, user interaction is not limited to the intrusive remote control paradigm. Even though in some cases such interaction is desirable, we propose enriching the user potential impact on the content. Some examples include the use of personal devices for personal

content interaction. In addition, other devices such as personal identifiers and a camera can register the identity and context of the user in a non-intrusive manner. Based on those variables, our system can derive actions on the multimedia content (e.g., to pause the show when there is nobody in front of the TV).

In addition to non-monolithic rendering of digital content and descriptions of the actions, this paper presents a solution for modeling interactive digital television content. The key question that this paper handles is how to model interactive television content in a rich and scalable manner. The solution provided is to use SMIL language linked to TV-Anytime metadata and to MLIF multilingual content. The major advantages of this solution is that the content can be annotated in a finer level of granularity, other resources can be included in the television packages, and further enrichments can be provided by professional and amateur users.

Finally, in order to validate the ideas presented in this paper we presented the implementation of an architecture and of a particular scenario. The implementation of the described examples required usage of diverse components, and consequently solving many software interoperability problems. Firstly, we had to support several integration interfaces. Components that we have used came from various (open-source) projects that use diverse integration interfaces, such as XML-RPC, OpenSound Control, HTTP, TCP, of which none is predominant. Adapting components to one common interface is not an easy task, and sometimes not possible, as components are developed in diverse implementation environments. Additional problem was that low-level components, such as sensors, and higher level components, such as web services, work with significantly different data structures and temporal constraints. For example, sensors, such as a face detector, can send dozens of UDP packages per second with simple data structures about detected events. Web services, on the other hand, use a more complex protocol (i.e., HTTP) and complex XML encoded data, with delay which is sometimes measured in seconds. To enable integration of components that work with significantly different data structures and temporal constraints, we had to abstract and map different data types and use temporal functions, such as frequency filtering.

Future work includes describing business models based on the ideas presented in this paper, and more importantly, carrying out a number of user studies for further validation. This studies will be performed in collaboration with other research laboratories with experience in usability.

Acknowledgements

This work was supported by the ITEA project Passepartout, by the NWO project BRICKS, and the IST-FP6 project SPICE. The development of Ambulant is supported by NLnet.

References

1. Baker, K.: Intrusive interactivity is not an ambient experience. IEEE Multimedia **13** (2006) 4–7
2. Blanco, Y., Pazos, J.J., Gil, A., Ramos, M., Fernández, A., Díaz, R.P., López, M., Barragáns, B.: AVATAR: an approach based on semantic reasoning to recommend personalized tv programs. In: Special interest tracks and posters of the 14th international conference on World Wide Web. (2005) 1078–1079 ISBN 1-59593-051-5.
3. Karanastasi, A., Kazasis, F.G., Christodoulakis, S.: A natural language model for managing TV-anytime information in mobile environments. Personal and Ubiquitous Computing **9** (2005) 262–272 ISSN 1617-4917.
4. Lie, H.W., Saarela, J.: Multipurpose web publishing using HTML, XML, and CSS. Commun. ACM **42**(10) (1999) 95–101
5. Jensen, J.F.: Interactive television: New genres, new format, new content. In: Second Australasian Conference on Interactive Entertainment. ACM International Conference Proceeding Series; Vol. 123, Sydney, Australia (2005) 89–96 ISBN 0-9751533-2-3.
6. Costa, R.M.R., Moreno, M.F., Rodrigues, R.F., Soares, L.F.G.: Live editing of hypermedia documents. In: Proceedings of the ACM Symposium on Document Engineering. (2006) 165–175
7. Chorianopoulos, K.: Virtual Television Channels: Conceptual Model, User Interface Design and Affective Usability Evaluation. PhD thesis, Athens University of Economic and Business (2004)
8. Cesar, P., Bulterman, D.C., Jansen, J.: Benefits of structured multimedia documents in iDTV: The end-user enrichment system. In: Proceedings of the ACM Symposium on Document Engineering. (2006) 176–178
9. Nichols, J., Myers, B., Higgins, M., Hughes, J., Harris, T., Rosenfeld, R., Pignol, M.: Generating remote control interfaces for complex appliances. In: Proceedings of the ACM Annual Symposium on User Interface Software and Technology. (2002) 161–170
10. Beaudoin-Lafon, M.: Designing interaction, not interfaces. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. (2004) 15–22
11. ISO: Multi lingual information framework – multi lingual resource management. ISO/AWI 24616 (October 2006)
12. Bulterman, D.C., Jansen, J., Kleanthous, K., Blom, K., Benden, D.: Ambulant: A fast, multi-platform open source SMIL player. In: Proceedings of the 12th ACM International Conference on Multimedia, October 10-16, 2004, New York, NY, USA. (2004) 492–495 ISBN 1-58113-893-8.
13. Cesar, P., Bulterman, D.C., Jansen, J.: An architecture for end-user TV content enrichment. In: Proceedings of the European Interactive TV Conference. (2006) 39–47
14. Aroyo, L., Nack, F., Schiphorst, T., Schut, H., KauwATjoe, M.: Personalized ambient media experience: move.me case study. In: IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces. (2007) 298–301