Next-Generation Interactive Broadcast Services

Uwe Rauschenbach
Siemens AG, CT IC 2
Munich, Germany
uwe.rauschenbach@siemens.com

Jörg Heuer
Siemens AG, CT IC 2
Munich, Germany
joerg.heuer@siemens.com

Klaus Illgner
Siemens AG, CT IC 2
Munich, Germany
klaus.illgner @siemens.com

Abstract

This contribution discusses some recent trends in interactive broadcast systems, namely personalisable rich media TV, scalability of services, the co-operation of broadcast networks with their fixed or mobile IP-based counterparts and the convergence of communication and entertainment services. For each of these, an example will be presented.

Keywords

Interactive Broadcast, Mobile Broadcast, Rich Media Broadcast, Scalable Services, Convergence.

INTRODUCTION

In the last years, digital TV broadcast services have seen rapid growth. Having started as a replacement of analogue television, recent developments add value on top of the classic time-linear digital video services. This paper discusses on some recent trends in the intersecting fields of television, multimedia, broadcast vs. IP network technologies, and new services. Examples from recent research and development projects are given for illustration.

The contribution is structured as follows: First, some trends of emerging and future interactive broadcast systems are described. After that, each trend is discussed in a separate section, and examples are presented. Finally, conclusions are drawn.

TRENDS

Digital television is a reality today. With the more efficient bandwidth use provided, the user will be offered more TV programs than ever before. That's why new ways of accessing this wealth of programs must be found based on personal preferences. Additionally, the digital nature of the new television services makes it possible to use additional media types in a service, to deploy personal computing devices in addition to the TV set to receive these enriched services, and to use data and communications networks to transmit service components.

Personalisable and Scalable Rich Media TV Services enrich the linear program by additional related audio, video and rich text material. Thus, these services create options for interactively browsing related content (e.g., a background report for a News story or an additional camera view in a Sports programme) and cater for the needs of special user groups by providing, e.g., voice-overs in a foreign language supporting different nationalities or a sign language video supporting the hearing-impaired. The paper will present a rich media service concept in the fields of News. Personalization supports the (semi-) automatic selection of content a viewer is especially interested in. This functionality usually requires a receiver which is capable of recording the TV program. Additionally, metadata have to be provided along with the program which structure, describe and classify the content and thus allow selecting the interesting parts of the recording. The paper will describe a metadata model based on the TV-Anytime standard [4].

Broadcast-IP network cooperation. Co-operating networks offer new functionalities and cost savings for the provider. Sending content to multiple subscribers at the same time is the domain of broadcast networks. In contrast to that, the strength of broadband IP networks is to transfer personalized content efficiently and to provide a return channel for "deep interactions" with the service providers, e.g., for votes. We will describe how a rich media service, which consists of various media items, can be efficiently delivered over a combination of a DVB and a DSL network to minimize transmission costs. A smart routing algorithm is selecting the most appropriate transmission channel for each service component. The media items are combined at the receiver into a seamless presentation such that it is transparent via which channel they have been sent. If required by the service scenario, a media stream transmitted via DSL (e.g., additional camera view) can be synchronized with the main TV program sent via DVB.

Another recent trend is to combine broadcast and mobile telephone networks. It is envisioned that this combination offers new opportunities for television-like interactive multimedia services on-the-move.

Service convergence. Looking at the technical trends described above, new convergent services will become possible. For instance, communications (e.g. voice, SMS, MMS) and entertainment (i.e. watching TV) can be integrated with each other, a combination which will be further elaborated in the last section of the paper. Many more convergent services are possible, like voting, multi-user games and virtual communities; all integrated with TV programs.

A PERSONALISABLE, SCALABLE RICH MEDIA TV SERVICE

Rich media TV services enrich the classic linear program by adding supplementary or background information using various media types. Within the project SAVANT [9], a News service concept has been realized which illustrates the potential of this new type of TV services. Each News show is broken along the time line of the main TV program

Published in: Proc. WSDB 2004 - 5th Workshop Digital Broadcasting, September 23-24, 2004, Erlangen, Germany.

into individual semantically coherent entities called *Program Items*. To each of the Program Items, additional related background information can be assigned: HTML pages from the Web presentation of the News show, Video clips featuring earlier reports on similar topics or audio clips featuring, e.g., reports from the radio related to a News story. This additional content is called *asynchronous* because it only has a loose coupling with the timeline of the program. Asynchronous content can also be assigned to a complete news show or to *Topics* – a concept which allows to group multiple related News reports (e.g., reporting on the presidential elections) and provide a rich set of background information to them with moderate editorial effort. The second class of additional content is of *synchronous* nature – i.e., it must be tightly synchronized with the main program's timeline. An example for that is an additional video stream carrying a sign language interpreter for hearing-impaired people.

Personalization requires a set top box (STB) which offers Personal Digital Recording (PDR) functionality and opportunities to filter media data. A service may be personalized in two ways: first, additional content is shown in a live situation only if the personal profile of the viewer indicates this (e.g., show a signer only if the viewer is known to be hearing-impaired). Second, the complete program may be recorded, and the individual Program Items are then filtered according to personal preferences. This allows creating a personal News program featuring Program Items originating possibly from multiple different News Shows.

The digital nature of the media data allows access to such a rich media service using a variety of portable devices. While a PDA or TabletPC may be used to consume the service in-house via WLAN access to the STB, access with mobile smartphones while on the move is possible, too.

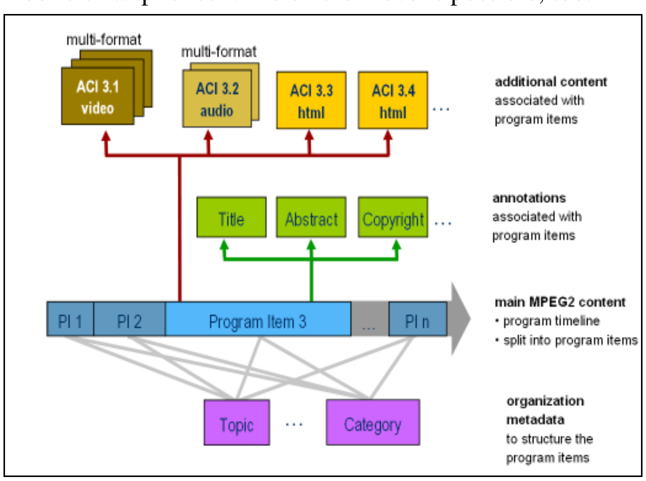

Figure 1. Metadata model for personalized scalable rich media TV services

In contrast to classic TV which consists just of audio and video, new TV services contain a variety of media objects with diverse relationships and meanings. To support the various ways of presenting, accessing and personalizing such a service, a rich set of metadata called a *service description* is necessary which contains the required information. Figure 1 illustrates the metadata model. The segmentation of the main program along the time line into program items is shown. For personalization and random access via the STB user interface, each program item is annotated with e.g. title, abstract and copyright. Furthermore, categories and topics provide means to group program items. Based on these metadata, a personalization component can select Program Items based on a user profile expressing preference for several categories or keywords. Such a user profile can be specified by the user, or it can be collected by indexing the metadata of Program Items actively selected by a user for watching.

Figure 2 depicts the user interface of a scalable News service which allows access to the individual News items plus background information by exploiting the annotations and organization metadata.

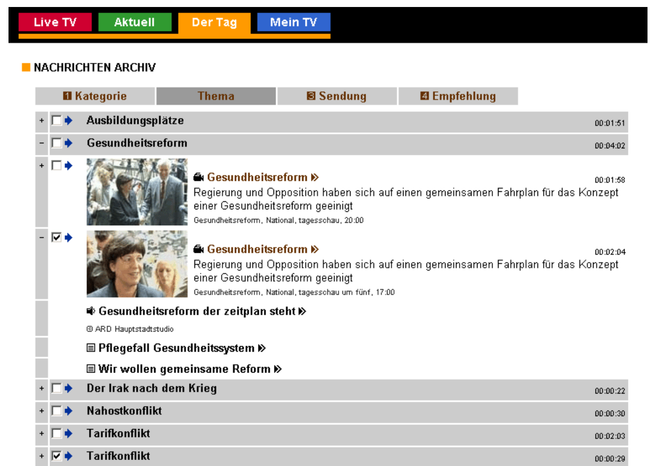

Figure 2. User interface of personalized, scalable News service

The second large group of metadata describes the additional content, supporting the rich media aspect of the service. For each additional content item (ACI), properties like title, synchronization with the main program or type have to be described. Each ACI references one or more *Media Items* which provide access to the actual media essences containing the additional content. The reference is realized by means of a media locator (URI), which may e.g. reference an HTML page via HTTP, a video clip via RTSP or via DVB MPE (multi protocol encapsulation).

Service scalability means that a service is designed such that it can be deployed on various devices with different capabilities. A scalable service can benefit greatly from a *scalable content format*. Such a format would allow encoding each video stream once and then adapt it to different devices by just decoding a well-defined part of the data packets to reduce resolution or bitrate. However, scalable content formats for video like MPEG4 Fine Granularity Scalability [10] have not seen wide acceptance yet. The current MPEG-21 activity on Scalable Video Coding [8] may provide a solution in a few years. Meanwhile, a scalable service can also be realized by a combination of *simulcasting* and *transcoding*, at the cost of lower bandwidth efficiency. Simulcasting means to transmit a media item simultaneously in various formats. Transcoding means to have a transcoder unit in the STB which converts a Media Item from one format/bitrate into another. The proposed metadata structure supports both approaches. By referencing multiple Media Items for one content item, simulcast is supported. For each Media Item, a description of the media properties is provided in the metadata. The most appropriate Media Item to present a content item on a specific device can be found by matching these media properties with the capabilities of the actual target device. To support transcoding, a Media Item can be marked as "virtual" by omitting the media locator but providing the media descriptions. This way, a transcoder engine can use a sibling media item from the same content item as the source for transcoding, turning the virtual media item into a real one and inserting a URI to the newly created media essence.

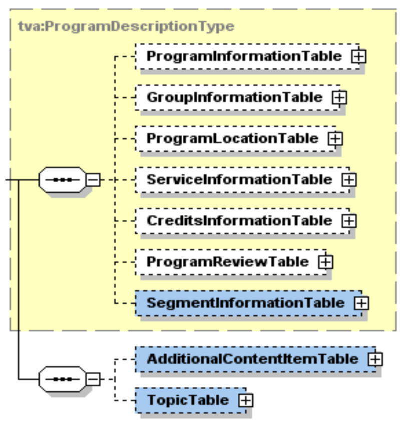

Figure 3. An extended TV Anytime representation of the metadata model

Several metadata standards for multimedia exist, namely MPEG-7 [6], MPEG-21 [7] and TV-Anytime [4]. As the latter standard most closely resembles the structure of a TV service, it has been selected as the basis for the personalized, scalable SAVANT News service [11]. Several elements have been added in order to support all the required functionality (see Figure 3). Because Program Items are modelled as segments of the main broadcast, the SegmentInformationTable has been extended to include all the metadata necessary. Further, two tables have been added: one containing the information about the additional content items and another one containing the Topics as an alternative way of accessing Program Items and additional content.

Figure 4 shows a system for presenting such a service to the user. The content is delivered to a set top box by a combination of DVB and DSL. The STB acts as a home media server and stores the content along with the metadata on its hard disk. The personalized rich media TV programme can be displayed live at a connected TV set. Furthermore, the recorded content can be accessed using the TV set or mobile devices (a TabletPC and a PDA) connected to the STB via WLAN. The user interface to personalize and select the content (cf. Figure 2) is generated on the STB by an MHP application, taking the metadata into account. The mobile devices have access to scaled-down versions of the stored content.

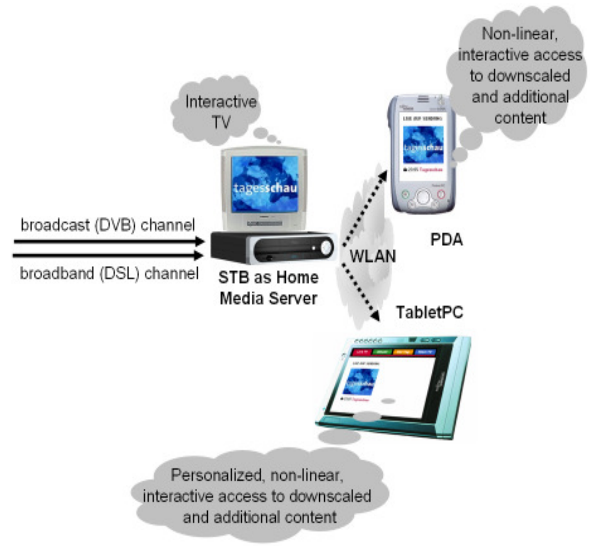

Figure 4. Content Access System to consume the personalized, scalable News service

After having described the metadata for a scalable, personalized rich media TV service which provide annotations of and relationships between the service components (see [11] for further details), the next section discusses ways how the actual service components can be transmitted to the terminal, possibly even by using different transport networks for different Media Items.

BROADCAST-IP NETWORK CO-OPERATION

Rich Media Content Delivery over DVB Networks

Within DVB, all content is sent packaged into an MPEG-2 transport stream (TS). A detailed discussion of the many functionalities of a transport stream is outside the scope of this paper. From the point of view of rich media services, it is important to point out that – besides the main audio-visual content that is transmitted in the transport stream – additional media objects may be sent using three different mechanisms: the *object carousel*, *private sections* and *multiprotocol encapsulation (MPE)*. The object carousel periodically transmits a hierarchical directory structure to the receiver and is this way well suited to send content structured into multiple files (e.g., HTML pages or interactive MHP applications (Xlets)), but also metadata to the receiver. Private sections provide another means to embed application-defined data packets into a TS. A client application has to be provided at the STB to extract them. This approach is well-suited for the one-time transmission of big files, e.g., asynchronous additional video content, in contrast to using the object carousel which may become overloaded as it repeatedly transmits its content.

Multiprotocol encapsulation allows carrying IP data packets in a DVB stream, making it possible to embed RTSP streaming video into a rich media service. This way, synchronized additional video streams can be sent along with the main programme (like a signer for the hearing impaired people or streams taken from additional camera angles in a Sports program).

Rich Media Content Delivery over Co-operating Broadcast and Broadband Networks

As described in the previous section, digital broadcast systems (DVB) are capable of carrying not only the main audiovisual content, but also additional media objects and even downstream IP traffic to a multitude of users at the same time. On the other hand, broadband IP networks (DSL) are becoming widely available, being ideally suited to carry personalized content on demand. We believe that a future personalized broadcast system will combine DVB and DSL networks for delivering a personalized service at optimized costs.

In order to make content transmission over co-operating networks a reality, two issues must be considered. First, the system should be able to select the channel via which a media item will be delivered. We call this feature *Smart Routing*. Second, the system must be able to *synchronize* content delivered through DSL with the main TV program delivered via DVB.

Smart Routing

The basic idea of smart routing is to save transmission costs by using the channel for transmitting the additional content which offers the lowest transmission costs. For DVB, the costs are independent of the number of users. For DSL, the transmission costs grow proportionally with the number of users because for each user there will be an individual stream transmitted. Making a routing decision means to select the N media items of a rich media program which are most likely to be consumed by the most users and to transmit them in the DVB channel, where N depends on the available capacity in the DVB channel and the bitrate of the individual media items. To get an estimate about how likely it is that a media item will be in high demand, the following criteria can be used:

- Estimation by the program author or playout operator and insertion into the *service description*
- Prediction using usage statistics from previous similar programs an heuristics based on media properties
- Actual measurements during the current program

While the first two methods provide a *fixed* routing decision and are suitable for both asynchronous and synchronous additional content (i.e. clips and streams), the last method allows *re-routing* and can thus only be applied to asynchronous additional content. The reason is that for synchronous additional content, the routing decision can not be changed while the content is playing – a seamless change would require too much overhead. In contrast, the routing decision for asynchronous additional content may be revised at any time. Revising means that all users currently consuming this content via the "old" channel will keep doing so, while new users will receive the content through the "new" channel. As asynchronous clips are usually short, the "old" channel will be freed quite fast.

This way, we can distinguish the following three ways of smart routing:

1. *Fixed routing of asynchronous content*: Asynchronous content is inserted into either the DVB or DSL channel depending on a pre-set field in the *service description*.
2. *Re-routing of asynchronous content*: if the number of users of an audiovisual Media Item accessed over DSL exceeds a threshold, the item will be inserted into the DVB stream, e.g. using private sections. If the usage figures drop again, the ACI is no longer made available via DVB but can still be pulled via DSL.
3. *Fixed routing of synchronous content*: synchronous content is inserted into either the DVB or DSL channel depending on a pre-set field in the *service description*.

A routing decision is executed by the system by triggering the playout system to insert the media item into the desired transmission channel. Furthermore, the system must change the media locator (URI) of a media item in the service description. This way, the Content Access System (cf. Figure 4) is instructed to extract the media from the correct channel. As a prerequisite for that, the service description must be updated regularly, and these updates must be signalled frequently to the Content Access System. As both DVB and DSL can carry IP traffic, the handling of the packets is the same after extracting them from the transmission channel.

Content Synchronization

Systems for rich media TV services must provide a new kind of synchronization: synchronizing additional content with the main TV program. While frame-accurate sync is required only in some very rare cases, synchronization with an accuracy of a few frames has many applications: additional camera angles in sports programs, quiz or talk shows; or a sign language interpreter to make TV programs accessible for hearing-impaired people. Synchronization has two facets: first, transmission delays must be compensated to ensure that the data packets carrying the additional content arrive at the right time in the decoder. Second, the presentation of the main and the additional content must be synchronized. Ideally, this should be possible using standard components at the receiver side.

When an additional video stream is streamed over RTSP and transmitted via DVB in MPE, transmission synchronization does not pose a major problem because no transmission delays between the main program and the additional content occur. When the additional stream is transmitted via DSL, however, both transmission delay compensation and presentation synchronization are necessary.

Figure 5: A signer synchronized with the TV program

In the SAVANT project [9], we have developed a combined approach to compensate transmission delays and ensure presentation synchronization:

- Additional video content is streamed as MPEG-4 via RTSP/RTP.
- A timing control component in the playout system ensures that an additional content stream is started at the right point in time, denoted in the service description and triggered by the clock driving the playout of the main video. If necessary (e.g. if the main content is sent over satellite), the start time may be delayed slightly.
- Each RTP packet is time-stamped with a reference to the clock of the main video (Normal Play Time, NPT).
- In the Content Access System, an *RTSP Proxy* intercepts the incoming data packets and buffers them.
- The transport stream demultiplexer in the CAS provides the RTSP Proxy with the NPT timing information extracted from the main video.
- Knowing the time stamp of an RTP packet (which contains the NPT value when the packet left the playout system) and the current NPT value of the main video in the CAS, the delay can be compensated by adjusting the time stamp in the RTP packet accordingly.
- The RTP packets with the adjusted time stamp are then passed to an unmodified MPEG4 player which presents the additional stream.

Figure 5 shows a News program enhanced with a signer which is synchronized with the main content using the method just described.

Convergent Mobile Broadcast Systems

Up to here, extensions to broadcast systems have been discussed which target the classic static TV set, with some in-house mobility added by using WLAN access. However, digital media technology can offer more: delivering interactive broadcast to mobile phones and ultra-portable devices. This issue is currently being addressed by the DVB forum [2]. A version of the DVB-T standard called DVB-H [3] especially designed for efficient battery use of mobile receivers will be used as the downlink to broadcast multimedia content to mobile phones. Additionally, personalized information, interactions and content protection / billing functions will be provided using the GPRS or UMTS channels of the mobile telephone network, creating converged mobile services. In alternative configurations of converged broadcast-mobile systems, DVB-H may be replaced by a multimedia version of DAB [5] or the Multimedia Broadcast Multicast Service (MBMS) [1] in cellular networks.

SERVICE CONVERGENCE

Mobile phones and digital TV receivers have in common that these device categories become highly widespread in daily life. However, a unique attribute of DVB set top boxes is the high display resolution compared to mobile devices. This especially allows increasing the ease and intuitivity of use of such devices due to enriched capabilities of the graphical user interface (GUI). Both aspects make it attractive to integrate other services than broadcast reception into a digital STB.

A very first trend towards this goal was the integration of web browsing capabilities over DSL into STBs. However, due to the lower display resolution of TVs and different browser characteristics of STBs compared to the capabilities of PCs in general, web content has to be designed specifically for the reception on TV. This recently results in ISPs providing web content in a “Walled Garden” concept which results in expensive information preparation. With a growing number of STBs providing the capabilities of web browsing, first business models of STB-specific web portals are discussed. Since these portals are accessed automatically when the box is turned on, the content of the portal appears to the user as if it has been pushed to the STB comparable to a broadcast service.

The convergence of such web and broadcast services is recognizable in that the so delivered web content allows interacting with the broadcasted content by means of programming PDRs via Electronic Program Guides (EPGs) or interactive games accompanying broadcasts. Even richer interactive web-based applications are enabled by synchronization mechanisms described in the previous section. A prerequisite to enable service convergence is that the service architecture has to converge. This is shown in Figure 6 which presents a high-level architecture view interfacing interactive TV servers with play out servers.

By adding DSL interfaces to the STB, it can act as a terminating point not only for broadcast but also for IP-based communication services. Due to its simplicity, SMS and MMS clients have been ported to first STB prototypes. The obvious advantage of these concepts is the further simplified use of the services. An example of an MMS Client on TV is shown in Figure 7. In this case the integration of MMS with broadcast allows new usage scenarios such as messaging of commented broadcast content. However, obstacles for this kind of service convergence are legal aspects of intellectual property rights regarding broadcasted content and the lack of a purely IP-based standardisation of SMS and MMS services as shown in Figure 6.

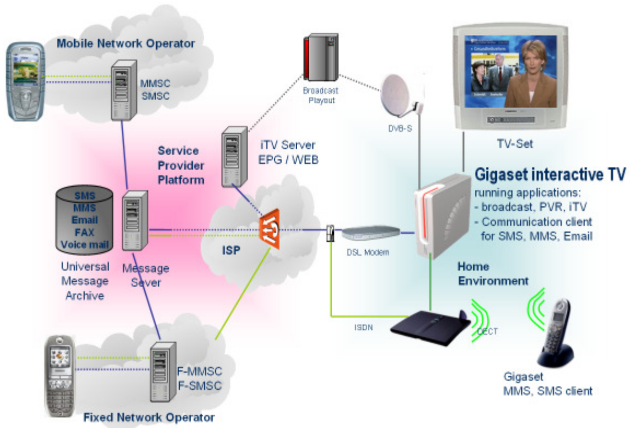

Figure 6: System architecture supporting service convergence

Besides being a device terminating external services provided for instance via DSL, the STB can be seen as component of the home network. Such an integration can be realized for instance via Universal Plug and Play (UPnP) to the STB to coordinate tasks between, e.g., a phone and the video recorder capability of a device as shown in Figure 6. In this example, after negotiating the offered services via UPnP, the phone is capable to send notifications of incoming calls to the STB which renders caller information on the TV screen. On reception of the call, the phone triggers the STB to time-shift the current broadcast. After termination of the call, the user has the possibility to resume watching the TV at the point he started the phone call.

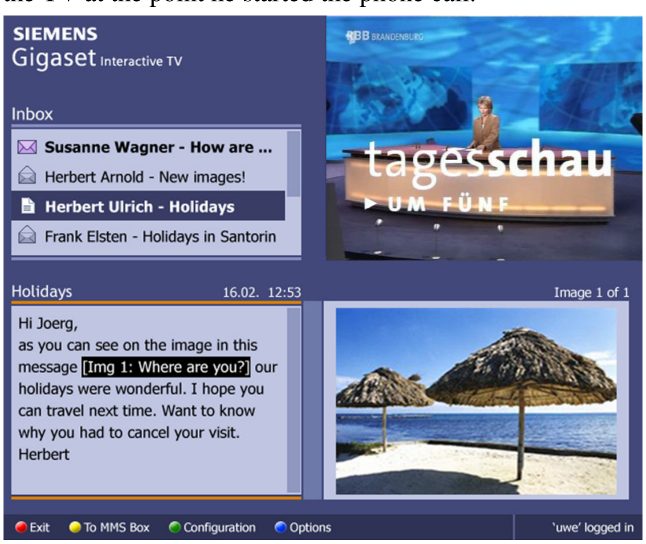

Figure 7: Gigaset Interactive TV as an example for convergent TV and communication services

According to the presented services in this section, convergence of services can be divided into two categories:

a) Convergence which requires the coupling of service provisioning.

b) Convergence which results from coupling of services on the receiving device and leads to new usage scenarios of services.

In both cases, the service convergence examples seen so far aim at the enhancement of ease of use and user experience at the same time.

CONCLUSION

In this paper, we have listed some recent trends in enhanced digital broadcast systems and have sketched technical problems, possible solutions and usage scenarios related to these trends. It can be expected that in the next few years a number of new, interesting services will emerge based on the new opportunities of digital media technology and converging networks.

ACKNOWLEDGMENTS

The authors express their thanks to the colleagues in the SAVANT project for many valuable discussions. Parts of the work presented have been funded by the European Union under contract number IST-2002-34814.

REFERENCES

[1] 3rd Generation Partnership Project; Multimedia Broadcast/Multicast Service; Stage 1, 3GPP TS 22.146 V6.5.0, June 2004.

[2] DVB Forum homepage, www.dvb.org.

[3] DVB forum, Draft DVB-H Standard, available from http://www.dvb.org/index.php?id=278

[4] ETSI TS 102 822-3-1: Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems ("TV-Anytime Phase 1"), Part 3 Metadata, Sub-part 1: Metadata Schemas. 2002.

[5] W. Hoeg and T. Lauterbach: Digital Audio Broadcasting – Principles and Applications of Digital Radio, 2^{nd} edition, John Wiley & Sons, 2003

[6] ISO MPEG-7, Part 5 - Multimedia Description Schemes, ISO/IEC JTC1/SC29/WG11/N4242, 2001.

[7] ISO MPEG-21, Part 7 - Digital Item Adaptation, ISO/IEC JTC1/SC29/WG11/N5231, 2002.

[8] ISO MPEG-21, Part 13 – Call for Proposals on Scalable Video Coding Technology, ISO/IEC JTC1/SC29/WG11/N6193, Waikoloa, December 2003

[9] IST project SAVANT, www.savant.tv.

[10] W. Li: Overview of Fine Granularity Scalability in MPEG-4 Video Standard. IEEE Trans. Circuits and Systems for Video Technology, 11(3), March 2001

[11] U. Rauschenbach, G. Stoll, W. Putz, R. Mies and P. Wolf: A Scalable Interactive TV Service Supporting Synchronized Delivery over Broadcast and Broadband Networks. IBC 2004 conference, Amsterdam.