Public
DELIVERABLE
ESPRIT 28798
AVIR

Deliverable # 14

Integrated demonstrator of the AVIR Consumer System

AVIR

Audio Visual Indexing and Retrieval
for non-IT-expert users

Table of Contents

1 Overview of the AVIR system

The objective of the AVIR project is to investigate and experiment end-to-end solutions for delivering new value-added services on top of broadcast television, aiming at a better exploitation of multimedia information resources by non-IT experts. The project creates and demonstrates an infrastructure for provision of distributed interactive services, intended to:

- enhance the quality and visibility of broadcast services by providing new attractive features and more interactivity at a low operational cost structure;
- provide intuitive modalities for home users to select and search audio-visual information by exploiting content descriptors.

Innovative applications for *Personal TV*, that is for content filtering, selection and search can be offered by means of content descriptors in the broadcast stream. Storage capabilities in the receiver will make the scenarios as described by the TV-Anytime Forum possible. In the AVIR project such functionalities have been realised through the usage of international standards in the area of multimedia applications, namely MPEG7 content descriptions and DVB-MHP applications.

The platform developed by the AVIR project supports the generation, distribution and use of metadata related to the television programmes and consists mainly of two subsystems:

- The **service provider system** generates a description of the audio-visual content according to the MPEG-7 standard, also using automatic routines for video content analysis, audio analysis and speech recognition. The content and associated description can be delivered in a broadcast stream according to the DVB specifications. Furthermore an application for programme selection and browsing can be distributed by the service provider using DVB-MHP specifications.
- The **consumer system** consists of a set-top box with storage capabilities. The broadcast content can be locally cached based upon the user's profile, as according to the TV-Anytime scenarios. The received descriptors are parsed and stored locally to support content filtering and content retrieval applications. Eventually the user can search through the recorded content with advanced retrieval techniques based upon the MPEG7 descriptors.

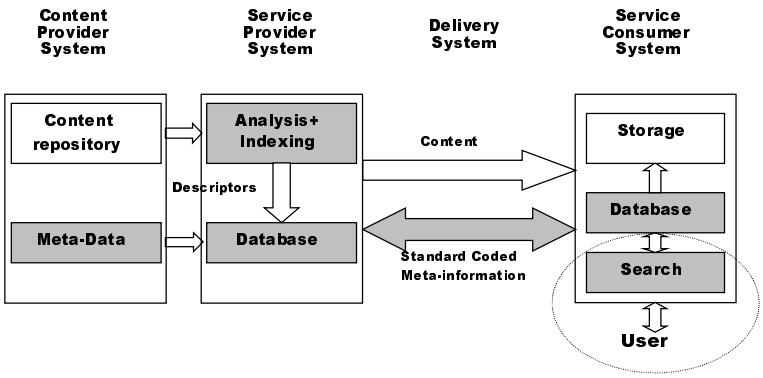

Figure 1 *Schematic drawing of the technology developed in AVIR (grey blocks) within the context of the broadcast chain*

2 The AVIR services and consumer platforms

The consumer system specified by AVIR provides a personal multimedia repository to the residential users. The AVIR consumer system aims at introducing considerable improvement over current TV receivers and video recorders, with new features and easy user-friendly ways to control the devices and the AV material. In order to realise this, appropriate descriptors of TV programme content are required. Therefore AVIR also specifies a platform for a service which supports the generation, distribution and use of metadata related to TV programmes.

Within the context of the AVIR project we focus on two different sorts of **services**:

- The first one, which we will refer to as ***TV content metadata*** service, is operated by a specific broadcaster and includes an EPG service customised for a specific bouquet and other descriptors of the audio-visual content that may be extracted automatically.

  The programme information and content metadata may be used during live viewing but can also be very useful for content browsing after a programme has been stored locally.

  The application may be resident on the box or preferably downloadable, so that a single receiver can run several applications from different service providers.

- ♦ The second one, which we will refer to as ***Personalised TV Guide*** service, is a full TV programme guide application, where the service delivers schedules and programme information of many TV channels. Further the application may allow for personalisation of the functionality and of the content of the box, e.g. allow for automatic recording based on user preferences. The STB may receive descriptors in a standard format from one or more service providers and the TV guide application allows for integration and personalisation of the displayed information and functions. The application may be resident on the box (developed by the manufacturer) but could also be downloaded from the service provider.

Both services offer considerable benefits to the home users by helping them in accessing and retrieving relevant AV material In both cases the chosen approach of distribution of descriptors by using open standards guarantees a large degree of flexibility for the receiver, being able to run resident or downloadable applications and exploiting descriptors from one or more service providers. As a reference, in figure 2 an overview of the AVIR services and platforms is shown.

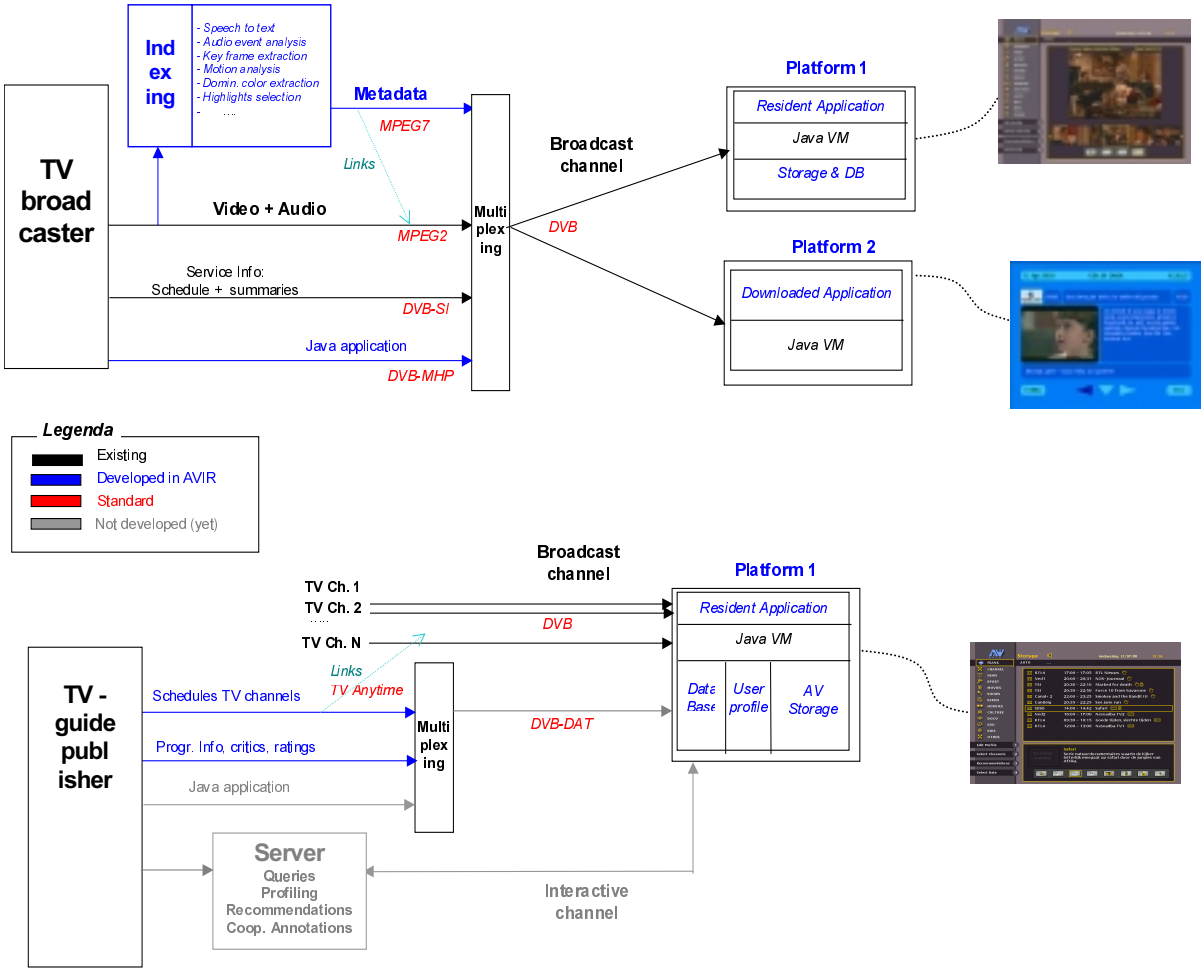

Figure 2 *The two kinds of services developed within the AVIR project*

2.1 AVIR consumer system platforms

For the AVIR consumer system two different platforms have been developed, one being based on a PC architecture is more flexible and powerful, the second one based on a STB prototype is much closer to a CE product realisation. Please note that the AVIR consumer system platforms are actually independent from the AVIR services. This means that they do not strictly correspond to each of the services described above. The MHP downloadable application with programme and segment information, operated by the broadcaster, has been implemented on the second platform, while most advanced functionalities, either requiring heavier processing or the support of a multimedia database have been implemented on the first platform.

2.1.1 AVIR CS platform 1

The first platform is a simple (and low-cost) thin computer, running a Linux operating system running. The hardware includes:

- Intel Celeron 400MHz processor
- 32 MB memory
- 20 GB HD

- MPEG PCI decoder board
- remote control (and possibly speech-mike for vocal control)

The software includes

- Linux 2.1.15 operating system
- Java virtual machine
- relational database with JDBC interface
- a Java GUI
- a set of dedicated routines for MPEG stream management and retrieval

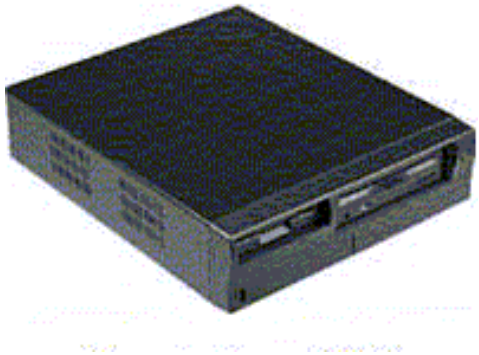

Figure 3 *AVIR consumer platform 1*

2.1.2 AVIR CS platform 2

The second platform is a consumer STB prototype developed at Philips Research Labs. This prototype differs from real CE products as it is based on the Trimedia processor, a powerful DSP for AV processing for consumer applications. Furthermore it features a 4 GB HD, 16MB RAM, a Philips DIVA MPEG2 decoder and different digital interfaces (a.o. IEEE 1394).

A picture of this receiver is shown in figure 4.

Figure 4 *AVIR consumer platform 2*

2.2 TV metadata service

Several programme descriptors, nowadays mostly in digital format, are made available to broadcasters from the content producers or the archives. Further, by applying automatic content analysis on the TV programme to be broadcast, several content descriptors or metadata can be obtained. In AVIR several routines have been developed for extraction of metadata such as editing effects, camera operations, key-frames, dominant colours, audio

characteristics, audio events, speech transcription. The analysis routines and indexing procedures for this service have been integrated in the AVIR service provider system (see deliverable #13). All in all the STB may receive:

1. the schedule information, normally of only the channels operated by the same service provider (called *bouquet*);
2. the specific programme information possibly with description of sub-programme segments;
3. "impulse" content metadata related to the audio-visual content (speech, key-frames etc.).

The descriptors of the audio-visual content can be distributed jointly with the content (in-band, i.e. in the same transport stream as the MPEG2 video) or separately in another transport stream or data channel. Using such descriptors the receiver can automatically select for recording not only entire programs but also segments of programs when the related metadata match with the user profile settings (for instance when a certain keyword is pronounced). After recording, thanks to the available metadata, it is possible to perform advanced search and navigation on the recorded material. The retrieval functionalities based upon the received metadata have been implemented on the AVIR platform 1.

The service can be operated by a broadcaster or by a third party service provider. The TV content metadata service could be actually offered by several service providers. The STB may receive a downloadable (DVB-MHP) application for TV programme browsing that is tailored to a specific channel or bouquet (by using logo and the desired look&feel of the broadcaster). The receiver should have the capability to download and execute the applications of each service provider that is required to visualize the information. For this purpose DVB-MHP is an appropriate standard as it defines an API suitable for downloadable applications.

In AVIR the role of this service provider is played by RAI. The Italian broadcasting corporation is interested in the new opportunities for advanced services offered by the DVB standards. The MHP application and related descriptors have been implemented and tested on the AVIR platform 2.

2.3 Personalized TV-guide service.

The distribution of TV guide magazines in electronic format will make program guides more interactive, more closely linked with channels and programs and more personalised to better cope with the increased number of available channels. This service can be operated by a TV guide magazine publisher or any other service provider which owns the rights for exploitation of information about programmes transmitted by all television channels available in the geographic area of distribution of the TV-guide service. Because of the different regulations in each country, such a service may assume different forms at national level. The AVIR Personalised TV-guide service allows the definition of different user profiles for each device. Each user may set his/her preferences in terms of favourite programs. The device will then generate a list of recommended programs to watch and depending on the available capacity, record the most interesting ones. Users can further set preferences or search programs by using vocal control and search for programs by using keywords.

Since the receiver must be able to support the delivered service (i.e. receive, parse, organise and present the received descriptors) a set of open standards must be employed to secure full compatibility. Else a co-operation between the set-top box manufacturer and the service

provider is required for the definition of the interfaces. At present only proprietary systems are in use in the U.S. market (TiVo, ReplayTV, MicrosoftTV) while no real service exist in Europe, yet. Currently the TVAnytime forum is active in the standardisation of formats ad protocols for the diffusion of PersonalTV services such as the one proposed by AVIR.

In AVIR the role of the service provider is played by TVSpielfilm/Tomorrow Internet, which is interested in the opportunity offered by current technologies for the provision of TV guide magazines in electronic format. Tomorrow Internet mainly took care of the design of the EPG, while Philips NL took care of the implementation and developed the underlying functionalities of the set-top receiver that have been integrated in the AVIR platform 1.

3 AVIR Consumer System

3.1 Main functions

In AVIR we envisage a situation where home users may be disoriented by the overload of video services available. Currently many channels can be received via satellite and the number is growing steadily. For example, in Germany people can receive already more than 70 channels in German, most of them are free to air. Thanks to the digitisation of television in the close future we will see a further growth of channels via satellite, terrestrial antenna, cable and even via phone lines (with ADSL technologies). Also new interactive services such as Video on Demand start to appear. The AVIR consumer system was built with this information in mind. It will help the user to cope with the information overload. This was realised by:

- easy control of a flat user interface
- filtering TV channels based on their genre or channel
- a personalized viewing advice
- profile based automatic recording
- voice control
- easy recording
- visual content navigation
- program search options

In the following sections the functionality of the AVIR consumer system will be presented.

3.2 EPG System

When the user needs to make a choice from over a hundred channels a good electronic program guide is essential. As a bare minimum the user needs to know the program title, the time of broadcast and the channel, the genre, and a short summary. Preferably the year of production, a list of actors and an image should be available too. Broadcasters recognize that if this information is made available it increases the likelihood that a viewer decides to watch a certain program. Broadcasters that offer little or no information run the risk of loosing viewers.

Figure 5 *Programme information displayed during viewing or zapping ("i" button)*

In the AVIR consumer system an EPG was implemented that allows users to browse through the available TV programs based on the broadcast channel or based on the genre of a program. The latter option allows the user to more easily find a program to his or her liking if the number of channels is very large. In this case the user first decides that he or she would like to see an action movie and then browses through the list of action movies by selecting the genre movie and subgenre action as shown in Figure 6.

Figure 6 *EPG user interface during genre selection*

3.3 Personalised Recommender

To help the user decide what to watch a recommender system is available. This system presents the user with a list of recommended programs. The recommendation is based on the user profile. This profile consists of a rating for each genre and subgenre telling the system how much the user likes or dislikes programs in this category. The user interface allows easy modification of the settings by the user by navigating through the genres and pressing + or – to increase or decrease the ratings. To avoid having to go through all the genres the user can first select a profile from a list of standard starting profiles and modify it later.

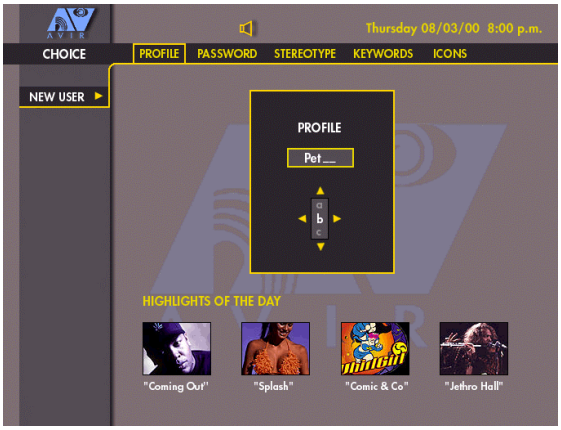

Figure 7 *User profile setting*

When the user presses the info button on the remote control, a screen appears that shows a list of users. By selecting a user, the corresponding profile is loaded. Recommendations will be made based on the user that was selected. A family profile can help finding nice programs to watch together. It is also possible to define specific characteristics of the programs or

keywords so that a program matching such criteria will always be recorded, or just the opposite, never recorded, depending on the user's choice. The user can enter words for which the system should look in the summaries. This way it is possible to notify the user e.g. when the word windsurfing is mentioned in a description of a sports show. In addition the user can search for programs where a certain actor is playing.

Furthermore, a user can select a favourite icon and modify the personal background settings in order to personalise the user interface and always recognise his/her profile. It is also possible to prevent others from seeing or removing the programs stored by a specific user by locking them with a pin code.

3.4 Easy Recording

When browsing through the programs per channel or by genre a jumping highlight can be controlled using the cursor keys on the remote control. To record a program all one has to do is push the record button. Programs that are recommended by the personalized recommender can be recorded in the same way. In addition to recommending programs the system can also automatically select programs for recording. The list of programs that are scheduled for recording can be easily adapted. This allows cancelling the automatic recording of programs that the user doesn't like. In Figure 8 the list of programs that are scheduled for recording is shown.

Figure 8 *List of programs scheduled for recording*

3.5 Storage management

In a device that features an automatic recording functionality it is extremely important to be able to get a good overview of the content of the disk and of all Programs that have been recorded can be browsed similarly to programs that will be broadcast. The user interface offers the possibility to either get a complete overview of recorded programs in order of time of recording, or to select them by genre. However more functionalities than in the standard EPG are offered as the program and all related information have already been captured. Therefore it is possible to get a summary of the programme with a representative picture, as in figure 10). It is obviously possible to delete the program, unless it has been recorded and locked by another user. It is possible to archive the program in order to make more room on the local disk or just to keep a copy of a movie in the personal video library. In this case the programme will be transferred onto an external drive such as a recordable DVD or a D-VHS tape or an analogue VHS tape. In all cases essential information such as title and genre will be preserved on the local database together with the recordable media identifier. In this way all the information of the home video library is centrally managed and available for future reference and search operations. Finally on recorded programs it is possible to visually

navigate the program and when content metadata are available (when provided through the *TV content metadata service* by the broadcaster) also to perform search operations on it.

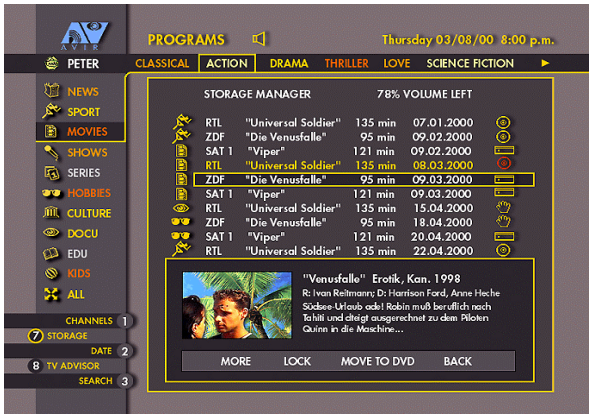

Figure 9 Browsing through the Storage manager

3.6 Searching programs using keywords

In the scenario envisaged by AVIR a user may be confronted with hundreds of available channels. Therefore even listings by channels and genre may not be practical when one is looking for something specific. For instance a user may be looking for a movie featuring Clint Eastwood or searching the channel where a ski downhill race event will be broadcast live or the stored films directed by Steven Spielberg. In this case it is very handy to be able to compose simple query. Thanks to an easy-to-use search field (see figure 10), the user may insert a keyword with the simple usage of a remote control. Furthermore all the keywords will be saved in the history, so that in the future new queries can be made by simply re-selecting the old keyword.

Figure 10 *Searching programs by keyword*

3.7 Voice control

In the AVIR platform a voice control system has been developed in order to increase the easeness of use of the device and investigate the advantages of a multimodal user interface to access large amount of data. The chosen strategy is to exploit vocal control (whole word recognition) in order to allow for short cuts in the navigation within the user interface (so-called 'Power Commands') as shown in figure 11. For example the user may jump to the most often used submenus of the user interface, avoiding in such a way tedious repetition of operations by clicking bttons on the remote control. The possibility to perform search by using vocal queries has been investigated as well, but phoneme-based recognition technology for keyword querying is not yet mature enough. A limited speaker-dependent vocabulary has been used since a speaker-independent vocabulary (to be trained with some 200 users) would take tremendous effort and time. Acoustic echo cancellation has been integrated into the user interface prototype as well.

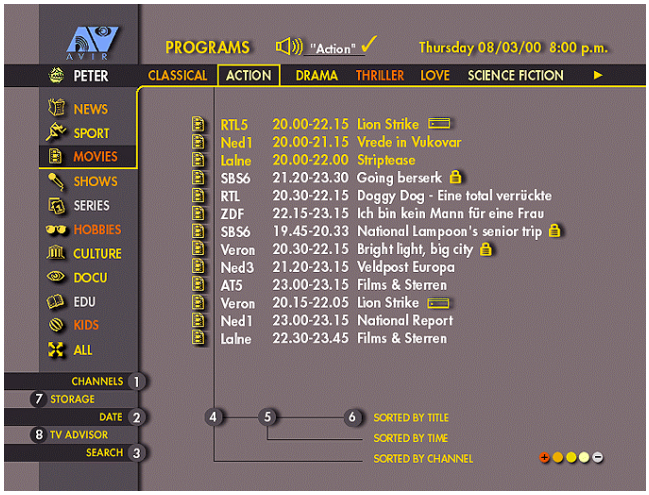

Figure 11 *Result of a short-cut operation made with vocal control, in response to the voice command 'Action'*

3.8 Content search and Navigation

In general it is very useful to be able to quickly navigate in a non-linear manner through video programmes at a much higher speed than fast forward or reverse feature allow on current home videorecorders. But in case a device automatically records programmes whose content is yet unknown to the user, intuitive navigation and presentation tools must be offered to the user in order to quickly inspect the content and decide whether or not to watch or keep the program or part of it. Browsing the visual content by means of so-called key-frames (see figure 12) allows users to get an idea of the content of a program or to find a certain segment the user is looking for. The key-frames are indicated by the broadcaster system and the references are sent along with the video stream in XML format.

Figure 12 *Visual navigation within a programme by key-frame browsing*

The broadcaster may offer key-frame signalling according to different criteria. Key-frames may be the result of a temporal segmentation of the program, in which case we have a shot navigation. The key-frames may be actually related to different sub-parts of the program (such as the chapter segmentation) or to the highlights of the program. In this case the indication of such frames may happen either manually (e.g. when a goal is scored in a football match) or automatically. Indeed in AVIR we have investigated on the possibility to identify relevant episodes in a TV program by analysing the audio track, in some case jointly with the video content. Indeed if we take again the example of the football match, the most interesting actions are underlined by a louder audio level, due to the excitement of both the speaker and the public. These scenes that by averaging the audio signal show the highest peaks are normally equvalent to the highlights of the football match, e.g. a goal or missed chance. In a smiilar manner, by detecting clapping and laughing, highlights in a talk-show can be found.

Another descriptor that can be obtained from the analysis of the audio-visual stream is the dominant color of a key-frame or even of a camera shot. When this information is provided, the key-frame browser can be enriched with an additional tool that we will call **Colour bar**. In the colour bar the dominant color of each key-frame is represented in a linear fashion. In our implementation (see fig. 13) we chose to show two bars, one which features a slider that can be used for fast browsing within a certain interval, and a second one which shows in which part of the program we currently are. This allows to effectively visualize the changes in the different parts of the program both at low and at high level of granularity.

Figure 13 *The colour bar allows users to easily identify different segments of the program or commercial breaks*

Yet an additional content descriptor might be associated to the extracted key-frames in order to enable simple similarity retrieval operations in the image database. The similarity criteria (the evaluation of the level of similarity between two images) can be based on low-level features extracted from the video signal (like color, texture, shape, and edges) or on higher-level descriptions manually or semi-automatically generated. In AVIR color coherence vectors have been used as a similarity measure.

When such a descriptor is available new functionalities could be implemented. For instance users may jump from an image in the video stream to the next (or previous) “similar” one. Users are also given the possibility to save some personal favourite images and use them at wish in order to retrieve favourite topics in the video database. Such a functionality offers users the possibility to jump to the next news item in a news program by looking for the next occurrence of the news reader, as shown in figure 14, or jump to the end of a commercial break by recognising the commercial logo of the broadcast station, or search for weather forecasts or sport reportage within a news program, etc.

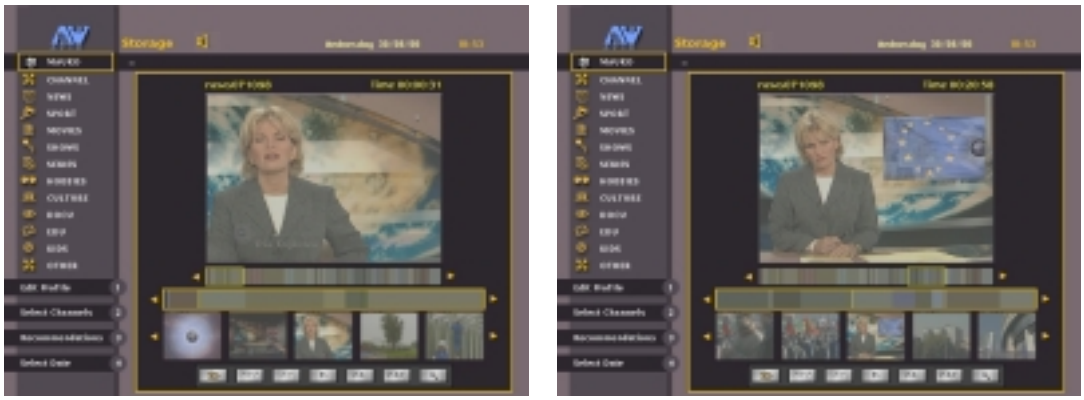

Figure 14 *The similar image search functionality in a news program allows to jump to the beginning of the following news items.*

Finally, when speech recognition is employed by the broadcaster to analyse the audio content of the program and generate a transcription of the dialogues, the received text can be stored locally and associated to the related key-frames. In this way it is possible for users to read the

speech transcription of a program and by this quickly analyse the content. For instance one could read the news and look at the associated pictures. This would allow a much faster scanning of the content and enable search for specific news. Another possibility that is offered by the AVIR system is to search for specific keywords occurring in the text, as shown in figure 15. For instance one may search for the latest news about the U.S. president and make a query for “Clinton” in the news programs recorded today.

Figure 15 *Search for news by keyword*

4 The broadcaster's EPG

4.1 Objective of the MHP demonstrator

In the field of TV broadcasting the transition from analogue to digital technology is under completion on the satellite segment and already started on the terrestrial one. Although public broadcasters have at their disposal the basic tools to deliver interactive content to viewers, they are still wavering before the choice of the user terminal, the weakest ring of this added value chain. The fragmentation of the present marketplace, with several proprietary technologies deployed by pay-tv operators, results in an obstacle to low cost development and distribution of interactive TV asset. As a response to this situation, recently the DVB Project published the first release of the Multimedia Home Platform standard, aiming to open the market to real competition, with benefits for all players, derived from economies of scale.

In the framework of the AVIR project, RAI developed an experimental EPG application, running on a prototype set-top box equipped with a partial implementation of the MHP APIs (IFA'99 profile), produced by Philips Research. The objective was to set up a transmission system enabling the delivery to the user terminal of high level metadata such as programme segmentation, produced by the content provider exploiting automatic indexing tools and low level metadata.

4.2 System Architecture

At the server side metadata are received from the service provider and stored in a database; for each programme the following segment textual information may be added to basic metadata:

- Title
- Description
- Duration (seconds)
- Offset (seconds)

To transmit this information to the user terminal, a private descriptor (segment descriptor) has been introduced, which is carried by EIT Present/Following sections, along with other standard descriptors (short event and extended event descriptors) defined by the DVB-SI specification. A scheduler retrieves metadata from the database, generates EIT sections and sends them to the transmission centre via ftp protocol, where they are multiplexed by dedicated hardware and software tools to the digital stream conveying the video services.

At the client side, an application was developed on the Philips set-top box, which displays segment information: the application, resident in the set-top box memory, exploits DVB MHP APIs to extract from the transport stream textual information about programming (Today & Tomorrow), including also segment information on the present and following programmes, and other APIs to manage graphics.

4.3 Demonstration

A transport stream composed by three services was recorded and segment documentation was produced for the present and following programmes, storing metadata in the transmission

database; the transport stream completed with metadata was recorded again. The user terminal was fed directly via its digital input, bypassing the tuner, in order to reduce hardware requirements (modulator and up-converter).

In Figure 5, some screenshots of the EPG are presented.

Figure 2 – *Screenshots from the EPG application*