Balancing the Power of Multimedia Information Retrieval and Usability in Designing Interactive TV

Hyowon Lee*, Paul Ferguson,
Cathal Gurrin, Alan F. Smeaton,
Noel E. O'Connor
Centre for Digital Video Processing &
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Glasnevin, Dublin 9, Ireland
*hlee@computing.dcu.ie

Heeseon Park
Samsung Advanced Institute of Technology
Samsung Electronics
Republic of Korea

ABSTRACT

Steady progress in the field of multimedia information retrieval (MMIR) promises a useful set of tools that could provide new usage scenarios and features to enhance the user experience in today's digital media applications. In the interactive TV domain, the simplicity of interaction is more crucial than in any other digital media domain and ultimately determines the success or otherwise of any new applications. Thus when integrating emerging tools like MMIR into interactive TV, the increase in interface complexity and sophistication resulting from these features can easily reduce its actual usability. In this paper we describe a design strategy we developed as a result of our efforts in balancing the power of emerging multimedia information retrieval techniques and maintaining the simplicity of the interface in interactive TV. By providing multiple levels of interface sophistication in increasing order as a viewer repeatedly presses the same button on their remote control, we provide a layered interface that can accommodate viewers requiring varying degrees of power and simplicity. A series of screen shots from the system we have actually developed and built illustrates how this is achieved.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: User Interfaces; H.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

Design, Human Factors, Theory

1. INTRODUCTION

The trade-off between the simplicity of an interface and the power of functionality which the interface can feature is a well-known design issue in many technology domains from using the web to visualising information. Thus while simplicity is generally advocated in any design field, the science or the art of a good design seems to lie in how to maintain a level of simplicity as more and more functionality is provided, this functionality being enabled by the ever-growing developments in technology.

The medium of interactive TV brings this trade-off issue onto another level: the need for simple interaction is far more crucial in the interactive TV setting than in any other new media. The viewer's lean-back position and attitude and the fact that viewing is done from quite a distance using a remote control which has a coarse level of control, all dictates a strict adherence to an almost simplistic interface.

The simple and easy-to-use interaction style of the conventional TV set we have been using at home for generations has been due partly to the limited number of features that a conventional TV set has. These are basically turning on/off, controlling the volume and changing the channel. Inevitably, this small set of features has been relatively easily facilitated through a small number of buttons on a remote control.

The ever-increasing possibilities afforded by digital media, processing power and network connectivity in interactive TV naturally add more functionalities to our TV and hence to our remote control. Emerging new technologies in the field of computer vision and multimedia information retrieval promise a whole new set of functionalities that will change how we see and interact with the digital media around us. Automatic face detection and identification, for example, could save our manual photo annotation effort dramatically when we upload hundreds of our travel photos to a website, allowing searching and browsing based on the people appearing in the photos without any prior manual annotation [11]. While of course it is desirable and necessary to develop new technologies, often the technically-driven perspective seems to make any possible end-user application scenarios less amenable and realistic. An attempt to integrate any emerging or new technologies into interactive TV, although the idea itself is appealing, is difficult, and considering the high priority on simplicity required for an interactive TV interface, seems a contradicting goal.

In this paper, we examine the unique characteristics of the interactive TV medium that requires its designers to set the usability priority on simplicity rather than advanced functionality and demonstrate how a number of emerging

multimedia techniques can be incorporated in such a way to support rather than contradict the designer's rule of simplicity on the TV interface. Our contribution in this paper is the series of interface solutions we designed that feature new multimedia management techniques and yet still adhere to a simplistic TV interaction paradigm. We illustrate these design solutions with examples taken from our own implementation on an interactive TV platform.

This paper is organised as follows. In the next section, we first briefly review the unique characteristics of the interactive TV medium that necessitate far simpler interfaces than those of, say, a desktop PC, and then we review a selective set of promising functionalities from the field of multimedia information retrieval. In Section 3, we present our design solutions in incorporating these multimedia techniques to an interactive TV interface in a series of screen shots with a consistent and coherent interaction paradigm. Section 4 concludes the paper with the status of our implementation and our view on interaction design for new media.

2. USABILITY VS. FUNCTIONALITY

2.1 Usability Camp: Simplicity for TV Interaction

In this sub-section, we summarise the most important characteristics and issues in interactive TV interfaces and their implications for design. These are taken as our basis for designing solutions to incorporate multimedia information retrieval techniques, which we describe later.

2.1.1 Multiple Levels of Viewer Engagement

TV is sometimes termed a "lean-back entertainment" medium where the viewer typically sits 2m to 3.5m away from the screen, sits back on a sofa and does not pay too much concentration. The "lean-back" nature of TV watching has been frequently addressed as a unique characteristic of TV [5],[31],[8], and can be addressed both physically and mentally: while the viewer physically relaxes and sits back in the evening hours while watching TV, his/her mindset also becomes relaxed and "leans back". Jenson [21] refers to the principle of "lazy interactivity" which requires a low level of attention while interacting, and Bonnici [5] suggests the term 'viewer' rather than 'user' in order to emphasise the passive nature of the user's attitude and to differentiate it from the more engaged and active PC user. The question has been asked early on whether viewers would *want* to be more interactive or not in using their TV [37](p458).

In an extensive survey conducted with 1,872 TV viewers [25], four different levels of viewer attention were identified; watching TV with full concentration; watching TV while doing another activity (such as eating or chatting); using TV as a peripheral activity with a different primary activity; and TV as background noise. The significance of these multiple levels of attention is that our interactive TV interface, even if designed to target for one particular level of attention, should not hinder activities at other attention levels. For example, an interaction where the TV does not proceed until the viewer responds to a pop-up dialogue prompt is poor design in that it does not support a user doing other activities while the TV is on [7].

Watching TV from a distance means any interactive elements such as text and widgets should be large, and details should be minimum. Generally most design guidelines sug-

gest a minimum font size for interactive TV of 18pt [5], [27], [1] and the maximum amount of text on a TV screen was suggested as 90 words [1]. A study on the balance of visual attention between playback and text elements found that when a long text (approx. 160 words) occupies half of the TV screen the viewer's attention was evenly distributed between the video playback and the text area, whereas a shorter text (approx. 80-90 words) with small inserted images around the text resulted in more visual attention to the video area [22], providing some clues on the amount of detail for interactive elements when the TV is playing content.

2.1.2 Use of Remote Control

The main input device for interactive TV is the remote control, now a ubiquitous gadget that everybody uses at home to control their TV and DVD player. Due to different affordances which a remote control of a TV and keyboard/mouse of a PC exhibit, the kinds of suitable interaction mechanisms and widget behaviours of the two platforms are very different. Most importantly, the keyboard-/mouse-friendly widgets such as radio button, scroll bar, slider bar, array of icons, and menu hierarchy become inappropriate when ported to a TV with a remote control — far more simplistic, large and obvious widgets and styles are required for the coarse level of control the remote control allows. A straightforward menu hierarchy navigation on a PC platform with a mouse becomes an extremely awkward and laborious experience when using the up/down/left/right buttons of a remote control while at the same time trying to digest what video is being played on the TV. Thus instead of a complex hierarchy of menus, it is important to provide a flat or shallow menu where a few remote control buttons can directly select frequently-used features without any menu navigation. For example, some ethnographic studies in a home setting reveal how people watch TV/DVR in their natural environment [3], [4], [36], [14], [18], [28] and [25]. In particular, one ethnographic study [14] using a naturalistic inquiry into 10 family homes found that the most frequently-used features included Pause, Skip, Rewind, Go to Electronic Programme Guide (EPG), and Go to Recorded List. The design implications from the above finding are clear: map frequently-used features so they are directly accessible via a remote control and thus reduce the menu navigation burden on the viewers.

Entering text using a remote control has been a major problem and has been addressed in a number of previous works. Having a virtual keyboard on the TV screen or an SMS text messaging style input have been suggested but currently the research community seems to agree that cumbersome text input with a remote control should be avoided if possible. Allowing each viewer's own mobile device (such as a mobile phone or PDA) as a text input device has been suggested as possible solution [33], [30] but the real utility and experience of such methods is still to be experimented with real users. Using a remote keyboard for a TV is already available in some commercial products, but it is unclear whether TV viewers will be willing to type text while watching TV.

2.1.3 Enjoyment-oriented Design

One of the most distinguishing factors in interactive TV design is that the designers should not assume a highly-attentive user like traditional usability engineering methods do [6], because a more *enjoyable* interaction is not neces-

sarily a more *efficient* one [15]. While supporting a non-primary task (as mentioned in Section 2.1.1) should be considered so that the TV application can blend in with people's everyday tasks better [7], there is growing evidence indicating that traditional desktop usability principles do not account for the pleasure of the user experience [20] and thus an enjoyment-focused service such as interactive TV requires different designer focus, mindset and priorities from the start.

Usability evaluation issues for interactive TV interfaces have been drawn and explored in [31] [8], and a structured evaluation framework for interactive TV has been suggested [9], but these are still based on the theoretical assumptions and past experiences from other media devices and need to mature considerably further. On the other hand, task-focused, performance-based evaluations of interactive TV have been going on in order to identify the usability problems of specific interface features [15], [32], [22], [29].

The aesthetic quality of a TV interface is closely related to user enjoyment. It is difficult to define or provide aesthetic measures or a method to go about designing an aesthetic interface, but the field of Affective Computing has been investigating this aspect. Priorities for interactive TV design are nicely summed up as design for quick decisions, a short attention span, a hand-held remote control, and instant gratification [21].

2.1.4 Divided Attention between Playback and Interactive Elements

Unlike other media interfaces, interactive TV incorporates two distinct elements on its interface: video playback (of either live or recorded) and interactive elements. It is the designer's responsibility to arrange how they are presented together, to what extent each claims the viewer attention, and how well the two should blend with each other. There is a balancing issue between the two in displaying, emphasising, and allowing switching between the two. *"L-shape" layout* and *overlay* have been suggested broadly as two main ways to arrange these two elements, and their pros and cons are compared in [24]. While the currently more dominant "L-shape" layout (where the size of video playback is reduced and presented on one corner of the TV screen with the other interactive elements displayed in an L shape around it) has been suggested as superior [26](p154), others suggest an overlay (where interactive elements are overlaid on top of the video playback partially blocking the playback but maintaining its full-screen size at all times) as superior [17] (p216). Whichever is chosen, the viewer needs to switch his/her attention between the playback and interactive elements while watching the TV. A good design would be one where this attention switching is natural and minimally constrains one while the viewer is interacting with the other.

In many cases viewers will spend most of their time looking at video playback rather than interactive elements. The view that the interactive elements are only transient and should not interfere with the main task of watching the playback, focuses on the *TV* side of interactive TV (rather than the *interactive* side), probably maintaining the assumptions about the conventional TV viewers as the premise for interactive TV. In our design (as will be illustrated in Section 3), the overlay method has been used with semi-transparent interactive elements sliding in and out as needed, minimally hindering the main video playback. A study on semi-transparent

overlays [19] explains how the user's attention switches between the two elements: the viewer's primary task of watching TV changes to the secondary task when interactive elements appear with which the viewer engages as a primary task, while still performing the secondary task of watching TV. While the level of transparency and consequent visual interference of the background playback become important factors in arbitrating the viewer's attention division, exploiting what a human user is inherently good at (i.e. doing something while absorbing contextual information at the background) is usually a good strategy [34] (p62).

2.1.5 Design Guidelines for an Interactive TV Interface

It is true to say that the amount and the variety of design guidelines is an indicator of how much we understand about the medium. Contrary to the relatively long history of the term 'interactive TV', the number of actual examples of truely interactive TV is very small. Empirical studies of interactive TV use is rare and our collective experiences, skills and understanding of interaction with interactive TV is shallow. As a result, there are not many design guidelines currently available for designers of interactive TV applications. Sources vary from the developer's experience in the broadcasting company [17], compiled after conducting focused usability testing [2], from a corporate perspective [1], and from a literature review [26]. References to most guidelines can be found in [2] and [23].

One single most important theme we have found in surveying these guidelines is *simplicity*. Simplicity is advocated in the design of most media (e.g. graphic design and web design) but in interactive TV it is crucially more important and is the deciding factor of the success or failure of an application. Other commonly agreed and recurring themes among the guidelines include graphic issues (safe colour, brightness, graphic-/text-safe areas, font type, etc.), designing content for interactivity, and clearly marking when a user is entering and exiting the interactive features.

2.2 Technology Camp: Emerging Multimedia Tools

In the field of computer vision and multimedia information retrieval, a number of novel technologies have been researched and are steadily progressing. One of the main foci in this field is to automatically index and manage multimedia content such as photos and video clips in order to reduce the amount of human intervention in organising visual media content and providing a means for retrieval. For example, the TRECVid [35] campaign has been driving the video retrieval research community since 2001. In this annual exercise coordinated by National Institute of Standards and Technology (NIST) in the US, several dozen participating groups around the world each develop their own video retrieval tools to automatically detect various features in videos such as the existence of faces (general or specific), buildings, fires, forests, etc. The participants subsequently structure a large amount of video material for subsequent interactive retrieval. Although a truly semantic indexing and retrieval based on analysing the video content alone is a significantly difficult challenge and still in its early stage of research, over the past few years the TRECVid exercise has resulted in a variety of novel techniques and ways to enhance the accuracy of feature detections. In relation to

interactive TV which we are interested in, we have identified the following techniques as useful for us:

- Shot/Scene boundary detection — by analysing and comparing the visual differences between adjacent frames in a video, it is relatively easy to determine camera shot boundaries, with currently over 95% accuracy for straightforward hard cuts achieved. By aggregating the detected shots into clusters in terms of their visual and temporal proximity, detecting scene boundaries can be achieved but its accuracy can be subjective and hard to determine for general genres such as films and soap opera. While camera shots are short, syntactic units of video content, scenes are considered higher-level, more meaningful units, and can provide a feature similar to the "chapter access" or "scene selection" feature of today's DVDs.

- News story segmentation — for the TV news genre which is highly structured in its content, a scene equates to a news story and detecting news story boundaries in a TV news show has been successful, as repeating patterns of anchorperson shots followed by reporter's footage provide strong cues for automatic indexing.

- Keyframe extraction — from an identified shot, scene, or a news story, the most representative image (frame) can be automatically extracted by comparing visual similarity among the frames within each shot/scene and calculating the most average frame. More sophisticated methods incorporate camera shooting heuristics such as selecting the frame where a camera motion stops, or selecting the middle frame where an object appears from one end of a frame and moves across to exit at the other end.

- Sports summarisation — important events in sports (e.g. a goal, penalty or foul in soccer) usually follow a particular shooting convention, such as camera movement following a ball, growing audience cheering, a close-up of a player followed by a shot of the cheering crowd. By detecting these sequences appearing in the sports video, in some pseudo ordering, automatically determining important events in such sports is now possible with good accuracy.

- Search — the indexed units of retrieval (shot, scene, news story or a whole TV show) can be compared and ranked in terms of their similarity. Search functionality based on an EPG description and visual content similarity can provide a powerful and directed user query feature, without requiring any human indexing effort.

- Personalisation — using content-based similarity measures and collaborative filtering, a system can automatically infer which objects an individual user is likely to want to find. This can be applied to books, movies, or products a user may wish to purchase. Collaborative filtering is a relatively mature technique for automatic recommender systems and has already been applied to a number of online shops to recommend products. The main usefulness in the context of interactive TV is to filter out some EPG content so as to dramatically reduce the viewer's navigation effort during and between watching.

In addition to the above multimedia technologies, a networked environment means a system (a TV in our case) can receive and send information on currently online viewers, dispatching messages to and fro, and enabling social interaction amongst viewers. A networked TV also means that an EPG can always be up-to-date, and extra information that is not available on an EPG can be collected via web spiders to enrich programme meta-data. The techniques listed above are currently still being investigated and developed in research labs, and are mostly still at too early a stage of their development to deploy in a commercial TV product. While waiting for these automatic content analysis techniques to mature and become more reliable for real use, we designed our interactive TV interface to incorporate these techniques, assuming their maturity.

3. DESIGN SOLUTION

Our design solution for an interactive TV interface featuring a number of emerging multimedia techniques addresses most of design issues as mentioned in Section 2.1. In order to reduce the menu navigation burden from the viewer, the most important features of the TV are mapped to five colour buttons on a remote control and screen elements are accordingly colour-coded to strengthen the colour association. For each feature, pressing the button invokes the simplest and most basic form of that feature; if the same button is pressed a second time, a more sophisticated and advanced form of the same feature takes over; pressing the same button a third time brings up an even more advanced form of the feature, etc. Thus for each feature, the level of interface sophistication and consequently the power of the advanced feature is determined by how many times a particular button on the remote is pressed. This "spiral" approach allows a passive viewer to stick to the most basic features without being exposed to potentially confusing, more advanced features; those who want to avail of more power and control over the interactivity can do so by pressing the same button repeatedly. For example, three levels of EPG browsing schemes are provided and accessed as Figure 1 illustrates.
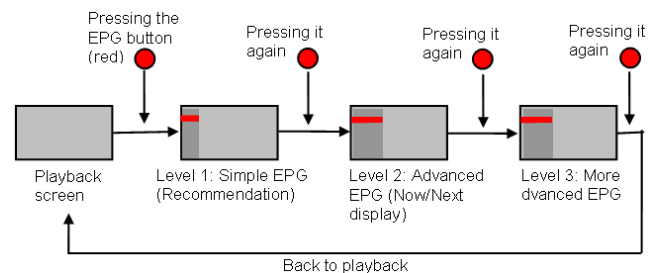

Figure 1: Multiple levels of EPG interface

For example, rapid double-pressing of a button will immediately brings in the Level 2 interface. This allows a viewer who prefers a particular level of interface sophistication to immediately invoke that level without having to go through the simpler levels. The invoked interactive feature slides in from an edge of the TV screen, partially covering the playback but still revealing what is behind due to the panel's semi-transparency. This minimally disrupts the main playback while a viewer is interacting with a slided-in panel. The following sub-sections describe each of the five main features

of our designed interface using screenshots from the system we have developed.

3.1 Browsing Genre-Specific Video Content

A number of browsing interfaces have been experimented within the field of multimedia information retrieval, which try to leverage the automatic indexing techniques for presentation to end users. Keyframe storyboards, interactive hierarchical browsers, dynamic slide shows and multiple Fast Forwards are some examples of this and mostly assume a desktop PC platform with intricate and delicate visualisations and user maneuvering using mouse and cursor. For a number of reasons explained in Section 2.1, most of these advanced approaches to content browsing are not suitable for interactive TV, and far simpler interfaces are required as the following screen shots we have developed, demonstrate.

In our application recorded TV shows go through a series of index processing depending on the genre of the show. For generic shows such as films and soap operas, shot and scene boundary detection followed by keyframe extraction occur and consequently a suitable presentation interface is prepared. Figure 2 shows a screen shot when the viewer presses the "Browse" button (colour-coded in blue) on the remote control.
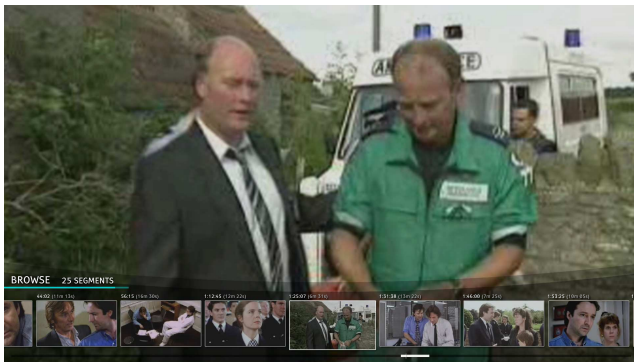

Figure 2: Browsing the currently played show

In Figure 2, a semi-transparent panel slides up to present 25 scenes of the current show (a film in this case) each represented by a keyframe. The currently played scene is in the middle of the strip of keyframes and slightly larger to indicate it is the current selection. The viewer can use left/right arrow buttons on the remote control to move the current selection, and pressing the "OK" button will jump the current main playback screen to that point. A thin timeline is displayed at the bottom of the panel, simply but effectively indicating the current point of browsing in relation to the whole length of the show. While this provides a convenient scene access feature as many DVD films do, the duration of each scene can be relatively long (in Figure 2, the currently selected scene is 6 minute 31 seconds long as indicated just above the keyframe). If the viewer is looking for a particular point in this scene, she/he would have to go to the beginning of the scene and start Fast Forwarding within the scene. A more powerful feature is to allow a within-scene browsing, by further segmenting the scene into shorter units. For this purpose, the viewer can press the same "Browse" button one more time to expand the panel to the double-layer browser (see Figure 3).

Figure 3: Hierarchical browsing for more fine-level navigation

In Figure 3, the current scene opens up a second layer of a keyframe strip that shows shot-level keyframes. This second strip lists all shots within that scene, in the case of the figure showing 12 shots within the current scene.

For a recorded news show, additional news story segmentation occurs as background processing, thus presenting the news show in story units. Figure 4 shows when the viewer pressed the "Browse" button while watching a recorded news programme.

Figure 4: Browsing the currently played news show

In Figure 4, what slides up is the same Browse panel as previously, but each keyframe is stamped with "News Story" sign, indicating that each keyframe represents a story. Selecting any of these will start playing that story. Here again, pressing the "Browse" button twice will expand this panel to show shot-level keyframes within the currently selected story (see Figure 5).

When a sports show is recorded, its processing includes our sports summarisation technique instead of scene detection. The result of sports summarisation is the detection of important events that happen in a sports match and their relative importance scores, as described in Section 2.2. Leveraging this technique, the "Browse" button can be pressed to see important events from a match (see Figure 6).
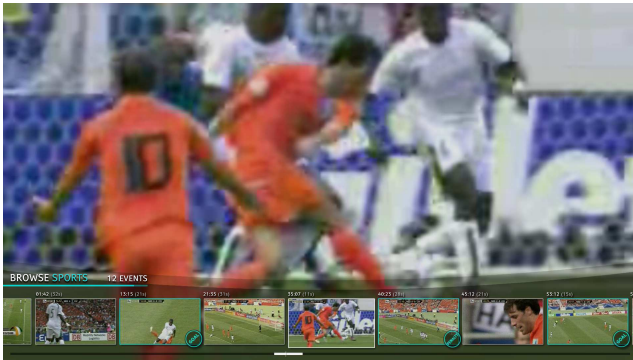
Figure 5: Hierarchical browsing of a news show

Figure 6: Browsing important events in a sports show

In Figure 6, total of 12 important events in a soccer match have been automatically detected and the keyframes representing each event are shown. Some of these events were identified as goals and a penalty, and accordingly stamped at the corner of the keyframes. Similar to generic/news browsing, the viewer can press the left/right arrow buttons to select an event, and pressing the "OK" button will start playing that event on the main TV screen. Pressing the "Browse" button one more time will expand the panel and show an overall visual summary of the match (see Figure 7).

Figure 7: Browsing a visual sports summary

The expanded panel lists all 12 important events at a glance, organised into 1st and 2nd halves. For each half, the keyframes each representing an important event are pre-

sented with different sizes: the more important the event is, larger the keyframe size is. For example, the two largest keyframes in Figure 7 are the "goal" events happening in each half of the match, shown as the largest two keyframes on the screen. The viewer can move the current selection and pressing the "OK" button on the remote control will start playing that event.

What is significant in our Browse feature here is that the viewer does not need to know anything about the double-layered hierarchical browsing or what buttons to press for different genres: a single "Browse" button on the remote control is used to invoke all of the panels shown so far, and there is no confusion as to how to use because different presentations as a result of different back-end processing are invoked depending on the genre of the currently played show. The multiple levels of simplicity/complexity of the navigation style are arranged in increasing order or complexity and can be easily ignorable if wished.

3.2 Search without Text Query

As mentioned in Section 2.1, text input usually required for querying is not suitable for interactive TV. Our solution is to use the viewer's currently viewing point (whether it be the show, scene, news story or shot) as the query itself. For example, while watching a film, the viewer presses the "Find Similar" button on the remote control (colour-coded as orange). This invokes a semi-transparent panel to slide in from the left (see Figure 8).
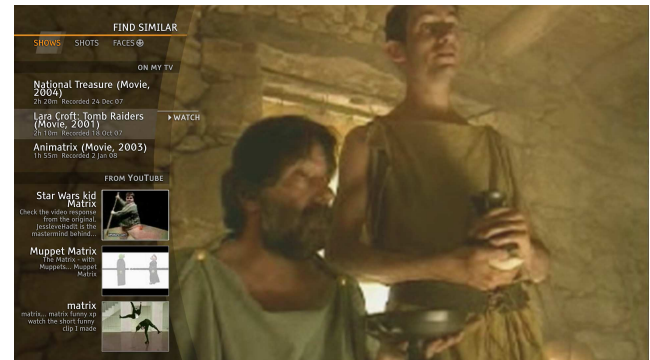

Figure 8: Finding similar videos with one button press

By default, similar shows/clips are selected (as indicated at the top of the panel) and thus on the panel a list of similar shows/clips to the currently watched film are listed. Retrieved entries are three similar shows previously recorded on the TV, followed by three clips from YouTube that are similar to the currently watched show. The viewer can press down/up arrow buttons to move the selection and pressing the "OK" button on the remote will start playing that show/clip on the main TV screen.

Pressing the "Find Similar" button one more time (or alternatively, moving the current cursor point to "Shots" just below the panel heading – this results in the same effect but is more cumbersome to maneuver) will retrieve a list of shots from all recorded shows that are visually similar to the currently watched shot (see Figure 9).

In Figure 9, the top five most similar shots to the current shot are presented, and for each entry a mini timeline in-
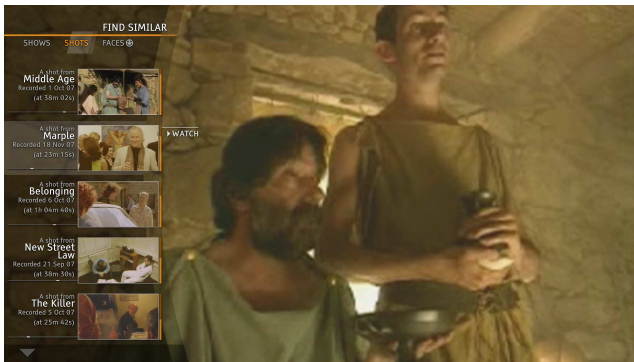
Figure 9: Finding similar shots to the currently watched shot

dicates where in that show the selected shot is placed. As before, the viewer can press down/up arrow buttons to move the current selection, and the "OK" button to play the show from that shot onwards. A feature such as this provides a novel way of jumping to completely different shows where a particular shot contain visually similar content.

3.3 Personalised Electronic Programme Guide

Designing an EPG has for some time been recognised as an important topic, as having a huge amount of channels with each show containing a variety of title, description and genre along with the time/date dimension presents a significant challenge for designers. Guidelines specifically addressing EPG design can be found in [5], [13] and [12]. In particular, Bonnici [5] mentions different needs for using EPG and the idea of designing different EPG interfaces depending on differing needs.

In our solution, the default EPG presentation when a viewer requests it is a filtered-out selection of EPG shows as a result of automatic recommendation based on the viewer's history of recording/watching and collaborative filtering from other viewers on the same service. When the viewer presses the "Programme Guide" button on the remote control (colour-coded as red), by default a personalised list appears as shown in Figure 10.
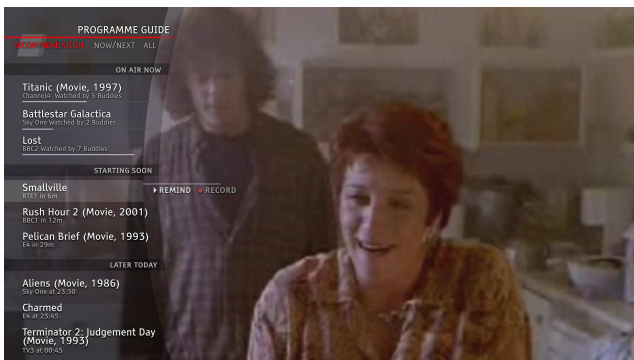

Figure 10: EPG: Personalised listing as default

In the listing presented in Figure 10, the first 3 shows are selected from the currently broadcast shows ("on air") each with a mini timeline indicating how far into the show it has been played. Selecting any of these will change the channel.

Below that, another list of three shows are presented that will start soon on the TV. For these, the viewer has an option of being reminded when the show starts or of requesting a recording of the show. The last three shows at the bottom of the panel are those that will be broadcast later. By using this selective list of shows, the viewer's need to find something else to watch can be efficiently satisfied.

One of the needs we have for an EPG is to see what is on now and what is coming up soon over all available channels [5]. In our solution we provide this feature as the 2nd level of EPG. When the viewer presses the "Programme Guide" button one more time (either a quick succession of double-pressing at the beginning or one press on the initial personalised EPG screen), the Now/Next panel is presented (see Figure 11).
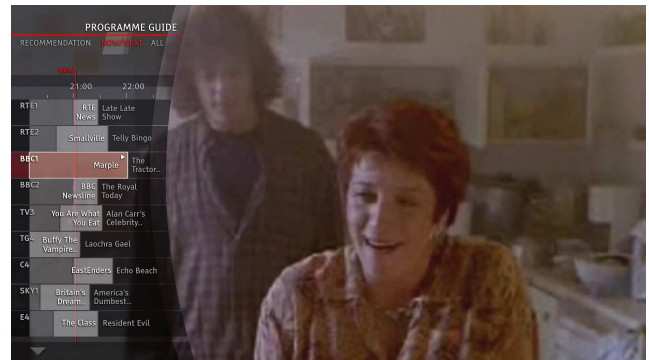

Figure 11: Now/Next display - quickly showing what's on now

Similarly to the "Find Similar" feature, the viewer can move the current selection (cursor) to "Now/Next" just below the panel heading and press the "OK" button, as an alternative to pressing the "Programme Guide" button twice. In this panel each row represents a channel and the currently broadcast shows are represented as bright grey boxes and the shows following them are in darker boxes. The current time is indicated by a thin and red 'Now' vertical line. More channels can be shown when the viewer brings the current selection point to the bottom of the list. Finally, the full EPG browsing is available when the viewer presses the "Programme Guide" button the third time (see Figure 12).
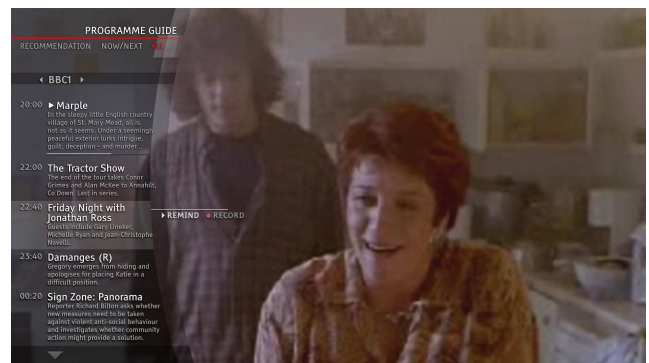

Figure 12: Browsing a full EPG

In Figure 12, the shows from the currently viewed channel are displayed with description of each show. The viewer

can press the up/down buttons on the remote to move the current selection of the show or the left/right arrow buttons to see different channels. In this way, what is displayed on the panel can be kept to a minimum and yet with four-direction arrow buttons on the remote control the viewer can easily and quickly navigate the full EPG. When a show is highlighted, appropriate options for that show appear beside it. For example, if the show is currently being broadcast, two options are available (Watch or Record) whereas if the show is scheduled later that night, the options are Remind or Record.

3.4 Browsing Recorded Shows

One of the most frequently used features on Digital Video Recorders is to go to the list of recorded shows [14]. When a viewer presses the "Recorded Shows" button on the remote control (colour-coded in green), a semi-transparent panel slides in with the list of recorded shows (see Figure 13). The default presentation is by date/time, but it could be re-ordered by frequency of watching, by popularity (by looking at all viewers of the service), or more elaborate visualisation techniques can be provided as options. In the default list, the currently selected show extends an information bubble beside it containing more textual information (from EPG) at the bottom half and visual summary video at the top half.
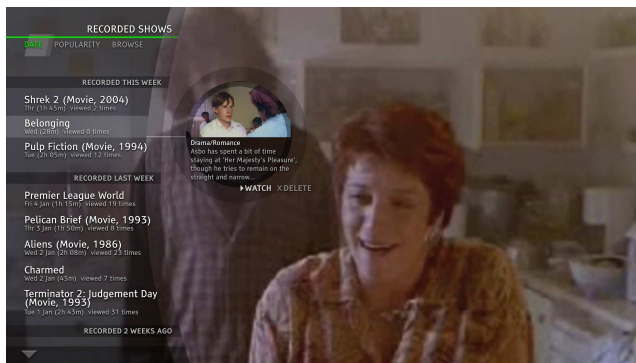

Figure 13: Browsing a list of recorded shows

The summary videos in the bubbles are generated differently depending on the genre. Currently in our design, a film summary video comes from querying YouTube with the film title and the first result is played back inside the bubble; a news summary video comes from detecting the news digest part of the news at the beginning and playing that part in the bubble; a sports summary video comes from our own sports summarisation where the important events are concatenated and played one after another in the bubble; for other genre videos we slide show the shot-level keyframes of the video in quick succession (3 keyframes/sec). There are a number of other ways of obtaining a playable video summary for the information bubble, which we are further experimenting at the moment.

3.5 Social Interaction

With interactivity and network connectivity built into interactive TV, the scope for a digital TV to support social interaction is huge. Simple voice connectivity among viewers in different locations has been found to be highly effective in facilitating sociability [18]. In our work, we focus on providing effective and useful social interaction without requiring extra physical equipment (e.g. wireless keyboard for chatting, camcorder for seeing each other's face, microphone for voice chatting, etc.). While watching TV, a viewer can press the "Buddy" button on the remote control (colour-coded in yellow) to see who else is watching the same channel as him/herself (see Figure 14).

Figure 14: Presence awareness: who is watching with me ?

On the right edge of Figure 14, the list of the viewer's Buddies (registered through a separate website where one invites the other to become mutual buddies, or in future versions leveraging a user's FaceBook or Bebo profile) who are tuned in to the same channel as the viewer, is presented. The list grows from the bottom of the list and newly-joined buddies slide in at the top of the list, while a buddy changing to a different channel slides out of the panel. The panel is thin and occupies minimal space on the TV screen in order to reduce the distraction of watching the main TV screen. When watching the TV, the viewer is always aware of who's tuning in and out, supporting a simple but effective *presence alert* [10]. Simple information on who is watching with me helps me decide which channel to watch, and the visibility of the group members watching the same channel enhances the feeling of belonging [16]. Pressing the "Buddy" button one more time will list all buddies who are watching other channels or are offline (turned off their TV) as shown in Figure 15.

Figure 15: Buddies currently watching other channels

The viewer can select a buddy from this list and notify him/her about the show the viewer is currently watching.

In Figure 15, the viewer selects his buddy Mary, to notify her about the show he is watching now. Simply pressing the "OK" button on the remote control will send the notification to Mary. Mary will receive a discreet notification on her TV (see the bottom right corner in Figure 16). Note that this social interaction requires no text input or other extra equipment. Our future work will be to extend this social interaction further in order to support more engagement.

Figure 16: Receiving a notification from a buddy

4. CONCLUSIONS

In this paper we addressed the balancing issue between simple and usable interface vs. potential usefulness of powerful functionalities afforded by emerging multimedia technology, and how getting the balance right becomes an even more critical issue in designing for interactive TV. We then presented our solution via a series of screen shots of the system we have developed and built, drawn from available design guidelines and from our own experiences in designing multimedia interfaces over a number of years.

As can be seen in our solutions, the TV screen is mostly occupied with full-screen playback with a rather ephemeral and transient interactive elements sliding in and out. Resembling more of a conventional TV than a PC interface, the passive viewer attitude of TV seems particularly well accommodated in our design. While the design aspect was the main focus in this paper, in order to demonstrate the interactivity we have designed we have fully implemented the interface in Macromedia Flex on a Windows-based PC connected to a 40-inch widescreen Samsung TV. This will allow us to conduct user evaluations in order to identify any usability problems introduced as well as help adjust various interface parameters (e.g. text and keyframe size, semi-transparency level, the number of displayed shows on a panel, etc.). Our system will need to be re-engineered to run directly on a networked TV set in order to fully leverage its design and capability as a commercial product.

When an interface to a new medium is being designed, it is natural to try to bring in our past expertise and skills from other more dominant media that we are used to. There may be some generic and transferrable design attributes and phenomena across digital media, but more often trying to re-use more established guidelines from other media fails because each medium avails of its own unique characteristics and a design that ignores these cannot bring out the medium's full benefits. Mobile interaction design went through this failure in which the re-use of the GUI paradigm from the desktop PC with its dense and delicate information visualisation and high level of visual attention, made small screen usability to the bottom. Mobile interaction designers are now aware of the importance of considering the unique characteristics of their medium and as a result they are in a much better position to develop good designs suited to the mobile environment and context.

In the same way, good design for interactive TV will come from the designer's awareness of the special characteristics of the medium. In our work, we started by fully understanding the characteristics of the TV and we tried to integrate potentially hazardous (in terms of usability) multimedia technologies while maintaining what is important in a TV interface.

5. REFERENCES

[1] Designing for Interactive Television v1.0, BBCi and Interactive TV Programmes. British Broadcasting Corporation, 2005.

[2] A. Ahonen, L. Turkki, M. Saarijarv, M. Lahti, and T. Virtanen. Chapter XII: Guidelines for designing easy-to-use interactive television services: experiences from the ArviD Programme. In *Interactive Digital Television: Technologies and Applications*, pages 207–223. IGI Publishing, 2008.

[3] R. Bernhaupt, M. Obrist, and M. Tscheligi. Usability and usage of iTV services: lessons learned in an Austrian field trial. *Computers in Entertainment*, 5(2), 2007.

[4] R. Bernhaupt, M. Obrist, A. Weiss, E. Beck, and M. Tscheligi. Trends in the living room and beyond. In *Proc. 5th European Interactive Television Conference (EuroITV 2007)*, pages 146–155, May 2007.

[5] S. Bonnici. Which channel is that on ? A design model for electronic programme guides. In *Proc. 1st European Interactive Television Conference (EuroITV 2003)*, pages 49–57, April 2003.

[6] K. Chorianopoulos. User interface design and evaluation in interactive TV. *The HERMES Newsletter by ELTRUN*, 32, May-June 2005.

[7] K. Chorianopoulos. Interactive TV design that blends seamlessly with everyday life. In *ERCIM User Interfaces 4 All (UI4ALL) Workshop*, pages 42–57. Springer, April 2007.

[8] K. Chorianopoulos and D. Spilnellis. Affective usability evaluation for an interactive music television channel. *Computers in Entertainment*, 2(3), 2004.

[9] K. Chorianopoulos and D. Spilnellis. User interface evaluation of interactive TV: a media studies perspective. *Universal Access in the Information Society*, 5(2):209–218, 2006.

[10] T. Coates. Social software for set-top boxes. In *Plasticbag.org*, 2005.

[11] S. Cooray, N. O'Connor, C. Gurrin, G. Jones, N. O'Hare, and A. F. Smeaton. Identifying person re-occurrences for personal photo management applications. In *VIE 2006 - IEE International Conference on Visual Information Engineering, Innovation and Creativity in Visual Media Processing and Graphics, pp144-149*, pages 144–149, 2006.

[12] O. Daly-Jones. The usability of electronic programme guides. In *Interactive Television Production*, pages 222–225. Focal Press, 2003.

[13] O. Daly-Jones and R. Carey. Navigating your TV: the usability of Electronic Programme Guides. Serco Usability Services, 2004.

[14] M. Darnell. How do people really interact with TV? Naturalistic observations of digital TV and Digital Video Recorder users. *Computers in Entertainment*, 5(2), 2007.

[15] S. M. Drucker, A. Glatzer, S. D. Mar, and C. Wong. SmartSkip: consumer level browsing and skipping of digital video content. In *CHI '02: Proc. SIGCHI Conference on Human Factors in Computing Systems*, pages 219–226, New York, NY, 2002. ACM.

[16] J. Fokker, H. de Ridder, P. Westendorp, and J. Pouwelse. Psychological background for inducing cooperation in peer-to-peer television. In *Proc. 5th European Interactive Television Conference (EuroITV 2007)*, pages 136–145, May 2007.

[17] M. Gawlinski. *Interactive Television Production. Chapter 5: Usability and Design*. Focal Press, Oxford, U.K., 2003.

[18] G. Harboe, N. Massey, C. Metcalf, D. Wheatley, and G. Romano. Perceptions of value: the uses of social television. In *Proc. 5th European Interactive Television Conference (EuroITV 2007)*, pages 116–125, May 2007.

[19] B. L. Harrison, H. Ishii, K. J. Vicente, and W. A. S. Buxton. Transparent layered user interfaces: an evaluation of a display design to enhance focused and divided attention. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pages 317–324, New York, NY, 1995. ACM Press.

[20] M. Hassenzahl, A. Platz, M. Burmester, and K. Lehner. Hedonic and ergonomic quality aspects determine a software's appeal. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pages 201–208, New York, NY, 2000. ACM.

[21] J. Jensen. Interactive television: new genres, new format, new content. In *Proc. 2nd Australasian Conference on Interactive Entertainment*, pages 89–96, November 2005.

[22] J. Kallenbach, S. Narhi, and P. Oittinen. Effects of extra information on TV viewers' visual attention, message processing ability, and cognitive workload. *Computers in Entertainment*, 5(2), 2007.

[23] T. Kunert, H. Kromker, and U. Kuhhirt. Navigation design guidelines for interactive television applications. In *Proc. 5th European Interactive Television Conference (EuroITV 2007)*, pages 7–23, May 2007.

[24] S. Lamont. An 8-step process for creating compelling enhanced television. In *Proc. 1st European Interactive Television Conference (EuroITV 2003)*, pages 133–136, April 2003.

[25] B. Lee and R. Lee. How and why people watch TV: implications for the future of interactive television. *Journal of Advertising Research*, 35(6):9–18, 1995.

[26] K. Lu. *Interaction design principles for interactive television. Chapter 5: Principles of interaction design for iTV (M.Sc dissertation)*. Georgia Institute of Technology, Atlanta, GA, 2005.

[27] S. Morris. A crash course in designing for the TV. In *Interactive TV Web*. http://www.interactivetvweb.org/tutorial/dtv-intro/tv-ui-design.shtml, 2004.

[28] J. O'Brien and T. Rodden. At home with the technology: an ethnographic study of a set-top-box trial. *ACM Transactions on Computer-Human Interaction*, 6(3):282–308, 1999.

[29] M. Obrist, R. Bernhaupt, E. Beck, and M. Tscheligi. Focusing on elderly: an iTV usability evaluation study with eye-tracking. In *Proc. 5th European Interactive Television Conference (EuroITV 2007)*, pages 66–75, May 2007.

[30] J. Park, M. Blythe, A. Monk, and D. Grayson. Sharable digital TV: relating ethnography to design through un-useless product suggestions. In *ACM CHI '06 extended abstracts*, pages 1199–1204, New York, NY, 2006. ACM.

[31] L. Pemberton and R. Griffiths. Usability evaluation techniques for interactive television. In *Proc. 10th HCI International Conference*, June 2003.

[32] M. Rice and N. Alm. Sociable TV: exploring user-led interaction design for older adults. In *Proc. 5th European Interactive Television Conference (EuroITV 2007)*, pages 126–135, May 2007.

[33] A. C. Roibás and R. Sala. Main HCI issues for the design of interfaces for ubiquitous interactive multimedia broadcast. *interactions*, 11(2):51–53, 2004.

[34] B. Shneiderman. *Leonardo's laptop: human needs and the new computing technologies*. MIT Press, Cambridge, Massachusetts, 2002.

[35] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.

[36] B. Svoen. Consumers, participants, and creators: young people's diverse use of television and new media. *Computers in Entertainment*, 5(2), 2007.

[37] J. van Dijk and L. Vos. Searching for the Holy Grail. *New Media & Society*, 3(4):443–465, 2001.