SURFACES: A NEW WAY OF LOOKING AT TV

Simon Parnall, Kevin Murray and James Walker
NDS

Abstract

The rapid evolution of home display technology offers the potential for an ever-more realistic and immersive experience of media and, within a few years, we will see large and yet also unobtrusive 'lifestyle' surfaces that could cover a whole wall. In the face of such capability the obvious question is "How might the television experience evolve?" and our vision is of a better, more integrated system that provides viewers with both a collective and personal experience and which adapts to a range of sources, including metadata, from both outside and inside the home.

INTRODUCTION

The choice of type and size of television screen for the home is so often a compromise between the extremes of an exciting viewing experience when the device is switched on and the wall or corner space occupied by a dark and dull object when the device is switched off. And, when the screen is on, the size of the picture may well be inappropriate for the type of content and engagement of the occupants of the room.

Science Fiction overcomes such concerns by assuming an invisible and scalable screen – often taking the place of the wall itself, or a window or indeed in mid air. Science Fiction has also assumed an intelligent management of presented material, following the individual and assimilating and prioritizing a range of sources.

Today's mobile phones make the Star Trek communicator look somewhat bulky as advances over the years have successively removed the novelty of such a concept. In the same way today's screen, projection and graphics technologies are slowly and steadily bringing us closer to a reality of the vision of Science Fiction. In fact, we are now very nearly at the point where key aspects of this vision can be realized and could be adopted by consumers in the not-so-distant future.

Walk into a consumer electronics exhibition today and you will find many example components of this vision. There are thin-bezel screens that can be treated as tiles to create larger and larger surfaces, or glass screens that transparently reveal the wall behind when off. We already have sophisticated companion devices offering touch control and each year we are seeing ever more sophisticated gesture and voice recognition.

Our role in this opportunity space will be to create the technologies that integrate such components to produce a sophisticated and intuitive user experience that matches content and mood, and which produces pictures of an appropriate size and position for each circumstance. Furthermore the presented audiovisual content will be supplemented with additional content and so-called domotic feeds (that is material concerning the home).

In this paper, believing in the inevitability of this trend in display technologies and the opportunities this creates, we set out our vision for how the television experience will evolve, some lessons learned from our first prototype implementation of this vision, and touch on our plans for the second-phase prototype which we are currently constructing.

VISION

Our vision of the future is of a viewing environment with one or more large display surfaces. Surfaces that are a) frameless, b) unobtrusive, c) ultra high-definition and d) ambient. These surfaces can be adapted to fill or partially fill one or more walls of a room, and will co-operate to provide an integrated experience. The opportunity is to open up possibilities way beyond the limits of today's devices though:

- content comprising multiple visual elements that can be adapted spatially and temporally, freeing the user from choosing a single element, or the system from having to impose overlays;
- shared, co-operative usage of the surfaces, with connected companion devices becoming personal extensions;
- supporting connected applications and services operating in a more streamlined, integrated manner, reflecting and effecting changes in viewer engagement in TV content;
- dynamic adaption to, and control over, the environment the surface finds itself in – such as physical size, resolution and the room in which it sits (e.g. adapting to the wallpaper or controlling the lighting); and
- introducing domotic content into the TV display in a sympathetic manner.

Based on this vision, a prototype was constructed and demonstrated at both IBC 2011 and CES 2012. This prototype has a single surface occupying most of one wall and a photograph of this is shown in figure 1. This shows a single surface constructed from six screens and one of several companion devices that may be used simultaneously to control and interact with the system.

Figure 1: Prototype System

IMMERSION

Many programs have a natural flow and pace – points at which the viewer or viewers are extremely immersed and engaged in the content. Examples of this may be a critical part of play in a sports game, a news story of direct relevance or a very dramatic scene in a soap. Likewise there may be times of lesser immersion or engagement. Examples of this may be waiting for players to take their positions, an uninteresting news item or a section of the soap that is recapping past happenings. In these areas of lesser immersion, the viewer's interest may naturally be taken by other related items, such as the current scores in related games, the next news story or what is being said about the soap by their social contacts.

In our system we have introduced the concept of 'immersion'. Immersion is key to the way that the surfaces are used and the way that the content is presented on them. Put simply, the more immersed in the content the viewer is, the greater emphasis that is placed on the core video, and the less immersed they are the more emphasis comes to be placed on related content which may then be introduced. This related material could be social media, advertising, program graphics, additional material, or virtually anything.

Examples of high and low immersion are shown in figures 2 and 3 respectively, which are screen captures taken from our prototype system. In figure 2, we see how the video roughly shares the surface with other social, voting and advertising graphics and content sources during the scene setting and build-up to the main performance. By comparison, figure 3 shows the high immersion example where the program in figure 2 has moved on to the main performance, and the related items have been removed, and the video increased in size and prominence.

Figure 2: a low immersion example

Figure 3: a high immersion example

In our prototype system, immersion is controlled in two ways – via "broadcast metadata" (as was used for the examples above) which indicated the broadcasters expected level of immersion, and also via a slider in the companion device which allows the user to modify the immersion (both up and down) as they wish. Clearly other mechanisms could also be employed, such as audio or video analysis of the room and the viewers, but the prototype shows that these two simple mechanisms work very effectively.

TECHNOLOGICAL MOTIVATORS

Displays

Display technology is continually improving. We have seen that relentlessly the average screen size is increasing year by year, as evidenced by (3). But there are two key technological changes which directly relate to our vision.

Firstly, screen bezel sizes are getting smaller. Our prototype system uses professional monitors with 5mm bezels, but LED backlit consumer displays are approaching similar, or better, bezel sizes and

OLED offers the prospect of a bezel width of near zero. Even with today's widths there is the real option of creating large ultra high definition surfaces out of tiled arrays of inexpensive displays.

Secondly, whilst still in the research laboratories, transparent displays which naturally allow the underlying environment to show through are starting to be developed. These would trivially allow the blending of displays into the room environment.

Video Content

We are also starting to see the first indications of the next jump in resolution beyond HD with the advent of 4K – both in displays and in content. At the same time as this higher resolution content is arriving, the importance of lower resolution content is not diminishing, whether from archives, citizen journalists or from challenging remote locations. Thus it is becoming hard to just assume that any content will look acceptable on any display size.

Non video Content

Outside the display arena, we are seeing ever more related data sources, from social media through games to dedicated websites. In the interconnected world, these are a crucial part of the entertainment experience, but today we are faced with the dilemma of either destroying the television experience by placing graphics over the video, or taking the viewer away from the lean back world of television into the very different and highly-interactive world of the internet.

BREAKING THE SCREEN BOUNDARIES

Today's television makes the basic assumption that "the display is always filled". Thus, video will fill the display, regardless of the size of display, quality of the video, or the resulting impact of an oversized face or object; and it also effectively does only one (main) thing at a time.

With larger, higher resolution display surfaces this implicit behavior and more can be challenged. Content need no longer necessarily fill the display surface, and the display surface can simultaneously be used for many different components.

In turn, these new capabilities mean that the traditional means of laying out video and graphics can be challenged. For instance we might:

- share the display between the content of more than one viewer, helping to make the TV a shared focal point rather than a point of contention;
- 'unpack' the constituent elements that are composited by a broadcaster in post-production, presenting these alongside the 'clean' audio-visual (AV) content, leaving it un-obscured. Obvious examples include digital on-screen graphics such as tickers, banners and sports statistics. To enable this, the composited elements would need to be delivered separately alongside the clean AV and then rendered in the client;
- 'unpack' all of the contextual assets that are composited in the Set-Top Box (STB), such as interactive applications and multi-screen content (e.g. multi-camera sports events);
- present contextually relevant online content alongside the video, for example, relevant web content, social comments (such as twitter hash-tags for the show), relevant online video etc;
- enable navigation and discovery user interfaces to be presented alongside video, going beyond today's 'picture-in-guide' presentation;
- present personal content, which whilst not directly related to the main television content, may still be desirable to end users to be seen on screen. Examples would

include personal social feeds, news feeds, images, discussion forums etc;

- present domotic content, such as user interfaces for in-home devices and systems, which can include video feeds from devices such as security cameras, door entry systems and baby monitors; and
- integrate visual communications, such as personal video calls, noting these may sometimes be used in a contextual way e.g. virtual shared viewing experiences between homes.

Thus, the way the TV experience takes advantage of the surface is by continuously managing a wide range of content sources and types that are combined appropriately for presentation.

Real Object Size

The tradition of a television picture scaling up to fill the display means that an object is effectively displayed at an unknown size. With this assumption broken, it now seems realistic to allow an object to be displayed at its real size, regardless of the display (as displays report their size though the standard connectors). For instance, in advertising it could be interesting to show just how thin the latest phone really is, just as is possible in print media today.

Content Opportunities

In the same way that the composition has always assumed a need to fill the rectangle, so has the creation of video content – which has followed the model of filling the proscenium arch of classical theatre. The proposed systems can offer new opportunities to the content creator.

One simple example of this is shown in figure 4. Here, the movie trailer is blended into the background to give the appearance that it tears its way through the wallpaper, dramatically conveying the unsettling nature of the promoted movie.

Figure 4: Non-rectangular content

There are numerous other areas where this technique opens up new opportunities. For example:

- editing could become more subtle with gentle fades, and several scenes can co-exist for longer and with less interference;
- content need no longer be fixed into a given size – if portrait content is provided from citizen journalists, then it can be displayed naturally in that form; and
- multiple synchronized videos could be used, in a fashion made popular in TV series such as 24, but without any requirement for their relative placement.

Implicit in this capability is the requirement to support an “alpha plane” style functionality that can be used both to describe arbitrary shapes and to allow for blending of the content into the background. This is, of course, not new and techniques such as luma and chroma keying are well known both in the professional head-end market place as well as supporting functionality in DVD and BluRay media. However, bringing this functionality into a traditional broadcast chain would represent a new usage.

A COMPANIONABLE EXPERIENCE

The growing importance of companion devices (tablets, phones, laptops etc) to the modern TV experience cannot be understated. Such devices permit us to construct an experience which is, at the same time, both collective (involving everyone in the room) and yet personal (allowing each person to interact with the various elements as they wish).

The companion device is key and integral to our prototype experience – and interactions with the companion device are directly connected with what is seen on the main surface(s). This is achieved through several means:

- The companion devices are able, within constraints, to adapt the content on display, including adding or removing components or re-arranging the layout. An example of this is interface is shown in the iPad screen capture of the web-browser in figure 5, where, for instance, the display can be re-arranged by dragging around the icons representing the parts of the content displayed on the surface.

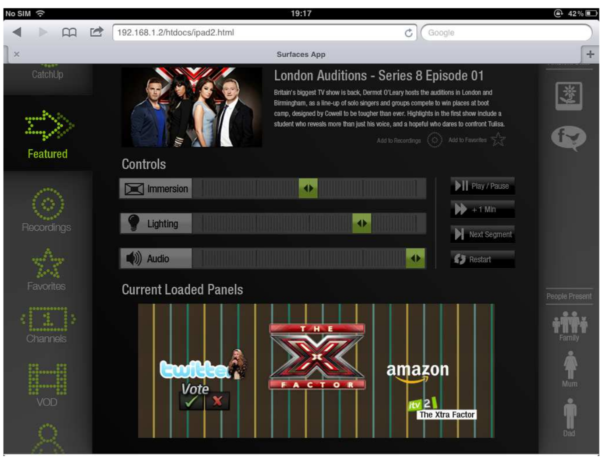

Figure 5: A Companion Application Interface

- Interactions, such as voting or feedback is done on the companion device, but this directly feeds back into the graphics displayed (in addition to the normal feedback one would expect).
- Control over the level of immersion. Although, as discussed earlier, a change in the level of immersion can be triggered through broadcast data and sensors in the room, the companion device is fundamentally able to control the final immersion experienced. In the prototype, as shown in figure 5, this is managed through a slider.

This approach results in interactions with the companion device that end back at the main display surface(s), rather than just with the companion device itself. For example scores from a game played by the whole family during a show could be displayed on the main surface.

A SURFACES SYSTEM ARCHITECTURE

The prototype system constructed to explore our vision was built using a single, six-output computer (an AMD Eyefinity graphics card in a powerful PC) with software that was itself built on standard HTML5 technologies (e.g. javascript and CSS transitions) in functionality largely contained within a standard browser. This approach enabled a fast and flexible development and exploration of the principles. Whilst the HTML-5 toolset proved to be an excellent platform, the use of a single six output graphics card places fundamental limits on scalability, the number of display surfaces that can be supported and, of course, on cost.

In our current work we are moving from the architecture of our first prototype. We are doing so because we will be using multiple display surfaces within a single room, and exploring how these can be combined for the presentation of a single entertainment experience, and co-operate to support multiple simultaneous entertainment experiences (e.g. the big game and the soap).

To achieve the required flexibility in the number of surfaces, scalability, cost and content presentation dynamism, we are developing a more advanced architecture, based around several concepts, including:

- rendering the graphics and video on more than one independent device;

- utilizing synchronization between the rendering devices, such as used in SAGE (1), but tailored for the specific use cases we are tackling;
- a separation of layout policy issues and rendering issues; and
- a single layout with a "world view" of the entire set of surfaces in use.

A high-level overview of the current architecture is shown in figure 6. This shows two separate surfaces, each driven by its own client. These clients then interact with the layout and synchronisation server(s) to ensure a consistent experience across the surfaces. In addition, the diagram shows that the audio is driven from only one surface, a deliberate choice to simplify the architecture.

Synchronization Architecture

It is important to be able to synchronize content spread between different clients. In a more traditional broadcast architecture, this would theoretically be possible using mechanisms such as the PCR values contained within a transport stream, but our approach does not assume either a direct transport stream feed to each client, or even that the content is made available in transport streams (e.g. it could be streamed over HTTP using any one of a number of mechanisms such as HLS or Smooth Streaming).

Instead, we have chosen to synchronize to a master audio playback clock on the main audio output. Where broadcast content is being consumed, there are many techniques that can be used to match this clock to that of the live broadcast content. This master audio clock is then replicated and synchronized via the synchronization server to other clients that are involved in playing back synchronized media.

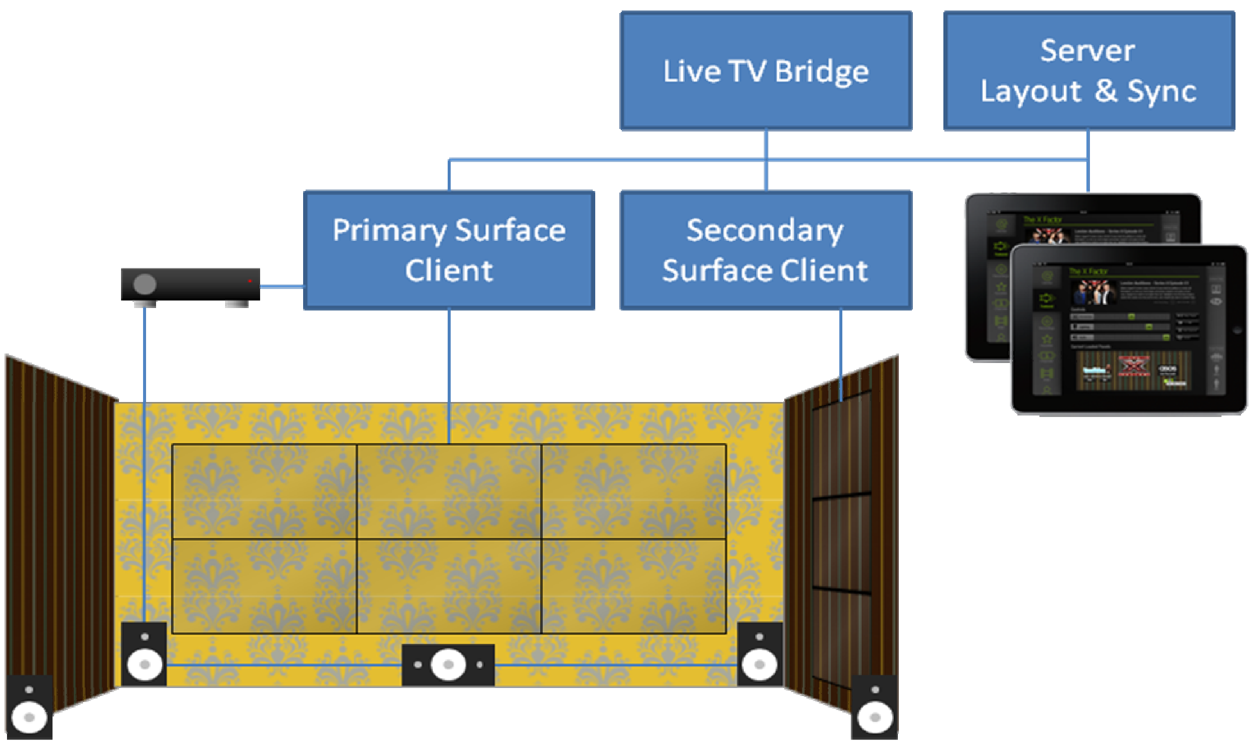

Figure 6: A New Surfaces Architecture

Our initial experiments with this architecture have shown that it appears to provide a reliable synchronization between different clients to a level that is acceptable for lip synchronization. Further experiments are underway to characterize and measure the accuracy that can be achieved.

Audio Architecture

Normally, audio is decoded and presented with simply a level control. However, in our proposed system the audio architecture becomes more complex than in a traditional approach, with various audio processing operations becoming an essential part of the overall architecture.

The most obvious audio processing requirement is positioning. From the proposed layout of surfaces in figure 6, it is clear that the secondary surface is not between the main speakers, and so any video that is presented on this surface with synchronized audio needs to have this audio repositioned. This repositioning needs to be dynamic, for instance as a video is moved from the primary to the secondary surface, the audio should be moved in synchronization. And, given the potential size of a surface, repositioning of the audio is desirable even when the content is moved within a surface. For example a video that occupies only the left third of the surface should have its sound stage correctly placed.

Earlier we discussed the concept of immersion, and how the video element of the experience can be balanced against other components to reflect the levels of interest both through a program's length and of a given viewer or viewers. This has a direct mapping to processing of the audio. Whilst the volume levels are one key part of this, this is best when combined with controlled compression – a reduction of the dynamic range of the content so that quieter parts become louder and the louder parts become quieter. Such processing allows the volume to be reduced in a fashion that retains access to the quiet sections of the content.

Much of the required functionality described above appears to be relatively easy to implement in the proposed Web Audio APIs that have recently become available on various platforms (2). This should make implementing the required audio architecture within an HTML5 environment relatively straightforward, and this work is currently underway.

Layout

One final component of the architecture deals with the layout of the media items to be displayed. Earlier in this paper we discussed how content typically packed together can be transmitted in an unpacked form, with the chosen and relevant components then laid out by the Surfaces system when the content is finally presented to the viewer. This process is not the highly constrained process we are used to where precise locations can be given for each item and, as the surfaces to be used might well be substantially different in each viewing environment, the process must be very flexible, and it is this flexibility that is an interesting challenge.

One aspect of the required flexibility comes from the number of inputs to the layout process to control what is displayed. These come from the local environment such as the range, sizes, locations and properties of the surfaces available and the immersion level of the viewer, and from the broadcaster, such as the list of potential components, their relevant priorities and a potential preferred immersion level. It is the layout engine that balances these inputs and selects a suitable set of components to display and locations for them.

In addition to the "what" of the layout is the "how', the appearance. More specifically, certain components may need to be adapted to the environment into which they are to be placed. For instance, if the room has white walls and the content item is white text, some means of making the text legible must be provided automatically. More generically, the design of an item should be able to adapt to the predominant background colors of the environment.

This introduces challenges at several levels that go beyond that of most current content presentation designs, such as may be found in many websites. Firstly we need an adaptive description of the requirements a broadcaster desires beyond those commonly in use today, and beyond even those of responsive web design (4). Next, we need a mechanism that can quickly and efficiently resolve these requirements in the face of a collection of local inputs. Finally, and perhaps most challengingly, we need the content producers and designers to understand that their content can and will be presented in many different ways, and a complete control over this presentation is potentially very counter-productive to the viewer's engagement.

CLOSING THOUGHTS

Our thinking started when considering the possibilities that the display industry will be offering in just a few years' time when the black boxes in the corners of out rooms disappear and unobtrusive, frameless, ultra high definition ambient surfaces take their place. In exploring the opportunities this technology will offer we have come to consider how content is presented, and the way in which its various components (current and future) will be assembled for the viewer. We have come to an appreciation of the way in which control and interaction with such an experience can work both in a personal and collective manner. And, in contrast to the 'lean forward' experience of today's connected TV we have seen how the 'lean back' experience of Surfaces requires a sophisticated automatic layout control engine.

As we have explored function, so we have explored form, and the PC based solution for a first stage demonstration now begins to give way to a believably scalable and cost effective hardware and software architecture.

It is often commented that the role of television in our lives has changed dramatically as other devices have fought for our time and won our attention. And yet, families and groups still wish to spend time together, sharing space and switching between personal and collective experiences. A developed television experience which embraces this truth, and which invites immersion and interaction at appropriate levels, must surely be for our industry a goal worth aiming for. Surfaces is, for us, a vehicle to explore this space and we are excited by the future we see before us, and the reaction we have received. The future is not one where the medium is marginalized, but a future in which people will truly find a new way of looking at TV.

REFERENCES

(1) High-Performance Dynamic Graphics Streaming for Scalable Adaptive Graphics Environment, *SuperComputing 2006, November 11-17 2006.* http://www.evl.uic.edu/files/pdf/SAGE-SC06-final.pdf

(2) Web Audio API, Chris Rogers, W3C, https://dvcs.w3.org/hg/audio/raw-file/tip/webaudio/specification.html

(3) Of Large LCDs, Unused Fabs, and Projector Killers, Pete Putman, Display Daily, April 9th, 2012. http://displaydaily.com/2012/04/09/of-large-lcds-unused-fabs-and-projector-killers/

(4) Responsive Web Design, Ethan Marcotte, 2011, ISBN 978-0-9844425-7-7, http://www.abookapart.com/products/responsive-web-design