Reflections on the digital television future

Stefan Agamanolis, Principal Research Scientist, Media Lab Europe

Authors note: This is a transcription of a keynote presentation delivered at Prix Italia in Catania, 19 September 2003. A lightly edited version of this transcript was printed in **Television Quality and Product Innovation in the Digital Age***, a book summarizing the proceedings of the workshop by the same name at Prix Italia 2003.*

I would like to begin by thanking the Chairman and all of the organizers of the Prix Italia conference for welcoming me here in Catania to say a few words on this topic that means a lot to me and on which I have many opinions. In my talk I would like to tell you about a few projects that I have worked on in the Media Lab organization as a way of illustrating some of the challenges and opportunities that we are all working with regarding digital and interactive television. In doing so, I will ask many questions, and I may not have all of the answers, but hopefully I will raise a few important issues that can feed into the later discussion.

Let me begin by telling you a little bit about my own background. As the Chairman mentioned, I come from the United States. I grew up in a small town in the state of Ohio and I went to a small college where I studied computer science together with philosophy and film. This background was seemingly attractive to the admissions board at the MIT Media Lab, and so I spent the next 7 years of my life going through the entire graduate school program there. During my time there I was a researcher in a group called Object-Based Media. This group was, and still is, trying to think about what happens when we think about media not in terms of pixels and scan-lines and frames, but rather in terms of *objects*, like chairs, backgrounds, 3-dimensional models of scenes, together with metadata that describe them and scripts that orchestrate how they appear in a presentation of some kind, like a television program.

One reason to think about media in this way was that we thought we could achieve better coding efficiency – that we could pack the same television program into a smaller number of bits. It was this kind of thinking that led to the development of object-based media encoding algorithms like MPEG-4. However, this way of thinking was also advantageous when we wanted to experiment with adding interactivity to television programs.

One of the first projects that I worked on when I arrived at the MIT Media Lab about 10 years ago was an interactive television program called **The Museum** [video]. This program, while it will undoubtedly look primitive by today's standards, is a good example of what I am talking about. While watching the program, the viewer had a physical knob that he could turn in order to slide the position of the camera around in the scene and watch the action from different directions. This was possible because the scene itself was not just a piece of video – it was an object, in this case a 3-dimensional model of a room in a

museum. Each actor is also an object. These actors were shot separately on a blue-screen stage from multiple camera angles. As the viewer moves around in the scene, the playback computer chooses the actor views that most closely correspond to the position of the virtual camera and layers them into the scene in real time. The overall result is a TV program in which the audience member isn't sitting still in a virtual theater seat but can move around and experience the action along a path chosen by the producer of the program.

This and other experiments that we did raised many questions. One of the biggest questions that emerged over time was: **Is 'interactive television' an oxymoron?** – or: **Is television really 'enhanced' if it is interactive?** We commonly think about television as a passive experience – something that we don't interact with, and more importantly, something that we don't *want* to interact with. Maybe this passiveness is a feature, not a bug – we *want* someone to take full control of our minds and tell us a captivating tale. Is it possible that interactivity is incompatible with this notion of traditional television content – in the same way that when you are asleep and dreaming, and then you realize you are dreaming and try to steer the dream, usually the whole thing falls apart and you wake up.

In any case, perhaps this idea of television being a passive experience is too ingrained in the minds of ordinary people and we should be looking for a different term that reflects the fact that the kinds of experiences that we refer to when we say "interactive television" now are really a completely different kind of content – a different genre, deserving of its own distinct category to cleanly differentiate it from that of television. The use of such a new term might help people understand the features of these experiences better, and this might help to increase the chances that these technologies will be accepted more widely over the coming years of the transition to a digital television infrastructure.

If we forget about television for a moment, there are many reasons why you might want to make a *story* interactive. For example, say you want to give an audience member the ability to discover a story world on their own or at their own pace, in a way that reflects the way we learn and discover things in everyday life, or in a way that reflects the content of the story in some way. Other reasons relate back to the tradition of oral storytelling, a rich mode of interactive communication in which a storyteller receives passive and active feedback from, adapts to, and interacts directly with an audience in real time while delivering content.

So, the question is, can we design compelling broadcast experiences that are interactive, but at the same time don't disturb a sense of deep immersion in an alternate reality – in other words, **Can we control our dreams and stay asleep?**

Along similar lines, another question I think we have to ask ourselves is: Can we create "television" programming in which the interaction is

not just an "add-on" or an "enhancement" but rather it is a core and essential part of the viewing experience? So again, it is not really "television" any more but a different kind of genre.

One possible example of such programming is **HyperSoap** [video], another project we worked on a few years ago at the Media Lab. This program is essentially a soap opera combined with a home shopping application. While you watch the drama and laugh at the extremely cheesy dialogue, if you see something in the scene or worn by one of the characters that you like, all you have to do is touch the object on the screen (or point to it with a pointing device), and the object will become highlighted and a small box will appear and tell you how much it costs and where to buy it. Or the system might offer a coupon for a discount on the item at the store, or it might even give you the opportunity to purchase it online immediately by clicking a second button (something we didn't implement in the prototype but is an obvious extension of the idea).

This HyperSoap project highlighted many interesting issues. It was a program in which interaction was an integral part of the experience, not an optional enhancement. It was also an exploration of what kind of programming we might expect in a digital TV world filled with personal video recorders (PVRs) in which viewers can simply skip over traditional advertisements. HyperSoap was an example of *100% product placement* – everything you could see in the program was for sale and was an advertisement, from the magazine on the desk to the sofa in the background, to the salon that dressed the hair of the main actress.

In HyperSoap, the viewer interacts with the television program, but the program itself doesn't change – only the information and decorations on top of it change. Now let's think about the opposite – where the viewer doesn't directly interact at all, but where the program content itself responds and changes in a passive way to the viewer. We could call this *personalized television*, or a term I like better: *responsive television*.

Today's TV sets are still largely "dumb terminals" – they just receive a signal and play it on the screen. But users of the World Wide Web are used to content that responds in a fairly rich way to preferences, profile information, history of interaction, and so on. The realm of video has largely been left out of this revolution of online responsiveness. We are still watching one-size-fits-all television programs tailored for large populations, and while that's fine and even preferred for many types of programming, there are others that might really benefit from a more finely-tuned customization of the content to specific viewing circumstances.

However, we are slowly entering an era in which there is a lot more processing power and bandwidth available throughout the chain from production to display, and these new capabilities enable a different

model of television. In this model, instead of editing the program prior to transmission, the broadcaster transmits all of the "raw materials" of a program directly into your receiver or set top box, and your set top box does the final edit of the program at the same time you are viewing it. This enables new kinds of programming that were not possible before, such as advertisements that are specifically tailored to the personal profile of the viewer, or educational or news programs that take into account the viewer's background. Real-time feedback is also possible – for example, if there is some kind of sensor on the set-top box that knows whether or not you are in the room, or even if you seem bored or confused, the program could be automatically re-edited instantaneously to take account of that information. Because this editing happens on the receiver, sensitive personal information does not need to be transmitted outside of the household or viewing environment.

Some new questions arise from these ideas. First of all, what kinds of tools do we need to make responsive television content, and more importantly, **how are we going to author broadcast content that is richly responsive yet maintains high production values**? Even more importantly, **can we build tools that will enable directors and producers to maintain full control over how their television programs look in different situations?** Speaking as an occasional director myself, if I want to make a "sex" and "violence" knob on my television program, then I want to have the ability to control exactly what happens when those knobs are turned – I don't want a third party or some kind of censorship organization making those decisions for me.

There is another interesting question that emerges: **What about the water cooler?** How will be able to stand around the water cooler at the office the next day and have a conversation about what was on TV the night before if none of us is seeing the same thing? This is a definite concern, but it may not be as big a concern as we think, because *seeing* the same thing is not necessarily the same as *experiencing* the same thing. Responsive television could be harnessed as a means for *equalizing* experience over a group or population of people. For example, a responsive news program could tell a story in a particular way that might help me to understand it most easily given my background, expertise, and current mental state, and this might be different from what makes a story most accessible for another person.

At the Media Lab, we have been working on an authoring tool for responsive television called **Viper**. This tool essentially allows producers to create television programs that can re-edit *themselves* on the viewer's set-top box. Unlike other automated editing systems that merely splice clips together end-to-end with simple cut transitions, Viper offers the ability to use a variety of more complex editing constructs commonly needed to make compelling programs, like L-cuts, inserts, dissolves, transitions of various sorts, graphic overlays, and so on. The characteristics of these edits can be tightly controlled and affected in complex ways by various "response factors" about the

viewer or his environment, such as preferences, profile information, or real-time sensing and feedback mechanisms.

Several productions have been made with Viper so far in order to test its functionality. For example, we made a political campaign advertisement that tailors itself to present the candidate in the most positive and compelling way for each individual viewer, based on information about the viewer's profile, preferences, and concerns. It's a provocative example of what could be possible to do with this technology, but nevertheless an important one to explore.

Thus far we have touched on many different kinds of "added value" that digital television technologies can bring to the world. We've heard about added *commercial* value – being able to make more targeted and interactive advertising and to sell things in new ways. We've alluded to added *storytelling* value – the potential for telling stories in new and different ways that might be more appropriate or compelling. We've also been hearing in other places about added *social* value – the ability to access government services through your TV.

There is still one important area of potential value that we have been missing, and to know what it is we need to remember one of the most important characteristics of broadcasting: **Broadcasting is all about creating a *shared experience*.** It's about millions of people seeing the same time, often at exactly the same time. It's also about the things that happen as a result of this shared experience, such as the formation of fan communities, or people gathering to watch and enjoy TV shows together, even if they are on the telephone separated by great distances. So how can we use these digital technologies to **enhance a sense of shared experience**, to **enable new modes of communication**, and to **support the creation of new communities** focused around television media. In effect, how can we use these technologies to ***broaden* the effect of broadcasting?**

As an example of this, consider the **Reflexion** project at the Media Lab [video]. In this prototype, a viewer sits in front of a television screen with a camera mounted above it. The software extracts a picture of the viewer from his background and creates a "magic mirror" in which you can see a reflection of yourself together with the reflections of other people in other places. The computer listens to when you are speaking and alters the layering and opacity of each participant on the screen in real-time to emphasize the persons who are most active while at the same time maintaining awareness of those who are less active. Because the real backgrounds are masked out, the virtual background can be filled with something like a slideshow, a document, or a movie. So it is possible to imagine watching your favorite television show, or perhaps a live televised sports event, together with your friends and family in other places. Because everyone appears at the bottom of the screen together with the program, a sense of shared experience is potentially enhanced.

Reflexion is an example of an application that would require a broadband "back channel" between all clients in order for it to work in the way demonstrated at our laboratory. While it may be possible to imagine variations on the theme that require less bandwidth, hopefully projects like Reflexion emphasize the importance of the back channel in digital television infrastructures, especially as they are trialed with the public over the next few years.

Let's take the bandwidth idea even further and imagine a future in which everyone has a very high bandwidth upstream connection to the rest of the networked world. In this case, we can begin to wonder about the answers to questions like: **What happens when it's possible for *anyone* to become a broadcaster?** The Internet and the World Wide Web enabled anyone to become a *publisher*; what is the next leap of technology that will enable anyone to become a *broadcaster*? And more interestingly, **what implications will this have for the traditional broadcasting business?** Is it going to jam all of the channels of the big media companies out of existence? What new forms of entertainment might we expect to see? What new forms of grass-roots teaching and learning? What new ways of sharing ideas and concepts will we have? **Will we need a new kind of "TV guide"?** Finally, and perhaps most importantly, **how can today's media companies evolve and survive in a future like this?**

There is a place where we are thinking intensely about things like this, trying to project ourselves out into the future to the greatest extent possible, and to create working prototypes that reflect aspects of this thinking – it is Media Lab Europe, the European research partner of the MIT Media Lab. This lab is independently funded by a combination of governmental organizations and corporations from all over Europe, and we are currently looking to expand our partner community. The research group that I lead at Media Lab Europe is called Human Connectedness, and we are especially concerned with exploring *the future of human relationships* as mediated by technology. There are other groups that focus on things like haptics, speech interfaces, biometrics, learning, storytelling, and so on. There are many other projects that I haven't mentioned from both my own group and the others that are available to browse online (links listed below), and you may also contact me directly for more information.

Once again I wish to thank the organizers of the Prix Italia conference for inviting me to speak, and I look forward to the discussion that will ensue.

Stefan Agamanolis – stefan@media.mit.edu

Principal Research Scientist, Media Lab Europe

http://www.medialabeurope.org/hc (Human Connectedness group)

http://www.medialabeurope.org (main Media Lab Europe web site)