2010 IEEE Consumer Communications and Networking Conference
Area: Multimedia Communication and Services

Paper Title: neXtream: A Multi-Device, Social Approach to Video Content Consumption

AUTHORS / AFFILIATION / ADDRESS

ReeD Martin*	MIT Media Lab	MIT Media Lab 20 Ames St. Cambridge, MA 02139 USA	reedm@media.mit.edu
Ana Luisa Santos	MIT Media Lab	MIT Media Lab 20 Ames St. Cambridge, MA 02139 USA	alsantos@media.mit.edu
Mike Shafran	MIT Media Lab	MIT Media Lab 20 Ames St. Cambridge, MA 02139 USA	mikeshafran@gmail.com
Henry Holtzman	MIT Media Lab	MIT Media Lab 20 Ames St. Cambridge, MA 02139 USA	holtzman@media.mit.edu
Marie-Jose Montpetit	MIT Media Lab	MIT Media Lab 20 Ames St. Cambridge, MA 02139 USA	mariejo@media.mit.edu

* Corresponding author

ABSTRACT

Video content has become available on an increasingly diverse set of devices and from an ever growing number of sources, creating a vast amount of choice for viewers. At the same time, the varying methods of viewing, interacting with, and sharing content have diverged. This paper introduces neXtream, a new approach to delivering video by integrating multiple devices, content sources, and social networks. This concept is developed following research in social television and converged applications, providing both personalization features and social interaction. NeXtream delivers video by dynamically generating streams of video customized to a viewer, while facilitating a common dialog between users around the content, creating both a user- and community-centric viewing experience. NeXtream integrates smartphones, PCs, and TVs to deliver video content to viewers. The paper presents the system concept, theory, and architecture, and describes the developed prototype.

KEYWORDS

multimedia; social TV; smartphone; personalization; video; IPTV; social networking

neXtream: A Multi-Device, Social Approach to Video Content Consumption

ReeD Martin, Ana Luisa Santos, Mike Shafran,
Henry Holtzman and Marie Jose Montpetit
MIT Media Lab
20 Ames St.
Cambridge, MA 02139 USA

***Abstract*–Video content has become available on an increasingly diverse set of devices and from an ever growing number of sources, creating a vast amount of choice for viewers. At the same time, the varying methods of viewing, interacting with, and sharing content have diverged. This paper introduces neXtream, a new approach to delivering video by integrating multiple devices, content sources, and social networks. This concept is developed following research in social television and converged applications, providing both personalization features and social interaction. NeXtream delivers video by dynamically generating streams of video customized to a viewer, while facilitating a common dialog between users around the content, creating both a user- and community-centric viewing experience. NeXtream integrates smartphones, PCs, and TVs to deliver video content to viewers. The paper presents the system concept, theory, and architecture, and describes the developed prototype.**

I. INTRODUCTION

The media viewing experience is becoming richer and more exciting than ever before as video content becomes available on an increasingly diverse set of devices, including TV sets, personal computers, game consoles, dedicated media boxes, and mobile devices. However, traditional broadcast networks continue to deliver content targeted to the masses and constrained to the networks' programming schedules, rather than tailored to the individual. Digital Video Recorders (DVR) have introduced the ability to record content to fit each viewer's schedule, but haven't changed the fundamental way broadcast TV content is delivered. At the same time, computers and mobile devices enable users to choose online content they are interested in and even share content recommendations through online social networks.

While these devices present new options, they also introduce complexity and frustrations for some viewers. The content available on a TV set is not universally available online, while online content is usually only accessible on a TV set by connecting it to a PC. The combination of the growing online content sources and the number of long-tailed broadcast networks creates a potentially overwhelming number of choices. This, along with significant differences in content navigation, creates a fragmented and confusing experience for many viewers and a widening of the generational and technological gaps.

While TV viewers usually expect a passive, lean-back interaction, mobile device and computer users actively seek out video they want to watch online. While the interactivity of PC viewing provides additional flexibility, it can also be more complex for viewers accustomed to the traditional TV experience. We are witnessing a media convergence: TV, PC and mobile devices are becoming one [1,2].

This paper introduces neXtream, a multi-device, socially-aware framework for video content consumption. The paper begins in Section II by exploring the neXtream concept with a focus on social interactivity and personalization. The system architecture is discussed in Section III, followed by the system features and implementation details in Section IV and socioeconomic aspects in Section V. Finally, we explore some of the future possibilities for neXtream in the conclusion, Section VI.

II. CONCEPT

NeXtream defines a framework and implementation for the next generation of media consumption and TV. It provides both social interaction and personalization features, creating a user- and community-centric viewing experience, where video content is delivered through dynamically generated streams. The design of neXtream is also very much aligned with recent work on contextual interfaces and devices, as it adapts to user behavior and preferred devices [3].

We define a "stream" as a sequence of videos clips, aggregated according to a specific theme or topic such as news, reality TV, comedy, cartoons, or education. Stream content can also be flexible: for example, viewers can create a stream of their favorite newscast, advertisements for their preferred products, or their friends' favorite sports clips. By creating theme-oriented content aggregation, neXtream aims to retain the now-familiar cable model of dedicated content channels.

A. Content Aggregation

Content creators deplore the fact that the attention span of the American TV-viewing public is becoming progressively shorter [4] and is increasingly attracted to content like YouTube shorts. NeXtream capitalizes on this trend by automatically assembling multiple short video clips into a longer viewing experience. By aggregating relevant content, it attempts to capture a viewer's attention for longer periods of time, allowing a lean-back experience with a wider range of content. Hence, neXtream's personal aggregated content can contain videos from various sources, whether commercially produced or amateur media, selected to retain user attention. Videos within a stream can be of any length, whether a short YouTube clip, a half-hour sitcom, or parts of a sporting event.

As social relevance plays an increasingly important role in the media experience [1], another goal of neXtream is to define a shared media experience and encourage users to participate socially in the content creation process. In neXtream, the online community determines the aggregate regional, national, and international popularity of specific video content. In addition, the user's social network is involved as a "virtual operator" [2] to determine content socially relevant to the individual. The role of the traditional TV broadcast network programmer is replaced by the online community at large, the user's social network, and the user's individual taste, determining the content to be delivered, at what time, and in what order.

Videos that have been viewed most often, commented on frequently, rated highly, or specifically shared between users within a social network, are more likely to be viewed by each user in that network. This creates an immediate connection with friends and community based on shared interests, and provides members of a social network with a greater shared viewing experience [5].

Streams can also be assembled to the user's specific requirements. For example, while a video about baseball may be very popular, both among the greater population and among a viewer's social network, if the viewer has shown repeatedly not to be interested in baseball, the video will not be included in the user's stream. This filtering algorithm is constantly adapting to better reflect the user's likes, dislikes, and general viewing habits, in addition to sorting and aggregating social recommendations. Fig. 1 shows this top-down aggregation of video content, ultimately filtered by the user's interaction with neXtream.

B. *Social Networking*

Another social feature of our solution is to allow users to share common dialogs via integration with major video content aggregators and social networking platforms. These connections bridge the gap between an individual's content viewing experience and their existing social network. This follows a model that was pioneered by work on connecting TV set-top-boxes to Facebook [6].

NeXtream takes advantage of integrating the capabilities of multiple devices. For example, a television and a smartphone are linked, and host elements of the user interface on the device that makes the most sense. By moving social network dialogs onto the mobile device, it is able to avoid interrupting the TV experience, while simultaneously facilitating asynchronous community dialogs around content [7].

C. *Delivery Platforms*

NeXtream is designed to be multi-platform, bridging devices from TV sets to PCs to mobile devices. It uses this whole ecosystem of screen-enabled devices to serve as interchangeable front ends to a relatively standard computational back end [8]. It combines the lean-back experience of the television and the interactive experience of a PC or a mobile device. It also allows that experience to be portable; neXtream's video content can be accessible to users wherever they go. Through the neXtream network architecture, which keeps track of the users, devices, and streams, contextually relevant information, such as UI elements and interaction information, can be provided on the appropriate device at the correct time.

D. *Interactivity*

A smartphone is neXtream's main user interface as well as a complementary secondary screen. It provides a controller, linked to the TV set or PC, for navigating the media landscape. When navigating these streams, the two devices act synchronously. The phone's display and input capabilities are also used for interacting with related content. For example, while continuing to watch videos on a TV set (or computer), a user can interact with friends, share opinions and ratings, and look up related information on the phone [9]. This approach preserves the big screen real estate for high definition video content. In these situations, the devices act asynchronously.

III. System Architecture

NeXtream aggregates multiple sources of content onto different devices which, while providing a simple user interface that masks this aggregation, requires a reliable and comprehensive system architecture for content management and device interoperability. Our architecture allows neXtream to provide the services of a Common Distribution Network provider or a Repurposed Content Aggregator. Both models enable the distribution of content over any IP-based network, whether cable, wireless, or mobile phone, as well as to any device. A summarized version of the architecture is available in Fig. 2.

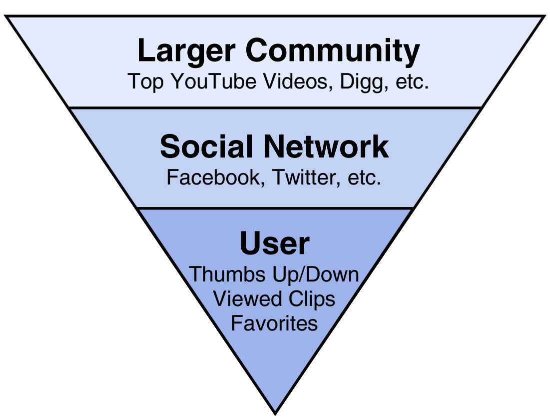

Figure 1. Top-down stream generation hierarchy from a larger community down to the individual

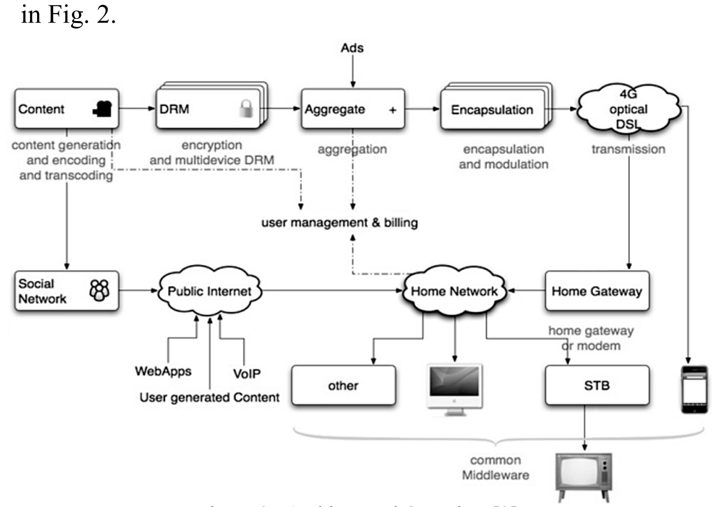

Figure 2. Architectural Overview [2]

A. *Common Distribution Network*

As a Common Distribution Network, neXtream can aggregate, transcode, and reformat content from a variety of partner sources, originally provided with a variety of encoding, encryption, and transmission features. In this model, we can manage and distribute content from online and traditional broadcast networks and operators, as well as directly from content creators. While this model requires partnerships with content owners in order to manage media viewing rights and distribution, recent developments in Digital Rights management like the Coral Consortium [10] point in the direction of seamless content consumption.

B. *Content Aggregator*

As a content aggregator, neXtream can remain somewhat source-agnostic. In this model, media is delivered to the user by linking directly to pre-existing content without rebroadcasting, re-encoding, or restricting the use of content in a way different from what is already in place from the source. While this model is simpler from a content management perspective, it becomes rapidly limited by the content's availability at varying online sources, as well as introducing potential rights management issues between sources and delivery. In addition, due to the rapidly growing combination of source coding/encryption and devices ecosystem, even in this model we may need to have some elements of content brokering and transcoding. As a content aggregator, our system provides a middleware platform for operators, existing aggregators, and content providers alike to provide their own branded experiences.

C. *Virtual Operator*

Independent of the two previous options, neXtream's architecture allows it to become a virtual operator via direct plugins to popular social networks such as Facebook, for general media, or LinkedIn, for business oriented streams. In that sense, neXtream is a next generation Facebook TV [6], leveraging the reach of social networks to create a more advanced TV experience. In this architecture, the neXtream platform uses tools like Facebook Connect[1] or specifically defined APIs to manage users and provide a single location repository for content.

IV. PROTOTYPE AND IMPLEMENTATION

To demonstrate the feasibility of the neXtream concept, a prototype was developed using a smartphone and a TV. The purpose of the prototype is to evaluate how the two devices should interact together: what information belongs on each screen, how the two devices discover each other, and how to best integrate a user interface that spans both. The prototype is implemented with an iPhone acting as a smart remote, synchronized with an Apple TV connected over HDMI to a television set, and connected to an online server providing "back-office middleware services" such as the storage, brokering, and retrieval of video content and links, user data, and social networking information.

[1] http://developers.facebook.com/connect.php

A. *Operation*

The prototype retrieves online video feeds over RSS from a server hosted at our laboratory, which populates content onto the interface of both the TV and iPhone. We built a database of links to online content on our server, which is based around basic learned user interests, a social network (Facebook), and the larger community (YouTube top videos.) Without losing generality, but for convenience in the prototype, using YouTube content "channels" enables us to quickly pull together video streams based around well known categories and genres.

The iPhone application is written entirely in Objective-C. The iPhone and Apple TV are discovered using the Bonjour[1] protocol over WiFi, which is then used to open a TCP/IP socket connection via Open Sound Control[2] (OSC). OSC brings a real-time data flow between the two devices that allows taking advantage of each device's capabilities to improve the user interface, particularly to navigate streams. Subtle movements of the iPhone impact the remote and the television simultaneously (Fig. 3) by mirroring the two interfaces. As a user scrolls through clips and streams, the elements are simultaneously mirrored on the television and the iPhone. This simultaneous replication on both interfaces compensates for a touch-screen device that contains little to no haptic feedback. Constantly looking at the remote to perform functions distracts the TV viewer and disrupts the immersive nature of TV [11].

The Apple TV uses Quartz Composer[3] to display content on the television, providing a rich 3D graphical interface. Quartz Composer directly communicates with the remote via OSC to relay fetched RSS and video streams back to the iPhone, as well as receiving user actions and displaying them on the TV.

The server architecture is based on an HTTP server running PHP to handle requests from the Apple TV and the iPhone, and a MySQL server to manage video URLs, "favorited" clipdata, and individual user ratings (Fig. 4).

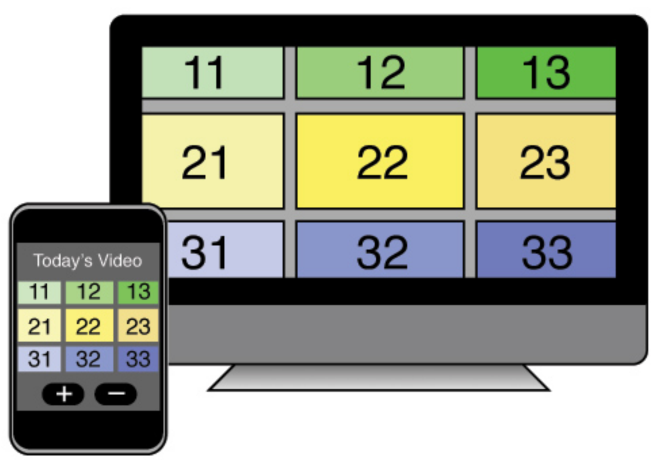

Figure 3. Replication of interface on both devices

[1] http://developer.apple.com/networking/bonjour/
[2] http://www.opensoundcontrol.org/
[3] http://developer.apple.com/graphicsimaging/quartz/quartzcomposer.html

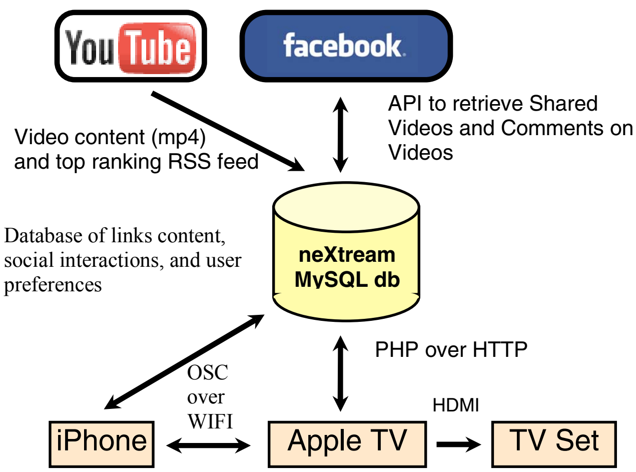

Figure 4. Server and device architecture of neXtream

B. Stream Formation

We use a simple algorithm to determine types of content a user might like, which is based on popularity within a user's social network, combined with metadata related to clips the user has indicated he or she likes or dislikes. This is related to work already done on community-mediated content [12].

Stored user preferences are based on a few basic actions, both implicit and explicit. When a user watches a video clip from beginning to end, there is an implicit assumption that the user enjoyed the clip, at least enough to continue to watch. Should a video clip be skipped, there is an implicit assumption that, based on when the skip occurs, the user may not have liked the clip to some degree. For example, if a user watches 4 minutes of a 5-minute clip, he or she may have enjoyed certain aspects of the clip, but became bored part of the way through. In this case, we assign a "*liked this*" probability of 4/5. If a user instead skipped that same clip after 30 seconds, we assign a probability of 1/10 that the user liked the clip. While this probability assignment is certainly simplistic, for the purposes of our prototype it proved reasonably effective.

The remote interface also offers the user the ability to specifically *like*, *dislike*, and *favorite* clips, where we can assign an even greater probability to the user's preferences. In this case, we set the "*liked this*" weight to 2 for *like*, -1 for *dislike*, and 3 for *favorite*, in addition to the sliding 0 to 1 scale for the implicit indications mentioned previously (Fig. 5.)

In order to determine the clips in a user's stream, the prototype takes the metadata that are associated with a video and averages the "likedness" value across each individual keyword to arrive at an overall average a. This average a is multiplied

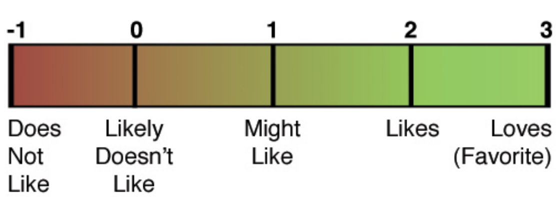

Figure 5. How individual preferences are ranked and stored

by 1.5, a value based on a simple assumption that any amount of the clip over 2/3 watched was assumed to be generally liked by the viewer (3/2 being the inverse of 2/3). This new weighted value will be exponentially weighted. Values over 1 will rapidly increase the result, while values below 1 will decrease it. This value could easily be adjusted by further user investigations on how to find a better correlation between how much of a video is watched vs. the probability of "likedness". The weighted average is then combined with a strength value S, which is based directly on the number of times a user has indicated a preference for a video clip. This yields the total "likedness" value v, as in:

$$(a \times 1.5)^{S} = v. \qquad (1)$$

Equation (1) is used to determine a threshold for a clip to be included in a viewer's stream. For the purpose of the prototype, we used a v value of 0.01 as a threshold. If a video contained a keyword (metadata) with a value of 0.01 or lower, the clip would not be included. This value was chosen by trial and error in order to yield the type of results we were looking for. For example, skipping a clip right at the beginning more than twice will flag the keywords negatively, while it would take 15 skips to do the same if the user watches over half of each clip. Again, a more advanced algorithm to weigh the likelihood of when a clip might occur (e.g., a user's preferences might change over time) could easily be substituted to create a more sophisticated metric. Our current model also does not allow for a video ranked low to ever be viewed—that is, once a keyword passes its negative threshold, its value is no longer adjusting, as it is no longer being shown to the viewer.

C. Social Interactions

In addition to real-time content interaction, our prototype also supports social networking and interactivity. The iPhone interface contains five tabs: *What's On*, *Media*, *Friends*, *Messages*, and *Me*, each divided into different subsections. The *What's On* tab is the primary interface for neXtream, acting as a remote control for the television as well as offering the ability to perform actions on the current clip. This tab also provides an indicator of communication with friends around the current video clip, allowing the user to comment asynchronously with friends about the currently playing video clip.

The other four tabs focus entirely on the community built around the video, and act independently of the TV interface. This allows a user to interact with his or her social network, find new media, and browse different clips, all without affecting the content shown on the TV. The *Media* tab contains a user's saved and "favorite" clips, an ability to browse an online network of featured and archived content, as well as a search mechanism to discover new clips through a variety of media sites. The *Friends* tab contains a live component to see what other friends are watching at the current moment, a feed of all friend video activity, and a list of all of the user's friends and their profiles, which are disclosed according to a user's privacy settings. The *Messages* tab contains an inbox of messages and videos that friends have sent directly to the user, with an at-a-glance indicator which alerts the user to new messages. The *Me* tab allows the user to see their individual preferences, modify settings, and create an outward-facing profile.

V. Socioeconomic aspects

When dealing with consumer electronics in a fast evolving landscape no one can ignore the socioeconomic impacts of a new offering. The joint rise of user-generated content (UGC) made popular by YouTube [13] and of social networking sites like Facebook is creating important disruptions in the way video content is consumed and distributed. The current shifts in the video industry are major, and the changes that we are witnessing are permanent. It is already predicted that the Internet will carry up to 91% of all video traffic by 2013 [14].

NeXtream has the potential to change the traditional video industry value chain by leveraging the disruption in the industry and embracing change which will create new value propositions for content providers and distributors alike. While it directly affects the economic value of social networks as video brokers, it also allows content providers, including advertisers, operators and end user device manufacturers to retain customer loyalty and grow their customers base. Social networks can directly affect the content that is provided to users, and content providers and advertisers, in turn, can profit from social features. Ratings, comment sharing and recommendations enable a viewer to easily watch content of interest, and further influence what friends will watch and buy. Novel social video consumption approaches like the one promoted by neXtream may finally give social networks the real monetization they seek. There are already examples of television shows driving traffic to second screens, such as those on PCs and mobile devices, for instant purchase of related content. For example, iTunes and the Fox channel have already experimented with the show “Glee”, with great success [15]. NeXtream goes a step further by allowing viewers to share content and ads with friends hence increasing the “virality” of popular streams and products.

VI. Conclusion

The rapidly changing and expanding media ecosystem for consuming video content presents both challenges and opportunities for using existing technology to locate and watch video content. The majority of service providers and content producers seem content with utilizing existing infrastructure and distribution systems, while the user is looking for more flexibility and innovation. The desire for a single, simple solution, to both explore and sort through the growing pool of video content, will only continue to grow.

This paper presented neXtream, a novel approach to social device-agnostic video content consumption. We propose an approach to consolidating, sorting, and prioritizing video content to resolve the mounting diversity of video offerings. By leveraging and integrating many existing technologies and platforms into one unified solution, neXtream allows users to continue to enjoy the lean-back approach to television that they are accustomed to, while providing new interactions with their social network, direct access to purchase products seen on TV, and most importantly, the ability to find content that is of interest to them and their community.

Moving forward, we plan to focus our research and development on many aspects of neXtream. We want to improve our dynamic stream building algorithm and to test neXtream’s usability with a significant number of users to determine its applicability, strengths and potential improvements. We will also refine the privacy and security that leverages our group's previous work, as well as other research in the MIT community [6]. Other areas of research include aggregating multiple sources of content, digital rights management, advertisement delivery, and viable revenue models.

The development of neXtream has allowed our team to navigate the ever-growing converged landscape of TV, online video and communications. Our research has shown that by simplifying access to the content a user is interested in, providing that content on a device of choice, and preserving the social aspect of television, neXtream provides a powerful approach for next generation TV.

Acknowledgment

We would like to acknowledge the MIT Media Lab MAS 962 Social TV community for comments and inspiration.

References

[1] N. Klym and M.J. Montpetit, “Innovation at the Edge: Social TV and Beyond,” *MIT Communications Futures Program*, October 2008, cfp.mit.edu

[2] M.J. Montpetit, N. Klym and T. Mirlacher, “The Future of IPTV: Adding Mobility and Social Networking,” *Proceedings of the 1st International Workshop on IPTV and Applications*, June 2009.

[3] P. Cesar, K. Chorianopoulos, and J. F. Jensen, “Social television and user interaction,” *Computers in Entertainment (CIE)*, January 2008.

[4] R. Kubey, *Creating Television: Conversations With the People Behind 50 Years of American TV (A Volume in LEA's Communication Series)* Lawrence Erlbaum, November 2003.

[5] P. Cesar, D.C.A. Bulterman, D. Geerts, A.J. Jansen, H. Knoche, and W. Seager, “Enhancing Social Sharing of Videos: Fragment, Annotate, Enrich, and Share,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2008.

[6] B. Mariana, “Television meets Facebook: Social Networking Through Consumer Electronics”, *MSc thesis, MIT Media Lab*, September 2008, http://pubs.media.mit.edu/pubs/papers/maiki-ms.pdf

[7] C. Harrison and B. Amento, “CollaboraTV: Using asynchronous communication to make TV social again,” *Proceedings of Euro ITV 2007, July 2007*, pp.218–222.

[8] G. Bell and A. Zafiroglu, “Teaching Intel to Love the Television” Industrial Case Study, *Proceedings of Euro ITV 2008*, July 2008.

[9] P. Cesar, D.C.A. Bulterman, A.J. Jansen “Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content,” *Changing Television Environments,* v5066, 2008, pp.168-177.

[10] http://www.coral-interop.org/

[11] I. Poupyrev, S. Maruyama and J. Rekimoto, “Ambient Touch: Designing Tactile Interfaces for Handheld Devices.” *Proceedings of UIST 2002*, pp.51-60, 2002.

[12] D.A. Shamma, R. Shaw, P.L. Shafton, and Y. Liu, “Watch what I watch: using community activity to understand content,” *MIR '07: Proc. of the International Workshop on Multimedia Information Retrieval*, 2007, pp. 275-284.

[13] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, “I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system,” *IMC ’07: Proc. 7th ACM SIGCOMM Conf. on Internet Measurement*, 2007, pp.1–14.

[14] C. Wilson, “Cisco: IP traffic to hit 667 exabytes by 2013,” June 2009. http://telephonyonline.com/global/news/cisco-ip-traffic-0609/

[15] http://www.fox.com/glee/