CollaboraTV: Using Asynchronous Communication to Make TV Social Again

Chris Harrison
AT&T Labs
chris.harrison@nyu.edu

Brian Amento
AT&T Labs
brian@research.att.com

Abstract
With the increasing use of digital video recorders and video-on-demand services, people are becoming less likely to watch shows synchronously. Many social aspects of television have become troublesome or have been lost entirely. For example, co-workers might talk over lunch about a show that aired the previous night. However, this ability is lost if people do not watch the show for days or weeks, or out of order with other programming the group has recorded. CollaboraTV provides an interactive and engaging way for users to participate in recorded shows, despite watching at different times. This is primarily achieved with a virtual audience, which captures events generated by viewers asynchronously. Initial results of a comprehensive usability study are presented, which indicates users have a strong affinity toward a virtual audience interface and that asynchronous interaction is enjoyable, practical and valuable.

Introduction
Television is undeniably a major component of modern society. It is not only the dominant media activity, but is considered the most exciting and influential media type (TVB, 2006; FCC, 2006; Putnam, 2000). TV is also pervasive, outnumbering people in the average home; less than a fifth of households have a single TV (AP, 2006; Children Now, 1995). Viewing habits are changing as well. Despite increasing competition from the internet, TV usage has been steadily increasing, and is now at it's highest level since viewing data was first collected, a 50% increase since the 1950s, and a 12% increase from 1996. The average person watches 4.5 hours of programming a day, with the average household tuned in for more than 8 hours (FCC, 2006; Putnam, 2000), consuming almost half of people's total leisure time (Bureau of Labor Statistics, 2006; Putnam, 2000). Television was once championed as the "electronic hearth," which would bring families together (Tichi, 1991). However, people are increasingly watching TV without their families, with some studies suggesting at least half of Americans usually watch alone (Putnam, 2000; Knowledge Networks, 2004).

Partly to blame is the increasing prevalence of Digital Video Recorders (DVRs), sales of which are expected to accelerate, approaching 250 million users by 2011 (ABI Research, 2006). Primetime shows used to bring families together. Now, and especially as more DVRs gain the ability to record multiple shows simultaneously, family members are opting to watch content based on their particular interests, often separating them from their family. Compounding this problem is the large volume of content catering to different demographics and interest. The average number of channels available grew from 18.8 in 1985 to 100.4 by 2003 (Nielson, 2004). The sheer quantity of content makes it less likely for people to watch the same shows. In the 1950s, two-thirds of all Americans tuned in to watch "I Love Lucy." The top two programs in the 90s, ER and Seinfeld barely achieved one-third (Putnam, 2000). Following this trend, the top watched show of 2006, American Idol, only drew 17.7% of viewers on average.

DVRs are not only impacting the time families spend together, but also the ability for friends, co-workers and extended family to discuss shows – the so called "water-cooler effect" (Putnam, 2000). People may watch shows days or weeks after they first air, and can watch newer shows before older ones. This considerably complicates how groups of friends can discuss TV, as it is likely that only a subset of the group has watched the same show. This causes people to be left out of discussions. Some might actively avoid conversation, fearing they might learn details that could spoil yet-to-be-seen episodes. Furthermore, by the time everyone in the group has seen an episode, parts of the group might be one or more episodes ahead and thus forgotten many discussion points, know new details that make earlier points less interesting, or are uninterested because too much time has elapsed.

This declining ability for people to interact, particularly with those external to the immediate family, is eroding once strong social ties. However, all indications seem to point towards a lack of ability to communicate, not a lack of desire. If systems can be built to facilitate communication, social capital need not be lost. However, in order to accommodate DVRs, a sophisticated asynchronous communication paradigm will have to be employed.

CollaboraTV is a simple, but powerful set of communication facilities, which support interaction between groups of people that do not watch shows concurrently. While individuals watch a show, they can generate certain events, such as gestures and

textual annotations, which are attached temporally to the media stream. When other users watch the same show subsequently, these events are replayed at the equivalent time they were generated and visualized using a small avatar-based, virtual audience. The system also provides a show-recommendation system, powered by social data, which helps users find (potentially new) content that other users are watching. Not only does this help filter the ever-increasing breadth of TV content, but also pools people together to increase opportunities for communication.

Related work

TV-based communication has been the focus of substantial research (Oehlberg, 2006; Coppens, 2004; Fink, 2006; Luyten, 2006; Geetrs, 2006; Regan, 2004). Typically, communication systems for TV provide text, voice and/or video chat for synchronous viewers. Many systems, including CollaboraTV, provide mechanisms that allow users to see what shows their friends are watching and the ability to watch synchronously with them.

Live TV has two important properties: 1) All users are constrained by a common start time, the show's airdate and 2) all users are forced to watch in parallel. This makes it easy for users to watch live shows together, as the variability has been removed (Although time zones are problematic). However, there are considerable obstacles for synchronous communication on DVRs. Foremost, users wanting to watch together will have to coordinate a common time. Of course, it is possible for people to join groups ad-hoc. However, it is unclear whether users are willing to tolerate waiting for others to join before starting the show or having to join mid-show in order to watch in parallel with other users. Another complication is that people will lose sync with other users if they pause or rewind, a common behavior for local disturbances like telephone calls. It is possible for control actions (i.e. pause) to affect all people in the viewing group, but it seems likely that users would get annoyed; the larger the group, the higher the likelihood of interruptions - a troublesome effect for a system designed to bring groups of people together. Inevitably, due to individual pressures, users will watch a substantial proportion of TV at different times than their friends. This means many users might end up watching TV with people they are less interested in hearing commentary from, or simply watch alone. In any purely synchronous system, where only simultaneous viewers are visible, valuable communication is lost. CollaboraTV sidesteps this problem by providing a common platform for synchronous and asynchronous communication.

The benefit of avatars has been widely researched (Diederiks, 2003; Smith, 2000; Salem 2000). Television is traditionally watched in groups, which makes avatars an obvious technique to simulate presence, humanizing users that are local or remote, or in the case of CollaboraTV, potentially teletemporal as well. Avatars are ideal for both annotations and non-textual communication, allowing for gestures and postures that provide a dimension of expression that people already find familiar and natural. CollarboraTV's virtual audience extends ideas and visual elements proposed by PARC's SocialTV, which uses avatars to provide presence cues to synchronous users. CollaboraTV's virtual audience uses avatars to represent both synchronous and asynchronous users, and more importantly, as a conduit for user communication, both for text annotations and emotion (via gestures).

CollaboraTV

CollaboraTV supports asynchronous annotation of shows, currently limited to text. These comments are generated at any point while watching a show. Future viewers will see these comments, linked temporally with the video, and can leave their own. If multiple users are synchronously watching, they are able to interact "live." This allows future viewers to see previous discussions adding their own thoughts to further evolve the topic.

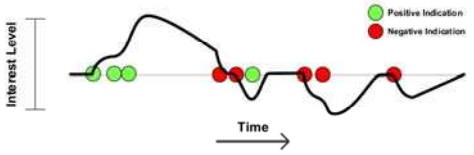

Figure 1. Discrete interest points forming a continuous interest profile (exaggerated for clarity).

Users can also generate temporally linked interest points in order to indicate a positive or negative reaction to a show's content. No meaning is associated with the action to avoid biasing the user. People are free to associate positive/negative to any characteristics they wish, i.e., thrilling/dull, witty/ cheesy and suspenseful/predictable. Several existing systems contain positive/negative feedback mechanisms, including the popular Tivo platform (www.tivo.com). However, this feedback is only used to assess the overall quality of a show. Since interest points in CollaboraTV have temporal locality, a continuous interest level can be created, called an interest profile. This is achieved by decaying interest back to a nominal level after the user generates one or more interest points (Fig. 1). This data is collected for each show an individual watches and can be merged to create group-level interest profiles (Fig. 2), which can then be used to

highlight where the user agrees or disagrees with the group (Fig. 3), potentially sparking additional discussion. Interest profiles can also be used for fine granularity collaborative filtering, synthesis of new programming by combining high-interest clips, and skipping to the most interesting points in the stream.

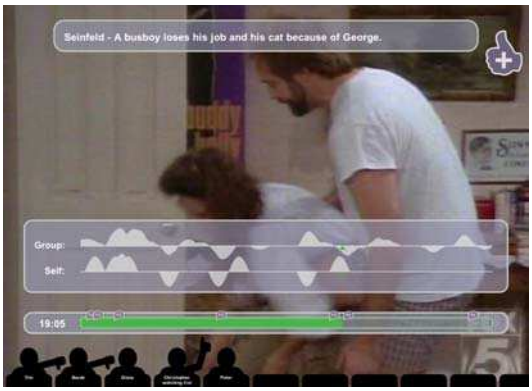

Figure 2. Screenshot showing group and user interest profiles.

Figure 3. Interest difference profile, highlighting significant points of disagreement.

In CollaboraTV, the virtual audience simulates the effect of a movie theatre, where avatars are seated and shown as silhouettes against the show (Fig. 4). The simple black figures are unobtrusive, immediately recognizable, and work well at a variety of scales. Each avatar is named and has a static location throughout the show, which assists the user in associating events with particular individuals. Annotations are shown in translucent comic-book-like speech bubbles, which appear above avatars. In addition, avatars raise and lower their arms to make thumbs up and thumbs down gestures, which correspond to positive and negative interest feedback respectively. All actions are smoothly animated and avatars subtly shift in their seats and look around. This effect increases the perceived interactivity and realism. The virtual audience naturally accommodates asynchronous and synchronous viewers, as well as mixed audiences. A "watching live" label is displayed below avatars that represent synchronous viewers.

CollaboraTV also features a novel electronic program guide (EPG). Data associated with television is multidimensional (shows, actors, genres, dates, guest stars, etc.). Traditional EPGs heavily rely on lists, which are unwieldy for navigating interconnected and cyclical datasets. CollaboraTV's EPG provides a dynamic spider-web-like view of recorded shows, and their associated meta-data. In addition, social elements, such as friends, the shows they watch, and items they tag as favorites can be added to the graph. Furthermore, a ratings visualization is provided to help users find content highly rated by their group. Full details about the CollaboraTV EPG will be provided in a forthcoming paper.

Figure 4. CollaboraTV's virtual audience.

Hypotheses

Foremost, users will participate in an asynchronous system. Secondly, the virtual audience will make television more engaging and enjoyable. Finally, when compared to a traditional (text) chat-like visualization, users will be more social when using the virtual audience, contributing more annotations and interest points.

Experimental Setup

In this study, users were seated in a presentation room with a large rear-projection screen. A Bluetooth keyboard was used for entering text annotations and an RF remote control for all other functionality. A labelled color printout of the remote control was provided for reference. There were 23 televisions shows, varied in content and genre, for users to select from. Advertisements were removed, and clips shortened to approximately seven minutes in length.

During a show, the CollaboraTV interface consists of four items (Fig. 2): an information window, displaying the title and description of the current show, an interest profile visualization, which toggles between the two states shown in figures 2 and 3, and a progress bar, displaying the current location and the location of any annotations, identified by chat-bubble icons.

The overlay interface elements faded from the screen after a preset period of inactivity. A button on the remote control allowed the user to redisplay this information if desired. When the virtual audience was enabled, user avatars remained permanently on-screen. However, annotations and interest feedback appeared for only 10 seconds.

All annotations and interest feedback generated during a show accumulated. This exposed participants to a variety of people, commentary, commentary styles, and audience sizes. However, most importantly, people knew their commentary would be saved and seen by others, a key component of an asynchronous system. The data collected from the varying circumstances will be analysed and presented in a future paper. The current scale and layout of the virtual audience meant that capacity was reached after ten viewers. In order to accommodate new avatars on-screen, users were temporarily disabled, five at a time.

Participants

Thirty-two participants (24 male and 8 female, with a mean age of approximately 42) volunteered for a small food reward. All but two participants worked in a large corporate research lab as researchers or administrative staff. The television, being ubiquitous and of universal interest, meant this group sufficiently represented the population at large - all but one person watched TV on a regular basis.

In order to investigate whether the virtual audience increased user participation, we asked subjects to select and watch television clips with two different on-screen interfaces. The first was the virtual audience and the second was a baseline chat interface modeled after a typical text chat application, where a sequence of textual comments is listed sequentially in a vertical layout. (Fig. 5)

Figure 5 .Baseline chat interface.

Procedure

Experimental conditions were randomized to remove any order effects.

Users were first shown a short video providing an overview of CollaboraTV. This was followed by a 10-minute, hands-on walkthrough, which familiarized participants with the system and its features. Users were then asked to select and start a show of their choosing. Based on the participant's experimental condition, either the virtual audience or the chat-style annotation visualization was enabled. Users then watched the show, participating through the annotation and interest feedback features as they saw fit. After finishing the first show, participants were again asked to select a show of their choice. However, this time the alternative annotations visualization was used. For the third and final clip, users were required to watch a common show to ensure some overlap. Just as in the previous two shows, users were allowed to use CollaboraTV's features. After the final show concluded, users were provided a desktop computer to fill out an 83-question survey. Following the survey, users were interviewed to gather any additional comments about the system.

Results

Initial findings indicate users both understood and appreciated the value of asynchronous communication. Users contributed 3.5 annotations and 3.9 interest points per clip on average. When asked if they would participate in an asynchronous virtual audience, 69% agreed or strongly agreed they would. When asked if CollaboraTV made watching TV more engaging and enjoyable, 53% of participants agreed or strongly agreed (m=3.56). Interestingly, the distribution of results was significantly different when the participants were separated into two groups, those who were familiar with DVRs and those who were not ($p<.05$) (Fig. 6). It is not unreasonable to imagine that people who are not accustomed to recording and watching shows asynchronously would understand the value of asynchronous communication.

Inline with the hypothesis, user participation was higher when the virtual audience was on, with users contributing 45% more interest points on average ($p<.05$). However, interestingly, the generation of annotations was not significantly greater, showing only a 4% increase on average when the virtual audience was displayed. It seems regardless of how annotations are visualized, there is a consistent desire to communicate. However, two-thirds of users agreed or strongly agreed that they preferred the virtual audience to the more traditional, chat-like presentation for annotations.

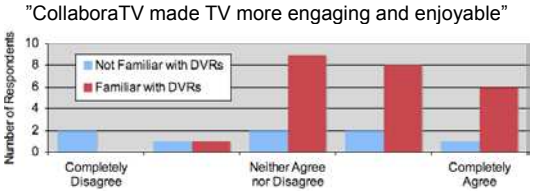

Figure 6. Participants familiar with DVRs had a strong, positive skew, while those who were unfamiliar with DVRs had a flat distribution.

Future work

The next version of CollaboraTV will mature the current feature set. Avatars will gain new gestures and abilities, and additional annotation mechanisms will be provided (sketch, audio, video, etc.). Data collected during the usability study and results from the survey will be analyzed in the upcoming months.

Conclusion

CollaboraTV presents a novel asynchronous communication paradigm for television, which is increasingly being watched asynchronously due to the growing popularity of DVRs. Initial feedback from the usability study indicates that we can achieve our goal of bringing friends together to communicate while engaged in this central part of modern life, and rebuilding much of the social capital that is now being lost.

References

ABI Research. (2006). *Worldwide DVR Market Analysis*. Oyster Bay, NY: ABI Research.

Associated Press. (2006, September 21). *Average home has more TVs than people*. Retrieved January 29, 2006, from USA Today: http://usatoday.com/life/television/news/2006-09-21-homes-tv_x.htm

Bureau of Labor Statistics. (2006). *American Time Use Survey*. Washington, DC: U.S. Dept. of Labor.

Children Now. (1995). *Sending Signals: Kids Speak Out About Values in the Media*. Los Angeles, CA: Children Now.

Coppens, T., Trappeniers, L., & Godon, M. (2004). AmigoTV: towards a social TV experience. In *Proceedings from the 2nd European Conference on Interactive Television: Enhancing the experience.* EuroITV '04. University of Brighton.

Diederiks, E. (2003). Buddies in a box: animated characters in consumer electronics. In *Proc. of the 8th international Conf. on Intelligent User interfaces*. IUI '03. ACM Press, New York, NY.

Federal Communications Commission (FCC). (2006). *12th Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming*. Washington, DC.

Fink, M., Covell, M. & Baluja, S. (2006). Social- and Interactive-Television Applications Based on Real-Time Ambient-Audio Identification. In *Proc. from the 4th European Conference on Interactive Television*. EuroITV '06. Athens, Greece.

Geerts, D. (2006). Comparing voice chat and text chat in a communication tool for interactive television. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction 2006: Changing Roles*. NordiCHI '06, vol. 189. ACM Press, New York, NY.

Knowledge Networks. (2004). *How People Use Primetime TV 2004*. Menlo Park, CA.

Luyten, K., Thys, K., Huypens, S., & Coninx, K. (2006). Telebuddies: social stitching with interactive television. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. CHI '06. ACM Press, New York, NY.

Nielsen Media Research. (2004). *Television Audience 2004*. New York, NY.

Oehlberg, L., Ducheneaut, N., Thornton, J.D., Moore, R.J., & Nickell, E. (2006). SocialTV: Designing for Distributed, Sociable Television Viewing. In *Proceedings of the 4th European Conference on Interactive Television*. EuroITV '06. Athens, Greece.

Putnam, R. D. (2000). *Bowling alone*. New York, NY: Simon & Schuster.

Regan, T., & Todd, I. (2004). Media center buddies: instant messaging around a media center. In *Proceedings of the Third Nordic Conference on Human-Computer interaction*. NordiCHI '04, vol. 82. ACM Press, New York, NY.

Salem, B., & Earle, N. (2000). Designing a non-verbal language for expressive avatars. In *Proceedings of the Third international Conference. on Collaborative Virtual Environments*. CVE '00. ACM Press, New York, NY.

Smith, M. A., Farnham, S. D., & Drucker, S. M. (2000). The social life of small graphical chat spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '00. ACM Press, New York, NY

Television Bureau of Advertising (TVB). (2006). *2006 Media Comparisons Study*. New York, NY: Television Bureau of Advertising.

Tichi, C. (1991). *Electronic Hearth: Creating an American Television Culture*. New York: Oxford University Press.

Zappia, C. (2006, December 20). *Eye of the Potato*. Retrieved February 8, 2006 from http://villagevoice.com/blogs/television/archives/2006/12/what_we_watched.php