Time Warp Sports for Internet Television

DAN R. OLSEN, BRETT PARTRIDGE, and STEPHEN LYNN
Brigham Young University

16

Internet-based video delivery offers new opportunities for interactive television. The creation and usability of interactive television is very different from desktop or web-based interaction. The concepts of frameworks and genres provides an approach to learnable interaction in an entertainment rather than task-oriented activity. The concept of a framework defines the tools required for both producing and viewing a particular style of interactive video experience. An interactive framework for televised sports is presented. This framework implements a sports television experience that support play-by-play navigation as well as viewer's interactive choice of camera angles. Tools for creating and viewing interactive sports are developed in parallel. In-home and in-lab experiments give indications of how sports fans will use interactive television in the future. The experiments demonstrate that fans will use the interaction rather than passively watching, can easily learn the interactive features and strongly prefer the new features over tradition rewind/fast-forward. The data indicates that many users will use the interactive controls to enrich and prolong their viewing rather than simply skipping as rapidly as possible through a game. However, there is also indication that some viewers will simply skip rapidly. There are also indications that the skip vs. review interaction depends on the interest level of current game play.

Categories and Subject Descriptors: H.1.2 [**Information Systems**]: User/Machine Systems; H.5.1 [**Information Interfaces and Presentation**]: Multimdia Information Systems

General Terms: Design, Human Factors

Additional Key Words and Phrases: Sports television, Internet television, interactive television

ACM Reference Format:
Olsen, D. R., Partridge, B., and Lynn, S. 2010. Time warp sports for internet television. ACM Trans. Comput.-Hum. Interact. 17, 4, Article 16 (December 2010), 37 pages.
DOI = 10.1145/1879831.1879834 http://doi.acm.org/10.1145/1879831.1879834

1. INTRODUCTION

This article describes a prototypical implementation and deployment of interactive televised sports. The sports experience is presented as one of a family of techniques that give viewers greater control over how they experience television over the Internet. In this article we not only consider the viewer's experience but the content creator's effort to provide such experiences.

Authors' address: Brigham Young University, Provo, UT 84602; email: olsen@cs.byu.edu.

DOI 10.1145/1879831.1879834 http://doi.acm.org/10.1145/1879831.1879834

Creation of the experience and viewing the experience cannot be addressed separately.

Television is a powerful cultural force, yet television is a relatively inflexible medium. Viewer choice is confined to deciding whether to watch at all, selecting from at most a few hundred choices, and the adaptation of their personal life to the broadcast schedule. The advent of the digital video recorder (DVR) has provided viewers with freedom of timing. In addition, DVRs provide viewers with pause, rewind and fast forward. DVRs give viewers more control but do not significantly change the nature of the viewing experience.

The advent of video over the Internet is starting to change the control relationship between viewer and broadcaster. Early Internet video systems were characterized by long waits for buffers to fill, low resolution and limited ability to select parts of a video stream to watch. However, bandwidth has improved sharply and video protocols have greatly improved. It is now possible to watch video over the Internet that has the following key attributes:

—high definition (HD) resolution
—video startup of just a few seconds
—delivery over a digital network to a computing device rather than an analog tuner
—the ability to jump to any point in a video stream in less than 2 seconds

The immediate benefit of these developments is that viewers can now watch what they want, whenever they want, outside of the control of large institutional gatekeepers. These are great developments which open a whole new range of possibilities.

Our research has focused on interactive television. We have built upon standard Internet video protocols. In particular, we base our work on the ability to rapidly skip to any point in any video stream. Such skipping provides a foundation technology that gives the viewing public complete second-by-second control of what they watch and how they watch. The challenge lies in how to expose this control to viewers and in understanding how viewers will exercise that control. Early attempts used Web browsers as a model for how people will interact with their televisions. However, the experience people expect at their desk while doing work or actively searching for information is very different from the casual "lean back" experience in their living rooms [Drucker et al. 2002; Lee et al. 2008]. Recliner-based interaction is not task-oriented. Much of what we have learned in building desktop productivity applications does not transfer here.

It is not possible to take on all of the future of interactive television in one paper or even one research group. As with command-line interfaces, direct-manipulation interfaces, and Web-based interfaces that have gone before, we will need to build and deploy numerous point solutions before an accepted model of interactive television emerges. This article describes our attempt to build and evaluate an interactive television sports experience. In this article, however, we develop the concepts of frameworks and genres. These do not produce the universal set of techniques for interactive video, but they do move the discussion beyond simple point solutions of one-off interactive video experiences.

There has been a great deal of work done to enhance the television sports experience with computer assistance. The vast majority of this prior work consists of 1) breaking the game into individual shots or segments, 2) automatically predicting which plays are important so as to produce a summary or highlight reel, and 3) characterizing the plays for subsequent retrieval. Prior work is also characterized by the use of archival video rather than the production of fresh and timely interactive experiences.

There has been little done to provide fans with an interactive experience where they personally are in control of what they watch. Time Warp Sports (TWS) is a system that offers this interactive control to sports fans. TWS is a generalization of Time Warp Football [Lynn et al. 2009]. TWS provides an interactive experience that is uniform across a class of two-competitor sports. This uniform interactivity simplifies production effort and fan acceptance. In TWS, we have developed tools for sports producers to create interactive experiences. We think it critical that interactive television research integrate both the interactive experience and the tools necessary to produce those experiences. These cannot be separated. We have produced demonstration experiences across a variety of sports and deployed them into homes as well as laboratory experiences to better understand how interactive television might work. In this article we report data on ease of learning interactive television controls, frequency of use for various interactive features, impact of actual game content on viewer's use of interactive controls and preliminary results on how interactivity will effect viewing time.

1.1 Interactive Television Frameworks

Television is not like desktop computing, where each experience has its own user interface. Being entertainment-driven rather than task-driven, viewers do not want to train or "come up to speed" on how to interact with their television. When sitting down on Saturday afternoon to watch football (American or any other variety) the fan wants a clear understandable experience that feels like every other football experience on every other Saturday. New features are nice. New information is nice, but football is football, and retraining for each game is not acceptable. Obviously some interactive training has occurred, otherwise nobody could watch television. However, that training is limited in scope and amortized over many television programs. The training to operate a television is several orders of magnitude less than that required for a desktop productivity application. For all DVR-based television programs, the possible interactive behaviors are identical. We do not believe there is a single monolithic approach to all interactive television programs. The truth of this assertion will depend on many more developments like this one. However, it seems clear that the desktop diversity of application and training burden will not be accepted as part of the normal television experience. The diversity is required because interacting with sports is different from interacting with drama and still different from interacting with the news or with reality shows. The diversity provides richness to the experience but may also create confusion. We should at this point note the exception of video games (which also use the television screen.) In video games, diversity and challenge are prized.

To address these issues, we present the concepts of *frameworks* and *genres*. A framework defines the interactive style for a piece of television material. We should point out that our use of the term framework is very specific in this article. A framework is a structure around which tools for producing interactive video and the viewing of interactive video can be organized. A framework is not a taxonomy and this article is not about taxonomies of interactive video. A framework is a specific structure around which tools can be built. Broadcast television has taught us the Schedule framework where viewers select time and channel based on a program guide. The DVR framework provides time-shifting, pause, rewind and fast forward. YouTube,[1] Hulu[2] and others have created interactive frameworks of their own. Each has its own style for how viewers will interact with the content. Frameworks imported from the web are quite clumsy in the living room but, as online video technologies improve and content is adapted, that will change.

A framework also provides structure to content providers. For example, the schedule-based broadcast framework gave content producers strict time boundaries in half-hour increments with a certain number of breaks for commercial insertion. Anyone producing a new TV movie or sit-com knows that if they produce material within those constraints it will fit on virtually any television channel and at any time slot and that viewers will know exactly how to consume that content. We see a framework as a mediator between content producers and content viewers. Content producers do not want special purpose software for each new volleyball game and fans do not want to learn new software for each volleyball game that they watch.

A framework also defines an interactive style for viewers to learn. Thousands of programs can be poured into a framework with clear knowledge that millions of fans can access that content in a familiar way. Progress in interactive television will require new frameworks but each framework must work for thousands of items of content.

A genre is a specialization of a framework. The general interactive approach of the framework remains unchanged but those concepts are specialized to a particular purpose. In this article, we will describe a framework for *two-competitor* sports. For such sports, we will show a powerful set of interactive functions that work across all such sports. These include football (American, soccer, or Australian rules), baseball, volleyball, basketball, hockey, etc. Each sport, however, is specialized in its own scoring, time of play, fouls, etc. Two-competitor sports is a framework with basketball or volleyball being genres within that framework. As we will show later, our two-competitor framework is more effective with episodic sports than with continuous ones.

We chose sports as our test bed for four reasons: 1) televised sports are very popular and lucrative; 2) unlike drama or comedy, there is a great deal of structure in a sporting event; 3) we see many opportunities for interactive

---

[1]http://www.youtube.com
[2]http://www.hulu.com

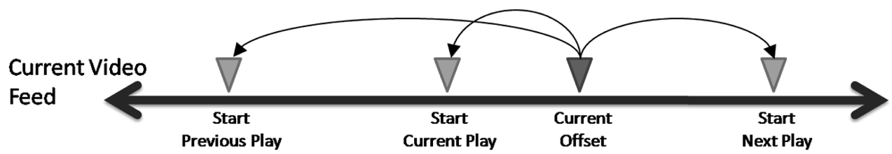

Fig. 1. Timeline representation of interactive video.

features that enrich the experience and 4) everyone on our development team liked sports. The range of two-competitor sports provides us with an opportunity to develop an interactive framework and specialized genres into which a lot of content can be placed. We believe that research into new frameworks beyond video-on-demand or YouTube will push the field forward before ultimately converging into a few dominant styles.

There is an alternative view that perhaps there is only one overarching interactive framework that can subsume all interactive television experiences [Lee et al. 2008]. We certainly have that now in standard cable or Web-based television. It is our position that it is too early for interactive television to settle for or even postulate the universal interactive framework. There are too few actual examples of truly interactive television. This article attempts to step beyond a single interactive style for a single interactive genre. We have generalized our tools across a whole class of sports and provided a uniform interactive style for all of them. However, in our work we have identified (but not yet developed) many other styles around which other frameworks could be built that would be more suitable for different topics. Perhaps when this field matures with numerous examples of interactive television experiences, then we will find the kind of universality that the Web currently offers. There is a danger that, in generalizing too far and too soon, important and compelling interactive differences might be blurred away. We propose frameworks and genres as a middle approach which explores generalization while leaving open the ability to adapt the interactive style to the content being experienced. This article describes the implementation of a specific framework in the realm of sports television.

1.2 Interactive Video

Our technical approach to interactive video is summarized in Figure 1. Video can be understood as a flow of images through time. At any point in a viewing experience, there is a *current offset* or current position in the timeline which represents the current frame being displayed. Such offsets can be represented as time in seconds from the start of the video or as number of frames from the start. We use seconds in our work. For a current offset, there are 1 or more other offsets in the video stream that are of potential interest to the viewer. Figure 1 shows a sports example where the offsets of interest are the start of the current play, the start of the previous play, and the start of the next play. The majority of our interactions consist of skipping from the current offset to some other offset of interest, possibly in a different video stream. These offsets are somewhat similar to those used in a DVD scene index.

The other dominant interactive technique is to overlay user interface elements over the top of video that is playing or paused. These two techniques are the technical foundation for most of what is described in this article.

Skipping from the current offset to some other offset of interest requires that the other points of interest be identified in some way. In the example of Figure 1, the points of interest are the start of plays. We have chosen not to use automatic scene/play detection as most of these techniques rely upon analysis of the video for interesting changes in the imagery. Suppose such a detector was accurate 90% of the time (a rate that is limit of the state of the art). That would mean that it would be wrong once every 10 plays, making for a very frustrating user experience. We already know from a wealth of work in graphical user interfaces that it is critical for a UI to behave predictably. If, for example, a fan selected “next play” and the system took them two plays later or half way into the next play because the automatic play detector was mistaken, the user experience would be degraded. It would not take very many such mistakes before the fans would become disgruntled.

We, instead, have created manual tools for people to mark the points of interest. An American football game takes about 3 hours to play. If we paid someone $50 (USD) per hour to mark the start/end of plays that would be a cost of $150 per game. Each game brings in hundreds of thousands to millions in television revenue. The automatic techniques would need to improve tremendously to compete economically with simple manual annotation. As we will show, manual annotation is highly accurate and very inexpensive. If, however, automatic techniques were to improve to a point where they were reliable, we will show where such information is easily integrated into our system.

1.3 Research Approach

Our research goal is to create the new interactive techniques that will form the basis for interactive television. Our research approach is to first define a viewing experience that we believe will be enjoyable. In the case of this work, we started with Time Warp Football [Lynn et al. 2009]. Having created an experience that home trials showed was successful, we then generalized that experience into a framework that supports multiple genres. In the case of sports, we generalized the football experience across all two-competitor sports. It is that generalization that is the topic of this article.

This generalization across sports led us to the *two-competitor* framework. This framework does not cover all of sports. It leaves out bicycle racing, car racing, track, sailing, steeplechase, gymnastics, golf and other sports that we term *parallel sports*. In parallel sports there are many things going on at the same time. There are individual stories (How is Tiger Woods doing today?) and competitive stories across the parallel activities. The ways in which viewers would interactively move back and forth among these stories as the event progresses would be very different from the two-competitor framework that we address. A framework embodies a particular style of interaction. Parallel sports would require a different framework for a different article in the future. Even in two-competitor sports, the behavior of individual players occurs in parallel. It is possible that the future study of parallel sports may inform the interactions provided in our current framework. Even within our two competitor framework, we see ways in which continuous sports, such as hockey or soccer, may not fit.

With a general design of an interactive framework in place, we turn to the tools necessary for creating content within that framework. The traditional schedule-based or video-on-demand frameworks have decades of tool design. Many researchers have attempted interactive video while ignoring the process of creating such content. Generally, a large corpus of preexisting video is collected, analyzed and, from that, an interactive experience is developed. The assumption in virtually all previous research is that the raw video is a given and computer technology must be applied after the fact. In contrast we have spent time in broadcast trucks watching sports television being created and talking with those who create it. We believe that viewer experience is driven by the creation process and in particular by the tools used to create that experience.

When the Macintosh was introduced, one of its key contributions was the consistency of the user experience across a variety of applications. This consistency allowed for a high degree of learning transfer and also simplified programming of new applications based on a common toolkit. We foresee a similar progression for interactive television. The Macintosh solution came after many years of experience building graphical user interfaces. At this point, there are few interactive television experiences on which to draw. This article generalizes its viewer experience across a number of sports and integrates that generalization into a set of tools. Hopefully, research like this will move interactive video forward, towards a more general understanding. However, this article is only a first or second step towards such general tool solutions. A key insight offered by the Macintosh and its predecessors was that a unifying interactive experience needs to be embedded in the tools that create such experiences. That is one of the goals of this research.

With a viewer experience in mind and the creation tools in place, we then turn to the software that will run the viewer's set top box to deliver the experience. This software must rarely change and must offer a consistent interactive behavior across many interactive video experiences. In an entertainment medium like television, a "gulf of execution" [Norman 2002] problem where the user cannot figure out what to do will be fatal to the experience.

With our tools in place, we then need actual content. This is both a technical and a financial problem. Television sports are big business. There is an intricate interlocking web of rights, contracts, technologies and organizations that go into producing televised sports and none of them want a small research project messing up their process. This makes research in the area quite difficult. Obtaining content for research purposes is an ongoing challenge and has limited our ability to perform evaluations of the technology. Solid evaluation relies upon realistic content which, in turn, is difficult to obtain. This article describes our compromise between research needs and business realities.

Because of these barriers, we chose to record the various camera feeds that arrive in a broadcast truck. Using these recordings, we then simulate the interactive content creation process in the lab, offline from the actual game.

This allows us to research the interactive issues and demonstrate value before addressing the other business hurdles.

With tools in hand for creating content in our framework, raw video footage from which to build the experience and a viewer implementation, we were able to deploy interactive sports experiences into homes and laboratory settings to elicit viewer response and to understand how people actually behave. The remainder of this article will follow this research outline and describe our development and deployment of "Time Warp Sports" (TWS).

2. RELATED WORK

The foundation for any Internet television technology is video transport. There are many such mechanisms and we will not reference them all here. The key properties that we require are high quality video and the ability to skip from place to place in that video and among separate video streams. Common Internet video platforms such as Adobe Flash[3] currently do not meet either of these criteria. Most Internet video mechanisms rely upon heavy buffering to overcome the timing problems introduced by TCP's congestion management algorithms. We know of two video platforms that have the desired characteristics. They are Move Networks,[4] upon which our implementation was done, and Microsoft's Smooth HD.[5] Hulu.com also exhibits this behavior but we are not aware of the video transport technology that they are using. There may be more in the future.

The next key technology for our work is the ability to annotate video with markers for plays and game events that occur during each play. There are a large number of algorithms that have been proposed for automatically preprocessing video to detect scene or play changes [Arman et al. 1994; Bobick 1993; Chorianopoulos and Spinellis 2004b; Drucker et al. 2002; Feng et al. 2004; Li et al. 2000; Qi et al. 2000]. These all have two problems. First, they are computationally expensive, which may be overcome in time. Secondly, they are inaccurate. Sports fans are not going to accept a "replay" control that varies widely as to the frame it selects for the start of play, sometimes skipping part of the play and sometimes going back two plays because the system did not detect a break in play. In continuous sports like basketball or soccer, it is impossible to determine play boundaries without a deep understanding of the game. The change of possession in continuous sports does create boundaries but their automatic detection is difficult.

There is a third problem for which the results of prior work are mixed. This problem is understanding what happened during a play. There are a number of articles that address "event detection." That is, the detection of a goal, score, foul or just an interesting bit of play. There are many articles on analyzing video of a particular sport for some small set of events [Duan et al. 2002; Han et al. 2002; Xu et al. 2003; Wan and Xu 2004; Xu et al. 2006; Xu et al. 2008]. Most of these techniques have accuracies of 50% to 80%, with some

[3]http://www.adobe.com/products/flashplayer/
[4]http://www.movenetworks.com
[5]http://smoothhd.com

reaching 95% in special cases. Such techniques may be useful in the future to augment our approaches but their current levels of accuracy and brittle reliance upon the characteristics of a single sport are not really suitable for our needs. An interactive experience that misses 15% to 20% of the shots made or goals scored will not be very satisfying. Most of these techniques are designed for the retrieval of interesting video clips from an archive rather than for supporting an interactive experience. An alternative source for play information is a direct feed from the scorekeeping system. We did not use this for software compatibility reasons, but it is a viable source of game information. All of these techniques yield varying amounts of information about the event that occurred but little or no information about the start and end of a play. The start and end information is critical to our interactivity.

There is also a thread of research on generating game highlights or summaries. Truong and Venkatesh [2007] provides a useful survey of this work. These use a variety of features such as video motion, whistles, crowd noise, event detection and other features to classify portions of the video with varying levels of interest. This work has its place in the spectrum of smart video features that viewers might consume. This article, however, is unrelated to such summarization efforts. We are interested in providing viewers with a personalized and interactively controlled experience rather than passive consumption of preselected video segments.

Our approach is to use manual annotation of the video stream. There are some systems that use human-generated information from the web [Xu et al. 2006] or close captioning to enhance their event detection. Media Streams [Davis 1993] offers an iconic language for highly flexible annotation of video for many purposes. VERL [Francois et al. 2005] offers a textual markup language. Their generality makes them less useful for our purpose. When a game is in play, the annotator must be able to make decisions and clearly annotate the game in real time. The annotation may be a few seconds to a few minutes behind actual game play but the annotation time cannot dilate to 1.5 or 2.0 times game time. General markup tools will not work. The annotation tools must be specific to the sport. An example of a sport-specific markup tool for soccer is Yu et al. [2008].

The last technology of interest is the actual interactive video experience. Li et al. [2000] produced a prototype of interactive video across several genres including sports. They clearly demonstrated that fans want a "skip ahead" functionality. They did not provide many of the controls that we do but their work is an important early step. This skipping of uninteresting material is also demonstrated for music videos by Chorianopoulos and Spinellis [2004a, 2004b].

Hypervideo [Girgensohn et al. 2004; Shipman et al. 2008; Chambel and Guimaraes 2002] is another interactive video technique where fragments of video are linked together and can be navigated much like the World Wide Web. While hyperlinking might be an interesting model for a sports news show or even an effective general framework, it does not work well for viewing an actual sporting event. We do not believe that viewers want to "navigate" the game. They want their controls to function in the terms of the game not in some generalized structure. Also, hyperlinking is very difficult to author in real time. This might be an interesting model for adding supplementary material, such as player background, related news items, similar plays from previous games, etc. We have not included such facilities in our prototype but they do offer an interesting future direction. Hypervideo is a strong candidate for a new framework for interactive television. Another candidate for an interactive framework is the Yo-Yo storylines of Hand and Varan [2007]. Their approach is geared towards dramatic narrative and thus is not appropriate for sports, but it provides a uniform model of interactive branching and merging of a storyline. In a sporting event, there is only one main story line. It is possible that branching of stories about individual players or coaches could be spliced into the experience but that would have a very different feel.

3. VIEWING INTERACTIVE SPORTS

In a research, one would like to begin with a careful ethnographic study of what people want of their interactive sports viewing experience. Based on carefully analyzed data, we would then design an experience. We did not do that. The members of our team were all avid sports fans who regularly talked to other sports fans. To be intellectually honest, we designed this experience to please ourselves and our friends. This strategy sidesteps the very challenging barrier of explaining the possibilities of a new technology to viewers who have never experienced anything like this before. We felt that the most effective strategy would be to deliver the best experience we could build and then let the fan-base respond to a concrete example.

In our discussions with sports fans, many of whom own and actively use DVRs, we identified four interactive capabilities that seem to cross all two-competitor sports. Most of these behaviors can be seen in the way people use their DVRs. Our framework made these behaviors much more usable and effective.

1. Skip dead time or fill that dead time with something.
2. Review or reexperience a play or other event, frequently from different camera angles.
3. Obtain score, statistics and other information about the game.
4. Skip advertising.

Many sports are episodic. A play happens, fouls are called, progress/score is accounted, competitors regroup, and another play occurs. American football is very much like this, as is baseball, bowling, and curling. In many cases, the time between plays is used by the competitors to secretly strategize among themselves for the next play. Standard sports broadcasts fill this time with commentator discussion, replays, analysis, and background information. Many viewers prefer to skip to the next play. A two-hour baseball game really has only about 20–30 minutes of actual play. In contrast, sports such as soccer (futbol), basketball, water polo, and hockey are not episodic. Play is continuous. The experience can be divided into changes of possession but there is little dead time between plays. We were surprised to find that volleyball is much more

Fig. 2. Multiple camera angles for play review.

episodic than we thought, with lots of hugging, slapping and cheering between plays.

A big feature that everyone wants is to review previous plays from as many camera angles as possible. Many sports fans want to personally advise the officials on virtually every call. Figure 2 shows three camera angles for a basketball out-bounds-call. Though the official is right next to the play and looking right at it, no true fan wants to accept his decision without personal review.

In standard television, the director in the broadcast truck arbitrates which views will be shown when. In the case of continuous sports such as soccer or basketball, the director controls how much of the ongoing play will be replaced by a replay. The viewer has no control. Our implementation provides viewers with the ability to restart every play and to request replays from every available camera angle at the viewer's discretion. Viewers can do this without any loss of ongoing play. In the post evaluation comments from our subjects, the personal control was the big attraction of our software. People like to control their own experiences rather than be controlled. Many viewers claimed that interactive sports let them watch a game faster when, in fact, they took more time to watch.

A very common use of a DVR is to view television content while skipping over commercials. There is obviously a great demand for this feature but, in the end, television content must be paid for either by advertising or by subscription fees. The role and nature of advertising in Internet television is a topic of great debate. Given the interactivity and targetability of Internet usage, there are many possible advertising/payment models being discussed. Because there is no clear resolution to this issue, we have omitted all advertising from our deployments. This does skew the experience somewhat but the confounds introduced by the diversity of advertising approaches were too great. The advertising issue needs resolution, but not in this article.

Many games have additional information that is of interest to fans. The score is an obvious choice, as well as number of fouls, yards gained, remaining time outs, status of the game clock, and many others. The standard television

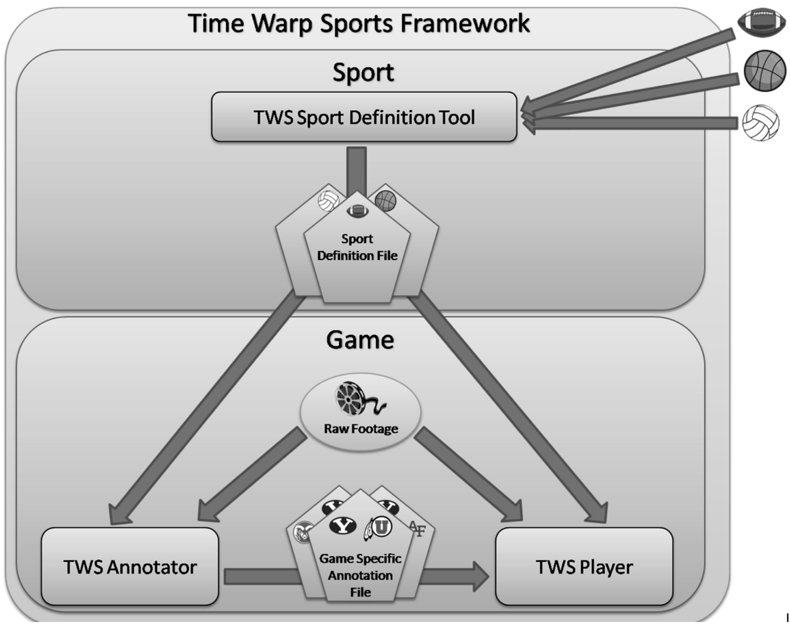

Fig. 3. Time Warp Sports architecture.

strategy is to make this information available through the audio commentary, during dead time, or overlaid on the screen. The first two choices mean that information appears when the director/commentator chooses rather than when the viewer wants to know. The overlay choice is limited in what can be shown because it obscures the game. Our implementation places all supporting information in the control of the fan and is displayed at their request at any time.

4. A FRAMEWORK FOR SPORTS VIEWING

A framework in this article refers to a specific structure for organizing an interactive viewing experience. It is the framework that drives the architecture of the tools for producing and viewing interactive video. A framework also defines the set of genres that are possible within the framework as well as the information required to specialize the tools to a specific genre. In this section, we will describe the architecture for our two-competitor sports framework as well as the specialization of that framework for specific sports (genres).

Figure 3 shows the high-level architecture of the Time Warp Sports (TWS) framework. For each sport, a *sport definition file* (SDF) is created that contains information about the structure of the sport. This file need only be created once for each sport and serves to specialize the annotation and player software to that particular sport. At present, the TWS Sport Definition Tool is a text editor. These definition files are rarely changed and easily edited.

Given an SDF, a game is created using the raw video footage from all of the camera feeds by using the *TWS Annotator* software. The annotator accepts the SDF and specializes its user interface to meet the specific needs of that sport. Using the annotator, personnel in the broadcast truck produce annotation information for a particular game.

The video streams are ingested onto an HTTP server using proprietary software from MOVE Networks. The MOVE networks technology provides us with encoding, transport and decoding software that will deliver high definition (HD) quality video over HTTP that can be randomly accessed to any point in the video timeline. The architecture described in this article is not dependent upon MOVE specifically. What we do require is a video transport that can rapidly (2 seconds or less) move from one point in a video stream to another. MOVE is a technology that demonstrates the possibility of this over normal home broadband connections. In our in-home trials, this performed beautifully. Video lag under interactive use was never mentioned by any of our participants. This is the foundation for our interactivity. The *game annotation file* (GAF) contains information about teams/individuals playing, play boundary times, and various events in the play. This file is created by the annotator software as the game is being played. After each play, new information from the annotator's user interface is posted to the GAF on the server where it is accessed by the player.

The *TWS Player* is a downloadable user interface written in Microsoft Silverlight[6] C#, and the MOVE video player. Together these form a platform that is easily downloaded over the web and will interactively present any TWS sporting event. The player uses HTTP to fetch the game annotation file and the relevant video to present to the viewer.

4.1 Generalizing TWS Across Two-Competitor Sports

Before building our sports framework, we analyzed the structure of 7 different sports. Figure 4 shows a summary of that analysis for three sports: football, baseball, and volleyball. Sports that fit in our two-competitor framework have the following attributes.

- —Exactly two opponents.
- —A hierarchic structure for game time, which we use for interactive navigation.
- —A set of discrete events around which statistics are gathered for the interest of the fans.

We looked at the set of sports contested in the Olympic games and identified the following sports as suitable to our framework: table tennis, badminton, racquetball, football, soccer, basketball, hockey, volleyball, fencing, wrestling, water polo, boxing, field hockey, judo, taekwondo, tennis, and curling. Most of the remaining Olympic sports would fit into a parallel sports framework that is not described in this article.

All of the sports in our framework have a hierarchic time structure. For example, a baseball game has 9 or more innings, each inning is divided into a top and a bottom half (one for each team); each top or bottom has 3 or more batters, and each batter gets 3 or more pitches. This hierarchic structure of

[6]http://silvrlight.net/

	Organization	Navigation	Scoring	Events/Info	Statistics
American Football	Play Series Quarter Half Game	Play Series	Touchdown -6 Field Goal - 3 Safety - 2 2 Pt. Conv - 2 Extra Pt. - 1	Penalty Fumble Interception Yard Marker 1st Down Marker Pass Complete Rushing Attempt Yards Gained	Score Offensive yards Penalties Turnovers Rushing Yards Passing Yards Pass Completions . . .
Baseball	Pitch Batter Top/Bottom Inning Game	Pitch Batter Top/Bottom	Run - 1	Error Out Stolen Base Hit Ball Strike Foul Ball Double Play Triple Play	Score Outs Count Errors Runs Batted In
Volleyball	Serve Sideout Game Match	Serve Sideout Game	Point - 1	Fault Spike Out Net violation . . .	Score Spikes

Fig. 4. Sport structure.

time is universal across our framework. The first thing specified in the SDF is the hierarchic structure for a particular sport where each level of the hierarchy represents a different *play period*. Some, not necessarily all, of these play periods are specified as units of navigation. For example, a viewer of a baseball game can ask for next pitch, previous pitch, next batter, previous batter, next top/bottom or previous top/bottom. For each play period, the SDF specifies: the name, how many of such periods there are, if there can be extra of these play periods and if this period is used for navigation. Figure 5 shows a partial SDF for the play periods in hockey.

Scoring is a bit of a problem. Most sports simply have ways to accumulate points that are summed as you go along. Tennis is odd in the way that points accumulate (0 or love, 15, 30, 40, ad in, ad out). Bowling also has a scoring mechanism that sums up in an odd way. Curling is similarly odd in that a team can lose points. To accommodate this, a sport's score either adds (with various events that earn points) or score is entered manually when it changes to accommodate odd scoring schemes.

There are also a variety of events that can occur during play. These events are of interest to fans and may in the future be useful to select interesting plays for game summarization. Each event type has a name and an optional value. The value may have a fixed value or an arbitrary value. For example, a

```
Sport
{     name: "Hockey"
      PlayPeriod
      {     name: "Period"
            navigation: false
            numPeriods: 3
            extra: true
      }
      PlayPeriod
      {     name: "Possession"
            Navigation: true
      }
}
```

Fig. 5. Partial sport definition file.

field goal in American football is always 3 points. However, the yards gained on a particular play is an arbitrary value that must be entered after each play. Event definitions in the SDF control the user interface that is presented in the annotator while encoding an ongoing game.

Events can also be accumulated into statistics. Statistics are calculated from events and other statistics. An additive statistic begins at zero, and after each play it adds the value of the corresponding event. Most scores, yards gained, pass completions, number of steals, number of spikes, etc. fall into this category. There are also average, percentage, and ratio statistics that aggregate two statistics or events into a single calculation. Their differences are only in the way they presented visually. The statistics facility was added to the tools to allow for their inclusion as part of the viewer's experience. This facility may be supplanted by the existing statistics software used by scorers in many collegiate or professional sports events. Though we have not done this in our prototypes, the inclusion of statistics from other software is a simple implementation problem and does not have a direct bearing on the nature of the interactive experience.

The tools of the TWS framework adapt to a specific sport genre based on a sport's SDF. A case could be made that it would be easier to treat each individual sport as its own framework to handle its own peculiarities. We chose the framework/genre strategy for several reasons. The first is to provide a uniform viewing experience across all sports within the framework. If a viewer has become adept interacting with baseball, the learning transfer for football season will be trivial. The user interface is completely consistent across sports within the framework because the viewer software is the same. This consistency, with its corresponding learning transfer from genre to genre, is very important in a relaxation/entertainment experience.

User interface consistency is also essential for the annotator software. The staff in the broadcast truck doing a basketball game tonight may have worked a volleyball game earlier in the week and will be doing a soccer game tomorrow afternoon. The annotation process is time critical for live events, and familiarity with the tools is essential. The framework approach gives us this desired commonality among sports.

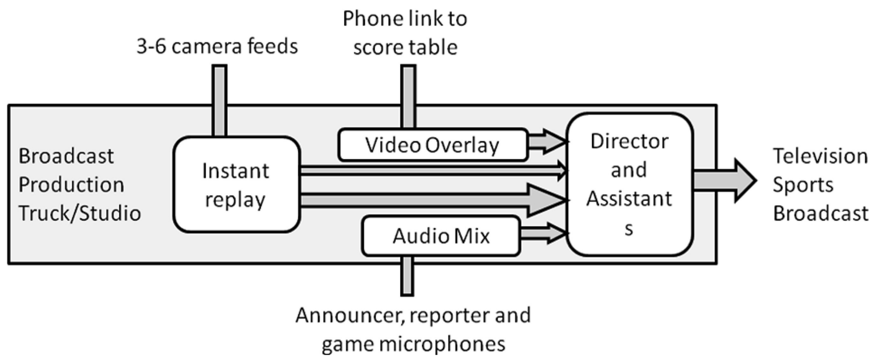

Fig. 6. Structure of TV sports production.

5. CREATING A TWS BROADCAST

Sports broadcasts are frequently created in a broadcast truck which is basically a semi-trailer filled with audio visual equipment. One of our research goals was to reach into the television creation process deep enough that the creation of interactive television is addressed on an equal footing with the viewing of interactive television. To address this goal, we spent several evenings and afternoons in a broadcast truck while basketball, volleyball, and baseball broadcasts were being produced.

Television production for all of these sports works in a similar fashion. Between 3 and 6 cameras are set up around the sporting event, all cabled into the broadcast truck, as shown in Figure 6. In addition, several audio feeds from announcers and scorekeepers are also fed into the truck. Each camera view is fed to the director's station and to the instant replay station. The instant replay staff searches all of the views for interesting camera angles that they deliver to the director on separate monitors. The director views all of the available feeds, including cameras, commercial/information breaks, and instant replay materials. There are also one or more people managing overlay information, such as the score, statistics, and news bits. The director then gives verbal directions to assistants, who control the switching among camera feeds and to camera people covering the sport. The result is a single video stream that is delivered to the broadcasting equipment (cable or satellite link).

The staff managing the overlay information are responsible for interacting with the scorekeepers to provide statistics. There is technology to digitally feed the scorekeeper's statistics directly into the overlay production. However, in the setups that we observed, this was not working and the engineers spoke slightingly of such digital feeds being useful to them. They preferred the more robust solution of pointing a camera at the scoreboard and then hand updating the score into the video overlay. They also had one person in continual phone contact with the scorekeepers to obtain statistics that were hand entered into the video overlay tool. This manual approach is not the only alternative. In the professional sports, we have also seen the use of direct feeds from scorekeeper software into the broadcasting process. Either approach is compatible with our interactive experiences.

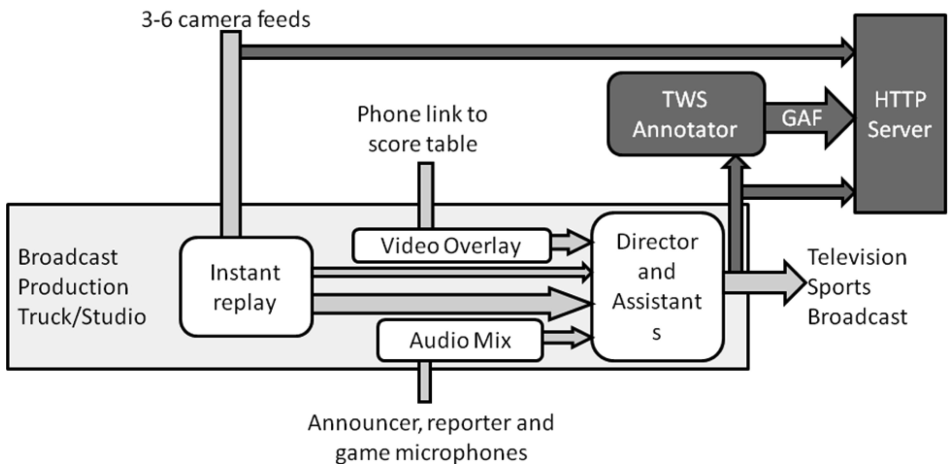

Fig. 7. Time Warp Sports workflow.

Figure 7 shows how we have modified this workflow to produce an interactive sports experience. We take all of the raw camera feeds and ingest them digitally onto an HTTP server. We do the same with the fully mixed broadcast feed that comes from the director's station. In addition, the broadcast feed is passed into the TWS Annotator software where one or two staff members add the necessary annotation information to create the game annotation file (GAF) which is also uploaded to the server. Not shown in Figure 7 is the fact that the audio track from the broadcast is also mixed into all of the other camera feeds so that all video tracks will have the same audio.

A minor phase in the game annotation process is the initial setup. The annotator asks for the names and logos of the teams/players as well as the names of the camera views available for this game. The number of cameras varies from game to game. Adding more cameras increases the potential for interesting shots but also increases equipment and staff costs. This can vary depending upon the importance of the game to the intended audience.

The primary role of the TWS Annotator is to add time markings for the start and end of each navigation unit described in the sport definition file. All of the camera feeds as well as the mixed broadcast feed are time synchronized. Because of this synchrony the game annotation information applies to all feeds. This is a primary difference between two-competitor sports and parallel sports. The resulting relationships are shown in Figure 8.

In intercollegiate sports, there are large amounts of money involved and, thus, numerous overlapping rights and contract arrangements. As such, it was impossible for us to achieve the architecture shown in Figure 7. For our first deployed demonstration using American football, we obtained recordings of the television feed and two camera feeds that were recorded for the coaches (high sideline from midfield and one end zone). This gave us three different views of the game. Media rights in sports are very difficult to work around. There are also problems with the competitive value of video among teams.

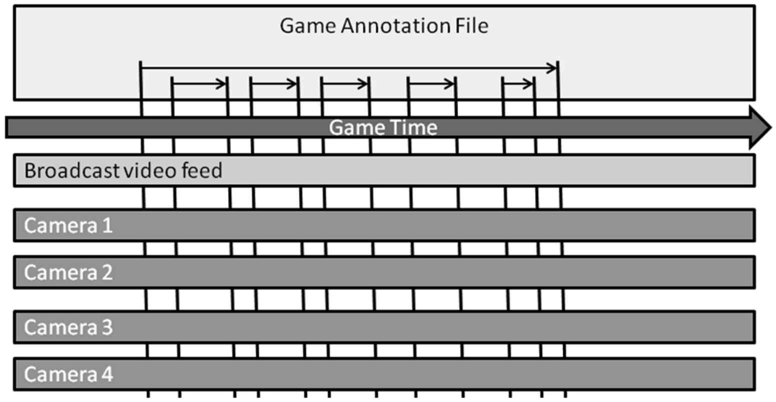

Fig. 8. Relationships among game annotation and multiple video feeds.

It seems that the coach's views are not shared among competing teams and are considered a strategic advantage. This would indicate that our time-warp sports initiative would need to be deployed across all games in a conference rather than selected games so that the relative coaching advantage would be uniform. The lack of available content that is reasonably free of restrictions makes research deployments difficult. This lack of content limits our ability to perform exhaustive user evaluations.

For basketball, baseball, and volleyball, we were allowed to attach 6 DVD recorders to the camera feeds coming into the university's broadcast truck. This allowed us to record all of the camera angles that are used by a director to construct a traditional broadcast. From this raw material, we were able to construct in-lab simulations of the production process shown in Figure 7. This approach allows us to evaluate tools and to deliver some games to viewers for evaluation. However, the quality of the available games is poor. The particular games we used are free of rights restrictions because they had little commercial value. The availability of content does bias the kinds of games we can use in our evaluations. We will discuss this bias later in the article. The DVD approach to obtaining content will not produce a live broadcast. A different technical solution is required that is beyond the scope of this article. However, what we have obtained has allowed us to perform evaluations of the tools across sports and with actual users.

For the American football demo, we used a single TWS Annotator application to encode all navigation boundaries and events. Several lab members encoded numerous games and all found it easy to accomplish in real time as the game video played. However, when we tried this strategy on a continuous sport, such as basketball, it failed. Football has a lot of dead time between plays in which event information can be recorded. There is little dead time in basketball, soccer or hockey. Based on this experience, we divided the TWS Annotator into two applications: one for marking navigation unit times and one for marking event information. These two applications work in series with navigation annotations feeding into the event marking tool as they are entered. With this modification, we were able to easily annotate any game in real time.

Fig. 9. Annotation interface for time recording in volleyball and basketball.

Figure 9 shows the navigation marking tool specialized for volleyball and basketball. The annotator simply uses the two buttons to mark when a navigation unit starts and ends. The user interfaces for these tools are automatically generated from the sport definition file by the TWS Annotator.

One alternative to this manual marking of the start and end of plays would be the automatic techniques that have been explored in so many articles. If such mechanisms were accurate, they could certainly be inserted into our process at

Fig. 10. Event annotation of volleyball and basketball.

this point in place of the manual technique. We do not expect this to happen any time soon. Most work in this area relies upon unique characteristics of a particular sport or easily recognized audio or video features. We know of no automated technique that can be generalized across multiple sports, as required in our framework. The visual and audio flow of baseball vs. basketball or football is very different. Techniques such as camera shot changes do not really meet play boundaries at all. This is particularly true in basketball or hockey. Motion activity may work in football (provided there is no hurry-up offense) or baseball but not in basketball or hockey. Listening for official's whistles might hit the end of play for football but be useless for baseball (no whistle) or other sports with infrequent whistles. This is all interesting work but not appropriate to our needs.

Figure 10 shows the event logging tool for both volleyball and basketball. Again, these user interfaces are generated from the sport definition file. Down the left hand side is a list of all of the events that can occur and across the bottom is the current state of play. The annotation operator simply clicks on the various events that have occurred within a play. On occasion an error is made. In such cases, the person annotating can restart the play (clearing out the old selections) and re-annotate the play. When the annotator logs the play clip, the video immediately skips to the next clip. If a pause was required for error correction, the time is generally recovered within the next play clip because it takes less time to record events than most plays require. Thus, the whole process stays very close to live time.

These tools have been used in our lab to annotate 5 games from 4 different sports (football(2), basketball, volleyball, baseball). Annotation is easily accomplished in the time necessary for a live sports event.

There is one technical challenge to our architecture that has not been fully addressed. A standard broadcast truck/studio has the facilities to take in many camera/audio feeds and to send out one video feed. Our architecture calls for as many as 7 video feeds to come out of the truck and onto the network. Satellite links and broadcast cable connections do not have enough bandwidth to accomplish this. In our prototypes, this was glossed over by making DVD recordings and physically carrying them to our lab.

The solution to this is to use the Internet rather than analog cable technology. There is wide variation in the reported bandwidth requirements for HD video. This variance is due to the amount of motion and/or detail in the video as well as the codec used. We have seen nice results at 2.2Mbps while others report that 4Mbps is required. At 4 megabits per second, all 7 feeds will require 28Mbps. This is well within the capacity of a 100 Mbps network connection. The solution then is to digitally encode all 7 feeds before they are sent out of the truck/studio. This would require the addition of approximately $50,000 in computing equipment to be added to the broadcast truck. The result is that all necessary information can leave the truck over a single Internet connection. We did not actually build this solution. Note that 28 megabits is only required for the broadcast truck, not for each home. Each home only requires 2.2–4.0 megabits per second. Modern video technologies will adapt the stream to the available home bandwidth. Standard quality was easily

Fig. 11. Interactive television controller.

attained in all of our trials. Homes will get a better view as more bandwidth is applied.

6. VIEWING A TWS BROADCAST

The viewing experience is centered around a standard wireless game controller as shown in Figure 11. We chose this controller because it is easy to work with using Java's JInput package, it has lots of buttons to use, it is familiar to a large audience, the control response is faster than infrared and it does not look like a television remote. We did not want prior experience with television remotes to confound the learning of new controls. As will be discussed in the Evaluation Section, this controller is not necessarily the ideal choice. Older viewers who do not play video games reported a preference for a one-handed controller that is more like a TV remote. Younger viewers who do play video games reported a preference for this controller. There seems to be no clear consensus on the style of controller. However, the interactive behaviors that we report later in this article are independent of the style of controller used. We did not do any comparative analysis of types of controllers because we were much more interested in the interactive experience. Device optimization studies can come later.

The viewer controls that we offered were:

— standard play, pause, fast forward, rewind
— audio volume and mute
— next/previous play controls
— switching of camera angles
— access to game statistics

Our deployment allows for a simple comparison between the VCR-style controls and the next play/previous play controls. The data shows that the play-specific controls are much preferred. One condition that we did not test against is the use of time-based skipping of a fixed number of seconds. This feature is found on many DVR. A forward skip is designed to skip commercials and a backward skip of 5 seconds is designed for simple resetting of an aggressive use of fast forward. The problem with such fixed-time techniques is that game play does not conform to fixed amounts of time. Skipping back a fixed amount of time for a replay would either land the viewer in the middle of the play, some time before the play starts (forcing them to watch dead time) or possibly in the middle of the previous play, which is visually very confusing. In working with interaction in sports, we found it very easy for a fan to lose track of which play they are actually viewing. Reliable controls are very important for the fan to keep context. Skipping forward a fixed amount of time would have similar problems.

Our data shows that the time between one football play and the next is an average of 26 seconds. If we were to use a constant skip time, then 26 seconds would be the best candidate. Our data also shows, however, that on average each individual play time varies from that by 18 seconds. This means then, on average, a fixed skip time in football is off by 18 seconds. That is a lot of viewing time to be wrong. In volleyball, the fixed time method would be wrong by an average of 29 seconds and in basketball, a fixed time skip would be off by 18 seconds. In basketball, the fixed skip would be very bad because 16% of the time the fixed skip would cross two or more plays. Fixed-time skipping is a clearly inferior solution and we did not include it nor evaluate it.

We did, however, encounter one notable exception. In the case of college basketball, which for historical reasons has a very long shot clock, the first part of a possession can be very boring and relatively unconnected to the play that occurs just prior to a change of possession. Skipping all the way back to the start of the play made for a very poor experience. Skipping back 10 seconds or the start of the play, whichever is shorter, worked out much better. We included options in the SDF to account for this if it appears again in other sports.

The alternative to user-control of the camera angles is the director's control of what angles are most appropriate. In our system, the standard TV broadcast is one of the choices a viewer can make. If viewers prefer, they can watch preselected replays or select their own. The data will clearly show the viewer's preference for making their own selections.

The alternative to interactive game statistics is to provide statistics as overlays on the screen or as text crawls across the bottom. We consider the overlay approach to be inferior because it limits the timing of available statistics and clutters the screen. However, with the content we had available to us, we could not remove the overlays. Therefore, it was impossible for us to evaluate the relative advantages. We do offer interactive access to statistics as a design alternative.

6.1 Implementation of the Viewer Software

The primary foundation of our technology was the MOVE Networks video player that runs as a plugin to most web browsers. The player can run in full screen mode so that the web browser controls are completely hidden. To provide interactive feedback, we implemented the user interface in C# using Microsoft Silverlight. This gave us the ability to process inputs and draw a wide range of graphics over the top of the video. Drawing with transparency was particularly important. A minor challenge was that Silverlight is a Web technology and therefore imposes a very restrictive sandbox to prevent importation of malicious code. The sandbox only allows for the simplest of mouse-based input, which excludes our controller. We resolved this by implementing a Web service in Tomcat running on the same machine. Whenever the sports viewer software wants to poll for user input, it contacts the web service at "localhost" and the web service polls the input device and returns its state. This is a gastly implementation but it allowed us to move forward with our experiments and easily achieved the necessary interactive speeds.

The play, pause, fast forward, rewind, and audio functionalities were implemented as simple calls to the MOVE player. Next play and previous play are among the most popular controls. These are founded on the MOVE player's ability to rapidly (1–2 seconds) change the play head of the video stream to any time offset. The game annotation file contains the time offsets for the start and end of any navigation unit defined in the sport definition file. For example, when a user selected "Next Serve" in a volleyball game, the TWS viewer software retrieves the time offset of MOVE player's play head and then looks in the game annotation file for the time offset of the next serve. It uses that time offset to instruct the MOVE player to go to that point in the video stream. All of the Next/Previous navigation controls for each navigation unit are implemented in a similar fashion. Other player implementations, besides that from MOVE, provide similar capabilities and could be used provided they can efficiently seek to a given point in the video (not true for most current protocols). The game annotation file for an entire sporting event contains, at most, a few hundred annotations making a linear search of this data structure more than fast enough to respond to user input in a timely way.

Each camera view is stored in its own video stream. Because the start time of each view is identical, the time offset data from the game annotation file works across all camera streams. When the viewer selects a new camera, the MOVE player is given the URL of that camera's video stream and the offset of the start of the previous navigation unit. To the fan, switching camera angles feels like instant replay from a new view.

6.2 Interactive Experience

One of the major issues we were concerned about was user training. We did not want user frustration with the technology to get in the way of the entertainment experience. Fans come to enjoy a sport, not to master a tool. The subjects were shown that pressing any of the trigger keys on the front of the controller (there are 4) would show them how to control the viewer. They were also told that the top-left button showed the default controls and that pushing any button on the face of the controller was the same as pushing that button when the top-left trigger was pressed.

Figure 12 shows the viewer software configured for the volleyball genre. The user has pressed the upper-left trigger and the control overlay has appeared. The control overlay looks very much like the face of the game controller and has all of the functions labeled. Pressing any of the buttons on the game controller will cause a corresponding echo on the screen overlay. The overlay disappears when the trigger is released. As the evaluation section will show, this duplication of the controller on the screen was very successful in training users how to control the game. They also quickly learned to control the game without the

Fig. 12. Volleyball viewer controls.

overlay because of the simple correspondence. We had very few issues with fans learning to control the game. The upper-left trigger displayed the main navigation controls shown in Figure 12. The bottom-left trigger showed standard play, pause, fast-forward, and rewind controls. The other two triggers would show game or team statistics.

In a separate project involving interactive news broadcasts, we used a one-handed control that had a similar overlay technique for presenting the meaning of the buttons. We achieved similar positive usability results to those reported for this two-handed control of sports viewing.

A second interactive problem is that sports viewing is frequently a group experience and there is only one controller. Traditional play, pause and rewind controls show visible artifacts on the screen that make it easy for other viewers in the room to understand when these controls are operating. The visual changes from camera and play navigation controls appear instantaneous to anyone that is not operating the control. This can be very confusing to other viewers. It is difficult to differentiate next play from previous play actions while only watching the screen and not personally operating the controller. We added a semitransparent overlay after each navigation action that would show the current camera view as well as key statistics, such as score, football down, or who controls the ball. This greatly assisted other viewers in tracking what was happening when someone else was operating the controller.

In our implementation, switching the camera angle would also restart the last play. The exception to this was basketball. College basketball has a very long shot clock and frequently the first 20 seconds of a play has little action. In preexperiment testing we found that returning to the beginning of the play caused viewers to lose context of the game and to become impatient viewing the dull part of the play. In this particular sport, the sport definition file directs the TWS viewer to jump back 10 seconds. This creates a much better experience. This modification was made before the experiments were performed.

7. EVALUATION DEPLOYMENT

We wanted to know how all of this technology will play as an entertainment experience for average viewers. The first challenge for an evaluation deployment is getting real content. Because of the networking issues described previously, no live game broadcast was possible. We instead chose to use recorded games to produce an experience as close to live as possible. For football, we used a 4 year old BYU vs. Notre Dame game. All of the subjects in our evaluation knew that BYU had played Notre Dame. Many could not remember the game outcome, and the details of the game were only dimly remembered. For volleyball, basketball and baseball, we used DVD recorders attached to the BYU broadcasting truck. All commercials and half-time breaks were removed from the videos. The remaining video streams were then uploaded to the video server.

We used two different setups for evaluating the viewing experience. For football we deployed the technology into people's homes using their own televisions, living rooms and their own Internet connection. For the remaining sports, we performed the evaluations in our lab. An in-home deployment is much more time consuming than a laboratory setting, but we wanted to test whether home networks and environments would support a quality viewing experience. These deployments and user tests are formative in nature. We did not attempt exhaustive comparative studies. The results that we have obtained are indicative rather than definitive. The numbers of subjects are relatively small when compared to the diverse populations of viewers or even only sports viewers. To achieve definitive results about what viewers are actually like and their detailed viewing habits, we would need to deploy a large number of games (30+) into a large number of homes (1000s). Such a study is simply not physically and economically feasible at this time. What we have done is to explore usage across several sports and have developed data that is indicative of how viewers might respond to such a deployment. The results we present should be considered as a guiding light for future efforts rather than a definitive understanding of sports viewer behavior.

For the football evaluation, we recruited 11 groups of viewers from the local population. Each homeowner was asked to recruit some friends or family to watch a football game in their home and we would provide the pizza. We used a laptop with video output to run the viewer software. The laptop was connected to the family's television and to their home network connection. We did require that they have a broadband connection but did not otherwise stipulate bandwidth or supplier. We set up a video camera next to the television pointed back at the viewers themselves so that we could evaluate their reactions to the game experience. The camera was clearly visible and the viewers forgot about it very quickly. After a brief introduction to the triggers on the controllers, the experimenters left the room for the duration of the experiment. All interactive actions were logged by the TWS viewer software.

We have discarded the data from 2 of the 11 groups watching football. These two groups worked in the software industry. The video tape showed that they spent most of the time fiddling with the controls, trying to break the software, stress-testing the boundaries, etc. In general, they did not watch the game but instead studied the technology. One of the groups, upon having played with the controls until bored, simply gave them back without ever watching the game. We did not consider this data to be representative of what actual fans with actual interest in the sport would do. The remainder of the results apply to only 9 groups of football fans. Admittedly there are many kinds of sports fans with many levels of interest. We will not have a complete profile of behavior from just 9 groups. However, a negative result from these 9 groups would be very informative. If many of the 9 groups do not like or cannot use our technology, that will be a very telling result. As the later data will show, there was strong interest and effective use of the technology by these 9 groups. We believe this is a strong indication of the value of interactivity and our techniques in sports viewing.

The selection of the football game was constrained by the availability of video from the football office and the need to choose a game that was interesting throughout. A game that is a blowout for one team quickly becomes uninteresting and thus would have given us poor indications of how much people like the technology. The only game that met these criteria was a BYU vs. Notre Dame game. Unfortunately, the first quarter of video and part of the second quarter had been lost before we got access. Therefore, the game that subjects viewed started part way through the second quarter. This did not seem to make a serious difference to the subjects. The total football video time was 102 minutes.

In addition to the in-home deployment of football, we also scheduled four groups of fans for basketball viewing in our lab. For basketball, we only tested the first half because the game became extremely lopsided after that and fan interest would have dropped to zero. The total available video time was 36 minutes. We also scheduled four groups of fans to view the first two sets of a volleyball game in our lab. The total video time was 57 minutes. This limited sampling of game viewing does not represent all possible viewing scenarios. We purposely restricted the game segments that were offered to those for which there was competitive interest. If a game becomes boring, there is no technology that will entice a viewer. In a boring sports event, the interactivity becomes irrelevant. What we did want to know, however, was whether interactivity would be used or neglected when the game was interesting. It is entirely possible that fans engrossed in a game might discard the controls and focus on the game. The sports segments that we used do provide us with some insight into this.

In all of these experiments, we wanted to duplicate the kind of experience we expected in a home. Informal discussions with numerous sports fans that use

DVRs indicated that many such fans will start recording the game and then not actually begin watching until 10–30 minutes after the start of the game. They intentionally build up this "slack time" so that they can skip commercials or dull parts of the game. Because all of our video was prerecorded, we created a simulated "live time." For football, this started 10 minutes into the video and for volleyball and basketball, it was 5 minutes into the video. The "live time" clock would move forward starting when the experiment started just as real game time would do. This would allow viewers to skip to next play, but eventually, they would run into live game time where no technology can skip into the future.

In looking at interactive viewing habits, there are three cases to consider. First, there is the viewer who wants to be as close to live as possible. They start watching when the game actually starts. For such a viewer, the replay controls and statistics controls are attractive but the skip forward play by play is impossible. We did not study this case because it would not give us any information about skipping forward behavior. The second case is where a viewer starts watching some time after the beginning of the game so that they can skip the boring parts. They want to stay close to live but want the freedom to skip or review as they wish. This is the case that we addressed in our studies. There is a third case where the viewer simply wants to watch the game as fast as possible. They do not start watching until the game is mostly over and skip forward at the end of every play. The studies reported in this article provide no insight into the relative distribution of these three cases across the viewer population. Such studies would require thousands of subjects over an extended time. By focusing on case two, we believe we extract the most useful information from such limited studies.

For all experiments, we conducted a post interview using a Likert scale as well as open-ended questions to assess user response to the technology. In addition, we coded behaviors from some of the viewer video to get a better feel for what they were actually doing during the experience.

Across all of our sports trials, there were 19 groups of viewers and 66 individuals. There were 72% males with 57% in the 18–24 age group. The remaining viewers were older with one viewer older than 55. There were 64% who reported that they had watched TV on the Internet once or twice and less than 1% had never watched Internet television. In terms of sports interest, 13% said that they never watch sports, 27% reported one game a month, 33% reported one game per week and 26% reported 2 or more games per week. Each of the sports had demographics similar to the overall profile.

7.1 Are Controls Useful?

The normal television experience has very little interactivity. The viewer selects a program and then watches. The advent of DVRs has raised interactivity by allowing viewers to skip advertisements and boring segments. The interactive controls for TWS are far more complex than standard viewing. There are many more choices than "pause" or "skip." We wanted to answer three big questions: 1) would anybody care about the additional controls, 2) could they learn them

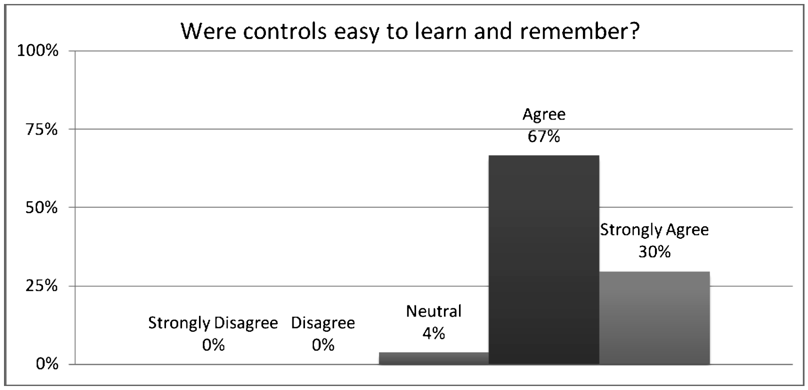

Fig. 13. Ease of use.

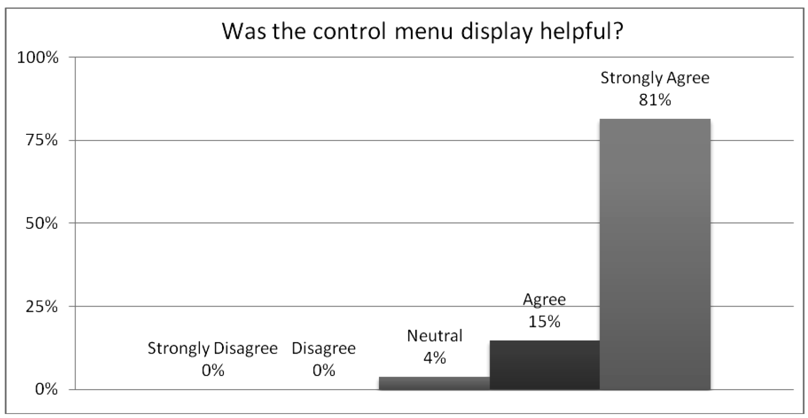

Fig. 14. Control menu.

and 3) was our pop-up overlay that documents the controls helpful? The learning tolerance for television viewing is much lower than for computer usage.

Among the total viewers in all groups there were 66 participants. Of these 66, there were 53 of them who actually manipulated the controls during the session. There were 13 others who merely watched. The answers to the two control questions from the postgame interview are shown in Figures 13 and 14. Both show good support for the learnability of our interface. From viewer opinion, the answers to all three questions seems positive.

Figure 14 shows the user's response to the pop-up overlay that documents the controls. The support for our technique is strong but we did not compare it to the alternatives. The current state of television remotes is to either provide no documentation or a large paper manual that regularly gets lost. Viewer frustration with complex remotes has become a cultural joke. We did not see this as an alternative to our technique. It is possible that there is another approach to educating the user about the controls. The important question to us, however, is whether there exists a technique that allows new users to be effective. Without such a technique TWS will

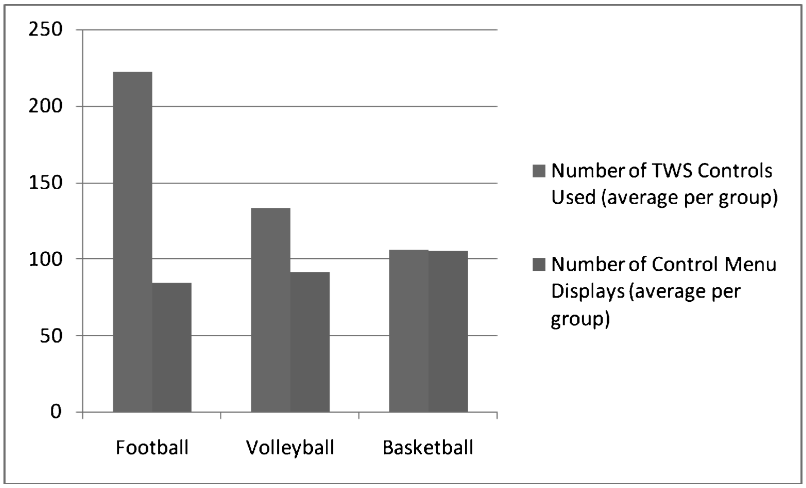

Fig. 15. Interactive control usage.

never be useful. The data clearly shows that our pop-up control overlay is effective.

Figure 15 shows the data from the log files about the number of actual control events that occurred during the experience. Some of the variation between sports can be attributed to the amount of time available for viewing. If we divide these by the available time, we get 2.2 controls/minute for football, 2.3 controls/minute for volleyball and 2.9 controls per minute for basketball. This is an interactive control event every 20-30 seconds. The viewers were obviously quite active in their control usage. One possible reason for this high use of the controls was that the viewers were just playing with them to try them out. This was definitely true for two of the football groups whose data we discarded as we have described. However, our review of the video logs show that the viewers were actually engaged with the game while operating the controls.

Self reports from response surveys can be skewed. We wanted to understand actual user behavior. Therefore, we logged all of the viewer's actions during the trials to measure what they actually did. Figure 15 shows the number of controls used but also shows the number of times the overlay menu was used to activate those controls. The key piece of information to consider here is the difference between the number of control commands issued and the number of menu overlays requested to explain those controls. If the overlay actually teaches control usage, then its use should eventually fade as they learn the controls. This would be shown by a strong difference between controls issued and menu requests. This difference is clear in football but not in volleyball and basketball.

In reviewing the logs and the viewer video, we see several indications that might address this issue. First, there were several groups that did not catch on to the fact that the controls would work with or without the overlay. Thus, they continued to unnecessarily bring up the menu. They never learned the shortcut. A second problem is that the interaction is designed so that you must hold the trigger to keep up the overlay. The approach is nonmodal. Many users expected the trigger to behave in a modal fashion where a click would bring up the overlay and another click would take it down. Thus many users would click the trigger many times before they understood how it worked. We do not have any data to indicate whether a modal or nonmodal trigger is best because other groups understood the behavior immediately. The most telling issue, however, is the camera angles. It was difficult for users to memorize which buttons would bring up which camera. Therefore, they frequently used the overlay to select a camera. If we remove camera angle selections from the data, the number of overlay requests per control request drops quickly. We conclude that the control usage (with the exception of remembering camera shots) was very easy to learn and use. The selection of camera angles was very popular as will be shown. The fact that viewers frequently needed help in selecting the desired camera indicates that more design work is needed in this area.

7.2 New Controls vs. Standard Controls

Our viewer interface provided the traditional play, pause, fast forward and rewind controls as well as the new play-based controls that go directly to next play unit and previous play unit. The traditional controls are more familiar while the new controls we believed to be more appropriate. Figure 16 shows the comparative data for play-based controls vs. the standard controls for all three sports. The standard controls got much less usage and that usage tails off over time as viewers become familiar with the new controls. In a head-to-head usage, it is clear that viewers prefer controls based on game play rather than generic rewind and fast-forward. The case that was not studied was controls that skip forward or backward a fixed time. We considered the play-based controls to be more in line with user intention.

7.3 Will the New Controls Reduce Game Viewing Time?

A critical question for sports broadcasters is whether new controls like "next play" will reduce game viewing time. An American football game takes 2.5 to 3 hours to play although actual game clock time is 60 minutes. A fair chunk of game time is consumed while no plays are in progress. This means that there is actually less than 40 minutes of action in an American football game. *The Wall Street Journal* [Bidcrman 2010] reports that there are actually only 11 minutes of action in a football game. If the viewers use the new controls to watch a game in 25% of the time required to play the game, then advertising revenue will shrink by 75% with no reduction in production costs.

The football experiment had 102 minutes of video for which we provided a 10 minute simulated delay from live time. This means that it was feasible for a group to watch the game in 92 minutes by aggressive use of next play. Of the 9 football groups, 2 groups actually watched in 92 minutes. The other 7 groups watched from 93 minutes to 114 minutes with a median time of 96 minutes. Most of the football groups caught up with simulated "live time" for

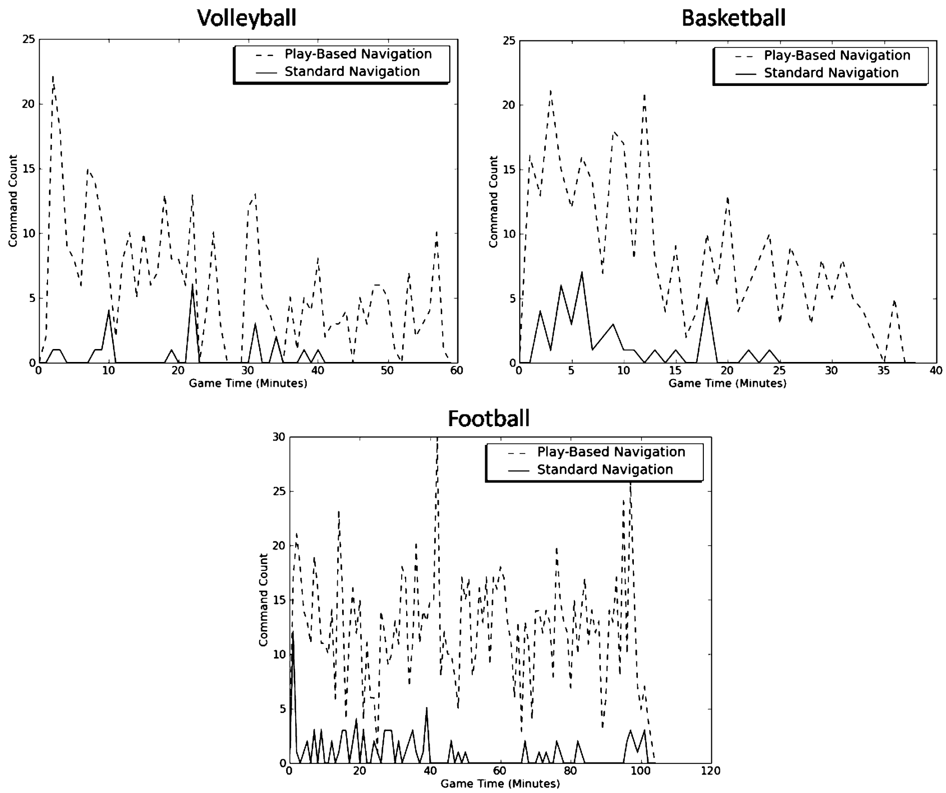

Fig. 16. Comparison of fast forward/rewind to next/previous play.

at least a few minutes during the experiment. This is easy to do in a sport that has so much nonaction time. The average group issued 81 next play commands and 60 replay commands. Many viewers commented that, if there had been better camera angles, they would have done replay more often. We did not have the closeup or goal line shots common for broadcasts. We only had the very high and wide sideline and end zone shots that coaches use. The indication here is that in a relatively close, high-interest game, viewers will spend as much time or more on additional interactive features. However, the data also indicates that 20% of the viewers may opt for the rapid skipping approach. The data shows that there is hope that multiple camera views and replay controls may mitigate the revenue effects of rapid skipping but more studies with much more diverse game and viewer profiles are needed for a definitive result.

The volleyball game had the standard broadcast camera angles and was a current rather than historical game. The total video time was 57 minutes with a 5 minute live time buffer. Two of the groups watched in the minimum 52 minutes. The other two groups never caught up to live time. Note that there is no visual indicator on our screen as to where the viewer is relative to live time. (This is a deficiency in our design that several subjects noted.) The two groups that did not catch live time were not self-regulating relative to live time.

They just got engrossed and used the controls to enjoy the game. Two of the groups expressed an enjoyment of volleyball in the pregame survey and two did not. The two groups who expressed interest in the game used the replay controls much more extensively than the other two. The average volleyball group skipped ahead 45 times and replayed 41 times.

The basketball experience had 36 minutes of video with 5 minutes delay from “live time.” Only one of the basketball groups ever caught up to live time and they immediately began using more replays to finish in 32 rather than 31 minutes. This is explained by the fact that basketball is much more continuous in its play than volleyball or football, which are rather episodic. A continuous game offers less opportunity to skip ahead. With the very long shot clock in college basketball it would be easy to skip from play to play skipping the dead time. However, only one group actually did this.

In the postgame surveys the viewers said they liked the ability to skip ahead over non-action periods. Many stated that they liked the fact that they could watch the game faster. However, the data shows that only 9 of the 17 groups ever caught up to live time and 4 of the groups that did catch up immediately started watching more replays. In reality, they were not watching faster. The sense of control made them feel like they were watching faster. This indicates the possibility of an interesting small study about whether viewer involvement shortens the perception of time without actually changing that time.

7.4 How Does Game Play Affect Control Usage?

One would assume that the way fans exert control over their viewing experience would be heavily influenced by the action in the game. Figure 17 shows the relative use of “next play” vs. “camera change” controls. A general observation is that when “next play” usage goes up, “camera change” usage declines in many cases. From our logs of user activity, we can examine the timestamps of high interactive activity and compare them with the game video. The general result is that high interest parts of the game with exciting or controversial plays have a higher incidence of camera angle control. As the game gets dull, there is a higher incidence of next play control usage.

Examination of the volleyball game usage shows that 30 minutes into the game the use of “change camera” starts to disappear. At this point, Team A jumped out to a strong lead and the game appeared to be a blowout. At about 40 minutes into the game, there is a small spike in camera angle usage. This corresponds to a strong comeback by team B which then faded until 50 minutes. At that time, Team B brought the score close but then finally lost. This corresponds to a resurgence of camera switching. This pattern of usage was uniform across all groups that viewed this game. It is clear that camera angle changes are indicative of interesting plays. This information may be useful in preparing game summarizations for later viewing by fans that have much less time.

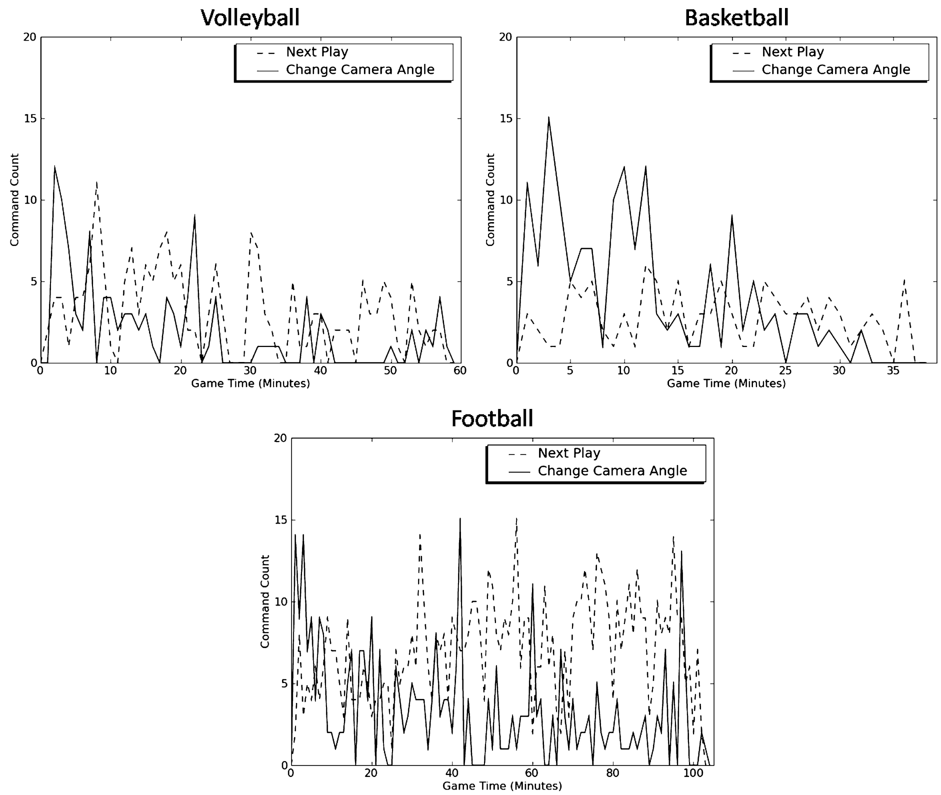

Fig. 17. Comparative use of game controls.

7.5 Postgame Interviews

We conducted open-ended interviews with viewers after their game experience. The following are the key insights from those discussions.

- —All 11 groups that watched football asked for more and better camera angles than the 2 coach's views that we had available. None of the volleyball and basketball groups who had all the standard broadcast shots made this request.
- —12 of the 19 wanted a slow motion facility. The video player we used has a very poor slow motion facility that we did not make available. This particular video player was designed for television movie and sitcom viewing where slow motion is unimportant.
- —There were some requests to freeze a frame and then flip among camera angles while on that frame. Everyone wants to be an official and make their own calls.
- —12 of the 19 requested a visual indicator of how much time difference there was between the frames they were viewing and "live time."
- —9 groups wanted to zoom in which is difficult with current technology and fundamentally limited.

—6 groups wanted special audio features. Many of BYU's football fans prefer the radio commentators to the TV commentators and they wanted the option of listening to the audio of their choice.

There was also postgame discussion about the game controller that was used. Younger viewers, particularly those familiar with video games, were generally happy with the controller. Older viewers wanted something more like a TV remote. There was also discussion about the desire to control the game with one hand rather than two so that the other hand would be free for snacks. Apparently the total experience is important. The techniques we have presented are easily adapted to a one-handed control. Although there seems to be mixed opinions about the style of control that is desired and some passion on both sides of the debate, this does not impact the nature of our viewing experience nor the results of our data.

8. CONCLUSION

It is very clear from our implementation and experiments that an interactive approach to sports is both technologically feasible and of great interest to the fans. The in-home experiments with football demonstrate that today's broadband networks and video transport are sufficient to produce compelling interactive experiences. At this point, we see only three technical hurdles to widespread deployment of this technology: 1) sports broadcasters will need to invest in video encoding hardware to produce an Internet-ready video broadcast that contains all camera feeds, 2) sports broadcast studios/trucks need at least 16Mb network connections to the Internet rather than just video, and 3) a uniform set-top box architecture on every television needs to be deployed that contains sufficient processor power and flexibility to support such technology. None of these hurdles are beyond today's technology. The challenge is in market deployment.

It is clear that controls such as "next play" and "change camera" are of great interest to fans. It is also clear that fans can very rapidly learn interactive controls using on-screen templates of the controls. There do not seem to be any usability barriers to fan adoption of the technology. There is evidence, but not conclusive, that fans will spend as much time watching time warp sports as they did the traditional broadcast and thus, revenue will be retained. There are two side notes to this result. The first is that skip-ahead behavior will likely increase with boring games. The volleyball data shows fans skipping ahead during the dull parts and then engaging in more camera angles as the score gets closer again. The second note is that about 20% of the viewers did engage in aggressive skipping ahead. As a side note, commercial Internet television software will not allow viewers to skip commercial advertisements. Sorry, there must be income to support the broadcast. The reintroduction of commercials that cannot be skipped should impact viewing times. We are convinced, as were our subjects, that time warp sports is an exciting new interactive television experience.

REFERENCES

ARMAN, F., DEPOMMIER, R., HSU, A., AND CHIU, M.-Y. 1994. Content-based browsing of video sequences." In *Proceedings of the 2nd ACM International Conference on Multimedia*. 97–103

BIDERMAN, D. 2010. 11 minutes of action. *Wall St. J.* (1/15/10).

BOBICK, A. 1993. Representational frames in video annotation. In *Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers*. 111–115.

CHAMBEL, T. AND GUIMARAES, N. 2002. Context perception in video-based hypermedia spaces. In *Proceedings of the 13th ACM Conference on Hypertext and Hypermedia*. 85–94,

CHORIANOPOULOS, K. AND SPINELLIS, D. 2004. Affective usability evaluation for an interactive music television channel. *Comput. Entertain. 2*, 14.

CHORIANOPOULOS, K. AND SPINELLIS, D. 2004. User interface development for interactive television: extending a commercial DTV platform to the virtual channel API. *Comput. Graph. 28*, 157–I66

DAVIS, M. 1993. Media Streams: An iconic visual language for video annotation. In *Proceedings of the IEEE Symposium on Visual Languages.* 196–202.

DRUCKER, S., GLATZER, A., STEVEN, D., AND WONG, C. 2002. SmartSkip: Consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 219–226.

DUAN, L. Y., XU, M., YU, X. D., AND TIAN, Q. A unified framework for semantic hot classification in sports videos. In *Proceedings of ACM Multimedia*. 419–420.

FENG, S., MANMATHA, R., AND LAVRENKO, V. 2004. Multiple Bernoulli relevance models for image and video annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1002–1009.

FRANCOIS, A., NEVATIA, R., HOBBS, J., BOLLES, R., AND SMITH, J. 2005. "VERL: An ontology framework for representing and annotating video events. *IEEE Multimed. 12*, 4 76–86.

GIRGENSOHN, A., WILCOX, L., SHIPMAN, F., AND BLY, S. 2004. Designing affordances for the navigation of detail-on-demand hypervideo. In *Proceedings of the Working Conference on Advanced Visual Interfaces.* 290–297.

HAN, M., HUA, W., XU, W., AND GONG, Y. 2002. An integrated baseball digest system using maximum entropy method. In *Proceedings of ACM Multimedia*. 347–350:

HAND, S. AND VARAN, D. 2007. Exploring the effects of interactivity in television drama. In *Proceedings of the 5th European Conference on Interactive TV: A Shared Experience,* Lecture Notes in Computer Science, vol. 4471, Springer-Verlag, 57–66.

LEE, H., FERGUSON, P., GURRIN, C., SMEATON, A. F., O'CONNOR, N. E., AND PARK, H. 2008. Balancing the power of multimedia information retrieval and usability in designing interactive TV. In *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video*. 22–24.

LI, F., GUPTA, A., SANOCKI, E., HE, L.-W., AND RUI, Y. 2000. Browsing DigitalVideo. In *Proceedings of the SIGCHI International Conference on Human Factors in Computing Systems*. 169–176.

LYNN, S. G., OLSEN, D. R., AND PARTRIDGE, B. G. 2009. Time warp footbal. In *Proceedings of the European Conference on Interactive Television*. 77–86.

NORMAN, D. A. 2002. *The Design of Everyday Things*. Basic Books.

QI, W., GU, L., JIANG, H., CHEN, X.-R., AND ZHANG, H.-J. 2000. Integrating visual, audio, and text analysis for newsvideo. In *Proceedings of the IEEE International Conference on Image Processing*. 520–523.

SHIPMAN, F., GIRGENSOHN, A., AND WILCOX, L. 2008. Authoring, viewing, and generating hypervideo: An overview of HyperHitchcock. *ACM Trans. Multimedia Comput. Comm. Appl. 5*, 2, 1–19.

TRUONG, B. T. AND VENKATESH, S. 2007. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Comm. Appl. 3,* 1.

WAN, K. AND XU, C. 2007. Efficient multimodal features for automatic soccer highlight generation. In *Proceedings of the 17th International Conference on Pattern Recognition*. IEEE, 973–976.

XU, C., WANG, J., LU, H., AND ZHANG, Y. 2008. A novel framework for semantic annotation and personalized retrieval of sports video. *IEEE Trans. Multimedia, 10*, 3, 421–436.

XU, C., WANG, J., WAN, K., LI, Y., AND DUAN, L. 2006. Live sports event detection based on broadcast video and web-casting text. In *Proceedings of ACM Multimedia*. 221–230.

XU, M., DUAN, L. Y., XU, C. C., AND TIAN, Q. 2003. A fusion scheme of visual and auditory modalities for event detection in sports video. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 333–336.

YU, X., YAN, X., LI, L., AND LEONG, H. W. 2008. An instant semantics acquisition system of live soccer video with application to live event alert and on-the-fly language selection. In *Proceedings of the ACM International Conference on Image and Video Retrieval*. 495–504.

Received July 2009; revised March 2010, August 2010; accepted August 2010