Watch What I Watch

Using Community Activity to Understand Content

David A. Shamma
Yahoo! Research Berkeley
1950 University Ave, Suite 200
Berkeley, CA, USA 94704
shamma@yahoo-inc.com

Ryan Shaw*
School of Information
University of California, Berkeley
ryanshaw@ischool.berkeley.edu

Peter L. Shafton
Yahoo! Research Berkeley
1950 University Ave, Suite 200
Berkeley, CA, USA 94704
pshafton@yahoo-inc.com

Yiming Liu†
School of Information
University of California, Berkeley
yliu@ischool.berkeley.edu

ABSTRACT

This paper presents a high-level overview of Yahoo Research Berkeley's approach to multimedia research and the ideas motivating it. This approach is characterized primarily by a shift away from building subsystems that attempt to discover or understand the "meaning" of media content toward systems and algorithms that can usefully utilize information about how media content is being used in specific contexts; a shift from semantics to pragmatics. We believe that, at least for the domain of consumer and web videos, the latter provides a more promising basis for indexing media content in ways that satisfy user needs. To illustrate our approach, we present ongoing work on several applications which generate and utilize contextual usage meta-data to provide novel and useful media experiences.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Video*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Synchronous interaction; Collaborative computing*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Indexing Methods*

General Terms

Design, Human Factors

*Author may also be reached at Yahoo! Research Berkeley: rshaw@yahoo-inc.com.

†Ditto: yiming@yahoo-inc.com.

MIR'07, September 28–29, 2007, Ausburg, Bavaria, Germany.

Keywords

Video, Content, Sharing, Community, Tagging

1. INTRODUCTION

People like video. In our homes, at the movies, or cradled in our hands, flickering, glowing screens have the power to draw us in, or at least divert our attention. But our love affair with the moving image isn't confined to the edges of those screens. The things we watch spawn conversations. These conversations extend to engage even those who never saw the original media, people who learn about recorded events in print, on the web, or from friends. Sometimes this conversation unfolds in real time, as people join friends or strangers in living rooms, bars, and movie theaters to share the experience of watching.

In the past few years, the web has multiplied and enriched these kinds of conversations around recorded media. Video is not new to the web; numerous video content providers have been streaming video content for almost a decade. Until recently, however, the web was considered to be simply another channel for distributing professionally produced content to the masses. Its potential for enriching the conversation around media was mostly ignored. Things have changed, and the web is being recognized as a medium for communicating about and doing things with live and recorded video, instead of just a pipe for delivering it.

Today there are hundreds of web sites which allow users to upload, watch, embed, share, and re-edit video. While they make some copyright owners very uncomfortable, these activities can be seen as extensions of the water-cooler conversations of previous eras. But where earlier conversations consisted mainly of spoken or printed words, these new conversations are materialized in links, embed codes, server logs and revision histories. As a result, what was previously ephemeral can now be archived and curated, enabling communities to create what Peter Lunenfeld has called "hypercontexts" for media [9]. These hypercontexts offer rich new ground for research into and applications of digital media.

Examining community usage of text in order to better organize and index that document is not a new idea. The fields of bibliometrics and informetrics pioneered this ap-

proach, looking at citation networks as a way of organizing scholarly works [4]. PageRank applied this idea to the textual web, using hyperlinks to web pages to measure authoritativeness [13]. Contemporary to PageRank, Bradshaw et al. introduced Rosetta; a system for indexing research papers using only citation text and not the source paper's content [2]. Systems like PageRank and Rosetta are successful because they use the communicative context (hyperlinks or citations) as a guide to significance, rather than focusing solely on trying to understand the content itself (which is meaningless outside of such context).

Much research into multimedia information systems, in contrast, has focused on content understanding. While some advancements have been made in using meta-data and closed captioning information, there is a primary concern about media which is *unpopulated* with rich annotation. Many researchers seek to "close the semantic gap" by bridging the low-level visual concepts identifiable using computer vision algorithms and the high-level concepts used by humans. The implicit assumption of this focus on semantics is that media is best understood by looking at what it literally represents. We suggest that media is best understood through the contexts in which it is used, and thus that research focus should shift in focus from semantics to pragmatics.

2. FROM SEMANTICS TO PRAGMATICS

Content analysis plays an important role in content understanding. Being able to detect who or what was shown in a given image, what was said in a region of audio, or where sentence or scene boundaries occur is useful for understanding the semantics of a piece of media. Having detailed meta-data about the content within a media item enables users to quickly navigate and segment the media into useful sub-items that can then be shared with or consumed by other users. The meta-data generated by content analysis will be necessary in creating systems that enable us to capture the context in which media is used. It is this usage that will lead us to pragmatics.

We believe that to be effective, investigation of applied media pragmatics must involve three related strands of research: qualitative and quantitative analysis of media practices "in the wild," design and prototyping of new kinds of tools for using media in ways that generates useful information, and research into algorithms for using this information to organize and index media.

2.1 Media use in the wild

Any effort to develop tools and systems for improving organization of and access to media must be grounded in a thorough understanding of how people are actually using media. Much current research in multimedia information retrieval has focused on the problem of identifying specific people, places, or things being depicted or discussed. While this may be useful for video editors searching stock footage libraries, intelligence analysts reviewing foreign news broadcasts, or security guards examining surveillance video, there exists a wide array of media uses which do not fit into this paradigm.

Take for example television news, a video domain which has received an inordinate amount of attention from MIR researchers. While there is high value in searching across a large corpus of video for an entity or concept, the larger utilization of these efforts are not applicable towards a greater community. The question "Why do people watch the news on TV or on the web?" must be investigated. It is not because they are looking for information on the specific topics being covered by those news reports. Rather they are interested in what recent events have been deemed "newsworthy," regardless of the actual content of those events. Newsworthiness is a property of news video that is inherently social, and cannot be determined through techniques focused on video content alone. The shared viewing and discussion of news video by a distributed community of viewers is what makes it newsworthy. Failure to take into consideration this social significance of news content has plagued designers of news information systems for decades [3].

To avoid such oversights, technical research into multimedia information systems should be accompanied by two kinds of non-technical research. First is the investigation of how specific communities are actually using media. This should involve both qualitative research aimed at developing interpretive understanding of how media content takes on layers of meaning as it diffuses through communities of users, and quantitative research aimed at developing more formal models of this diffusion process. Second is qualitative examination of the media objects themselves, in order to better their physical features and the kinds of meaning-making processes they might afford. For example, much work in MIR assumes that speech-to-text technology will be generally useful, but examination of how language is actually used (or not used) in popular videos on the web calls this assumption into question.

2.2 New tools for using media

A thorough understanding of people's relationships with media is a perquisite for successful design of tools for using media. The creation of new tools is critical for putting media pragmatics to use. New tools provide new sources of contextual information about media usage that can then be used to organize and index that media. The invention of web-scale hyperlinking provided not only a new tool for authoring and navigating documents, but also a powerful new source of information that could be used to rank search results.

Likewise, tools on social media sharing sites that allow users to comment on, rate, share, and create collections of media are more than just attractive features; they are also important ingredients for filtering and organizing media content. Even if only a small percentage of users actually utilize these features [5], these few "power users" can generate value for all users of the site through their activity.

But comments, ratings, shares, and collections are only the tip of the iceberg. Better support for multimedia by web browsers and plug-ins such as Adobe Flash has created the opportunity to create systems that enable both rich interaction and a constant stream of data about how media is being consumed, shared, reused.

For example, where editing video was once an opaque process confined to a single computer, with web-based video editing that process can be made transparent, enabling systems to learn how different users make creative decisions and how different kinds of content get reused or not. Ideally, a virtuous circle forms: new tools make possible new ways of using media, generating new kinds of information about the social functions of media content, which can in turn be fed back into the design of new systems.

2.3 Media Organization

Perhaps the most difficult challenge for research into media pragmatics is figuring out how to usefully harness the contextual information generated by new media tools. There are three areas in which we expect this kind of data to be useful: search, filtering, and presentation.

Improvements in search technology come with better understanding of the searcher's intentions. While intentions will never be transparently available, the information about when, where and what people are doing when they are looking for media can help quite a bit. For example, a user who queries for images while using a web-based tool to author a slide presentation is likely looking for images of a certain resolution, that are not too visually complicated, and that are thematically related to the rest of the presentation. This kind of contextual information, as well as data on what kind of features are shared by images that have been selected by other slide presentation authors in the past, could be exploited to improve search results given this particular task. Content-based retrieval techniques can provide some much assistance into this particular search result tweak, but it does not provide the whole solution.

Contextual and collective information can be used to filter media collections even in the absence of a query. This is already occurring on sites like Flickr and YouTube, using features like comments and shares. As these features expand to include new and different features that provide more nuanced and finer-grained measures of interest in media items, we expect this kind of filtering to improve.

Finally, logs of authoring processes and corpora of reused media objects might prove useful for automating the presentation of queried or filtered media results. By identifying syntactic patterns in the way certain kinds of media are used by human authors, presentation engines could attempt to emulate these patterns in automatically constructed presentations. For example, it may be possible to classify certain kinds of audio as "good background music" or certain segments of video as "suitable for text overlays." Of course, better semantic understanding of media content is necessary for fully automatic generation of high-quality presentations, but simple syntactic patterns may prove to be very useful.

3. NEW EXPERIENCES

Advances in pragmatics in MIR means the creation of new experiences with video. Currently, video that is found online is mostly short form clips. In a four month study[1]of 1515 users sharing videos via Instant Messaging (IM), we find that the clip durations range from 70 to 500 seconds (averaging around 230.2 seconds per clip). This, in large part, is due to YouTube's popularity and their 10 minute maximum video length. However, there is more video online or in the world which is not on popular web services like YouTube and is often much longer and unstructured. While, looking for content structure or aligning meta-data (like closed captioning) has some advantages, it is orthogonal to the web itself. The web is about communication. Video as a medium of communication remains unexplored as there exists few tools to enable these new forms consumption and communication. We will describe how we are changing the consumption and usage of video from reuse to organization to sharing.

[1]Data collected from the Zync Plug-in for Yahoo! Messenger. See section 3.4 for more details about this study.

3.1 Remixing

Cheaper bandwidth and disk space have fueled the growth of web video. Faster processors and simple editing tools have contributed to the mainstreaming of a "remix culture" in which amateur video editors appropriate and re-create pop culture media. Though these sorts of activities are not new, what is new is the scale on which they are occurring. This presents an excellent opportunity for research into community behavior with respect to media reuse, and to explore the creation of services that leverage this behavior.

To investigate these possibilities further, we developed a web-based platform that allows users to select, annotate and remix material from a shared media archive. An initial deployment in association with the San Francisco International Film Festival (SFIFF) provided a useful data set for analyzing user behavior, which in turn led to many insights into user behavior and the potential for leveraging community annotation and remix data for a range of purposes, including intelligent authoring assistance and identification of reusable media. These findings are presented in detail in [14].

One advantage of allowing users to select and trim segments by hand (as opposed to doing this automatically) is the resulting fine-grained statistics on media usage, unprejudiced by a priori shot boundaries. To facilitate analysis of the data, we generated reuse histograms for each source media object. For each 0.1 second interval of the source media object, we counted both the number of different remixes in which the interval was used, and the total number of times the interval was used (to reflect looping and other repetitive usage).

Qualitative evaluation of the histogram curves yielded a number of interesting patterns. As expected, peaks in the histogram were correlated with points of high emotive energy. The histogram slopes indicate how reuse builds and tapers off as energy builds and wanes. Figure 1 shows a typical example. The source media shot depicts a long zoom toward three flirtatious girls, eventually focusing in on the center girl, who suggestively puts a lollipop in her mouth. The reuse histogram clearly shows an early peak at the point where the three girls flip their hair, and then builds to the main peak at the point where the lollipop is inserted.

As proxies for the emotive impact of media content, these reuse histograms have a number of potential uses for source media summarization and browsing. The amount of reuse can serve as an importance score for selecting representative thumbnails. A scalable video skim could be produced by only including frames above a certain usage threshold, which could then be adjusted for zooming in on scenes of interest. We also believe that these statistical patterns will have utility for automatically trimming shots. An automatic trimming algorithm that takes community reuse statistics into consideration could trim in such way as to preserve emotive peaks.

The platform developed for the film festival has provided a solid base for further research into human-centered multimedia. The data we gathered using the platform provided detailed views of usage patterns beyond the basic searching and browsing roles to which users have traditionally been relegated. Furthermore, we were able to demonstrate that this data can be used to develop emergent pragmatics for the media being explored and reused, with implications for new and improved media retrieval, browsing, and authoring applications.

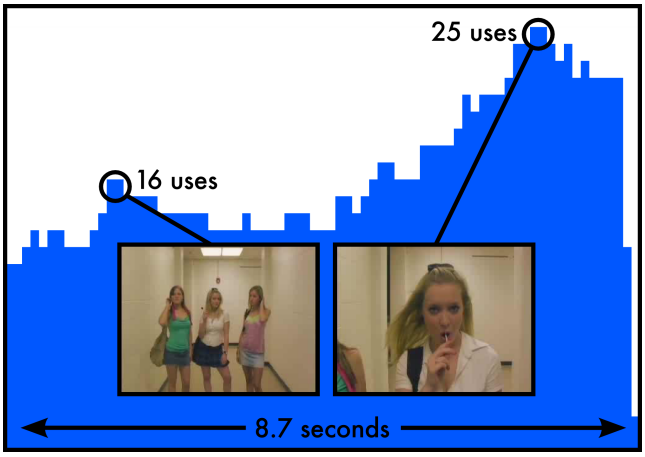

Figure 1: Histogram peaks (counting frame re-usage in a community remix application) are correlated with points of high emotive energy. The source media shot depicts a long zoom toward three flirtatious girls, eventually focusing in on the center girl, who suggestively puts a lollipop in her mouth.

3.2 Community Chapters

Most online video is unannotated and unedited. While many people are reluctant to edit a 10 second camera-phone clip [7], there are still many archives of long duration content that remain unsearchable. The major challenge, especially with long form video, is in media organization. There has been much work in automatic chaptering of content [8, 20]; we were interested if we could collect segmentation information in a WIKI style collaboration.

In a study of video webcast tools, Toms et al. [17] noted users perceived the greatest usefulness from a table of contents (TOC) when assessing relevance and during summarization. This was contrasted with the timeline, which was generally noted as not useful. However, during summarization, users spent significantly more time using the timeline over the other components. Users spent approximately 59 seconds spent using the timeline during a summarization task versus time spent using the TOC (23s), the Power Point Slides (31s), the video (30s), and the search box (30s).

At Yahoo!, we have a series of events called HackDay: an innovation contest where participants are given 24 hours to build a working prototype which uses web technology. At the end of 24 hours, all the participants are given 90 seconds to present their 'hack' to a jury. This presentation is recorded and made available online. However, the resulting video is one monolithic video often running over four hours in length and containing over one hundred presentations.

We noted that the individual hackers would generally know how to locate their presentation (people remember if they were in the middle or towards the end of the set of people who presented around them). This list could be used to construct a TOC of the hacks which could segment and chapter the timeline. From this observation, we built HackDay TV[2] (see figure 2), a system to support community chaptering of content. To reduce the amount of time spent on the timeline

[2]See `http://tv.research.yahoo.com/hackdayuk` for a non-editable demo of this application

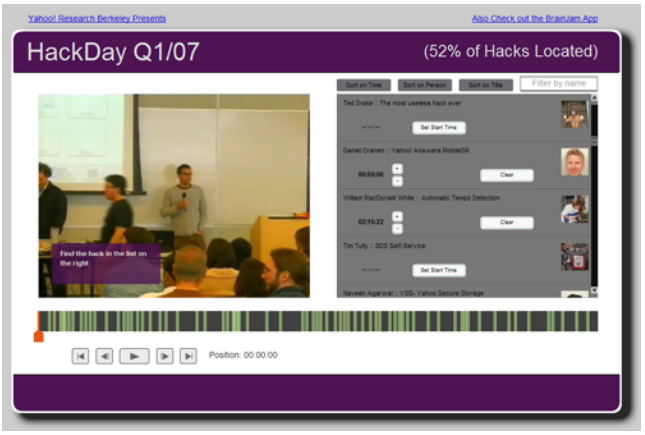

Figure 2: HackDay TV: a community based chaptering tool for a 4 hour video.

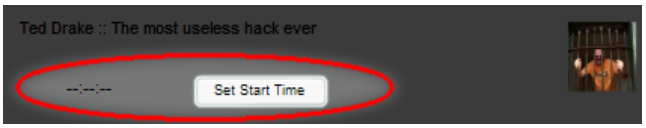

Figure 3: Clicking Set Start Time adds the Hack to the timeline for a 90 second duration.

searching for content, we designed a timeline which could reflect the TOC.

To make HackDay TV, we first acquired a list of hacks and hackers. The list was put next to the 4 hour video. Using an employee database, we posted each hacker's photo next to their hack. With 4 hours of video and a searchable list of 130 hackers, we created two problems. Hackers need to find their hack in the list as well as on a timeline. To remove some of the barriers in finding a hacker in the list, the list of hacks and hackers can be sorted by title and author and easily searched. Each hack entry has a **Set Start Time** button (figure 3). When clicked, that hack is 'dropped' or populated onto the timeline with a 90 second duration mark. Since 90 seconds represents a small portion of a 4 hour timeline, call-out balloons (as seen in figure 5) were used to increase visibility. A start time is listed in the hack's list entry (figure 4) that can be "nudged" using the +/- buttons next to the hack's start time. The hack's duration is set and cannot be edited. A button marked **Clear** removes the hack from the timeline (but leaves it in the TOC).

In HackDay TV, anyone can add listed hacks to the timeline. We saw users who added several hacks (in many cases, they added other people's hacks but not their own). Addi-

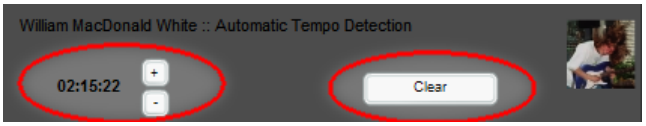

Figure 4: The hack's time can be 'nudged' using the +/- buttons next to the hacks start time. The button marked Clear removes the hack from the timeline (but leaves it in the TOC).

tionally, we had no security restrictions in the application. Anyone could edit, nudge, or remove from the timeline a populated hack. In effect, the timeline becomes a social reflection of the TOC—enabling navigation and increasing its usefulness during summarization.

The highest perceived usefulness of any component in an interactive video system is the TOC, however research shows users spend more time on the timeline [17]. HackDay TV gives the starting point of the TOC but then directly links that to the timeline. In our internal deployment, 60 percent of the hacks were located within 1 week of launching the application. A total of 22 users located 82 percent of the hacks. 178 users have come back to view the application.

3.3 Social Sharing

Community chaptering solves a general indexing problem with long form video content online. However, the aforementioned work of Page [13] and Bradshaw [2] described indexing content on how it was being referred to and not what it contains. Recently, tagging has made several advances in the multimedia research community [10] and on popular photo sharing websites (like Flickr). The effectiveness of applying tags to an entire video file (regardless of length) remains to be seen. In a video, people talk about certain sections or *moments* where an event occurs: the memorable part of a speech, the perfect goal kick, or when the cat falls off the table. The annotation of segments of a video, generally referred to as a time-tag or a deep-tag, has been a proposed solution. The time-tag is the first step towards the taming of media in the wild, but it only addresses part of the issue at hand; It is the sharing of the time-tag (and their media segments) which will create a network of social communication.

3.3.1 Time-Tags

With time-tags, annotation tags are applied to a point in time in a video (like a bookmark) or to a segment of video (setting a start time and a duration or a start and end time). However, basic or time-tagging cannot be treated as communication (as we have seen with hyperlinked structures). A time-tag is simply a personal annotation. If the time-tag is publicly visible, implicit communication can occur.

While tags are popular online [6], tagging time based media has been given little attention. Tags are generally applied to an entire video and not necessarily only a segment. Time-tagging segments of media can be particularly powerful for explanation (a common usage of tagging in other media [1]). Deep tagging is needed to meaningfully annotate video content, particularly long form content. More importantly, time-tagging is essential for sharing.

In several informal conversations with lecture attendees, we noticed viewers generally wanted to share only a small segment of a lecture. In a one hour lecture, the part they wish to share is under 5 minutes (often under 2 minutes). From this, we built BrainJam TV: a video viewer which allows people to trim out a small segment and share it with a colleague. Instead of HackDay TV chaptering, individuals can add tags to the timeline and set the intended duration.

3.3.2 Deep Sharing

BrainJam TV facilitates the sharing of segments of video. The application prototype, figure 5, presents a standard player and timeline with two new added features. First,

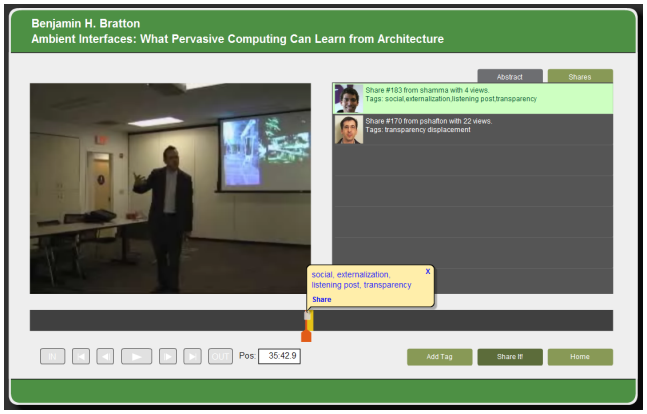

Figure 5: The BrainJam TV application displays a list of personal tags (as labels) on the timeline as well as a ranked list of shared segments.

there is an *Add Tag* button which puts a label and segment on the timeline with a default duration (3 minutes). The duration can be edited and tags can be added. These tags remain persistent and serve as personal notes and bookmarks per user.

Each personal annotation, figure 6 is displayed with several actions. A tag can be edited (changing its labels, duration, or location) or deleted entirely. A tag may also be shared. Sharing a tag is an explicit action that takes a private annotation, makes it public, and sends it to another user. Sharing prompts for a recipient email address(es) and a comment to send along with a URL. The sent URL is a link to the exact spot (and duration) in the source video. This URL is much like a *permalink* in a blog—it provides a direct access point to jump into a video and play the tagged segment from a start point to an end point (both user specified). We refer to this process as *deep sharing* which enables a small portion of a long video to be shared effectively. The video segment becomes the mechanism of communication between two individuals. This communication collects additional meta-data about the segment, retains the meta-data of the entire source video, and collect meta-data from other shares on the same source.

When viewing a video (or a share), we display all the shares on the source video (as seen to the right of figure 5). This list serves as a summary of the video. Not a summarization of content, but a summarization of how people are communicating this video with their colleagues. Previous research efforts would communicate areas of interest by marking *footprints* [12, 11] (areas people have watched) without annotation. Footprints only show you what region has been seen by the community of users which provides a summary via an indication of viewed regions.

BrainJam TV creates a succinct list of regions to watch with small tag-based summaries—this is a change in the how we think about community consumption and annotation of video. The collection of time-tags and explicit shares creates a link structure of video segments. This structure shows connections within and across media. In a body of technical talks, a simple query can retrieve all the clips relating to a single 'tag' query. Moreover, we are beginning to investigate how the overlapping segments of tags can be used to disambiguate semantic meaning (during clustering and retrieval)

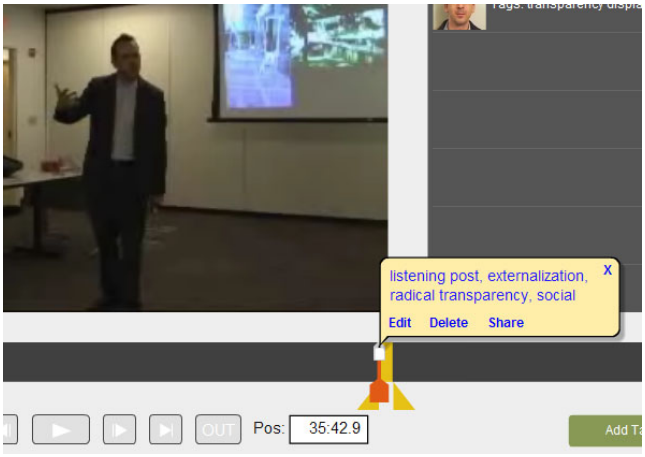

Figure 6: A time tag showing a set of text annotations and the duration on the timeline. The tag's owner can edit, delete, or share the tag.

in video content as we have seen with other tagged content online. [15]

3.4 Collaborative Viewing

The growth of online video has lead people to upload and share videos. In some cases, the video is sent via link (like an email or an instant message). Online chatting along with a live stream or broadcast has become popular. While there is only a single broadcast audio feed, distraction still occurs. In a study, even with this added distraction, Weisz et al. [18] showed people feel closer to each other and will enjoy the media more if they can chat with others at the same time.

Other times, people will wait until their friend drops by their home and show them the video in person. Many people leave video clips on their cell phone to share with people they might run into [7]. Aside from the distractions inherent to listening to two concurrent audio streams at once [19], face-to-face sharing provides a rich social engagement, one where people discuss while the video is playing or pause and rewind to review or talk about the funny or critical moments.

There are many applications which allow for group chat during live streaming (like `http://justin.tv` and `http://operator11.com`). In these applications, a stream is always moving which prohibits long conversations. Additionally, a few applications allow for chat with unsynchronized video playback (YouTube Streams and `http://meebo.com`). Here, conversation occurs but is not correlated with a given time in the video stream. We built a multi-user video player that stays in sync across locations and is integrated into Yahoo's Instant Messaging system, where millions of people already converse across remote locations.

Zync (`http://zync.research.yahoo.com`) captures these moments when people synchronously share video; one where both viewers share the 'remote control' and can play, pause, and rewind the video while chatting. During a conversation, a video player is 'docked' next to the text chat in an IM window. Either participant can enter a video by typing a video URL from a popular video sharing site (Yahoo! Video, YouTube, etc). The video loads for both participants and starts playback. This player always stays in sync with the remote side. The control is completely open and

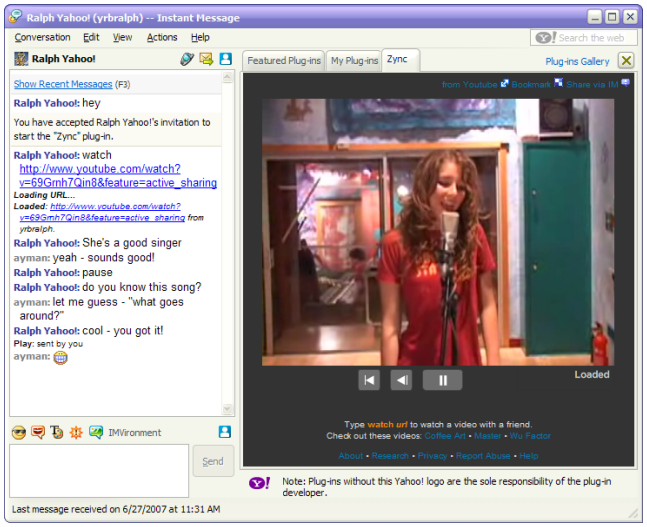

Figure 7: A synchronized video chat using Zync. Here, the remote user (Ralph) pauses the video when he asks a question. The local user (Ayman) answers and resumes playback himself.

is not a master/slave relationship. In figure 7, the remote user (Ralph) pauses the video when he asks a question. The local user (Ayman) answers and resumes playback himself. Zync's model of interaction facilitates an interaction that is similar to a face-to-face video sharing experience.

Unlike our previous prototypes discussed in this section, Zync's community collection is purely implicit. No sharing, chaptering, or tagging component is present. Zync collects[3] chat volume, emoticons, and scrub behavior from its IM users. With this data, we have begun to examine sequence behaviors across users and video URLs.

While this work bears similarity to that of Syeda-Mahmood and Ponceleon [16] (who analyzed video browsing patterns for key-frame generation), we are looking at these behaviors in an IM context to see how the synchronous communication of shared video affects people's sequence behaviors. In addition to sequence patterns, Zync leads towards an understanding of overall sharing and communication behaviors in this new medium. Zync examines both browsing behavior and communication. Again, sharing and communication is integrated in the pragmatics of our approach. Figure 8 shows the aggregate activity (as a percentage) of 391 people sharing (in pairs) a single, particular video using Zync. We can see scrub behaviors (play, pause, and rewind) happen more frequently in the first part of the video, while chat behaviors tend to become more prevalent in later parts. Figure 9 shows the same data as raw counts (so we can see that activity and users actually decreases over time). In these figures, we depict chat as a boolean (chat or no chat). We use chat volume as an additional feature, but it is not represented here. With this data, we are looking towards modeling the sequence behaviors across users to determine areas with a high level of interest, which can be used for remixing applications and summarization.

[3]Users are presented the option to opt-in to data collection upon first run.

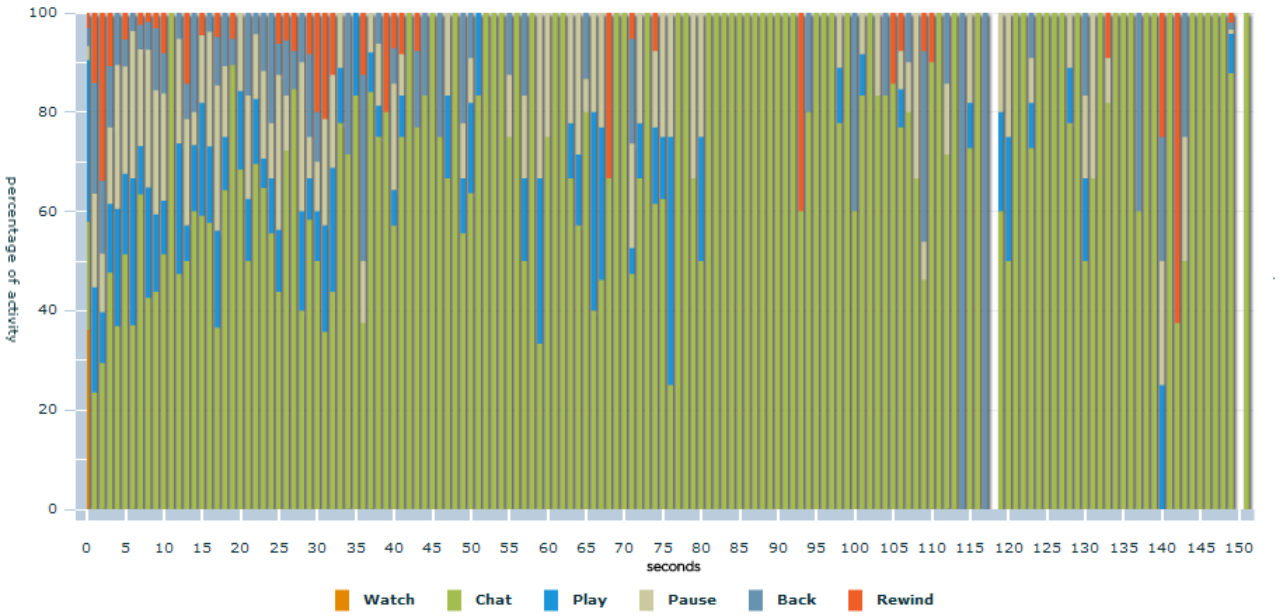

Figure 8: This chart shows the activity (as a percentage) of 391 people sharing a single video (in pairs) using Zync, broken down by activity type.

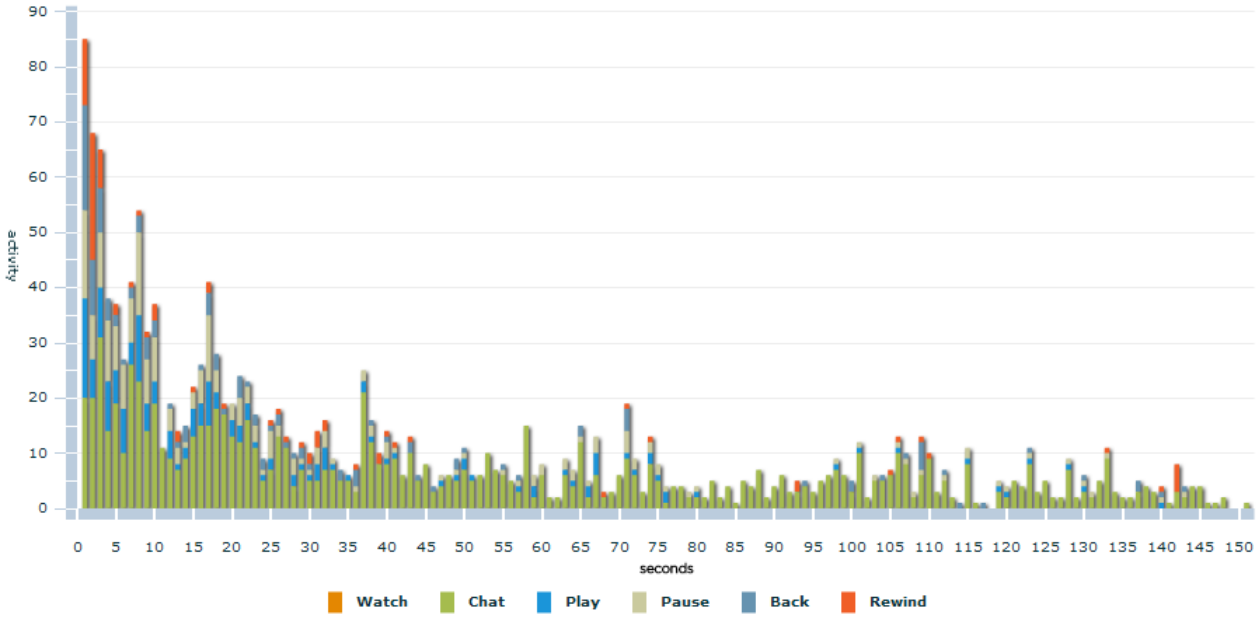

Figure 9: Here we see the total volume of activity by second of the same video as seen in figure 8. These counts are the number of users 'doing something' (like: chatting, pausing, etc.) at each second, not how many were viewing at the second. (The volume of each atomic chat is not represented here.)

4. THE FUTURE OF VIDEO

The new experiences we have presented are the beginning of online social video. The tools we presented in this article can benefit by adding existing techniques (using content analysis) and can enable new forms of indexing (using social networks) for future retrieval.

4.1 Content is People Too

While meta-data that comes from the user interaction with the video is an important part of many of our systems, there is still a role for content analysis. Many tasks people perform when interacting with our systems could be greatly assisted by the addition of meta-data that comes from content analysis. For example, the deep sharing task requires the user to define a segment to be shared. While it is cumbersome for the user to navigate to a specific frame that defines a scene change boundary, this is a task where content analysis has an established solution. When the user is selecting a segment that they are interested in sharing, the system can use the meta-data from a scene change analysis to suggest in and out points to the user. Likewise the system could use moments of detected silence as suggested segment boundaries.

Since our system architectures use meta-data from content analysis in a similar fashion to meta-data that comes from other users, the user is free to use or ignore the information as it suits their task. The way we handle analysis meta-data allows the user to determine what meta-data from which analysis will best assist them in what they want to achieve. This works in a similar fashion to how the user chooses to accept or ignore the suggestions from other users in the system and has large design implications for the future of video applications.

4.2 Sharing People

People sharing media can be used to further 'tune' the meaning of a segment of video. Within any larger organization, university or social network, there is a diverse set of people which comprise the greater community. For each video or segment sent, we take the sender and receiver into account as another valuable piece of meta-data. We have begun to investigate what it means to know the sender and their recipients. Unlike collaborative filtering, our representation changes the indexed meaning of the video and it segments with usage. Integration into other online communities (like del.icio.us or Facebook) will be integral in this approach. For any online community this will steer the media in new directions.

We have presented a new focus for how we think about multimedia; one where communication is the center of semantic context. This focus brings a new approach with the study of pragmatics in the forefront—pragmatics will bring semantics.

5. ACKNOWLEDGMENTS

The authors would like to thank Ellen Salisbury, Mor Naaman, and everyone at Yahoo! Research Berkeley for their continuing guidance, support, and general coolness.

6. REFERENCES

[1] M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM Press.

[2] S. Bradshaw, A. Scheinkman, and K. Hammond. Guiding people to information: providing an interface to a digital library using reference as a basis for indexing. In *IUI '00: Proceedings of the 5th international conference on Intelligent user interfaces*, pages 37–43, New York, NY, USA, 2000. ACM Press.

[3] J. Brown and P. Duguid. Borderline Issues: Social and Material Aspects of Design. *Human-Computer Interaction*, 9(1):3–36, 1994.

[4] E. Garfield. New international professional society signals the maturing of sciento-metrics and informetrics. *The Scientist*, 9(16), August 1995.

[5] M. J. Halvey and M. T. Keane. Exploring social dynamics in online media sharing. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1273–1274, New York, NY, USA, 2007. ACM Press.

[6] T. Hammond, T. Hannay, B. Lund, and J. Scott. Social Bookmarking Tools (I). *D-Lib Magazine*, 11(4):1082–9873, 2005.

[7] D. Kirk, A. Sellen, R. Harper, and K. Wood. Understanding videowork. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 61–70, New York, NY, USA, 2007. ACM Press.

[8] C. Y. Low, Q. Tian, and H. Zhang. An automatic news video parsing, indexing and browsing system. In *MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia*, pages 425–426, New York, NY, USA, 1996. ACM Press.

[9] P. Lunenfeld. The myths of interactive cinema. In D. Harries, editor, *The New Media Book*. British Film Institute, 2002.

[10] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA, 2006. ACM Press.

[11] R. Mertens, R. Farzan, and P. Brusilovsky. Social navigation in web lectures. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 41–44, New York, NY, USA, 2006. ACM Press.

[12] R. Mertens, H. Schneider, O. Müller, and O. Vornberger. Hypermedia Navigation Concepts for Lecture Recordings. *Proceedings of E-Learn*, 2004.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web, 1999.

[14] R. Shaw and P. Schmitz. Community annotation and remix: a research platform and pilot deployment. In *HCM '06: Proceedings of the 1st ACM international workshop on Human-centered multimedia*, pages 89–98, New York, NY, USA, 2006. ACM Press.

[15] S. Sood, K. Hammond, S. Owsley, and L. Birnbaum. TagAssist: Automatic Tag Suggestion for Blog Posts. In *International Conference on Weblogs and Social Media*, 2007.

[16] T. Syeda-Mahmood and D. Ponceleon. Learning video

browsing behavior and its application in the generation of video previews. *Proceedings of the ninth ACM international conference on Multimedia*, pages 119–128, 2001.

[17] E. G. Toms, C. Dufour, J. Lewis, and R. Baecker. Assessing tools for use with webcasts. In *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, pages 79–88, New York, NY, USA, 2005. ACM Press.

[18] J. D. Weisz, S. Kiesler, H. Zhang, Y. Ren, R. E. Kraut, and J. A. Konstan. Watching together: integrating text chat with video. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 877–886, New York, NY, USA, 2007. ACM Press.

[19] A. Welford. Single-channel operation in the brain. *Acta Psychol (Amst)*, 27:5–22, 1967.

[20] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying scene breaks. In *MULTIMEDIA '95: Proceedings of the third ACM international conference on Multimedia*, pages 189–200, New York, NY, USA, 1995. ACM Press.