Art or Circus?
Characterizing User-Created Video on YouTube

Brian M. Landry and Mark Guzdial
School of Interactive Computing & GVU Center
College of Computing
Georgia Institute of Technology
85 5th St., Atlanta, GA, 30308 USA
{blandry, guzdial}@cc.gatech.edu

ABSTRACT

Video and networking technologies have advanced such that posting and viewing video online is practical. Everyday people now post video online to communicate asynchronously with remote audiences. This paper explores the forms in which people communicate on the popular video sharing website YouTube. It also examines whether end-user video creators on YouTube use plot-based storytelling as a communication strategy. We analyzed popular content on YouTube and found the majority of that content showcases everyday people engaging in uncommon activities. Furthermore, a small minority of popular video actually tells a story. Based on our findings, we propose the *compostion gap* as a means of conceptualizing the disparity between video content on YouTube and professional content. We then discuss opportunities for designing technologies to support communication through performance-based video as well as story-based video.

ACM Classification Keywords

H.5.1Information Interfaces and Presentation (e.g., HCI) Multimedia Information Systems[video, evaluation/ methodology]

Author Keywords

user-created content, content analysis, composition gap, YouTube, remote asynchronous communication

INTRODUCTION

In recent years, posting and viewing video on the Internet has become not only practical, but a popular culture phenomenon. Continual increases in network bandwidth and processing power paired with advances in video capture and compression technologies have democratized the video authoring and distribution processes. While video production and distribution was once limited to those with the technical competence and financial means, tools for video editing and distribution are now available to the masses. As a result, video sharing websites (*e.g.,* YouTube [36]) have emerged as the primary means of publishing end-user created video.

Receiving over 100 million views per day and over 65,000 uploads per day (as of November 13th, 2006), YouTube is the most popular video distribution mechanism among end-user video authors [37]. YouTube facilitates asynchronous communication between end-user video creators and a global audience. This research aimed to determine how users of YouTube communicate with their audiences using video. Specifically, we were interested in the communicative and technical methods end-user video creators employ to engage in dialog with the YouTube community.

By developing an understanding of the current practices of end-user video creators, we can uncover opportunities for CSCW researchers to better support communication through video in online communities. Accordingly, this paper examines: 1) the genres and constituent elements (*e.g.,* soundtrack, storyline, etc.) of YouTube video, and 2) the extent to which end-user video creators use a particular form of communication, plot-based storytelling. The first question provides insight into end-user video creation practices. This insight can serve as a foundation for designers of video authoring tools.

Plot-based storytelling pervades the history of communication as well as popular mass media (*e.g.,* movies and television commercials). As a result, the second question explores whether an established and effective approach to communication (*i.e.,* plot) has been adopted by end-user video creators. It is clear from YouTube that creators can record and upload video. The objective of the second question was to determine if they take full advantage of the communicative power of the medium.

To evaluate these research questions, we analyzed a sample of video to develop a snapshot of the types and properties of videos hosted by YouTube. Our analysis categorizes the sample into genres and identifies the elements present in the video including whether plot was used as a communication strategy.

One view suggests that end user video creators are largely engaging in voyeurism, and that storytelling with visual media (*e.g.,* photographs and video) by everyday people is not occurring on its own. In a piece on digital moviemaking, the *Wall Street Journal* solicited the opinion of George Lucas, the creator of *Star Wars* regarding Internet video [22]. Lucas answered by describing the difference between circus and art. He equates circus to voyeurism and suggests "you don't have to write anything, you don't have to do anything, you just sort of watch it happen and it's interesting." In con-

trast, he considers art to require the telling of a story and "hopefully that story reveals the truth behind the facts." Lucas places the state of Internet video, particularly the content on YouTube, in the former category.

If Lucas is correct, storytelling is not taking place on YouTube despite the role storytelling plays in our culture as a foundation of communication across place and time. From a cursory glance it seems that George Lucas is correct (see the video of one band performing their music while engaging in a synchronized dance routine on treadmills[1] or Tay Zonday's rendition of *Chocolate Rain*[2] as examples).

In this research we systematically studied YouTube content and found an overwhelming amount of the video on YouTube is voyeuristic. The videos often resembled content broadcast on television shows based on content submitted by viewers (*e.g.,* car chases or embarrassing moments). We also found that end-user video creators include many of the technical aspects of video production (*e.g.,* sound effects, credits, etc.), but commonly do not use plot as a means of communication. Based on our findings we define a research agenda based on supporting communication through end-user created video.

In the next section, we situate this work within the larger context of end-user content creation. We continue in the following section by describing our study design after which we present our findings. Based on those findings, we define the *composition gap* — a concept describing the state of remote asynchronous communication with video. Finally, we reflect on potential contributions CSCW researchers can make to minimize the composition gap and conclude with future plans for this work.

RELATED WORK
YouTube is a fertile ground for learning about the behaviors of end-user content creators. Halvey *et al.* present an analysis of YouTube content to determine whether online video search behavior mirrors online text search with the goal of understanding whether knowledge of text search can transfer to video search [12]. Cha *et al.* present a statistical analysis of video popularity over time among other quantitative evaluations of user behavior [3]. In this work we choose to examine the nature of the content itself versus the search behaviors of users to determine what people post, and how they communicate with video in an online community.

Communicating with Personal Media
A number of CSCW researchers have explored communication (particularly storytelling) in the domain of digital photos. Frohlich *et al.* highlighted the importance of exploring the time/space matrix where photo sharing is concerned [10]. We extend this work in communication to digital video. By exploring current practices of end-user video creators, we can discover how CSCW researchers can leverage technology to better support creators. As a result, we analyzed the videos posted on YouTube to explore communication in the

remote, asynchronous case. In particular we were interested in the elements that constitute end-user video content and what creators choose to communicate about.

Crabtree *et al.* in their study of photoware consider storytelling an "essential, intimate feature of photograph use" [6]. They found that sharing is possible only when users can successfully navigate the activities required to produce a final artifact. These necessary activities include the production of accounts (*i.e.* storytelling). As a result, they implore researchers to "augment and support the production of accounts, the telling of stories, and thus support the conveyance of experience." Though this work focused on photograph use, we believe these findings apply to video, and in part motivated our interest in storytelling.

Kirk *et al.* defined videowork as a workflow for user interaction with video to expose the behaviors of end-users regarding video [14]. Research in the domain of end-user video typically focuses on supporting interactions in the workflow up to but not including sharing. Traditionally, a focus on developing solutions for better browsing [1, 20, 31, 32], search [26, 15], summarization [35, 33], and analysis [13, 9] of media content has prevailed as the means of providing users with support in the media production process. We chose to explore what people actually share to gain insight into what support (if any) could be provided.

A number of researchers have explored supporting the authoring process. Adams *et al.* developed a system to help creators produce artifacts with professional quality [2] . The system was based on a professional workflow adapted for amateurs. Riedl *et al.* sought to decrease the difficulty of authoring machinima by providing intelligent support tools for plot construction and 3-D scene generation [25]. An underlying theme in these and similar works is the desire to raise the quality of end-user productions to a professional level. While this is a reasonable and interesting goal, we believe that the target of end-user creators is not always a professional outcome. We complement this approach by exploring the nature of content end-user video creators share. Studying current practice allows us to identify opportunities for designing technologies that support communication through video in a group setting. Shaw and Schmitz took steps in this direction by deploying an authoring system which supported video content remixing [29].

The Power of Story
In the context of CSCW, storytelling is a useful means for remote asynchronous communication especially where digital media is concerned. Frohlich *et al.* documented "reminiscing talk" — the act of recounting an experience with people who were present — as a typical and natural process for telling stories using photographs [11]. Reminiscing talk tends to occur between people who share the experience documented by photographs. It also tends to occur around various types of artifacts such as video, scrapbooks and family mementos. However, telling stories using physical or digital artifacts to people who do not share the experience is more difficult, particularly when the person is not co-located

---

[1] http://www.youtube.com/watch?v=pv5zWaTEVkI
[2] http://www.youtube.com/watch?v=EwTZ2xpQwpA

and synchronous communication is not an option. In either case, the artifacts themselves are embedded in the context of storytelling and are an essential part of the communication process.

Storytelling also has impact beyond effective communication. In addition to relaying experiences to others not present, storytelling encourages introspection, affording storytellers the opportunity to derive meaning from their experiences [21]. Davis documented his use of storytelling to help urban youth develop a sense of identity [7]. Pennebaker showed that writing about emotionally traumatic experiences has a positive impact on recovery [24]. In the oral tradition, stories are used to teach life lessons and pass family history down through generations. Ellis and Bruckman constructed an online community to support the creation and sharing of oral histories [8]. Similarly, Shen *et al.* constructed an interactive tabletop display for constructing and sharing group histories [30].

Given the power of storytelling, we endeavored to determine the extent to which it occurs. A lack of storytelling could expose opportunities for encouraging the use of story to communicate with the YouTube community. Storytelling with video is fraught with challenges however. Landry and Guzdial described the challenges to story construction using digital media (including video) [19]. These include story development, process support and access to feedback. The difficulties introduced by this process hamper attempts to overcome distance and inhibits the formation of networked communities [23] based around online storytelling. Few current technologies attempt to provide holistic support for storytelling in the digital medium. This work represents a step toward establishing whether storytelling occurs without intervention. Our results suggest that it does not, which indicates research in this direction is provident.

STUDYING AN END-USER VIDEO CREATOR COMMUNITY

We studied YouTube content to ascertain the types of videos end-user video creators post and the elements that characterize those videos. We also investigated whether plot-based storytelling has permeated the digital medium where end-user video creation is concerned.

We collected a representative sample of videos to perform our analysis. However, methodically choosing a sample was difficult due to limitations imposed by the site. Filtering YouTube video is limited to popularity, time and predefined categories. For such a study, one might consider sampling videos of most, least, and average popularity to evaluate the properties of YouTube video. However, the categorizations provided to YouTube users for browsing the site do not allow for selecting videos in this manner. As a result, we drew our sample from the content the community values (we define value according to the various means of registering opinions provided on the site). Sampling highly valued content produced a more manageable set of data, and provided a view into the type of content viewers find most satisfying. Measuring the presence of storytelling on YouTube by using popular content might seem illogical initially. However, story-

telling has been the cornerstone of not just human culture but popular culture. If end user video creators have embraced storytelling in the digital medium it is likely to show up in the video that has reached the pinnacle of popularity.

Method
We extracted the top 100 videos[3] from each YouTube ranking category to create a representative sample. The ranking categories included: Most Viewed, Top Rated, Most Discussed (text comments), Top Favorites, and Most Responded (video responses). Each category represents a method by which a user can express their interest in a video (*e.g.*, by choosing to discuss positively, negatively or not at all). Videos can then be explored by these categories across various time scales (*i.e.*, today, this week, this month and all time). We chose to look at the best videos over all time to base our analysis on the most temporally stable collection possible.

We began by combining the videos from each category (100 per category) which produced a set of 500 videos. We removed duplicates which left a set of 380 unique videos from which to draw our representative sample. From the set of 380 videos, we randomly selected 182 videos to serve as our representative sample. Because we were only interested in end-user created video, we then removed professional content from our sample. Professional content was identified by determining whether it airs in main stream media (*e.g.,* on broadcast and cable television channels). Examples of excluded videos include professional music videos and television shows. After removing professional content, our sample included 100 videos. The results are based on our analysis of these videos.

For each video we recorded the duration, number of views, comments, responses, times selected as a favorite, number of times it was rated, along with the average rating value. We also analyzed each video using two coding schemes: genre and component. During data preparation, similarities between videos emerged. We constructed a coding scheme to classify videos into respective genres by generating descriptive labels for each video. We combined the labels into larger categories until categories could no longer be combined without compromising the category's meaning. We then used the final labels to classify the videos in our sample into genres. Note the coding scheme was generated using a different set of video from our sample.

In addition to developing a set of criteria to classify the videos according to genre, we also developed a set of criteria to explore the characteristics of each video. Storytelling and video production are well understood activities and many resources explain how to engage in these activities effectively. The codes for the component scheme were selected by aggregating the guidance provided by resources on screenwriting, video production and storytelling.

Genre Coding Scheme
While preparing our sample for analysis, a set of themes common across the corpus surfaced. These themes are codified

---

[3]Videos were extracted August 1, 2007.

Table 1. Genre Coding Scheme

Code	Description
Activism/Outreach	Presents a topic and solicits participation from viewers either by taking action or expressing their viewpoint by responding with a video
Performance/Exhibition	Showcases the talent of a person or an object of interest for the audience to admire or experience
Instructional	Presents directions for completing a task
Experiment	Presents the results of an experiment

Table 2. Component Coding Scheme

Code	Description
Plot	Video includes an observable conflict, climax, and resolution
Emotional Content	Emotions are directly observable in the video
Soundtrack	Music is used in a supporting role
Sound Effects	Sounds are used in a supporting role
Voice	Voice is used to engage with the audience
Intro	A sequence is used to introduce the video
Outro	A sequence is used to conclude the video
Credits	Acknowledgement or promotion of people, software, organizations, etc. that appear in or contributed to the video

in our genre coding criteria and are summarized in Table 1. We analyzed our sample using the following codes:

- *Activism/Outreach:* YouTube hosts a community of people who interact to learn about one another and engage in conversations on various topics. Creators of this type of video typically implored viewers to perform activities related to a topic. They also requested video responses related to a topic (*e.g.*, one creator asks the YouTube community to remember United States soldiers in the line of duty).

- *Instructional:* Videos classified as instructional presented directions for engaging in a particular activity (*e.g.,* table etiquette).

- *Performance/Exhibition:* This genre describes videos that showcased the talent of a person in music, storytelling, dance, martial arts and other abilities. It also included videos that presented an object or possession for the rest of the YouTube community to experience (*e.g.*, footage of a dog playing with a ball).

- *Experiment:* Videos included in this genre featured people performing experiments (*e.g.,* filming the result of mixing Mentos candies with Diet Coke). The video creator had to explicitly outline and perform the steps in the experiment for it to fit into this category.

Component Coding Criteria
The component codes were extracted from a selection of diverse resources. This set of criteria allowed us to explore the technical as well as the communicative aspects of each video. Writings on video production [5], screenwriting [34],

and digital storytelling [16] offer similar advice about communicating through digital media. We capture this advice in the following criteria:

- *Plot*: Professionals typically recommend developing a strong sense of purpose for telling a story. This is typically accomplished by developing a storyline or plot around which to base the editing of footage. Though the plot is not an absolute and unchanging construct, it is strongly suggested as a guide for the production process. We determined a video to have a plot if it included an observable conflict, climax, and resolution.

- *Emotional Content*: Lambert [16] suggests the use of emotion as a means to keep the audience consistently engaged. Determining the emotions the author intended to evoke within potential viewers could only be ascertained by interviewing the actual author of the video. In light of our desire to objectively inspect a considerable number of videos, we decided to determine whether there was any observable emotion with which users could connect. For example, a video was classified as having emotional content if we could observe behaviors that represented emotion (*e.g.* someone smiling could be connected with happiness). While we cannot say whether the observed emotion was the intent of the author, we can say emotions are being displayed with which an audience could potentially identify.

- *Soundtrack*: Music is often used to set the tone of particular segments of a media production. We coded each video for the presence of a soundtrack used in a supporting role. Many of the videos we viewed used music as a primary feature of the video (*e.g.*, playing an instrument

or dancing to a song). We were explicitly concerned with the extent to which creators used soundtracks in the manner advised by the resources we consulted. Videos which used music solely as a primary feature were not coded as having a soundtrack.

- *Sound Effects*: Sounds are often used to emphasize a particular moment (*e.g.* the sound of glass breaking or footsteps). Sound effects were considered present if they were used in a supporting role.

- *Voice*: In film, characters commonly communicate through speech. Lambert [16] suggests the use of voice to connect with an audience. We coded videos for voice based on whether any subject in the video used voice to address the viewer or others in the video. We included communication through song, voiceover, and directly speaking to the audience.

- *Intro & Outro*: Introductory and closing sequences are often used in many genres of television and film to transition the user into or out of a viewing experience. We coded each video for the absence or presence of an intro. Each video was also coded for the absence or presence of an outro.

- *Credits*: While not quite as important as other aspects (*e.g.,* plot) of a production, credits acknowledge contributors to the media production. We coded each video for the absence or presence of credits.

We coded all 100 videos using a total of 12 codes (See Tables 1 and 2). Each video was coded according to the presence or absence of the property represented by the particular code. To ensure that our coding schemes could be consistently applied by people outside those conducting the research, we established the reliability of the coding schema.

Establishing Inter-Rater Agreement
We performed our content analysis using multiple raters to ensure our coding schema could be consistently applied to our dataset. To establish inter-rater agreement we enlisted two coders to assist one author of this paper. The three coders met initially to review the proposed analysis criteria and discuss how the codes would be applied to a set of sample videos. The outcome of this meeting was the set of criteria presented above. A set of 25 videos (excluding professional content) were then randomly selected and provided to the raters along with the rating criteria. Each rater independently rated each video in the sample.

To determine the agreement between the three raters, we computed Cohen's Kappa (κ) statistic [4]. For the component scheme, a total of 200 observations per rater were made (25 videos $\times$ 8 codes). For the genre scheme, a total of 100 observations were made (25 videos $\times$ 4 codes). The agreement between the raters is presented in Table 3. According to Landis and Koch [17], 0.70 or more is considered an acceptable κ value. With our inter-rater values at acceptable levels we proceeded with coding our sample.

Table 3. Agreement κ values

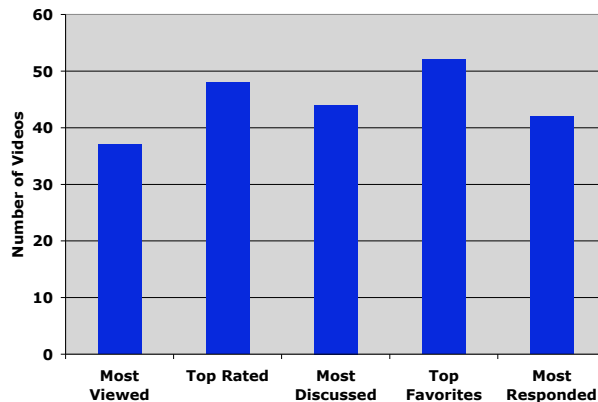| Pair	Professional	Genre
Rater1 * Rater2	0.89	0.75
Rater1 * Rater3	0.82	0.74
Rater2 * Rater3	0.75	0.75

Figure 1. Sample Distribution by YouTube Category

RESULTS
In addition to performing a content analysis on the YouTube videos in our sample, we also compiled descriptive statistics using the information we collected about each video. Approximately 6 hours of videos were viewed and coded for the properties listed in Tables 1 and 2. The sample ($n = 100$) was viewed by YouTube visitors 388,209,408 times, rated 1,611,588 times and commented on 1,237,119 times. Figure 1 depicts the distribution of the videos we sampled across the categories on YouTube.

The average user rating of our sample was 3.59 out of 4 (σ = 0.7398) which translates to a qualitative rating between "pretty cool" and "awesome" on YouTube's rating scale. The average video length was 3.71 minutes with σ = 2.81 minutes and $\tilde{x}$ = 3.36 minutes. See Figure 2 for the distribution
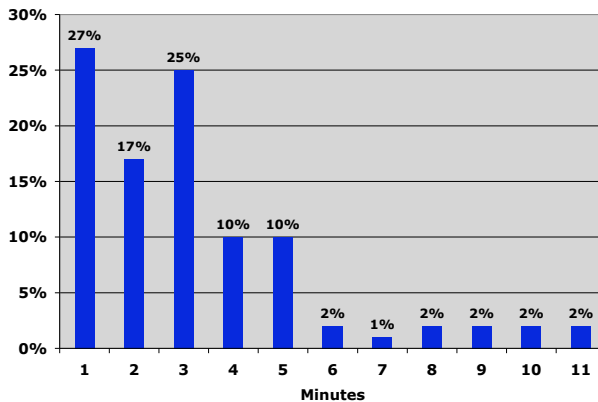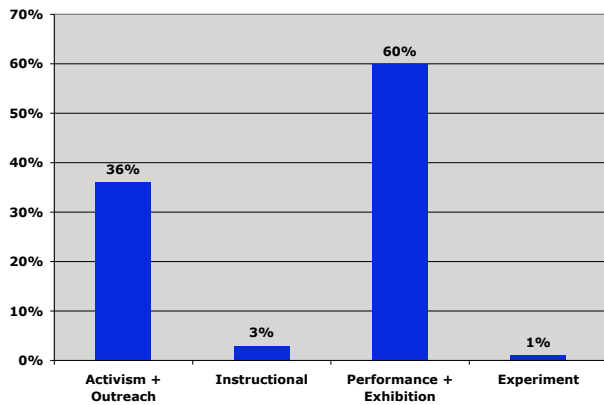

Figure 2. Distribution of Video Lengths

Figure 3. Videos by Genre Code

Figure 4. Distribution of YouTube Categories by Video Genres

of video lengths in the sample. Although creators can upload up to 10 minutes of video (with a file size restriction of 100MB), the majority (89%) of the sample remained under 6 minutes. This potentially suggests a threshold for the length of videos users are willing to create and watch (at least for the type of video found on YouTube).

Genre Coding Results
Recall we classified the videos in our sample using the coding scheme that emerged during data preparation. These results enumerate the frequency with which users post video from a particular genre. Each video was classified into only one category.

The Performance/Exhibition genre encompassed 60% of the videos while the remaining 40% was distributed across the other categories (See Figure 3). The Performance/Exhibition genre is reminiscent of mainstream variety television shows that showcase everyday people performing not so everyday acts. In particular it parallels the content on the American television program *America's Funniest Home Videos*. Videos in the Performance/Exhibition genre included a posting of a puppet show, someone playing the Super Mario Bros. video game theme song on a flute while beat-boxing, a video created to the musical theme of Mortal Kombat and a video of the singing group OkGo performing a synchronized dance routine on treadmills. The amount of videos that represent this category not only underscores the popularity and prevalence of performance pieces, but suggests end-user content creators will likely continue to create performance pieces. This genre was the only one to appear in all five YouTube categories (See Figure 4) indicating the value of this type of video in the community.

Activism/Outreach videos accounted for 36% of the sample. Examples included a recording of a campaign to distribute free hugs, an appeal to collect the worst videos ever made by YouTube video creators, and an invitation to users of the site to expound on why they participate in the community. This genre describes the second-largest portion of the videos in our sample and solicited the highest number (68.42%) of responses and second-highest number of comments (29%).
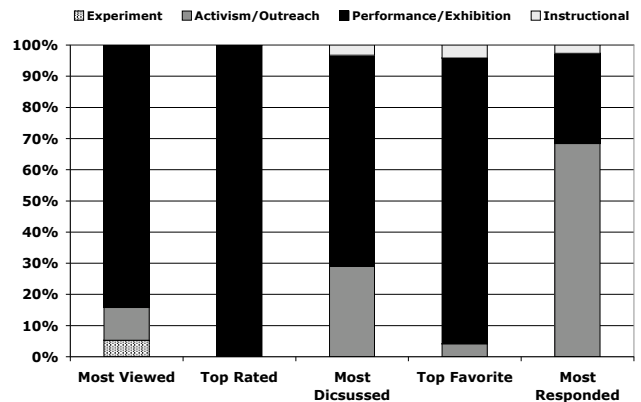
This is likely due to the tendency of Activism/Outreach videos to implore viewers to engage in discussion or reply with a video of their own. Videos in this genre were not selected as a favorite (4%) or viewed (10%) frequently.

Instructional videos constituted 3% of the sample. An instance of this genre provides rules of etiquette for using the men's room. When inspected by individual YouTube categories, instructional videos appeared only in the Most Discussed, Top Favorites, and Most Responded categories. The videos were selected as a favorite (4%), discussed (3%), and responded to (2%) though not in great proportion. This finding suggests the content is valuable and generates discussion despite not being viewed or highly rated.

Examining the absence of particular genre codes across the YouTube categories unearths some interesting findings. Videos coded for Activism/Outreach were not highly rated though they accumulated many responses (68%). They did elicit some discussion (29%) and views (10%) but not an overwhelming amount. Strikingly, no genre outside of Performance/ Exhibition received high ratings. All of the videos in the sample taken from the Top Rated category were classified into the Performance/Exhibition genre. This finding further underscores the value of this type of video to the YouTube community. Overall, it seems that particular types of videos evoke particular types of responses from the YouTube community. For example, Activism/Outreach videos were discussed the most, while Performance/Exhibition videos were viewed the most.

Component Coding Results
The results of applying our component schema reveal the frequency with which end-user video authors use communicative and technical elements in their videos. Additionally, our findings provide insight into our question of whether end-user video creators are employing plot-based storytelling in their videos. The results are visually described in Figure 5.

The most prominent result is that 90% of the videos were not plot-driven by our definition. One interpretation of this
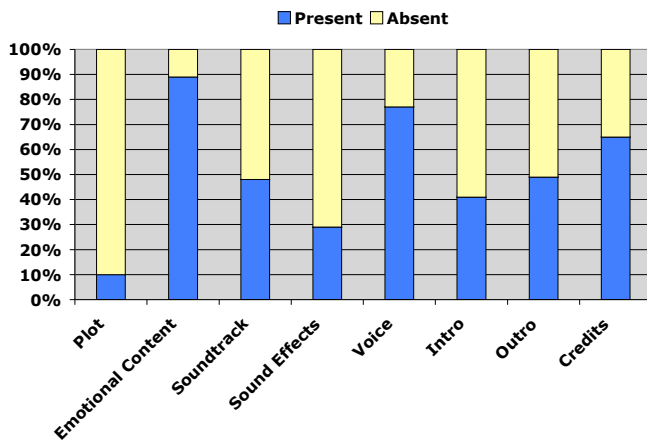
Figure 5. Videos by Component Code

result might claim the community simply does not view YouTube as the place to publish plot-based artifacts. Given the popularity of YouTube and the fact that even major broadcasting companies post their content to YouTube, we suspected this was not the case. Ten percent of the videos in our sample incorporated plot and were also valued enough to reach "top content" status. We interpret this result to mean viewers value content from creators that use plot as a communication technique. However, there may not be a significant amount of that type of content on YouTube. Another possibility is that plot-based content may not be executed well. Poorly executed plot-based content, or a lack thereof, can be explained by the numerous barriers to composition of complex artifacts containing plots. Certainly we do not think that everyone desires to create plot-based video. However, even for those that are appropriately motivated, Landry and Guzdial report that remote asynchronous communication through plot-based video is laden with challenges [19]. It seems a considerable opportunity for designing technologies to support the creation of plot-based video exists. We return to this in the Discussion section.

Though a sizable portion of our sample did not contain a plot, 89% did contain emotional content. This is not necessarily surprising given the fact that a significant part of what makes us human is our propensity for displaying and communicating through emotion. Emotion could be leveraged to support the development of plot-based video by end-user video creators. Designing user experiences that assist users with moving from emotional exhibitions to engaging stories seems a fruitful endeavor. Research in supporting the video practices of users typically focus on organization [1] and other video workflow activities [14]. Traditionally there has been less focus on supporting the creation process for sharing video artifacts (particularly with a remote audience). We suggest that eliciting emotion can serve as the bridge from supporting creators with organization to supporting them with composition.

Less than half of the videos (48%) contained a soundtrack, while 29% percent of videos contained sound effects. While

a soundtrack or sound effects may not be appropriate in some genres, they are certainly common in professional content across many genres. Sound potentially enhances the experience of viewers. At the very least a soundtrack can be used as part of introductory and concluding sequences of the video. Sound effects could be used to add emphasis to particular moments in the video. These findings suggest end-users do use sound (though not the majority) to support the experience they present to viewers. Lambert suggests sound be used to enhance, not overpower, the story [16]. While we did not assess how effectively sound was used in our sample, further study could elicit opportunities to support effective soundtrack and sound effect use.

Videos in our sample tended to thrust the viewer directly into the action. An intro sequence was present in 41% of the videos and and an outro sequence was used in 49% of videos. While inserting a viewer into the middle of the action can be a useful technique, it can also be a nuisance if not effectively executed.

Television shows, movies and other media forms use credits to acknowledge everyone who contributed to the final outcome. Today, authorship is much more difficult to define, especially in today's remix culture where copyrighted media assets are easy to procure and appropriate for different purposes. However, acknowledgment seems relatively common with 65% of the videos coded using credits to recognize contributors as well as technical tools used to produce the video.

DISCUSSION: THE COMPOSITION GAP

In an ACM Special Interest Group on Multimedia (SIGMM) report, leaders of the community present an assessment of the current state of the field and suggest new research directions [27]. One of the grand challenges posed to the community is "to make authoring complex multimedia titles as easy as using a word processor or drawing program." This is also an apropos goal for the CSCW community. A significant amount of group communication technologies now exist solely online (*e.g.,* blogs). These technologies facilitate asynchronous communication between remote, distributed groups of people. Online video websites support communication in this manner as well, though using a different modality (*i.e.,* video versus text). Instead of discussing work, creators engage in a number of different social or politically motivated conversations through video.

The proliferation of digital video cameras paired with the emergence of online communities devoted to sharing media has led to the democratization of authoring and publication of video content. However, currently available tools typically provide users with a means to create video, while neglecting the fact that many people have little video production experience. Furthermore, Lambert [16] found people generally give up in the video process. Addressing the barriers to video production can thereby impact the type and level of communication that takes place in online video communities.

We consider it prudent to think about how technology could

scaffold the video production process for end-user video creators. Efforts in this direction could make authoring engaging video as easy as authoring a letter. Again, supporting authoring could expand the ways in which people can harness video to communicate. A reasonable first step is examining the video artifacts end-users currently create. Our work has taken this step by beginning to characterize the types of video YouTube authors create and the elements the videos include. We continue the discussion by exploring the disparity between end-user creator practice and professional practice, as codified in the help resources we used to develop our component criteria. By analyzing this disparity, we can discover occasions when technology might facilitate more expressive uses of video. We then propose possibilities for minimizing the disparity.

Defining the Gap

Expert advice is useful because it is concrete and proven to be effective. Based on our findings, it seems a considerable disparity exists between the advice (and thus the properties of the content experts would produce) and the video end-user video creators produce. We define this disparity as the *composition gap*.

The composition gap defines two concerns: vision and technical skill. Vision relates to Lucas' observation of a lack of storytelling in Internet video; however, it extends beyond storytelling. Vision encompasses the activities required to craft the message a video is intended to communicate. Rather than simply describing storytelling, vision describes the process of forming an intent and selecting a structure for communicating that intent.

Technical skill refers to the ability required to execute the vision effectively. In any creative domain, artistic vision as well as technical skill is required. For example, a painter must make artistic decisions while also skillfully applying paint to the canvas. A writer must determine why and what to communicate while producing comprehensible text. Similarly, creating a video production necessitates vision and technical skill. End-user video creators must consider the purpose of their video and how to best communicate that purpose. In addition, they must also acquire the skills to create the video and apply those skills effectively.

The two concerns of the composition gap (*i.e.* vision and technical skill) present potential barriers to people who wish to author videos that tell a story. Historically, research has focused on supporting video browsing, search, and similar concerns as a means of supporting end-user media production. These contributions have decreased the technical skills needed by creators. Likewise, commercial applications (*e.g.*, iMovie[4]) provide a simple, yet feature-rich experience.

These technological accomplishments have helped decrease the composition gap where technical skill is concerned by making media composition a technically and financially accessible process. Furthermore, Kirk *et. al* suggest the technical quality of media content is not a major concern for pro-

---

[4]iMovie. http://www.apple.com/ilife/imovie/

ducers and viewers [14]. However, progress toward decreasing the gap also depends on the video creator's ability to develop and refine a vision for a video project. We posit research directed at supporting creators with articulating their visions will make considerable strides towards minimizing the composition gap.

Minimizing the Gap

We now revisit our findings to further define the composition gap and suggest avenues of research for progressing toward minimizing the gap. We will focus more on the vision aspect because we believe the most progress can be made in this area.

Vision

The ACM SIGMM retreat report [27] concedes that "creating multimedia content and using it in everyday applications (*e.g.*, email, documents, web titles, presentations, etc.) is still not possible for most users." We believe users are capable of creating video, but as of yet do not leverage the communicative power of video. For example, the videos in our sample tended to fall into the Performance/Exhibition genre (68%), and typically do not contain plot (80%). More surprisingly, every video sampled from the "Top Rated" YouTube category was classified into the Performance/Exhibition genre. George Lucas' assertion that Internet video lacks story seems to hold true. Though a majority of the videos could be considered "circus," it is not necessarily the case that they are not entertaining. The number of videos that fall into the Performance/Exhibition category not only underscores the popularity and prevalence of performance pieces, but emphasizes the likelihood that end-user video creators will continue to create performance pieces. This genre was the only one to appear in all five YouTube categories (See Figure 4). Clearly people find value in the videos based on the amount of views, comments, ratings and responses they receive.

As a result, the vision aspect of the composition gap must encompass more than just plot-based storytelling. It must describe and address the process of determining the desired audience experience. It may be the user's goal to tell a well crafted story or to simply present their favorite card trick. In either case, articulating the point of the video before engaging in the authoring process is useful. For the former (and arguably more complicated case), using sensemaking as a lens could offer further clarity. Russell *et al.* [28] define sensemaking as "the process of searching for a representation and encoding data in that representation to answer task-specific questions." Sensemaking adapted to user video content creation refers to the exploration of digital collections and extraction of meaning from experiences in the world to develop an understanding of what those experiences mean in the larger context of life. This understanding can then be conveyed through media (in this case video). Based on these observations, we propose a few different avenues for research.

The vision portion of the gap could be reduced by technologies that summarize content according to meaning (*e.g.*, emotion or reflections) rather than contextual or temporal meta-

data. Viewing content summaries by theme (*e.g.,* family or success) could assist users with establishing a vision for their video productions. It could also provide a new way of browsing an ever-growing media collection. A loftier objective would entail generating summaries of video and image collections that actually communicate a particular message. For example, with some help from the user, a tool might generate a summary of a family's summer vacation while expressing the importance of family.

We found few of the videos in our sample implement plot or some identifiable higher-order organizing strategy to create an engaging experience for the viewer. Another approach to minimizing the vision portion of the gap could involve providing users with wizard-like, suggestive interfaces. These interfaces would guide the user through the process of establishing the purpose of their video. Research in this direction has used question prompts to invite users to think about a strategy for communicating with their audiences [18].

In addition, technologies could provide potential organization strategies as a starting point. For example, formulaic templates could serve as the basis for creating videos the community clearly values. One type of video included in the Performance/Exhibition genre was the lip-synched music video. The user would select a popular song and act out a video while lip singing to the soundtrack. A music video template could be provided for users who desire to create a music video for their favorite song. The template would define the length of the video based on the chosen song, set up the song in a timeline and provide directions on how to execute the video. This would decrease the effort required by the user to establish a vision and shift the process to a more technical one.

Technical Skill
Soundtracks (48%), sound effects (29%), intros (41%), outros (49%) and credits (65%) were included in more videos in our sample than we anticipated. Though these elements were not used in the majority of the videos, some end-user video creators implement many of the technical elements of professional video productions. Effectively using technical elements (*e.g.,* a soundtrack or sound effects) require time and effort. Ineffective use of such elements can often detract more from a video than they add. Developing interfaces and experiences that reduce time and effort required is desirable. Interfaces that provide advice on how to effectively use sound and other techniques could provide benefits as well. Although the use of technical skills by no means occurs in the majority of our sample, we expected a smaller amount of users to use audio and visual effects in supporting roles. This suggests that the technical portion of the composition gap may not be as disparate as we suspected. On the other hand, when considering our sample comprises the "best" of YouTube, the gap is likely to be much bigger when considering videos that never make it to the pinnacle of popularity.

In summary, we conjecture that the greatest opportunity for closing the composition gap is in introducing technologies that encourage and help creators to develop productions which employ a presentation structure (*e.g.,* a plot). Obviously, applying a plot to every type of video described by our genre coding scheme is neither comprehensive nor necessarily appropriate. However, the notion of defining a higher level understanding of the intended audience experience seems a worthwhile pursuit for YouTube video creators. We believe urging users to consider the purpose of the videos they plan to create will help close the gap between novice and professional practice. Other opportunities for closing the composition gap include helping users effectively use titles, credits, soundtracks, and sound effects to further enhance the experiences of their viewers.

CONCLUSION
We have presented an analysis of YouTube content, which classifies the type and properties of the site's most popular videos. Ultimately we were interested in determining what end-user video creators communicate about, and identifying the constituents of the videos. We also explored the extent to which YouTube authors use plot-based storytelling as a communication strategy. Based on our findings, we presented a series of opportunities to support and enhance end-user video creators regarding vision and technical implementation.

We realize this study by no means classifies all of the content on YouTube. However, we do believe we have uncovered some of the practices of end-user content creators, which contributes to our understanding of communication through video on video sharing websites. It also contributes to our understanding of how end-user video creators differ from professionals and where opportunities exist for supporting creators who are motivated to produce more advanced videos. As future work, we plan to shift from analyzing end-user video artifacts to interviewing actual creators. We hope to learn more about their authorial intents and desires for support by technology.

REFERENCES
1. G. D. Abowd, M. Gauger, and A. Lachenmann. The family video archive: an annotation and browsing environment for home movies. In *Proc. MIR 2003*, 1–8. ACM Press, 2003.

2. B. Adams, S. Venkatesh, and R. Jain. Imce: Integrated media creation environment. In *ACM Transactions on Multimedia Computing, Communications and Applications*, volume 1. ACM Press, 2005.

3. M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system. In *Proc. IMC 2007*, 1–14. ACM Press, 2007.

4. J. Cohen. A coefficient of agreement for nominal scales. *Educational and Pyschological Measurement*, 20(1):37–46, 1960.

5. M. D. Collier. *The ifilm Digital Video Filmmaker's Handbook*. Lone Eagle Publishing Company, Hollywood, 2001.

6. A. Crabtree, T. Rodden, and J. Mariani. Collaborating around collections: Informing the continued development of photoware. In *Proc. CSCW 2004*, 396–405. ACM Press, 2004.

7. A. Davis. Co-authoring identity: Digital storytelling in an urban middle school. *THEN*, 1(1), 2004.

8. J. B. Ellis and A. S. Bruckman. Designing palaver tree online: supporting social roles in a community of oral history. In *Proc. CHI 2001*, 474–481. ACM Press, 2001.

9. J. Fan, Y. Gao, and H. Luo. Multi-level annotation of natural scenes using dominant image components and semantic concepts. In *Proc. MULTIMEDIA 2004*, 540–547. ACM Press, 2004.

10. D. Frohlich, G. Adams, and E. Tallyn. Augmenting photographs with audio. *Personal Ubiquitous Comput.*, 4(4):205–208, 2000.

11. D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss. Requirements for photoware. In *Proc CSCW 2002*, 166–175, New Orleans, Louisiana, USA, 2002. ACM Press.

12. M. J. Halvey and M. T. Keane. Analysis of online video search and sharing. In *Proc. HT '07*, 217–226. ACM Press, 2007.

13. H.-B. Kang. Affective content detection using hmms. In *Proc. MULTIMEDIA 2003*, 259–262. ACM Press, 2003.

14. D. Kirk, A. Sellen, R. Harper, and K. Wood. Understanding videowork. In *Proc. CHI 2007*, 61–70. ACM Press, 2007.

15. A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gwizdka. Fotofile: a consumer multimedia organization and retrieval system. In *Proc. CHI 1999*, 496–503. ACM Press, 1999.

16. J. Lambert. *Digital Storytelling: Capturing Lives, Creating Community*. Digital Diner Press, Berkeley, 2002.

17. J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data, 1977.

18. B. M. Landry and M. Guzdial. iTell: Supporting retrospective storytelling with digital photos. In *Proc. DIS 2006*, 160–168. ACM Press, 2006.

19. B. M. Landry and M. Guzdial. Learning from Human Support: Informing the Design of Personal Digital Story-Authoring Tools. *Journal of the International Digital Media and Arts Association*, 3(1):106–119, 2006.

20. F. C. Li, A. Gupta, E. Sanocki, L. He, and Y. Rui. Browsing digital video. In *Proc. CHI 2000*, 169–176. ACM Press, 2000.

21. K. Madej. Towards digital narrative for children: from education to entertainment, a historical perspective. *Computing Entertainment*, 1(1):1–17, 2003.

22. W. S. Mossbert and K. Swisher. All things digital. *Wall Street Journal*, June 18, 2007.

23. E. D. Mynatt, A. Adler, M. Ito, and V. L. O'Day. Design for network communities. In *Proc. CHI 1997*, 210–217. ACM Press, 1997.

24. J. W. Pennebaker. Telling stories: The health benefits of narrative. *Literature and Medicine*, 19(1):3–18, 2000.

25. M. O. Riedl, J. P. Rowe, and D. K. Elson. Toward intelligent support of authoring machinima media content: Story and visualization. In *INTETAIN '08*, 2008.

26. K. Rodden and K. R. Wood. How do people manage their digital photographs? In *Proc. CHI 2003*, 409–416. ACM Press, 2003.

27. L. A. Rowe and R. Jain. Acm sigmm retreat report on future directions in multimedia research. *ACM Trans. Multimedia Comput. Commun. Appl.*, 1(1):3–13, 2005.

28. D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proc. CHI 1993*, 269–276. ACM Press, 1993.

29. R. Shaw and P. Schmitz. Community annotation and remix: a research platform and pilot deployment. In *HCM '06*, 89–98. ACM Press, 2006.

30. C. Shen, N. B. Lesh, F. Vernier, C. Forlines, and J. Frost. Sharing and building digital group histories. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 324–333. ACM Press, 2002.

31. F. Shipman, A. Girgensohn, and L. Wilcox. Hypervideo expression: experiences with hyper-hitchcock. In *Proc. HYPERTEXT 2005*, 217–226, 2005.

32. J. R. Smith and S.-F. Chang. Visually searching the web for content. *IEEE Multimedia Magazine*, 4(3):12–20, 1997.

33. H. Sundaram, L. Xie, and S.-F. Chang. A utility framework for the automatic generation of audio-visual skims. In *Proc. MULTIMEDIA 2002*, 189–198. ACM Press, 2002.

34. D. Trottier. *The Screenwriter's Bible*. Silman-James Press, Beverly Hills, CA, 1995.

35. S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky. Video manga: generating semantically meaningful video summaries. In *Proc. MULTIMEDIA 1999*, 383–392. ACM Press, 1999.

36. YouTube. `http://www.youtube.com`.

37. YouTube. Google closes acquisition of youtube. `http://www.youtube.com/press_room_entry?entry=AwPf9c9qJDc`.