Malcolm Slaney
Yahoo! Research and Stanford CCRMA

Tell Me a Story

Aisling Kelliher
Arizona State University

Malcolm Slaney
Yahoo Research and Stanford CCRMA

"Daddy, tell me a story. A good one." An engineer pauses at the request. We know how to measure and build many things. But imagining, crafting, and telling a story? That is hard. If you are building or researching any aspect of multimedia, however, then you should be thinking about how your work helps both you and the user experience a story. Telling stories is an important part of how we communicate and make sense of our experiences. Even this column tells a story (hopefully nonfiction)—see the sidebar for a map of our "journey."

Tension in This Story

All content is best when it tells a story—even this column. As we mention in the main text, a child casually demanding a story might provoke anxiety in his engineer dad. Scrabbling for a solution, any of us might dig deep into cultural tradition, finding comforting familiar structures in long-established forms such as Homer's *Iliad*. In this article, we relate this experience to the hard challenge of creating movie trailers. Tension is then released as we discuss current sports summarization techniques and conclude on a note of optimism for the emergence of compelling computationally supported stories (see Figure A).

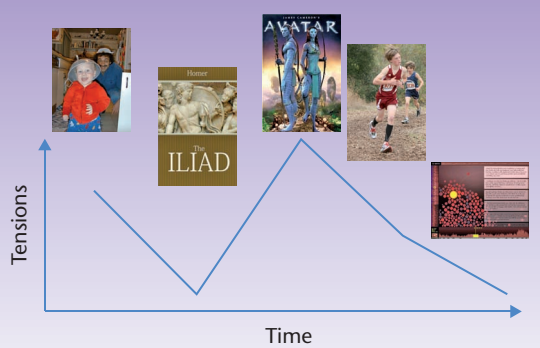

Figure A. Tension in this story over time. The third image is from an* Avatar *movie poster, by Twentieth Century Fox. The last image is from "The Dumpster," http://artport.whitney.org/commissions/thedumpster/about.html.

A previous column in this department talked about the importance of entertainment versus information in the multimedia world.[1] Most multimedia users are looking for entertainment, not information, so measures such as precision-recall are not necessarily relevant for multimedia. In this article, we wish to argue that storytelling is a much more integral aspect of multimedia than most of us consider.

Multimedia offers researchers developing tools, methods, and frameworks an intuitive and cross-cultural application space with a diverse set of motivated end users. As we progress from shoeboxes of Hi8 tapes to terabyte drives filled with rarely viewed media directories, the challenges continue to grow. What is truly significant in this collection? What meaningful patterns could emerge to help me shape my argument as I present my slideshow to the board? How best can I craft a heartfelt eulogy for a family member to share with relatives abroad? For now, such scenarios remain a largely human-driven endeavor, but the opportunity for stimulating, augmenting, and enhancing these efforts with computational support structures presents an intriguing possibility space.

What Is a Story?

Stories are a fundamental component of the human experience. Creating and sharing stories help us understand and relate who we are as individuals and as social beings. When considering stories, we might typically think of *Little Red Riding Hood* or the *Da Vinci Code*, leading us to ponder, what do these things have to do with me, a multimedia researcher? Perhaps it helps to expand our notion of stories beyond myths and fiction novels to a broader realm that includes problem solving, strategizing, summarizing, sensemaking, or just things that make us smile.

Storytelling can be a simple exchange of experiences. In sharing stories, we communicate events we find memorable and valuable.

1070-986X/12/$31.00 © 2012 IEEE

Published by the IEEE Computer Society

The practice of storytelling has always used available media and technologies—from the earliest cave drawings and stained glass windows to Polaroids and iPhones. The basic properties of a particular medium determine what type of stories can be told, how they are presented, why they are communicated, and how the audience experiences them.[2] In essence, we make our thoughts, experiments, and future plans visible by crafting them into coherent narrative structures over time. We then put these mediated stories to work in a variety of ways, depending if our goal is to persuade, comfort, entertain, terrorize, or inform the intended audience.

In general, multimedia is nothing without a story. The original storytellers—think Homer's *The Iliad*—memorized long stories, with plenty of emotional content, and delivered them in a highly entertaining (and possibly interactive) manner. Music, movies, and videogames are perhaps today's most popular forms of storytelling. The multimedia field is contributing to the evolution of these forms by summarizing long video sequences, discovering patterns in related content, or developing techniques for generating stories. The best question we can ask is, how can multimedia systems create better user experiences?

Storytelling is one important part of this process. The richest multimedia experiences are not created randomly—even fair or good experiences are not entirely accidental. Instead, there is a compositional arc. Muzak, the now-defunct company that provided environmental (elevator) music to stores and offices, was astute about the emotional arc. Their guiding principle for content delivery included varying the activity (emotional) level of the music over time. Playing only one kind of music is boring. Instead, the music's content, and thus the listener's emotional state, went up and down in a regular cycle within an hour and over the day.

There's demand for better ways to consume copious amounts of multimedia data. People want to find compelling content, subsample longer versions, and string together short content into an enjoyable experience. Enabling this is what we as multimedia researchers and engineers should strive to achieve. Many people are professional domain experts in this area. Think of movie and music producers, composers, DJs, and gallery and museum curators. What can we hope to achieve with machines?

Summarization

All of this is why today's automatic movie summarization systems are frustrating—there is a high bar. Movie trailers for popular movies are created by specialized companies for targeted audiences. Their goal is to advertise, the goal being to persuade, even tease, the viewer so they want to see (and pay for) the longer-form content. This is a level of sophistication that we cannot yet hope to mimic with machines. This is beyond AI-complete, and there is more than enough value to pay creative humans to do the task. Machines will struggle to do better than humans.

We often evaluate text summarization by measuring whether readers get the same information from the original and the short-form text. Readers are asked a series of questions and their responses are tabulated. The summarization is good if the answers from those that read the long and short versions are similar. This is purely an information-theoretic view of the problem. It is a valid measure, but not necessarily what people are looking for when they sit down to watch a video or listen to music. Much multimedia summarization today, by comparison, is primitive. A good system will look for significant events in the video clip and then string them together. One can imagine looking for the applause points or the word ''Goal!'' This creates a summary of the exciting points in the video, but it does not necessarily create a compelling and sophisticated story arc that leaves you craving more.

The opposite approach is necessary with groups of images. We could generate a slide show by ordering the images so similar images are close together or by sampling the most important images first.[3] Yet, neither of these approaches necessarily create engaging stories or address the viewer's emotional needs. A similar problem exists with video recommendations. Today's videos are often short clips, and the content provider wants to keep the user's attention by providing suggested content to view next. We can treat this as a content-recommendation problem, ala Netflix, or a related content problem like a search engine. Again, neither of these approaches measure or cater to the consumer's long-term pleasure.

This is where multimedia researchers should come in. It is difficult to determine users' emotional desires. How long should a video be? Are users looking for something to cheer them up,

inspire them, or to help them relax? While the story arc of the hero's journey[4] in *Star Wars* is mined from a rich narrative tradition stretching back to Homer's *Odyssey*, contemporary writers get paid big bucks to create new forms of stories to keep users entertained. And none of us really want to choose whether our story is better told as a Dickensian episodic reveal or a *Memento*-style nonlinear puzzle.

First Steps

Automated systems can put together summaries of sports stories. We could argue that these are primarily information templates, with a modicum of emotional sugar coating. Still, it will be hard for an automated system, without human data, to understand that the quarterback's play today is especially uplifting because his mother died last night. Or that the last three losses have left the home team particularly hungry for a win. Those are the details that attract new people and bring audiences back for more.

Personalization is one area where computers and multimedia researchers can succeed. As mentioned at one talk at the 2011 ACM Multimedia Conference (www.acmmm11.org), just stringing together all the clips that include the grandmother's progeny is likely to be a winning solution. But can we do something better for professional teams? Would some people pay to receive a summary at the end of the day showing their favorite players key plays on the baseball diamond in a historically contextualized way?

Music recommendation systems try to create a coherent stream of songs. Perhaps by finding related genres or tempos. Musicians lament the shuffle button because it throws out the careful thought they put into creating a linearly structured album. Still some people enjoy the serendipity of a random shuffle over all their music. Clearly, the artists believe there is value in the composed sequence of songs. Can we create a coherent stream of all a user's music?

We really don't know how to meaningfully get from a pile of content into a story arc. Detecting events[5] or faces is but a start. We need to find ways to empathetically cheat—give people something that satisfies their emotional needs without understanding all of human emotion. How can we do this?

How Do We Get There?

It's encouraging that you have read this far. We are not the first to say that stories and emotions are important to the multimedia field,[6] nor are we alone in this story-integration endeavor. Researchers in data visualization are beginning to explore new classes of visualizations that adopt a narrative approach,[7] start-ups such as Storify are embracing a curatorial approach to social media sharing, and large funding initiatives such as the Knight News Challenge are awarding grants to innovative future news technologies that digitally inform communities.

When you design and build your multimedia ideas, please ask yourself how your work will help users enjoy a story or connect with others. That is the magic of multimedia, and we shouldn't lose track of the power we have. For the storytellers and artists that have read this far, what are the baby steps that engineers can take to help users tell and consume better stories? What are the big mistakes that have been made? For the engineers, what is possible? For everybody, what are outstanding examples of multimedia storytelling? **MM**

References

1. M. Slaney, "Precision-Recall Is Wrong for Multimedia," *IEEE Multimedia,* vol. 18, no. 3, 2011, pp. 4–7.
2. M.L. Ryan, ed., *Narrative Across Media,* Univ. of Nebraska Press, 2004.
3. D.K. Mahajan et al., "A Classification Based Framework for Concept Summarization," *Proc. 20th Int'l Conf. Companion World Wide Web* (WWW), ACM Press, 2011, pp. 83–84.
4. J. Campbell, *The Hero with a Thousand Faces,* Princeton Univ. Press, 1949.
5. R. Jain, "EventWeb: Events and Experiences in Human Centered Computing," *Computer,* vol. 41, no. 2, 2008, pp. 42–50.
6. A. Satish, R. Jain, and A. Gupta, "Tolkien: An Event Based Storytelling System," *Proc. VLDB Endowment,* vol. 2, no. 2, 2009, pp. 1630–1633.
7. E. Segel and J. Heer, "Narrative Visualization: Telling Stories with Data," *IEEE Trans. Visualization and Computer Graphics,* vol. 16, no. 6, 2010, pp. 1139–1148.

Aisling Kelliher is an assistant professor of Media Communication Systems and Media Theory with the School of Arts, Media, and Engineering at Arizona State University. Contact her at aisling.kelliher@asu.edu.

Malcolm Slaney is a principal scientist at Yahoo Research and a consulting professor at the Stanford Center for Computer Research in Music and Acoustics. Contact him at malcolm@ieee.org.