First Monday, Volume 1, Number 3 - 2 September 1996

fi ® s t m ¤ ñ d @ ¥

PEER-REVIEWED JOURNAL ON THE INTERNET

CAPTIONING VIDEO CLIPS ON THE WORLD WIDE WEB

By GEOFF FREED

Abstract

The Corporation for Public Broadcasting (CPB) and WGBH National Center for Accessible Media (NCAM) develops strategies and technologies to make media accessible to millions of Americans, including people with disabilities, minority language users, and those with low literacy skills.

The Web Access Project, announced by NCAM in February 1996, was initiated to research, develop, and test methods of integrating access technologies (such as captioning and video description) and new Web tools (like Java and RealAudio) into a World Wide Web site, making it fully accessible to blind or deaf computer users. NCAM uses WGBH Online - public broadcasting's Web site visited by more than 2,000 users a day - as its test-bed for research and field-testing of solutions. The Web Access Project is made possible by the support of the Telecommunications Funding Partnership for People with Disabilities, and The Boston Foundation.

This paper will describe NCAM's prototype method for captioning QuickTime video clips. The current technology employed is Apple Macintosh-based only but the captioned movies operate on either Macintosh or PC platforms. It is important to note that the procedures described in this paper represent solutions still under development. They are likely to change as technology improves.

Contents

Introduction and Backgrund

Captioning for television programs has been available since 1971, when the WGBH Educational Foundation in Boston, Mass., created The Caption Center with the purpose of making television programs accessible to deaf and hard-of-hearing viewers via open captions. Open captions require no decoder in order to be seen by the viewer. In 1972, Julia Child's *The French Chef* became the first program to be broadcast with open captions. Closed captions, which require a special decoder in order to be made visible, made their debut in March 1980.

The growth of closed captioning throughout the 1980s was steady yet slow. A chicken-and-egg conflict was part of the problem: the set-top decoders which were necessary to see the captions were fairly expensive - about $US170-$US200. Sales were low, making broadcasters reluctant to support closed captioning because statistics indicated that the audience was relatively small. It was obvious that something was needed both to simultaneously enlarge the audience as well as to increase the number of closed-captioned programs available.

In the late 1980s, legislation was introduced in the United States to require television sets to include decoder circuitry. The United States Congress realized there was a need for captioning but did not want to mandate captioned programming. Passage of such a law could, however, potentially increase the number of captioned programs by eliminating the need for a separate decoder. Supported by deaf advocacy groups, educational organizations, captioning agencies and others, this movement culminated in 1990, when President George Bush signed into law The Television Decoder Circuitry Act. The law required that after July 1, 1993, all new television receivers, with screens 13 inches or larger manufactured for sale in the United States, must contain circuitry for decoding closed captions. Annual sales of televisions with screens in this range are estimated at 20 million units per year. By the end of this century, nearly every home in America will have at least one caption-capable television.

An additional piece of American legislation which has had significant impact on closed captioning is The Americans with Disabilities Act (ADA), which took effect in January 1992. The ADA requires that businesses and public accommodations take steps to ensure that disabled individuals are not excluded from or denied services due to the absence of auxiliary

aids. Such aids include (but are not limited to) open or closed captioning. While the ADA does not specifically require all television programming to be captioned, all public service announcements produced or funded by the federal government for television must be closed-captioned.

Greatly aided by these two laws, the amount of closed-captioned programming in the United States has grown tremendously. Currently, all prime-time network programming is captioned on ABC, CBS, NBC and FOX. Over 7,000 captioned commercials are broadcast on the networks each year as well. Virtually all American news and sports programs are captioned, as well as 100% of all programming for children. The Public Broadcasting Service (PBS) in the United States provides captioning on virtually 100% of its prime-time programming, as well as its news and children's programming throughout the broadcast day. Cable television features approximately 400 hours per week of captioned programming. Finally, over 10,000 home video titles are available with captions.

The Broadcast Caption Process

Before discussing the captioning of video clips on the World Wide Web, it would be helpful to understand how the captioning process works in the broadcast environment.

There are two forms of captions: off-line and real-time. **Off-line captions** are created for non-live, pre-produced programming such as dramas, documentaries, or instructional videos. These captions are also known as "pop-on" captions because of the way they display by popping serially onto the screen in the manner of subtitles. Pop-on captions may be placed virtually anywhere in the video picture to indicate who is speaking or the source of important sounds. **Real-time captions**, on the other hand, are created during live programs, such as news, sports, or special live broadcasts. Real-time captions most often are displayed as "roll-up" captions, so named because they display by rolling up continuously from a baseline. A maximum of four lines of text may be visible at one time. While roll-up captions usually appear at the bottom of the screen, they can be placed in the middle and at the top. Older decoder models, however, are limited to displaying the text only at the bottom. Since off-line techniques are more applicable to captioning on the World Wide Web, they will be the focus of the remainder of this explanation.

Using a cassette copy of the original (uncaptioned) master videotape, a captioner transcribes the audio portion of the show into a PC-based word processor, running special captioning software. On one audio channel of this cassette, longitudinal SMPTE timecode has been recorded from the original master. The other audio channel contains a full program mix. Additionally, a visible representation of the same SMPTE timecode has been placed in the upper third of the picture. The captioner's computer reads the longitudinal timecode track and, when instructed by the captioner, assigns a frame-accurate timecode to each caption, thus synchronizing the text's appearance and/or disappearance from the screen. The captioner also edits the text for reading speed and, if necessary, assigns appropriate codes for screen placement, which may include speaker identification and sound effects.

When the captions have been prepared and reviewed, the text file is converted into coded data and sent to a videotape facility for encoding. The caption data, along with the video from the original uncaptioned master, are fed through an encoding device, which inserts the caption data into line 21 of the vertical blanking interval (VBI). The newly captioned video emerges from the encoder and is recorded onto a second videotape, which is called the **closed-captioned master**. Once the encoding session is finished, the closed-captioned master is ready for broadcast, duplication, or distribution.

Captioning on the World-Wide Web

There are a relatively small number of Web sites making use of video clips at this time. One reason is that the size of a video clip, even a short one, can be unwieldy. A 60-second clip, for example, can contain approximately 2.5 MB of data. A file of this size can take 20 to 40 minutes to download using a 28.8 bps connection. Downloading time can increase during times of heavy Internet traffic or if the user has a slower connection. Cost is another factor. Most people pay Internet Service Providers (ISPs) a fee for Web access, and this fee is often based on the number of on-line hours accumulated on a monthly basis. However, there are some ISPs which provide unlimited access for a flat monthly charge, and their numbers are growing. Finally, the quality of video on the Web, while improving, is not ideal. There is as yet no way to transmit smooth, full-motion video over the Internet comparable to video or film.

With these factors in mind, there is often little motivation to download a digital movie. However, technology is rapidly improving to decrease download times and provide higher quality images. One advance could be the implementation of "streamed" video. Streaming is the process of feeding video and/or audio to the user directly from the server, as opposed to forcing the user to download a clip before playing it back on a local hard drive. Streaming provides nearly immediate access to the clip. This technology has been used for audio transmissions already for about two years; visit the RealAudio site for an example. One software manufacturer, VDOnet Corporation, is already providing video player and server software from their Web site. For an example of streamed video using VDOnet technology, visit the Public Broadcasting Service (PBS) Web site, PBS OnLine.

Creating Captioned Video Clips

As of this writing, NCAM's captioning method is applicable only for QuickTime movies created on the Macintosh. To the best of our knowledge, current movie software for the PC is not capable of adding text tracks to video clips. However, it is possible to convert a Macintosh-captioned movie to one which is playable on the PC. I will provide the details shortly.

To caption video clips on a Macintosh, NCAM used a Mac Quadra 610, operating under System 7.5. Software included MoviePlayer version 2.1 or higher, two plug-ins "Authoring Extras" and "Goodies", and a word processor with drag-and-drop capability, such as SimpleText. Also on hand was a reference source *QuickTime(TM): The Official Guide for Macintosh(R) Users* (authored by Judith L. Stern and Robert A. Lettieri and published by Hayden Books in 1994).

A QuickTime movie is made up of separate video and audio tracks. Depending on the player being used, these tracks may be turned on and off by the viewer. Because the tracks are discrete, there may be multiple audio and video tracks, any number of which may be selected by the user. For example, a QuickTime video clip may contain one video track and separate audio tracks in English, French, and German. A user would select the appropriate language track at the time of playback.

In addition to video and audio tracks, a separate **text track** may be added to the clip. This text track becomes, for our purposes, a caption track. Again, depending on the player being used by the viewer, the captions may be turned on and off, thus simulating broadcast closed captioning. If a player is not capable of turning the tracks on and off, the text track becomes, in essence, open captions which cannot be turned off.

A captioned video clip, therefore, contains the normal video and audio tracks plus the additional text track. Unlike broadcast captions, which obscure a portion of the visible picture, captioned video clips display their text track in a small window *below* the video. Also, while broadcast captions are limited to a maximum of four rows in a single caption, the number of rows available in a single text-track caption is virtually limitless. In its experiments, NCAM was able to fit 19 rows of text below a video clip before running out of space on the computer monitor. However, displaying more than three rows of text at once may prove impractical as the viewer may have difficulty reading the captions and keeping up with the video.

To the right is a single frame of a captioned video clip. The whole clip, plus five other clips (for both Macintosh and PC), may be viewed at the Web Access Project's captioning page. All the video clips are accompanied by a transcript of the audio track.

The procedure for creating a text track using MoviePlayer is relatively simple. Full, step-by-step details of the process may be found at the Web Access Project's captioning page. In brief, the process involves first transcribing text into a word processor, such as SimpleText, and then breaking the text into captions which closely follow the audio. The captions are then inserted into the text track using a drag-and-drop method between the word processor and MoviePlayer; that is, the text is selected and then dragged directly into the text track. A first-time captioner may take more than an hour to caption a one-minute clip, but with practice this effort can be reduced in half to 30 minutes or less.

While Web-based captioning is, in the most general sense, similar to broadcast captioning, there are several important differences. For example, broadcast off-line captioning permits the caption writer to use timecode for frame-accurate synchronization of the text with the audio. Currently, there is no way to use timecode this way on a Web-based video clip. Therefore, all captions must be inserted and "timed" manually. This is not a difficult procedure, but it can be time-consuming. In addition, broadcast captioning allows for the text to be placed virtually anywhere on the screen. This makes it possible for the caption writer to indicate who is speaking by placing the captions under or near the person talking. Current Web-based captioning procedures allow only for bottom center-placed captions. Thus, in order to make clear who is speaking, captions should ideally contain an identification line above the text. For example:

Mary Anne:
You know, Bill,
I always say

"He who hesitates is lost."

Bill:
True, but I prefer

"Fools rush in
and get the best seats."

Other considerations in captioning of any kind include the number of characters that will fit on a given text row and the font which may be used. Broadcast line-21 technology allows a maximum of 32 monospace characters only. Also, the font is not selected by the caption writer. It is determined by the circuitry in the viewer's built-in or set-top decoder. Broadcast captioning also does not accommodate styles such as bold, outline, and shadow but allows for the use of underline and italic. Web-based captioning is more flexible, however. The maximum number of characters per row varies depending on what font, size, and style are employed. The proportionally spaced 12-point Palatino font used in NCAM's captioned clips accommodates approximately 30 characters per row. The caption writer has access to all styles and font sizes available in conventional word-processing software with one limitation: there can be no mixing of styles in the text

track. Changing a style, however, is a straightforward matter. Simply create one appropriately styled caption in the word processor; then select it and drag it into MoviePlayer. The software will automatically alter the rest of the text track to conform to that caption's style.

Platform Conflicts

Even though captioned clips must be created on the Macintosh at this point, it is possible to convert the clips, with the text track intact, to the AVI or MOV formats, which may then be played on PCs.

To convert clips to the **AVI** format, use QuickTime to AVI Converter. For playback of AVI files, use Media Player, which comes with Windows, or another Windows-compatible AVI player, such as AVIPRO, Net Toob or Video Launch Pad. If the new AVI file must be transferred from the Macintosh to the PC on floppy disk(s), use Compact Pro to compact and segment the file on the Macintosh, and ExtractorPC to reassemble the file on the PC.

To convert clips to the **MOV** format, first use **Text Movie Converter** (available on the CD which accompanies *QuickTime(TM): The Official Guide for Macintosh(R) Users*) to "burn" the text into the video. Next, open the file in MoviePlayer and use the "Save As" menu option to save the file as a flattened movie in the MOV format. For playback of MOV files on the PC, use QuickTime for Windows or Net Toob.

A useful feature of QuickTime for both the Macintosh and PC is that the text track is searchable. Using the "Find" feature, the user may type in a key word or phrase and search the text track for any captions containing that text. The software will go to the specific point in the movie where that word or phrase is used. The viewer may then play the video from that point forward, or search for a new word or phrase. Based on our experiences, no other players support a "Find" feature at this time.

Both Macintosh and PC-compatible captioned clips may be found at the Web Access Project's captioning page. NCAM will continue to work to develop a method to caption video clips in their native platforms.

Benefits of Captioned Video Clips

Deaf and hard-of-hearing Web users are the immediate and obvious beneficiaries of captioned video clips. As with broadcast closed captioning, however, the benefits extend beyond this audience. Those using computers without sound, for example, can view captioned clips and obtain the same information as is available from those with sound capability. Also, as educators have already discovered, captions used in conjunction with audio and video can be a valuable tool for improving reading skills of children and adults.

Another useful feature of captioned video clips are the transcripts which are inherently generated by the captioning process. Posting a transcript with the video file allows the user to read the text before deciding if it is worth the time to download the file. At the minimum, transcripts may be used by those who do not have any video-playback capability, as a partial substitute for the clip itself. For maximum accessibility, transcripts should always be used in conjunction with audio-only clips.

Conclusion

Captioning on the Web is still in an early, developmental stage. It is useful but not widespread with plenty of room and opportunity for improvement. Yet it will probably not take Web captioning very long to become integrated into the production process. Web captioning is less complex, and today's video software can be adapted relatively easily to accommodate captions, as NCAM has shown in its initial experiments. The methods presented in this article should be considered volatile, like the Internet itself; as the technology evolves, so will the basic method for captioning. Guidelines for Web captioning should be developed now while the technology is still young and relatively flexible. As technology improves, these guidelines can be adapted and improved to accommodate new techniques. NCAM will continue its research into Web captioning and will regularly post new techniques and technologies, as well as basic Web-captioning guidelines, at its Web site.

About the Author

Geoff Freed is Manager of External Projects for the CPB/WGBH National Center for Accessible Media. He is currently heading the Web Access Project, which studies ways to make the World Wide Web more accessible to users with disabilities. For more information on this and other NCAM projects, visit the NCAM Web site or send e-mail to geoff_freed@wgbh.org.

Copyright © 1996, *First Monday*

Captioning Video Clips on the World Wide Web by Geoff Freed.
First Monday, volume 1, number 3 (September 1996),
http://www.firstmonday.org/?journal=fm&page=article&op=view&path[]=481