

First Report of the Video Programming Accessibility Advisory Committee on the Twenty-First Century Communications and Video Accessibility Act of 2010

Closed Captioning of Video Programming Delivered Using Internet Protocol

July 13, 2011

Table of Contents

I. INTRODUCTION/OVERVIEW	3
II. BACKGROUND ON CAPTIONING	6
A. History	6
B. Development of CEA-608/708	8
C. Caption Use on Television	9
D. Background on Captioning for Internet-Delivered TV	10
III. PERFORMANCE OBJECTIVES OVERVIEW	13
A. Completeness	13
B. Placement	13
C. Accuracy	14
D. Timing	14
E. Specific Technical Capabilities	14
IV. TECHNICAL REQUIREMENTS	16
A. Overview	16
B. Use Case #1: Delivery of Video Programming Content Directly to a Consumer Video Player	18
C. Use Case #1a: Support of Legacy Analog Devices	19
D. Use Case #2: Unmanaged Delivery of Video Programming Content to a Web Browser	19
E. Use Case #3: Delivery of Managed Video Programming Content to Managed Applications or Consumer Devices	20
V. TECHNICAL CAPABILITIES AND PROCEDURES NEEDED	21
A. Introduction	21
B. Functional Requirements for Encoding, Transmission, and Display of Closed Captioned Programming	21
C. Standards Needed	22
D. In-Band and Out-of-Band Delivery of Closed Caption Data	23
E. Criteria Used to Recommend Caption Encoding Formats	24
F. Automated Format Translation	26
G. Recommended Standards to Accomplish These Goals	26
1. Interchange Format	26
2. Delivery File Formats	27
3. Standardized Linkage to User Captioning Display Controls	28

VI. NEW DEVELOPMENTS	28
A. Considerations for the Evolution of the Field	28
B. Emerging Protocols	28
C. Advancing Innovations for User Experience	29
D. HTML5 and Direct In-Browser Support for Captioning	29
VII. SCHEDULE OF DEADLINES FOR THE PROVISION OF CLOSED CAPTIONING.....	29
APPENDIX A – SUMMARY OF DTV RECEIVER REQUIREMENTS.....	31
APPENDIX B – EQUAL CAPTIONING EXPERIENCE	33
APPENDIX C – UNRESOLVED ISSUES.....	34

I. INTRODUCTION/OVERVIEW

Introduction and Legislative Background

On October 8, 2010, President Obama signed the Twenty-First Century Communications and Video Accessibility Act of 2010 (“CVAA” or “the Act”), which amended certain sections the Communications Act of 1934 relating to communications access and video programming.¹ The purpose of the bill was “to update the communications laws to help ensure that individuals with disabilities are able to fully utilize communications services and equipment and better access video programming.”² Among its provisions was the establishment of an advisory committee known as the Video Programming and Emergency Access Advisory Committee (later known as “VPAAC”).³ The Chairman of the Federal Communications Commission (“Commission” or “FCC”) was charged with appointing members to the VPAAC that included representatives from a wide range of organizations and other entities with an interest in the delivery of video programming via the Internet who have the technical knowledge and engineering expertise to fulfill the VPAAC’s duties.⁴

One of the VPAAC’s charges was to develop and submit to the Commission a report concerning online closed captioning to help inform it as it implements the captioning requirements of the CVAA in future rulemaking proceedings, including the revision of its regulations to require the provision of closed captioning on certain video programming delivered via the Internet.⁵ The Act specified that the closed-captioning report should include the following:

¹ See generally PL 111-260, as amended by PL-265 (Oct. 8, 2010).

² See S.R. 111-386 (Dec. 22, 2010), at 1. See also H.R. 111-563 (July 26, 2010), at 19.

³ See CVAA § 201(a). In order to avoid confusion with the Emergency Access Advisory Committee, the Commission changed the name of the advisory committee to the Video Programming Access Advisory Committee, or VPAAC. See Video Programming Accessibility Advisory Committee By-Laws (“VPAAC By-Laws”), at 1.

⁴ The statute specified that the VPAAC should include “(1) Representatives of distributors and providers of video programming or a national organization representing such distributors. (2) Representatives of vendors, developers, and manufacturers of systems, facilities, equipment, and capabilities for the provision of video programming delivered using Internet protocol or a national organization representing such vendors, developers, or manufacturers. (3) Representatives of manufacturers of consumer electronics or information technology equipment or a national organization representing such manufacturers. (4) Representatives of video programming producers or a national organization representing such producers. (5) Representatives of national organizations representing accessibility advocates, including individuals with disabilities and the elderly. (6) Representatives of the broadcast television industry or a national organization representing such industry. (7) Other individuals with technical and engineering expertise, as the Chairman determines appropriate.” See CVAA § 201(b).

⁵ See CVAA § 201(e) (1). The Act requires the Commission to “revise its regulations to require the provision of closed captioning on video programming delivered using Internet protocol that was published or exhibited on television with captions after the effective date of such regulations.” CVAA § 202(b) (revising Section 713(c) of the Communications Act of 1934 (47 U.S.C. 613)).

- A recommended schedule of deadlines for the provision of closed-captioning service.
- An identification of the performance objectives for protocols, technical capabilities, and technical procedures needed to permit content providers, content distributors, Internet service providers, software developers, and device manufacturers to reliably encode, transport, receive, and render closed captions of video programming, except for consumer-generated media, delivered using Internet protocol.
- An identification of additional protocols, technical capabilities, and technical procedures beyond those available as of the date of the enactment of the Twenty-First Century Communications and Video Accessibility Act of 2010 for the delivery of closed captions of video programming, except for consumer-generated media, delivered using Internet protocol that are necessary to meet the performance objectives identified under subparagraph (B).
- A recommendation for technical standards to address the performance objectives identified in subparagraph (B).
- A recommendation for any regulations that may be necessary to ensure compatibility between video programming, except consumer-generated media, delivered using Internet protocol and devices capable of receiving and displaying such programming in order to facilitate access to closed captions.⁶

The statute specified that the VPAAC report on closed captioning should be submitted within six months of the VPAAC's first meeting.⁷

Formation and Work of VPAAC Working Group 1

On December 10, 2010, the Chairman appointed the members of the VPAAC.⁸ Subsequently, the FCC announced the appointment of the VPAAC co-chairs and members of the Commission staff to assist the VPAAC. See attached member and staff list. At the first meeting of the VPAAC in Washington, D.C. on January 13, 2011, the VPAAC co-chairs divided the members into four advisory working groups to assist the VPAAC.⁹ Working Group 1 was created to examine issues involved in transferring closed captions provided on television programs to the online environment, including “identification of protocols, technical capabilities, and technical procedures needed to encode, transport,

⁶ See CVAA § 201(e) (1) (A)-(E).

⁷ See CVAA § 201(e)(1).

⁸ FCC Public Notice, *Video Programming and Emergency Access Advisory Committee Announcement of Members*, DA-2320, 25 FCC Rcd. 17094 (released Dec. 7, 2010 and erratum released Jan. 7, 2011).

⁹ See VPAAC By-Laws, at 1.

receive and render closed captioning of video programming delivered” via the Internet.¹⁰ The leadership of Working Group 1 consisted of Roger Holberg, FCC Policy Co-Chair; Henning Schulzrinne, FCC Technical Co-Chair; Vince Roberts, Industry Co-Chair; and Shane Feldman, Public Interest Co-Chair.

At the January 13 initial VPAAC meeting, Working Group 1 was briefed on the work done by the Society of Motion Picture and Television Engineers (SMPTE) broadband standards committee, which has published a standard (timed-text standard) and a recommended practice for converting analog captions authored using CEA-608 for Internet distribution. See <http://www.smpete.org/standards/st2052-1-2010.pdf>. Working Group 1 also discussed some examples of how captions initially created for broadcast television can be converted for display on an online video player or other Internet-based viewing method. Working Group 1 was encouraged to further designate subgroups to address specific topics that the statute had identified for inclusion in the closed-captioning report.

For the next six months, Working Group 1 deliberated primarily through weekly conference calls, which were transcribed contemporaneously and were intended to ensure full participation and access to all working group members. The conference calls were supplemented by periodic email exchanges among Working Group 1 members. The discussion included a presentation by the World Wide Web Consortium’s (W3C) Web Accessibility Initiative (WAI) about formats for Web display of captions, including W3C Timed Text Markup Language (TTML), on which SMPTE-TT is based; SMPTE-TT; WHATWG (Web Hypertext Application Technology Working Group) WebVTT; and user requirements for accessible media associated with development of HTML5. Google provided a presentation and demonstration of the WHATWG’s WebVTT caption format. Certain written work product was posted on the group’s workshare Web site (or “wiki”), which was accessible to the public. See <http://vpaac1.wikispaces.com>.

Subgroups were created within Working Group 1 to examine the following issues:

- The “performance objectives” for protocols, technical capabilities, and technical procedures needed to encode, transport, receive, and render closed captions of video programming, except for consumer-generated media, delivered via the Internet;
- The definition of certain types of programming, such as live, near-live, and pre-recorded programming and programming edited for Internet distribution, and the appropriate schedule for phasing in closed-caption obligations for that programming;
- Responses to a series of technical questions for consideration, which were provided by Commission staff, and which ultimately formed the basis of the technical portions of this report.

At a second meeting at the Commission on May 5, 2011, Working Group 1 discussed a range of issues, including the three topics listed above. Commission staff provided periodic guidance about issues or questions Working Group 1 could address. These

¹⁰ Id.

suggestions were contained in a presentation by Karen Peltz Strauss at the January 13 VPAAC meeting and in technical and policy considerations subsequently sent to the working group in writing. Several of these issues identified by Commission staff were discussed at various points.¹¹

II. BACKGROUND ON CAPTIONING

A. History

In the era of silent films, much of the story of the film was told via graphical text describing the spoken word or sounds. This was an early form of open captions. When “the advent of sound films in 1927 suddenly deprived the deaf of one of their chief sources of information and entertainment,” the need for an accessible technology, which would eventually evolve into closed captioning, was born.

While there were several attempts in the intervening years at utilizing then-current technologies, through the efforts of Dr. Edmund Burke Boatner, superintendent of the American School for the Deaf (Connecticut) and Dr. Clarence D. O’Connor, superintendent of the Lexington (New York) School for the Deaf, in 1949, the Captioned Films for the Deaf program was formed. It was later incorporated in Connecticut with an office at the American School for the Deaf. This program established the earliest version of what is now called a captioning agency. As a not-for-profit, it worked hard to secure marginal funding over the ensuing years, making open-captioned theatrical films available on a limited basis.¹²

In 1958, Congress authorized the establishment of Captioned Films for the Deaf as an agency of the United States Office of Education, Department of Health and Welfare. The new agency had great success under its first director, John Gough. He was followed by Dr. Malcolm Norwood, who until his retirement in 1988, was Chief of Media Services and Captioned Films for the Deaf, and is widely credited as one of closed-captioned TV’s pioneers. The work Gough and Norwood started continues today as the Described and Captioned Media Program, with funding provided by the U.S. Department of Education (ED), and administered by the National Association of the Deaf (NAD).

In the early 1970s, the National Bureau of Standards (NBS)¹³ had an idea about providing precise time information nationwide over television broadcasts.¹⁴ They planned on using a part of the video television signal that is not normally seen by the viewer to carry the time information. When the television networks were approached to consider

¹¹ Several of the issues contained in FCC staff guidance were referred to in Section 202(b) of the CVAA, which described certain requirements for the revision of closed-captioning regulations by the Commission. Those issues included how “video-programming distributors” and “video-programming providers” should be defined and how to define “good-faith effort to identify video programming” using a mechanism established by the Commission to make information available to video-programming distributors and providers about programming subject to the Act.

¹² <http://www.dcmp.org/caai/nadh93.pdf>.

¹³ Now the National Institute of Standards and Technology (NIST).

¹⁴ See the National Captioning Institute website <http://www.ncicap.org> for more history on closed captions.

participation in the work, the ABC television network agreed to be part of the project. For a number of reasons, the project did not work as planned. However, Julius Barnathan, then Vice President of Broadcast Operations and Engineering at ABC, proposed that the technical ideas put forth could be used to carry captions for the deaf and hard of hearing audience instead of time information.

Following this revelation, various demonstrations of the technology were held in several locations, and on February 15, 1972, at Gallaudet University, ABC and the NBS presented the television show *Mod Squad* with captions embedded. As a result of these technical demonstrations, an appreciation grew for the importance of providing this service within the television broadcast.

In 1972, PBS, under a contract with HEW's Bureau of Education for the Handicapped, began refining this technology and system. In 1974, under special temporary authority from the FCC, PBS began broadcast tests of the system. In December 1976, the FCC granted approval for broadcasters to present captioned material on a limited portion of the television picture.¹⁵ In 1976, the FCC reserved line 21 of the analog television signal exclusively for the analog waveform that carried the caption information.

In February 1977, President Carter requested information from the three commercial broadcast networks regarding the viability and practical application of closed captioning.¹⁶ Shortly thereafter, development of decoding equipment began.

ABC, the NBS, PBS, and the newly formed National Captioning Institute (NCI) continued the development work on captioning. In the early 1980s, set-top boxes used to decode caption information became available for purchase by the consumer. When connected to the antenna and the television receiver, these decoder boxes allowed the viewer to see the captions that were broadcast on programs containing the caption data. More programming became available over time. Some examples of early captioned programs in 1980 were *The ABC Sunday Night Movie* (ABC), *Disney's Wonderful World* (NBC), *Masterpiece Theatre* (PBS), and the children's program, *3-2-1 Contact* (PBS).¹⁷

While the early closed-caption set-top decoder boxes and the singular consumer television with a built-in closed-caption decoder (the Sears 19" TeleCaption Color TV) made the new technology available, the Television Decoder Circuitry Act of 1990 changed everything. The Act mandated, as of July 1, 1993, "that apparatus designed to receive television pictures broadcast simultaneously with sound be equipped with built-in decoder circuitry designed to display closed-captioned television transmissions when such apparatus is manufactured in the United States or imported for use in the United States, and its television picture screen is 13 inches or greater in size."¹⁸ The only manufacturer to

¹⁵ See Leonard E. Maskin, Statement before the Science and Technology Subcommittee on Science, Research and Technology, United States House of Representatives, November 9, 1983.

¹⁶ See Richard W. Stubbe, A Chronological History of ABC Television Network's Contributions to the Development of Closed-Captioning in America, 1983. This information was compiled as support documentation for Leonard E. Maskin's testimony before the U.S. House of Representatives.

¹⁷ See footnote 14, supra.

¹⁸ See http://transition.fcc.gov/Bureaus/OSEC/library/legislative_histories/1395.pdf.

support the Decoder Circuitry Act of 1990, Zenith was the first to develop and market analog TVs with built-in closed caption capability in 1991.

Approximately one million closed-caption set-top decoders and specially equipped televisions were sold during the first 13 years of closed captioning (1980–1993). As of July 1, 1993, the nearly twenty million new televisions with screens 13 inches or greater sold each year in the U.S. were now caption ready. By 2003, virtually every TV in America had the capability to display captions. Captions had made the leap from obscure to the mainstream, visible and recognized on televisions everywhere in the U.S., from homes, to airports, to eating establishments, to public spaces; virtually anywhere in America television programming was watched or exhibited.

B. Development of CEA-608/708

1. CEA-608 Captions (Analog)

The technology to encode closed captions, generate the waveforms, and decode the captions in receivers was formalized by the Electronics Industry Association (EIA) as standard EIA-608. When EIA changed its name to the Consumer Electronics Association, the associated standard became known as CEA-608.

CEA-608 was developed to provide technical guidance and assurance that equipment used to generate captions transmitted by the broadcaster would be received and properly displayed on consumer devices designed to decode and display captions. Because of the limitations of technology at the time, there was a limit in the number of channels of captioning and the rate at which the caption data could be transmitted. In addition to the characters that are displayed on screen, there is a requirement to transmit control characters to the decoder for such things as position, font, underline, italics, erase, etc., within the same bandwidth. These non-displayed characters take up some portion of the total 120 character-per-second transmission rate of CEA-608.

Here are some of the features and limitations of CEA-608 captions;

- fixed font size, with italics, color options, underlining, etc.;
- fixed-block black background;
- multiple channels for different languages;
- roll-up, pop-on or paint-on modes;
- lack of support for many character types, including multilingual.

2. CEA-708 Captions (Digital)

With the advent of digital television in the 1990s, it became possible to do more with broadcast television captions than was possible with analog television technology. Since the signal was inherently digital and the data rate in the channel was sufficient to provide more bandwidth for captioning, a more advanced standard for captioning was planned and developed. In addition to the wider bandwidth for more caption channels, receiver technology had greatly improved character-generator and memory capabilities that could be incorporated in the development of a new captioning standard for broadcast.

The digital standard developed for closed captioning is known as CEA-708. This standard provides for a rich set of features and capabilities above and beyond those supported by CEA-608 captions. In addition, CEA-608 captions can be transported within 708. This is especially important in HDTV programs, where a consumer may have an HDTV capable of displaying CEA-708 captions, but may also have legacy analog televisions where down-conversion of the HDTV program for analog display will require the use of the CEA-608 captions. For over-the-air, cable, and satellite consumers who only have analog receivers, the HDTV signal has to be down-converted to standard definition to be viewed on their receivers, and the CEA-608 captions must be extracted from CEA-708 in order to be displayed.

Because the complete transition from analog television to digital television is not yet complete, and millions of households continue to use analog television receivers, CEA-608 captions will continue to play a vital role. Analog receivers cannot decode native CEA-708 captions and, because of the technical differences, CEA-708 captions are not backward compatible with CEA-608-capable receivers. The developers of CEA-708 were therefore careful to ensure that CEA-608 captions could be carried within the CEA-708 transport.

Improving upon CEA-608, CEA-708 captions support a wider range of features:

- a data rate that is ten times greater than CEA-608 captions;
- characters for multilingual support;
- fonts and backgrounds with variable translucency;
- fonts with edges;
- multiple language channels;
- CEA-608 captions transported within its structure.

An important feature of CEA-708 captions is that viewers can customize the caption display in the receiver.

C. Caption Use on Television

In the 1970s, WGBH introduced captioned television for viewers who are deaf or hard of hearing. Since decoder technology was not yet available, programs were presented with open captions, which were visible to all viewers. The first program to be captioned in this manner was PBS' *The French Chef* with Julia Child. In December 1973, the *ABC Evening News* was rebroadcast on WGBH with open captions, and then made available to all PBS stations on August 6, 1974, under special temporary authority from the FCC.¹⁹

In December 1976, the FCC granted permanent authority to PBS and other broadcasters to televise closed captions, utilizing line 21 of the TV picture – a line that did not carry picture information. Line-21 closed captioning debuted on broadcast television in March 1980 with initial participation from ABC, NBC and PBS, which broadcast a total of 16 hours a week of captioned television. In April 1983, CBS began captioning using the line 18 NABTS (North American Broadcast Teletext Specification) teletext standard in a

¹⁹ See <http://www.wgbh.org/wgbhtimeline/index.cfm>.

competing service called Extravision. In an early glimpse at the future possibilities of the Internet, Extravision offered viewers a constantly updateable electronic magazine filled with information such as CBS program information, news, sports, weather, and pages which could be customized by the local affiliate station carrying it, for such things as program schedules, local community announcements, and station promotions. In 1984, CBS started to transmit “dual-mode” captioning on both lines 21 and 18. CBS later discontinued the Extravision line 18 teletext service after line 21 closed captioning became the industry standard.

In the early years of television captioning, captioning services were only available from two not-for-profit entities, WGBH's The Caption Center (now known as The Media Access Group at WGBH) and NCI. The relatively high cost of captioning services initially limited the growth of available hours and types of captioned television. To help spur captioning growth, funding partnerships were established between the public and private sectors. In a natural extension of the captioned-films program, the U.S. Department of Education (ED) began to award daypart-specific (news, sports, children's, primetime, and daytime) grants for television closed captioning. The ED grants made funding available to the captioning agencies and their clients, broadcast, and cable networks, resulting in an ever-growing volume and variety of captioned programs. Over time, substantial funding also came from other sources, including the broadcast and cable television networks, foundations, and advertisers.

As the marketplace grew, so did the emergence of for-profit captioning agencies. Today there are a wide variety of choices and vendors of this product, which has become a fully integrated part of television industry production,

Throughout the 1980s and into the early 1990s, the amount of available captioned television hours continued to grow steadily. By the mid-1990s, almost 100% of traditional over-the-air broadcast television was captioned on a voluntary basis, as were many additional hours of cable, syndicated, and local programming.

In 1996, Congress passed the Telecommunications Act of 1996, which directed the FCC to adopt rules requiring closed captioning of most television programming. The FCC rules on closed captioning became effective January 1, 1998. They required entities that distribute television programs directly to home viewers to make sure those programs are captioned. Under the rules, 100% of non-exempt programs shown on or after January 1, 1998, had to be closed captioned by January 1, 2006. Also, 75% of non-exempt programs shown before January 1, 1998, had to be closed captioned by January 1, 2008. The rules do not apply to videotapes, laser disks, digital video disks, or video-game cartridges. Additional rules regarding Spanish Language and other programming are also part of the mandate.²⁰

D. Background on Captioning for Internet-Delivered TV

²⁰ See <http://transition.fcc.gov/Reports/tcom1996.pdf>.

On the Internet, closed captioning evolved in more than one format, driven by the multiple video formats already in use on the Internet. By 1999, captions were available both in-band (contained within the media file) and out-of-band (provided as a separate resource and synchronized in a media player). For example:

- in-band captions can be provided as text tracks in QuickTime;²¹
- out-of-band captions can be provided using SMIL (RealVideo and RealText; QuickTime and QTText);²²
- out-of-band and in-band captions can be provided using ASX (Windows Media video and SAMI).²³

During the early days of the Internet, captioned videos were simply open captioned through video production since player support was not widely adopted. In 2000, WGBH released MAGpie 1.0, the first professional and free tool for authoring captions for the Web for QuickTime, RealPlayer and Windows Media Player. This permitted conversion between formats, allowing video creators to choose the caption format best suited to their preferred video-delivery format. Later releases of MAGpie added support for the W3C TTML format.

In the last few years, there has been a dramatic improvement in video-streaming technology that has enabled delivery of long-form full-length movie and TV content to a wide range of Internet-connected devices including television sets, set-top boxes, game consoles, disc players, phones, tablets, and other mobile devices. Recent measurements show that over 50% of Internet bandwidth during prime time is devoted to adaptive streaming of long-form television content, and that percentage is growing rapidly. Adaptive streaming has become the dominant delivery method of video measured in megabytes and probably measured in viewing time, recently exceeding the size of peer-to-peer distribution (which includes large amounts of unauthorized TV and movie content).

Historically, streaming video evolved first on computers running Web browsers playing short, partial-screen video clips using file-download delivery. But advances in adaptive streaming, video compression, and network bandwidth have enabled fast-start streaming and reliable viewing of high-quality, full-screen television programs on all types of Internet-connected devices and high-resolution displays. Although these devices have been able to build the necessary audio and video decoders based on widely adopted standards (i.e., ISO/IEC/ITU MPEG video and audio), there isn't widespread support in content or devices for one or more standard caption formats to make captions equally decodable.

SMPTE addressed the problem by applying the W3C Timed Text Markup Language to in-band video streaming and adding other features required for television content in the U.S. and internationally. As a result, SMPTE Timed Text (SMPTE-TT) is specified as the caption and subtitle format for the Digital Entertainment Content Ecosystem's "UltraViolet" format for commercial movie and television content, and is specified in draft standards for Internet television delivery in the U.K., France, Germany, Italy, and other European countries. SMPTE-TT can be decoded by a standalone decoder in a device

²¹ See <http://webaim.org/techniques/captions/quicktime/>.

²² See <http://webaim.org/techniques/captions/real/>.

²³ See http://webaim.org/techniques/captions/windows/add_captions.

(similar to CEA-708) or in a browser, and can be delivered as a single document by download, or in small segments in video streams, compatible with adaptive streaming.

On the Internet, developers also started also inventing open and proprietary text formats for captions and subtitles. Among the formats that are used are the SRT SubRip format and SSA Substation Alpha format, both of which are simple text formats authored by free subtitling and captioning tools. Initially, these formats were rendered directly onto videos by burning in the subtitles. Later, as synchronization became more widely supported by media players, these formats contributed further to the creation of an active online subtitling culture.

In addition to formats developed by Web users, more formal industry-driven standards emerged. These included:

- W3C TTML (started in 2003 as DFXP¹, and released by the W3C Timed Text Working Group in November 2010²⁴)
- SMPTE-TT (TTML extension features not covered by the W3C standard, published by SMPTE December 2010)²⁵.
- MPEG-4 Timed Text / 3GPP TTX (developed in 2004 by the 3GPP working group²⁶ for cellular networks and adopted by ISO/MPEG in 2006²⁷).

Adobe Flash is also a popular platform for video delivery. Flash has supported captions since 2002, when video display was added to the Flash Player. With the release of Adobe Flash CS3, support for a portion of the full TTML specification was added, and improvements to that support have been added in subsequent releases. In 2007, Microsoft introduced Silverlight, an application framework for writing and running browser plug-ins or other rich Internet applications, with features and purposes similar to those of Adobe Flash. The run-time environment for Silverlight is available as a plug-in for most Web browsers and enables closed-caption support of streaming media and rich Internet applications.

In subsequent years, site-specific video players such as those on YouTube.com, Hulu.com, Netflix.com, Amazon.com, and iTunes provided built-in support for standard, community-generated and proprietary formats. In 2009, Google introduced machine-translation, auto-timing and auto-captioning using speech recognition for YouTube and Google Video, aiming to simplify and speed up the labor-intensive task of manual caption and subtitle creation.

²⁴ See <http://www.w3.org/TR/2010/REC-ttfl-dfxp-20101118/>.

²⁵ See <http://www.coataccess.org/files/2010-1210%20Letter%20to%20COAT.pdf>.

²⁶ See <http://www.3gpp.org/ftp/Specs/html-info/26245.htm>.

²⁷ See http://en.wikipedia.org/wiki/MPEG-4_Part_17.

III. PERFORMANCE OBJECTIVES OVERVIEW

Given the goal of providing closed captioning for television programming delivered over the Internet, the fundamental performance objective is that regardless of how the captioned video is transmitted and decoded, the consumer must be given an experience that is equal to, if not better than, the experience provided as the content was originally aired on television using the CEA-608/708 system.²⁸ The CVAA requires television programming exhibited with captions after the FCC rules go into effect to be distributed on the Internet with captions. The VPAAC expects, therefore, that if there is a problem with the caption file for one of these programs, the program will still be captioned when distributed on the Internet and that these captions will meet the performance objectives laid out in this section.²⁹

By “experience” we mean:

- the presentation format of the captioning; e.g., within one or separate caption “windows,” with text that appears all at once (pop-on), with text that scrolls up as new text appears (roll-up), or where each new letter or word is displayed as it arrives (paint-on);
- semantically significant formatting, such as italics, colors, and underlining;
- the timing of the presentation of caption text with respect to the video;
- the consumer’s ability to control the caption display, including the ability to turn it on and off, and to select font sizes, styles, and colors, and background color and opacity.

A. Completeness

Regulations should require that transcoding of captions based on CEA-608/708 for use on the Internet preserve the functional capabilities and features included in the elements listed in Sec. V.B, and offer the same user-control options as are defined by the FCC in 47 C.F.R.§15.122 (summarized in Appendix A of this report). Nothing must be lost in transcoding when converting captions between conventional broadcast captioning formats and Internet. This includes not only the caption text, but the timing and positioning information, and presentation format (roll-up, pop-on). In the case where video has been edited before being placed online, the appropriate corrections to timing must be done.

B. Placement

²⁸ These requirements are detailed in Appendix A.

²⁹ The VPAAC also points to the best practices listed in Appendix B as further guidance for entities providing online captioning.

For Internet-delivered caption content, the positioning information as originally authored shall be made available to the consumer device.

C. Accuracy

When captioned television content is repurposed for Internet use, such captioning must be equal to or greater than the accuracy of captions shown on television. Efforts towards improving the overall quality of captioned content are encouraged. Accuracy of captioning shown on broadcast, cable, and satellite television is not within the scope of the VPAAC. When captioned live-broadcast television content is repurposed for Internet use, such captioning may be processed to improve the user experience.

D. Timing

All processing through the distribution chain, including transcoding, must provide a timing experience that is equal to or an improvement to the timing of captions provided in the captioning shown on television.

E. Specific Technical Capabilities

The user's Internet-connected media players must support, at a minimum, the performance objectives below. Innovation, experimentation and augmentation of user controls are encouraged to achieve these objectives.

Extension of responsibilities throughout the delivery chain

All entities in the delivery chain must each do their part to ensure the user's Internet based experience meets or exceeds the equivalent of the broadcast experience as specified by FCC's DTVCC R&O 00-259. This represents a straightforward extension of delivery responsibilities for broadcast television to apply to Internet-delivered content.

Performance-objectives support by player tools displaying captions for the end user

Players operating on a platform which cannot possibly support features needed to meet the criteria listed below may request an exemption from the FCC (e.g., a device with a gray-scale screen does not need to support color choices for caption text, or a platform that does not support all character edge attributes only need meet the criteria for the attributes which are supported by the platform).

User settings are new to players which support Internet-delivered video, and will require time and effort to implement.

User Settings

Player tools must support the display of captioning information as defined in the interchange format, as outlined below, including caption-placement information intended to help identify speaker location.

- Player tools must support the ability of end-users to customize the display of captioning information
- Players shall permit the user to choose a setting that displays captions as intended by the caption provider.
- Once the viewer chooses a set of customized caption display features, such as font and/or color, these settings remain until subsequently changed by the user.

In addition, the following capabilities must be supported.

- Character color
 - Players must provide the end user with the ability to override the authored color for characters.
 - Players must provide end users with a selection of no less than eight colors which must include the following: white, black, red, green, blue, yellow, magenta and cyan.
 - Players must support the ability to choose from the full complement of 64 colors defined in CEA-708.
- Character opacity
 - Players must provide customization support for 100%, 75%, 25% (opaque, semi-transparent) opacity.
- Character size
 - Players must provide customization support for character font size, with the ability to increase the size up to 200% or decrease the size to 50%.
- Fonts
 - Players must provide customization support for selection of seven fonts which must be mapped to the seven font styles defined in CEA-708.
 - End users must be able to assign fonts available on their systems to act as the defaults for each of the seven styles.
- Caption background
 - Players must provide the end users with the ability to override the authored color for caption background.
 - Players must provide end users with a selection of no less than eight colors which must include the following: white, black, red, green, blue, yellow, magenta and cyan.
 - Players must support the ability to choose from the full complement of 64 colors defined in CEA-708.
 - Players must offer customization support for 100%, 75%, 25%, and 0% (opaque, semi-transparent, transparent) background opacity.
- Character Edge Attributes

- Players must provide support for character edge options: none, raised, depressed, uniform, or drop shadowed.
- Caption window color
 - Players must provide the end user with the ability to override the authored color for caption background.
 - Players must provide end users with a selection of no less than eight colors which must include the following: white, black, red, green, blue, yellow, magenta and cyan.
 - Players must offer users the ability to choose from the full complement of 64 colors defined in CEA-708.
 - Players must offer customization support for 100%, 75%, 25%, and 0% (opaque, semi-transparent, transparent) caption window opacity.
- Language
 - The ability to select caption tracks in additional languages must be provided, when available.
 - The ability to choose “easy reader” captions must be provided when such content is available; i.e., the same content simplified for those who need or prefer to read less text. Such a choice must be clearly marked as “easy reader.”
- Player tools must provide the ability for the user to preview settings. Once chosen, the settings would remain as the default until changed by the user.
- Player tools must support end user’s ability to turn the captions on and off as easily as muting the audio or adjusting the volume.

IV. TECHNICAL REQUIREMENTS

This section describes the technical requirements related to the delivery of closed captioning in television programming delivered on the Internet.

A. Overview

Captioning created for television content today is authored using tools based on the ANSI/CEA-608-E Line-21 Data Services standard. CEA-608 captions, as described in the section on the history and background of captions, were originally designed for use with the analog NTSC television system. With the advent of digital television, a new standard for use with digital television was developed, known as CEA-708. CEA-708 captions have a much richer set of caption capabilities than CEA-608, and also allow for the carriage of CEA-608 captions. In CEA-708, the authored CEA-608 data is carried all the way through to the consumer receiver as CEA-608 compatibility bytes for use if the receiving device needs to create an NTSC waveform, needed for compatibility with legacy analog TV sets.

Per current FCC regulations, caption data must be carried in digital television broadcasts “pursuant to the technical specifications set forth in part 15.”³⁰ In part 15, section 15.122, requirements for consumer receivers are specified: “Digital television receivers and tuners must be capable of decoding closed-captioning information that is delivered pursuant to EIA-CEA-708–B: “Digital Television (DTV) Closed Captioning.”³¹ To meet this requirement, caption data in CEA-608 format is transcoded into the digital television closed-captioning (DTVCC) format defined in the CEA-708 *Digital Television (DTV) Closed Captioning* standard.³² Today’s television broadcasts therefore carry both CEA-608 “compatibility bytes” as well as DTVCC data conforming to the CEA-708 standards.

Decades of video programming have been captioned in CEA-608/708 format, therefore a standard format must be specified for these captions to be delivered via Internet protocols in such a way that the consumer’s experience is in no way degraded. To reduce cost and facilitate the availability of captioned television programming on the Internet, it is highly desirable that there be a single standard interchange format for content providers to encode closed captions into programming before they distribute it, maintaining at least the level of quality and consumer control that CEA-608/708 enables. With this single standard, content providers can caption video for the Internet one time. Otherwise, they might have to re-caption Internet video, incurring additional cost and delay. The recommended single standard interchange format may be revisited in the future if an updated version becomes available which meets the applicable consumer and industry requirements.

The use of a single interchange format does not imply that there should be one single standard for delivery of the captioned programming to the devices and applications that consumers use to display the content. The Internet is continually evolving in the number and diversity of applications and devices that can display video content. It would be highly undesirable to preclude this flexibility and diversity in display applications. However, availability in a common encoding format assures the widest audience.

Therefore, distributors of programming services and applications must be required to (a) receive the captioned content from the content provider encoded in the standard interchange format, and then (b) ensure that any reformatting performed before delivery to end users (consumers) is supported by the applications and devices (e.g., Web browsers, proprietary downloaded applications, and generalized video players) used for playback so that the caption-viewing experience is at least equal to that which CEA-608/708 enables, and that the additional delivery formats used are based on standards developed within an open process by recognized industry standard-setting organizations.³³

To clarify the process of distribution and the functionality of different formats used in captioning for Internet video, we propose the following definitions:

³⁰ 47 CFR § 79.1 (a) (4).

³¹ 47 CFR § 15.122 (b).

³² In 2003, CEA redesignated EIA standards, and those EIA standards under CEA auspices, as CEA standards.

³³ The terms “standards” and “recognized industry standards-setting organizations” as used in this report are intended to have the same meaning as the terms “voluntary consensus standards” and “voluntary consensus standards bodies” as defined in OMB CIRCULAR NO. A-119, Revised (http://standards.gov/standards_gov/a119.cfm#4).

- Interchange format: The encoded caption data that preserves all of the original semantic information and text (including information which may not be used in display, such as edit decision lists) and allows easy conversion to other formats. These formats are typically used by professional captioners to archive captions for later re-use. SMPTE-TT and W3C TTML are examples of formats used for interchange. Historically, formats such as SCC (Scenarist Closed Captions), which is CEA-608 specific (an open format), Cheetah CAP (a proprietary format), and Swift Interchange Format (also proprietary) have often been used for interchange.
- Delivery format: The encoded caption data contained within a download or stream of content to a consumer device in either the standard interchange format or a different network-specific or video-player-specific format, such as SRT, 3GPP TTX, W3C TTML, and others, which are transmitted to and read by a video player. SMPTE-TT is currently being used by several video services and Internet video players (Netflix, Adobe, Microsoft), and is specified in Internet video formats such as the Digital Entertainment Content Ecosystem (DECE), The Digital Television Group in the U.K, and the HD Forum in Europe. WebVTT, currently implemented in early demonstration builds of WebKit³⁴, is one of the potential delivery format for HTML5. While the delivery format is typically downloaded (as a file) or streamed (real-time) to a consumer device, the method(s) for delivering the caption format is/are not necessarily specified by the delivery format itself.

For interchange purposes, captions may be encoded in the single, defined interchange format; for delivery purposes, captions may be encoded either in interchange or delivery formats as long as captions are always available to all video users.

Three use cases corresponding to different ways this television content can be distributed from the content provider through the Internet to be received and processed in the consumer device are described below.

B. Use Case #1: Delivery of Video Programming Content Directly to a Consumer Video Player

In this example, content is sent directly, with a standardized delivery format, from the content provider (or through an intermediary service) to an Internet-connected consumer device that includes standardized video-player functionality (decoding and rendering of audio, video and captions), allowing the user to access, stream or download, record and/or play files containing television content. Note that in this example, the video player is implemented as a combination of built-in hardware and software, where the software component is a general-purpose embedded/resident decoder rather than an application that is downloaded to the player in order to provide the needed decoding functionality. In simplified form, the system consists of:

³⁴ See for example, the talk archived at <http://www.youtube.com/watch?v=tua3DdacgOo#t=10m45s>.

- the content provider, who makes video programming available for delivery to the consumer via the Internet;
- a standardized format for representing the captioning within the file or stream;
- the programming distributor, who transmits the content and passes through the captions;
- the consumer's Internet-connected device containing the video player, which includes functionality for decoding and displaying closed captioning;
- the user interface offered by the consumer's device for the purpose of controlling the display of closed captioning.

To support this use case, the file or stream must include the captioning both encoded and transported in a standardized format. Therefore, the programming distributor must be able to transmit the content with captions encoded in the standardized format. Ideally, this same format is usable as an interchange format, allowing the content provider to utilize programming that can be offered to the end user without modification. However, if the standard interchange format is used for delivery, the video player need only implement those features which are consistent with CEA-608/CEA-708 display. Interchange formats may support many additional elements which are never used by TV decoders for closed captioning and are solely present to facilitate conversion. It is only necessary for the player to support those features which are already established for DTV receivers and specified in Title 47 C.F.R., Part 15, § 15.122 (see Appendix A).

Some consumer video players of this type will have platform-level controls for closed captioning. For example, the device may be a regular DTV receiver which includes Internet-enabled features. Users would be expected to control the rendering of closed captioning on this device using the same control functions present to support the receiver's regular television functions. For example, the "CC" button on the remote control or menu must function to turn the captions on and off, regardless of the method used to deliver the video (broadcast or Internet). Likewise, controls over font size and color, background color and opacity must operate similarly.

C. Use Case #1a: Support of Legacy Analog Devices

A use case that is closely related to Use Case #1 is the reception of video/audio/captioning content via the Internet by a device which decodes it into a standard-definition analog NTSC output for delivery to a legacy TV display, DVD recorder or VCR. This type of device is analogous to the DTV converter boxes in use today. The support of closed captioning on the NTSC output requires that the stream carries the CEA-608 caption data needed for the receiver to re-create line 21 in the NTSC waveform. These "CEA-608 compatibility bytes" are delivered in ATSC-compliant broadcasts today to support the converter boxes. Television content delivered over the Internet must also be capable of carrying these CEA-608 bytes for the same reason.

D. Use Case #2: Unmanaged Delivery of Video Programming Content to a Web Browser

In this example, the captioned video content is provided to an Internet-based video service, such as YouTube, which may reformat it for accessing through a consumer device that includes a Web browser. The user may be expected to have downloaded the necessary browser plug-in(s) to render the video programming. When the consumer visits a certain

Web site and selects content to view, the server creates the desired user experience using the tools resident in the browser or downloaded to it. Elements of this system are similar to the first example:

- the content provider, who makes video programming available for delivery to the consumer via the Internet;
- the programming distributor, who transmits the content;
- the Web server, which supplies the video programming in a format that is acceptable to the user's browser configuration;
- the consumer's Internet-connected device containing a Web browser that natively supports video or has been configured with plug-ins necessary to render audio, video, and captions;
- the user interface offered by the consumer's device for the purpose of controlling the display of closed captioning.

Note that in this case, the closed-caption data need not be delivered to the consumer device in a specified standard format. As long as the user's experience is at least as good as what he or she would have experienced in viewing the original television broadcast, any method based on standards developed within an open process by recognized industry standard-setting organizations, can be used, as long as the standard caption format is accessible to any users whose player may only decode that caption format. This flexibility permits Web-based video programming distributors to innovate and potentially improve upon the captioning enabled by the standardized interchange format.

As in Use Case #1, the consumer device may have platform-level controls for closed captioning. As before, these must be usable to control the rendering and display of captions for content played through the web browser. A standard is needed to link the browser's caption rendering functions to the platform's caption control functions.

Note that in both Use Cases #1 and #2, it is highly desirable that one common industry-standard format be specified as the interchange format, to avoid the need for a content provider to create multiple formats for distribution.

E. Use Case #3: Delivery of Managed Video Programming Content to Managed Applications or Consumer Devices

In this example use case, the television programming content is provided to a programming distributor, which reformats it for transmission over the Internet to be decoded and rendered by a managed device or application that has been installed or downloaded by the user for this purpose. This application or device accesses content from servers on the Internet that are designed to be compatible with it, and other applications and devices outside the managed system cannot access the video content. The formats used to encode the video, audio, and captions and the protocols used to deliver the content may be non-standard and/or proprietary but must provide a caption-viewing experience at least equal to that which is enabled by CEA-608/708. This flexibility to reformat allows for innovation and improvement upon the basic captioning capabilities. As with the other example use cases, any platform-level closed-caption controls must operate on the captions for content played through the application. In addition, the server should be able to accept content from the content provider in a common industry-standard interchange format.

V. TECHNICAL CAPABILITIES AND PROCEDURES NEEDED

This section describes the technical capabilities and procedures needed for entities to reliably encode, transport, receive and render broadcast-television closed captions over the Internet. As a general matter, the VPAAC believes that methods based on standards developed within an open process by recognized industry standard-setting organizations must be required.

A. Introduction

TV content providers generate thousands of hours of closed-captioned programming each year, and there are many thousands of hours of previously aired captioned content. This content is primarily delivered to hundreds of millions of TV/set-top-box receivers which require 100% of captions to be authored based on CEA-608 and CEA-708 standards. Any technology used to repurpose and display these programs over the Internet must properly translate and transport the CEA-608 and CEA-708 captions carried within the content to the end user.

Repurposing 608 and 708 captions over the Internet will not limit the future development of technologies that improve captioning feature sets or improve the viewer experience.

B. Functional Requirements for Encoding, Transmission, and Display of Closed Captioned Programming

As discussed above, multiple methods are feasible for transmitting video programming with closed captions via the Internet. This diversity fosters innovation and is good for consumers, so long as the consumer's device is appropriately equipped (e.g., with built-in hardware and software, specialized Web browser plug-ins, and/or downloaded software applications) to decode captions with at least the same quality and control as the CEA-608/708 system enables for broadcast TV.

The major elements needed to deliver closed captions in video programming content over the Internet to the end user include:

- signaling of caption services available (type and language);
- signaling to link caption data with video data, for delivery methods where caption data is not embedded in video;
- methods of encoding and transporting caption essence and metadata;
- transcoding of one format to another, as needed, in the distribution chain;
- functional capabilities and features in the authoring environment, including:
 - positioning information in all parts of the screen, including centering, justification, and multiple lines of text; and including display of multiple regions of text (separate "caption windows") simultaneously;
 - presentation timing information, including assignment of timestamps, support for frame-accurate synchronization;

- presentation type (roll-up vs. pop-on);
- specification of a text encoding, with support for left-to-right and right-to-left segments within a vertical run; different natural languages within the same caption resource, and common typographical conventions and glyphs of that language;
- ability to pre-deliver text to a non-visible window and the ability to make a non-visible window visible at a desired timing;
- support for a range of font faces, sizes, colors, mixed display styles, caption styles.
- functional capabilities and features of the caption decoding and display system, including:
 - user control over caption display (captions on/off, preferred font faces, size, color, styles and background);
 - includes the need to tie the user’s control over caption display to overall preferences for the device, such as for example a TV’s captioning controls (“CC” button, size preferences, etc.);
 - required support of a minimum set of protocol features and capabilities (see 47 C.F.R. § 15.122), for example:
 - the ability to render a minimum number of simultaneous caption windows;
 - support for certain character sets and special symbols;
 - support for specific display modes such as roll-up, pop on, paint-on, modes.
- requirements related to distribution of content:
 - Closed-captioning data must be carried through the content distribution chain intact (e.g., in a lossless manner) and with no change in timing relative to video.
- requirements related to pre-captioned video programming content:
 - Decades of video programming have been captioned in CEA-608/708 format; therefore a lossless method must be specified for these captions to be delivered via the Internet.

C. Standards Needed

Methods based on standards developed within an open process by recognized industry standard-setting organizations are required. Standards are needed in the following areas:

1. Interchange Format. A standard interchange format is needed to reduce the cost and complexity of content authoring. By “interchange format” we mean the format of closed-captioning data carried within television content as it is distributed from the content provider to programming distributors. A distributor may be a file-streaming or download service which sends the file in the interchange format, not re-formatted, to a consumer’s device which plays the video. A distributor may be a Web site, where the content

is offered in a variety of ways to consumers who access it via a Web browser. Another type distributor may be a programming distributor who reformats the content as appropriate for delivery over the Internet to a proprietary application residing in a consumer's device.

2. **Delivery Formats.** In addition to the delivery formats, the interchange format should also be supported within streaming and file download delivery methods to end user video players/recorders, as outlined in Use Cases 1-3 in this document. Interchange and delivery formats based on standards developed within an open process by recognized industry standard-setting organizations are required. This is analogous to the way captions are defined with digital terrestrial broadcast and digital cable today. Some transcoding may be required. This should employ standardized methods of conversion, such as SMPTE RP 2052-10:2010. Such transcoding should not be onerous or require human intervention. Formats must include caption metadata (descriptions of the available service or services) and essence (the caption text and formatting commands themselves).
3. **Linkage to User's Captioning-Display Controls.** Whatever method is used to deliver the captioned programming to the end user's device, the end user must have at least the level of control over the captioning experience that the CEA-608/708 system enables. The linkage between the caption decoding and display system and the device's user interface functions (as they relate to control and display of captioning) must be standardized. For example, if captioned content is being rendered through a Web browser on an Internet-connected DTV receiver, there must be a way to control the closed captioning on Web-delivered video that is equivalent to the controls for broadcast TV captioning. The closed caption on/off function must operate on captioning rendered through the Web browser and the user's preferences for font styles, sizes and colors, and background colors and opacity must apply as well.

D. In-Band and Out-of-Band Delivery of Closed Caption Data

Closed-caption data can be delivered between systems either in-band (embedded in the video data stream or file) or out-of-band (as a separate data stream or file from the video). This is an important distinction, as the use of these different delivery techniques has a significant impact on production workflows. Consider the following use cases:

On-demand streaming (pre-recorded programming): To minimize the barriers to providing Internet closed captions, it is desirable to have platforms and applications that support both in-band and out-of-band delivery of closed-captioning data to the video player.

Out-of-band delivery is more flexible, in that it allows for the provisioning of captions and the translation to the required delivery format to occur without the need to modify the actual video container file. This flexibility is extremely important in Internet video

delivery, where video files are typically distributed and cached via content delivery networks. Out-of-band delivery also can readily accommodate provisioning of captioning data in multiple languages, facilitating workflows where the alternate-language captioning files are delivered after the video file has been transcoded and published for Internet distribution. Another benefit of out-of-band management and delivery of closed-captioning data is that captions can be more readily corrected without the need for video files being repackaged and redistributed (only the closed-captioning file would need to be updated).

There are also some workflows where **in-band delivery** of closed-captioning data for on-demand streaming will be useful, particularly in scenarios where a video and all its metadata must be represented in a single file (such as an mp4 container).

Unlike the on-demand streaming scenario for pre-recorded content, the most optimal delivery method for live simulcasting of a television channel is more likely to be the in-band method. The process for live simulcasting would flow as follows:

- The video signal for the television channel (with embedded CEA-608/708 closed captions) is used as the input for an Internet streaming encoder, which transcodes the video signal to one of several possible Internet video-streaming formats.
- In addition to video transcoding, the Internet streaming encoder will need to extract the CEA-608/708 captions from the input video signal, transform the CEA-608/708 data to the standard interchange (or delivery) format, and embed the standard formatted data in-band into the Internet video stream.
- On the receiving side, the player platform/application will need to extract the standard formatted closed-caption data from the video stream and render the captions on the display.

As there are a wide variety of workflows and processes for both pre-recorded and live simulcast streaming, VPAAC recommends that platforms and applications accommodate in-band and/or out-of-band delivery techniques as appropriate. As a precedent, the W3C proposed HTML5 specification indicates in section 4.8.10.12 that both in-band and out-of-band text tracks shall be supported. It is recommended that both in-band and out-of-band captions be supported for Use Cases 1 and 2. Managed ecosystems, (Use Case 3) need not support both.

E. Criteria Used to Recommend Caption Encoding Formats

A recommended caption-encoding format must meet the following criteria:

1. be based upon open standards from recognized industry standard-setting organizations and accommodate existing practices.
2. be convertible to and from a standard interchange format, or ideally the caption format should be recognized as a common interchange format itself as well as a delivery format (as defined in Section IV.A, above).

3. consider the needs of broadcasters and content providers who are required to caption television programming in CEA-608/708 format and who possess large libraries of content already captioned in CEA-608/708 format:
 - a. The format(s) must allow for a machine-based transcoding method to convert captions created with existing broadcast standards (CEA-608/708) to the format used for transport via the Internet and Web display.
 - b. The transcoding method must be well specified and robust.
 - c. The transcoding method must allow transfer from and to existing broadcast formats without loss, i.e. to preserve all text content, positioning and formatting information present in captions as originally authored, to give consumers the option of having the same experience on the Web as they do on television.
4. use a stable standardized format, recognizing that some classes of consumer receivers (e.g., DTV receivers) are generally not upgradable after purchase.
5. be usable on all devices that a user may use to consume television programming content (e.g., different playback/recording devices, apps, browsers).
6. enable text-based capabilities such as translation, content indexing, consumer searching, and style transformations. In practical terms this implies using character codes such as Unicode rather than relying on purely graphical images.
7. either be extensible or be able to be readily modified to support different production workflows, regional requirements (e.g., international) and future requirements as technology evolves (e.g., 3D television).
8. enable modification to the caption-format specification via an open process by recognized industry standard-setting organizations.
9. employ a format that supports captions/subtitles and other international methods, including those which deliver data that rely on graphics (e.g., DVD, European formats).

In addition, the standard interchange format and delivery solutions must address the following needs:

1. Accommodations of consumer devices which decode the video and wish to output or record an NTSC analog video signal, with standard interchange and delivery formats that provide a method to carry the original CEA-608 data, much the same way that the CEA-708 standard defines the method for carrying the field 1 and field 2 CEA-608 bytes in the MPEG-2 digital video stream in terrestrial broadcasts today.³⁵

³⁵ Note that only a subset of types of consumer devices offers an NTSC analog video output. Other types of devices disregard the CEA-608 “compatibility bytes.”

2. Continuation of existing practices for caption authoring and distribution in television systems ensuring their effectiveness on the Internet:
 - a. in recognition of and to take advantage of existing television captioning workflows and infrastructure;
 - b. to allow television content to migrate to the Internet quickly and without the need to re-author captions;
 - c. to allow consumer recording of Internet-delivered video content with closed captions and pass-through of such content with captions intact to other devices or for archive on removable media; and
 - d. to allow content owners/providers to migrate existing libraries of archived television content to the Internet, some of which may contain captions authored in legacy formats.
3. Devices should support a consistent and integrated representation of user preferences for control of their closed-captioning display options. For example, a single user setting to prefer English captions when available should apply to content arriving from any source (terrestrial broadcast, IP streaming, IP file playback, or rendering by Web browser).

F. Automated Format Translation

Existing television content containing captioning in CEA-608 format must be machine-translatable to the standard interchange format. Translation from CEA-608 (legacy analog) to CEA-708 (DTVCC) formats are routinely performed in today's production flows. Any translation (transcoding) from source material captioned in CEA 608/708 format must preserve the functional capabilities and features included in the elements listed in Section B. Not all features of CEA-608/708 are supported by decoders; see 47 C.F.R. §15.122. In their recommended practice RP 2052-10:2010, SMPTE has defined a translation model which maps CEA-608 caption data to SMPTE Timed Text. Follow-on work in SMPTE will define a similar mapping from CEA-708. For compatibility with CEA-708 source formats, the FCC should consider the provisions of 47 C.F.R. §15.122, which define the required capabilities of the DTV receiver (including cable and satellite) for rendering of closed captioning. This functionality encompasses most of the capabilities of CEA-608 and key features of CEA-708. Since DTV receivers must include caption decoders complying with the FCC rules, this is a good model.

G. Recommended Standards to Accomplish These Goals

1. Interchange Format

VPAAC considered several technical solutions including SMPTE Timed Text (SMPTE-TT), W3C Timed Text Markup Language (TTML), and WebVTT. VPAAC recommends that the industry use SMPTE 2052-1:2010 *Time Text Format (SMPTE-TT)*. Of the solutions available, SMPTE-TT best meets all the requirements listed above. We further note that SMPTE-TT is already being employed in production environments to repurpose television content for Internet use. These environments employ equipment that has implemented SMPTE Recommended Practice 2052-10, *Conversion from CEA-608 Data to SMPTE-TT*, which describes the

process whereby captions authored in CEA-608 format may be machine-translated to the XML format employed by SMPTE-TT.

In addition, the Entertainment Technology Center (ETC)³⁶ has recently released an “Interoperable Master Format” specification employing SMPTE-TT for support of captioning and subtitling. The format will be used for business-to-business interchange of media assets. The specification was driven by the major Hollywood studios and post-production houses and has been contributed to SMPTE for standardization.

On May 3, 2011, SMPTE issued a press release titled “SMPTE Makes Closed-Captioning Standard Freely Available, Widening Access to Broadband Video for Individuals with Disabilities.” The standard may be downloaded without charge at <http://www.smpete.org/standards/st2052-1-2010.pdf>.

2. Delivery File Formats

Case #1:

To support Use Case #1, where the video player is resident in the consumer device, a common file format is needed in which closed-captioning essence and metadata are transported. This file format must be suitable for both streaming (where the content is played, after some small amount of buffering, as it is received) and file download (where a complete file is downloaded, stored, and played later).

SMPTE-2052 (SMPTE-TT) is recommended as the standard caption-data encoding format for delivery of captions in content intended for consumption by a consumer video player. Captioning in Internet-delivered content created for playback on a standard consumer video player should be created under the guidelines established in SMPTE Recommended Practice 2052-10, *Conversion from CEA-608 Data to SMPTE-TT*, operated in “Preserved” mode (see Sec. 5.5).

The method whereby SMPTE-TT is encapsulated in a file wrapper should be left to the industry to define. One large industry consortium³⁷ has recently released a common file format specification employing the Base ISO File Format (sometimes known as MP4), and SMPTE-TT for support of captioning and subtitling functions in Internet-delivered content.

Cases #2 and #3:

Use Cases #2 and #3, by definition, require a specific standard distribution format based on standards developed within an open process by recognized industry standard-setting organizations, because it is envisioned that the programming distributors, Web site operators, and application developers involved in those use cases will provide all the necessary software (e.g., applications and browser plug-ins) to decode whatever format is used in distributing the captioned content.

³⁶ <http://www.etcenter.org/imf/>.

³⁷ The Digital Entertainment Content Ecosystem (DECE) LLC, <http://www.uvu.com>.

3. Standardized Linkage to User Captioning Display Controls

It is recommended that recognized industry standard-setting organizations propose voluntary standards for the needed functionality acceptable to the stakeholders in the affected industries. These standards would apply to cases where captioning is rendered via a Web browser (Use Case #2 above) and where captioning is rendered via a downloaded application (Use Case #3 above). Note that for Use Case #1, the resident player has direct access to the platform's captioning-display controls, hence no further standards are needed in that case.

VI. NEW DEVELOPMENTS

A. Considerations for the Evolution of the Field

As described above, Use Cases #2 and #3 allow the programming distributor to deliver captioned content using non-standard or proprietary protocols and methods, as long as the user (at his or her choice) can duplicate the experience of watching captioned television as it would have been experienced had the content been received on a broadcast DTV receiver. Web servers can be configured and applications can be designed that offer alternative experiences and potentially improved user experiences via advanced features for the display and consumption of captioning.

B. Emerging Protocols

As discussed in other sections of this document, protocols and formats exist today that can be utilized to achieve the goal of migrating content originally captioned for TV to the Internet. While this means there are no immediate needs for new standardization efforts in order to meet the requirements of the CVAA, this is not intended to prohibit or constrain any future development or innovation in this area. As with all technology, it is certain that technical capabilities for content distribution and consumption will continue to advance beyond the use cases set forth in this document. We encourage continued and open innovation in the field of accessibility, and this innovation should be unrestricted. Where technology advances are incorporated in an advanced standard developed within an open process by recognized industry standard-setting organizations, the VPAAC recommends that the FCC consider use of that advanced standard.

C. Advancing Innovations for User Experience

Innovation and feature enhancement shall be incorporated when achievable. Likewise, nothing in this document shall be construed to prohibit the use of advanced features and/or advanced user controls.

D. HTML5 and Direct In-Browser Support for Captioning

The development of native browser support for video, captions, and subtitles marks a significant shift in technology on the Web that would allow captions to be rendered directly by a browser. In 2007, the W3C and WHATWG introduced a new <video> element for HTML5³⁸, and in July 2010, a <track> element was added as part of the <video> element to allow linking to subtitle and caption files.³⁹ (HTML5 does not currently specify that browsers must support any particular caption-file format.) The WHATWG also proposed a new subtitling format called WebVTT.⁴⁰ Support for the <video> and <track> tags in HTML5 are under active development by major browser organizations.

VII. SCHEDULE OF DEADLINES FOR THE PROVISION OF CLOSED CAPTIONING

The VPAAC is offering these definitions and deadlines without knowing how technology may change in the future. There are a large number of possible ways to deliver captioning via the Internet given the many forms of video programming contemplated by the Act. The VPAAC has not identified and is divided as to whether these practices should be modified. Therefore, we offer these definitions and deadlines based on certain known scenarios, and recommend for new and emerging technologies that these definitions should be used in a way that best encourages the captioning of all programming. These rules should be interpreted broadly as to include emerging technology and ways of delivering programming on the Internet.

The VPAAC proposes the following definitions for certain categories of content:

- Live programming: programming created and presented on television and simulcast for Internet distribution to the end user as it airs on television.
- Near-live programming: near-live content is defined as any programming that was produced from start to finish within 12 hours of being published or exhibited on television. Production work is typically completed too close to air time for offline captioning workflows. Near-live programming remains an issue; see Appendix C.

³⁸ See <http://lists.whatwg.org/pipermail/whatwg-whatwg.org/2007-February/009702.html>.

³⁹ See <http://lists.whatwg.org/htdig.cgi/whatwg-whatwg.org/2010-July/027386.html>.

⁴⁰ See <http://www.whatwg.org/specs/web-apps/current-work/webvtt.html>.

- Prerecorded and edited for Internet distribution to the end user: any programming that is prerecorded and has been substantially edited for Internet distribution to the end user. Examples of substantial edits include alternate music scores or scene deletions (typically due to rights restrictions) that prevent broadcasters from directly re-purposing the broadcast captions for Internet distribution to the end user.
- Prerecorded and unedited for Internet distribution to the end user: any programming that is prerecorded and has not been substantially edited for Internet distribution to the end user. For the purposes of this definition, changes in the number or duration of advertising pods from those in the material as broadcast are not considered as program edits. Thus a program with the exact same content version on television for Internet distribution to the end user, but with different advertisements, would be classified as an unedited program.

Each of the content categories defined above shall have a compliance date that identifies the date for which content in that category will require captioning for Internet distribution to the end user, in cases where the content has aired on television after the compliance date. The VPAAC proposes the following schedule:

- Effective six months after rules are published in the Federal Register: programming that has been prerecorded and unedited for Internet distribution to the end user.
- Effective twelve months after rules are published in the Federal Register: live and near-live programming.
- Effective 18 months after rules are published in the Federal Register: Programming that has been prerecorded and substantially edited for Internet distribution to the end user.

APPENDIX A – SUMMARY OF DTV RECEIVER REQUIREMENTS

The following are recommended requirements for receivers capable of rendering closed captioning for television content delivered via the Internet. The requirements are those established for DTV receivers and specified in Title 47 C.F.R., Part 15, § 15.122:

- 1) Font (character) colors:
 - a) Required character foreground and background colors are white, black, red, green, blue, yellow, magenta, and cyan.
- 2) Font opacity:
 - a) Requires transparent, translucent, solid and flashing character type attributes.
- 3) Font sizes:
 - a) Standard, small, and large font sizes shall be available.
- 4) Font styles:
 - a) Eight font styles shall be available.
- 5) Character edge attributes:
 - a) Edge options shall be none, raised, depressed, uniform, or drop shadowed.
- 6) Color attributes of the edges of character foregrounds may be specified separately from the character background foreground and background.
 - a) Edge opacities shall have the same attributes as the character foreground opacities
- 7) Caption window colors:
 - a) Required colors are white, black, red, green, blue, yellow, magenta, and cyan.
- 8) Caption window opacity:
 - a) Transparent, translucent, solid type attributes are required.
- 9) Caption services:
 - a) The number of possible caption services shall be six: Caption Service 1 through 6. These services are used for alternate languages and “easy reader.”
 - b) Users shall be able to identify caption services available for a given program.
 - c) Users shall be able to choose at least one of these services to display onscreen.
 - d) It is not necessary to display more than one caption service for a given program at the same time, though manufacturers are free to permit such operation.
- 10) Character code set:
 - a) The character code set shall support the set of languages permitted in CEA-608/CEA-708 as reflected in SMPTE RP 2052-10, Table 3.
- 11) User settings:
 - a) Users shall be able to choose settings that display captions as intended by the caption provider.
 - b) Users shall have the ability to choose among eight font sizes.
 - c) Users shall have the ability to choose and select among eight font styles.

- d) Users shall have the ability to choose and select among the eight foreground and background colors.
- e) Users shall have the ability to choose and select window opacity.
- f) Once the viewer chooses a set of customized caption display features, such as font and/or color, these settings shall remain in effect until subsequently changed by the user.

APPENDIX B – EQUAL CAPTIONING EXPERIENCE

Representatives of deaf and hard-of-hearing consumers provided many useful suggestions about practices that would help ensure a quality experience when watching captioned TV programs delivered over the Internet. While there is not consensus about whether these practices should be mandated or only offered as suggestions, it is believed the following do provide helpful guidance about what consumers want for entities that provide closed-captioned programming on the Internet.

1. Programming broadcast with captions that indicate there is no audio for people seen talking shall also provide this information in captioning over the Internet.
2. Programming broadcast with captions that consistently identify speakers or sound effects shall be delivered over the Internet with the same captioning consistency in identification.
3. Broadcast programming with captions that convey meaning and emotion through non-verbal elements such as background sound or changes in a speaker's voice through pauses, pitch, volume, pace, and inflection are delivered in the same manner over the Internet.
4. Programming broadcast with captioning that identifies sound or speech in the foreground, background, or off screen shall deliver this information over the Internet.
5. Programming broadcast with captions that provides a word-for-word transcription of background audio shall also be provided over the Internet.
6. Captioning of song titles in broadcast programming that includes thematic music shall also be provided over the Internet
7. Programming broadcast with captions that include non-speech sounds such as "gasp," "grunt," and "groan" shall also be provided over the Internet.
8. Poems should be rendered visibly on the screen as they were originally written.
9. Programming broadcast with captions that positioned captions in a way to help identify multiple speakers shall also be provided over the Internet.

APPENDIX C – UNRESOLVED ISSUES

Working Group 1 and the entire VPAAC worked diligently to provide the best guidance to the FCC with the above document, attempting to achieve a rough consensus on most issues. However, there are a few items that, due to the constricted timeline and/or strongly held positions, will need to be resolved by the FCC through its upcoming NPRM and eventual R&O. It is recommended that these issues be surfaced for further input in the upcoming NPRM.

1. User-Controlled Placement of Captions and Positioning

Mandated user control of the placement of captioning, so that captions may be moved to a preferred location determined by the viewer, is indicated as a highly desired feature by some, while others felt that there may be technical limitations making it difficult to accomplish on some devices. Some objected to allowing users to reposition captions to locations that would overlap specific portions of the video image.

In addition, some preferred to have an option to have all captions placed in a region below the video; for example, placing the captions below a widescreen/letterbox formatting of the video. Others were concerned that for some devices, this could not be readily accomplished, if at all.

2. Timeline for Implementation of User-Controlled Features

A related issue is the timeline for roll-out of to-be-agreed-upon user-controlled required features (those that match the present set of user controls available in HDTV/CEA-708). One group suggested that a minimum of 24 months from issuance of final rules is needed for build-out of the software and hardware to accomplish this task. Others felt this was too long a period to wait and could result in an unacceptable gap for the consumer between basic captions and the eventual full-featured set. An agreed-upon schedule for the implementation of basic captioning is included in this report; only the timing for availability of the user-controlled feature set is in dispute.

3. Responsibility to Assure Caption Delivery

Determination of the responsibility for assurance of delivery of captions is a complicated issue that the groups did not have time to discuss in full; that is, what roles and requirements should be apportioned to various stakeholders?

4. Near-Live Programming Defined for Use in the Schedule of Deadlines

In Section VII, near-live programming is defined as a content category to determine the schedule of deadlines for the provision of closed captioning. There is, however, a difference in perspective between industry and consumer groups on the definition of near-live programming and how the definition may

affect eventual rules; this difference manifests itself in a variety of ways, including the use of the word "substantively" in alternate definitions. As the time between production and distribution of videos and their captions shrinks to fewer than 24 hours, production workflows, processes and deadlines will all affect when captions are delivered and in what format (e.g., real-time or off-line captions).

It is understood that this definition of near-live programming is only to be used for determining the schedule of deadlines for the provision of closed captioning.

5. Scope of Rules and Requirements

The scope of the rules and specific requirements for different devices and platforms is an issue that demands further discussion and attention. There are a variety of possible ways for the new rules to be applied to mobile devices, desktop and mobile computers and even newly emerging HDTVs with embedded browsers. These variations will need to be determined during the upcoming rulemaking process.

6. "Good-Faith" Effort and "Economically Burdensome" Considerations

The VPAAC did not address two additional subjects: a definition of good-faith effort to identify video programming; and the determination of economically burdensome relative to services, programs and equipment. Working Group 1 considered them to be beyond its scope. It was felt these were best left to the FCC's NPRM process.