BEGINNER'S GUIDE TO CAPTIONING
Contents

WHAT
What are captions?

HOW
How do you create captions?

WHERE
Where do you publish captions?

WHY
Why should you caption?

WHO
Who is 3Play Media?
What Are Captions?
What are captions?

Defining Captions

Captions are text that has been time-synchronized with the audio track and appear on-screen while the video plays. They originated in the 1980s as an FCC mandate for broadcast television in the US. Captions display dialogue and describe relevant non-speech elements like sound effects and speaker identification that are necessary to the viewer’s understanding of the video – especially if the viewer cannot hear the audio.
Captions vs. Subtitles vs. Transcripts

Captions assume the viewer can’t hear and display words in the same language that is spoken in the video. Subtitles assume the viewer can’t understand the language and display a translation of the spoken words and exclude non-speech sounds. Transcripts are merely the text version of the audio and are not time-synchronized with the video but are the first step to creating captions. In some countries, like the UK, the word “subtitles” is used for both captions and subtitles as described here.
How Do You Create Captions?
How Do You Create Captions?

DIY Method: YouTube

The best way to create captions yourself is with YouTube’s caption editing interface. YouTube uses automatic speech recognition to transcribe the audio of videos you upload. Since they’re computer generated, they will only be around 60–70% accurate, but can save you a lot of time by generating a rough draft of the transcript. You can then edit these autocaptions in the Creator Studio and add or correct punctuation, misspelled words, timings, and other errors to make proper, accurate captions. Finally, you can download the captions for use elsewhere when you’re done.
Quality Standards and Best Practices

To create an equivalent experience for viewers who cannot hear the audio, captions and transcripts should be as easy to read as possible. At the minimum, that means including proper punctuation, grammar, and accurate spelling. The industry-wide standard for word-to-word spelling accuracy is at least 99%.

Non-Speech Sounds & Speaker IDs

Besides accurate transcription of the text, captions should also include relevant non-speech sound effects and speaker identification.

Caption Appearance

Fonts should be non-serif, only 1 to 3 lines of text should appear on-screen at a time, and each line should have a maximum of 32 characters. Captions should appear for a minimum of 1 second but not linger too long after the speaker is finished. Captions should appear in the lower third of the screen but should move if they are blocking any text or other important information.

Verbatim vs. Clean Read

For broadcast and entertainment, every scripted or intentional “um,” pause, stutter, and stammer should be included. For other content like lectures and live presentations, remove these sounds for a “clean-read” transcript that is easier for the viewer to follow. Punctuation should be included for ease of reading.
On Accuracy Rates

As previously mentioned, captions and transcripts must be at least 99% accurate. But what exactly does “accuracy” mean when it comes to captioning? The chart below demonstrates this concept more clearly. For example, with a word-to-word accuracy rate of 95%, 1 out of every 20 words will be inaccurate. That means that *every* word in a sentence will have a 1 in 20 chance of being wrong. At a word-to-word accuracy rate of 95%, a 10-word sentence will only be 60% accurate!

<table>
<thead>
<tr>
<th>Word-to-Word Accuracy</th>
<th>1 of x Words Incorrect</th>
<th>8-Word Sentence Accuracy</th>
<th>10-Word Sentence Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>1 of 2</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>67%</td>
<td>1 of 3</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>75%</td>
<td>1 of 4</td>
<td>10%</td>
<td>6%</td>
</tr>
<tr>
<td>85%</td>
<td>1 of 7</td>
<td>27%</td>
<td>20%</td>
</tr>
<tr>
<td>90%</td>
<td>1 of 10</td>
<td>43%</td>
<td>35%</td>
</tr>
<tr>
<td>95%</td>
<td>1 of 20</td>
<td>66%</td>
<td>60%</td>
</tr>
<tr>
<td>98%</td>
<td>1 of 50</td>
<td>85%</td>
<td>82%</td>
</tr>
<tr>
<td>99%</td>
<td>1 of 100</td>
<td>92%</td>
<td>90%</td>
</tr>
</tbody>
</table>
On Speech Recognition Software

Automated speech recognition (ASR) software can transcribe spoken audio extremely quickly. This technology is commonly used in some live TV captioning and to transcribe video for some online video players, like YouTube. Unfortunately, even with the latest technology, accuracy rates with ASR rarely exceed 80-90%. Commonly, ASR will make errors that make sense acoustically but not linguistically. It will also have problems making judgment calls because it cannot understand the context of the spoken audio, making it more likely to mix up words like “can” and “can’t.” Videos with multiple speakers, speakers with accents, fast speakers, background noise, and other audio that is hard to understand will have more errors when relying solely on ASR.

Other Things Lost with ASR

Spelling is huge, but it’s not the only issue to look for when using ASR for captioning. Most ASR programs are not going to understand every proper noun, name, and brand, so a lot of words go uncapitalized or are capitalized by accident. Speaker labels and punctuation, like commas and periods, will also commonly be left out.
Where Do You Publish Captions?
Where Do You Publish Captions?

Captions are supported on most devices where you can publish video. Most online video players and platforms have caption functionality, and many have advanced features that allow you to customize caption appearance on-screen via user controls.

**Side Car File:** With online video players like YouTube and Vimeo, captions are published as a side-car file. A side-car file is a separate file that is uploaded along with your video to your video platform. When captions are uploaded as side-car files, they play along with the video and can be turned on and off.

**Encode:** Sometimes, you’ll need to encode captions. Common use cases for caption encoding include kiosk videos and other offline videos. In this case, captions are encoded into the video and can be turned on and off.

**Open Captions:** Open captions are burned directly into the video to create a single file. These are captions that cannot be turned off.

**Integrations:** When publishing video online, integrations between captioning vendors and video platforms can automate and simplify the caption publishing process.
Common Caption Formats

The kind of caption file you need to publish depends on which video player you’re working with. Below are examples of two of the most commonly used caption file formats: SRT and SCC. The images below are different formats of the same transcript file.

**SRT**

The SubRip Subtitle (SRT) caption file format is very common for online video players and is very easy to read. It’s essentially a basic text file with caption frame numbers, time codes, and lines of text.

**SCC**

SCC files use hex codes, combinations of letters and numbers, to represent text. Caption format converters are recommended when working with this format.
Why Should You Caption?
Why Should You Caption?

Accessibility

Over 48 million Americans are deaf or hard of hearing\textsuperscript{11} and captions make videos accessible to these individuals who cannot hear the audio. This is the primary and initial reason captions were introduced, but there are numerous other benefits, as well.

Better Comprehension

Captions ensure any viewer – not just those with hearing loss – can understand the dialogue spoken in a video, especially when thick accents, background noise, and unfamiliar words or phrases are involved. In a joint study we conducted with Oregon State University (OSU), it was found that 80% of students – including those who have no difficulty hearing – said they found captions helpful as a learning tool.\textsuperscript{14} People who
speak English as a second language (ESL) also use captions because they aid with English comprehension.

**Video Search**

In the age of Google, we’re all getting used to being able to “search and go.” You can do the same with video when you associate an interactive transcript with your video. This functionality allows users to search a keyword within a transcript and jump to the point in the video when that word is spoken. An MIT survey discovered that 97% of students using a video search feature found it enhanced their learning experience.

**Flexibility in Noise Sensitive Environments**

Noise sensitive environments – the library, the gym, the subway, the office – make captions necessary, especially without headphones. We don’t always want to play the audio of a video out loud on our devices. Facebook found that 80% of users react negatively when “feed–based mobile video ads play loudly when people aren’t expecting it” and that captions increase view time of video ads by an average of 12%.
Search Engine Optimization (SEO)

Search engines cannot “watch” your video. Captions and transcripts provide metadata Google can crawl and index, allowing it to “read” spoken audio content and give a video the ultimate SEO advantage. Web television distributor Discovery Digital Networks found an overall increase of 7.32% in lifetime views after adding captions to YouTube videos, and radio program This American Life found that 6.26% of all unique visitors who came from search traffic landed on a transcript page.\textsuperscript{viii} You can also get these SEO benefits by injecting the transcript into the HTML of a webpage, making your video more discoverable without having to publish a transcript on the same page.

Reusability for Other Content

After a video recording is transcribed that transcript can be used to easily create other media in addition to captions. At the University of Wisconsin, 50% of students repurposed transcripts of in-class videos as study guides for their courses.\textsuperscript{x} Professors can also use these transcripts to create course materials. Additionally, many
businesses will repurpose transcripts of recorded events like webinars to create marketing content like infographics, white papers, case studies, and other media.

**Translations**

In addition to captions, video transcripts are the first step towards creating video subtitles. Subtitles require a transcript in the source language before being translated into the target language. Having subtitles makes your video content accessible to audiences on a global scale.

**Required by Law**

In many cases, captioning is required by disability and civil rights laws to ensure video content is accessible to those who are deaf or hard of hearing.
Captioning Laws & Lawsuits

It’s important to stay on top of the legal requirements surrounding web accessibility in order to provide the necessary accommodations to individuals with disabilities, as well as to be in compliance with the laws and avoid potential lawsuits. With several different laws surrounding accessibility, you may find it hard to figure out what exactly you should be doing. Here’s a rundown of laws that apply to captioning.

Anti-Discrimination Laws

In the US, anti-discrimination laws require closed captioning for video. These include the 21st Century Video Accessibility Act (CVAA), the Americans with Disabilities Act (ADA), and Section 504 and Section 508 of the Rehabilitation Act.
The Americans with Disabilities Act (ADA)

The ADA guarantees equal opportunity for individuals with disabilities in employment, state and local government services, public accommodations, commercial facilities, and transportation. Title II – Public Entities – and Title III – Places of Public Accommodation – most directly affect web accessibility.

While the ADA doesn’t specifically name video captioning as a requirement, its vague wording has been disputed in several lawsuits against organizations (such as Netflix, Hulu, Amazon, FedEx, and institutions like MIT, Harvard, and UC Berkeley) for failing to provide captions on videos, thus violating Title III of the ADA, which prohibits disability discrimination by “places of public accommodation.”
Rehabilitation Act: Sections 504 and 508

Section 504 of the Rehabilitation Act protects the civil rights of people with disabilities by requiring that federal or federally funded organizations make accommodations for equal access. For users who are deaf or hard of hearing, video captioning must be provided.

Section 508 of the Rehabilitation Act was signed into law in 1998 and requires that Federal agencies make their electronic communications and information technology accessible. Therefore, all film, video, multimedia, and information technology produced or procured by Federal agencies must include captions for audio. It is also worth noting that many organization and state laws directly require compliance with Section 508, so those same requirements can extend beyond Federal agencies.

In early 2017, the language in Section 508 went through a refresh, which updates requirements for information and communication technology covered by Section 508. The Section 508 refresh, which takes effect in January 2018, assigns WCAG 2.0 success criteria to each existing section of the 508 Standards in an effort to modernize the requirements for captioning.
Captioning Under the New Section 508 Refresh

Changes to Section 508 include:

- Categorization by functionality instead of by product type
- Revisions to improve ICT usability, including interoperability with assistive technologies
- Clarification on the types of ICT covered
- Correlation of WCAG 2.0 Level A and Level AA standards to all Section 508 requirements

The Section 508 refresh assigns WCAG 2.0 success criterion to each existing section of the 508 Standards. The standards modernizing the requirements for captioning of prerecorded video content are as follows:

- Success criterion 1.2.2
- WCAG 2.0 Level A
- 508 section 1194.22(b) and .24(c)
- Synchronized captions are provided for non-live, Web-based video

WCAG 2.0 Guidelines for Closed Captioning

The Web Content Accessibility Guidelines (WCAG) 2.0, the international standard for web accessibility, requires captioning under guideline 1.2 for time-based media. WCAG 2.0 is the legal requirement under the recently revised section 508 standards and is often required by state laws. Most web accessibility experts recommend striving for
WCAG 2.0 Level AA compliance, which requires that captions be provided for all prerecorded and live audio content in synchronized media.\textsuperscript{11}

**21st Century Communications and Video Accessibility Act (CVAA)**

Signed by Obama in 2010, the CVAA sets clear rules for internet video programming and clips that previously aired on TV in the US. It is enforced by the FCC and has been phased in for video creators over several years (see timeline below).
As of September 30, 2013, all broadcast video that previously aired on US television with captions must include captions if published on the internet. Stricter CVAA compliance deadlines have passed for video montages and live footage.iii

**Captioning Lawsuits**


In 2012, the NAD argued that Netflix barred deaf users from enjoyment of their streaming video service by not captioning videos. Netflix unsuccessfully argued that ADA Title III did not apply to their online business. Instead, the court ruled that Netflix is a “place of public accommodation” and that the ADA does apply. The case was settled and Netflix agreed to caption all videos by 2014.xiii

**EEOC, NAD vs. FedEx**

In October of 2014, the US Equal Employment Opportunity Commission (EEOC) filed a lawsuit against FedEx Ground for allegedly violating the ADA over failing to provide sign language interpreters, closed captioned training videos, and other reasonable accommodations to their numerous deaf and hard of hearing employees across the country. In 2015 the NAD joined in support of the EEOC and in January of 2016, a federal
judge denied FedEx’s motion to dismiss the case and motion to strike.xiv The case remains in litigation.

**National Association of the Deaf, et al v. MIT/Harvard**

The National Association of the Deaf argued that both schools are violating the ADA and Section 504 of the Rehabilitation Act for the lack of quality, comprehensive video captioning in the massive open online courses they make available to the public. A motion by both schools to dismiss the case was opposed by the DOJ and denied by Federal Magistrate Judge Katherine J. Robertson in February of 2016.xvi The case is currently being litigated.
Other Actions by the DOJ and OCR

Sometimes, after receiving letters of complaint about website accessibility issues, the DOJ or the Department of Education’s Office of Civil Rights (OCR) will investigate those claims and urge the organizations to make the necessary changes. These are often referred to as “dear colleague” letters and are very helpful to organizations in that industry because they list in detail exactly what the DOJ and OCR consider satisfactory compliance criteria. While these letters don’t necessarily set a legal precedent, they do help to inform the industry of the DOJ and OCR’s stance on major issues, as well as the standards, requirements, and solutions that they believe organizations should be following with regards to web accessibility. Other documents, known as consent decrees, arise from legal settlements and outline the steps the defendant organization agrees to take to remedy the situation, which gives other organizations a checklist with which to compare their own accessibility compliance.

UC Berkeley

Recently, the University of California, Berkeley chose to remove their entire library of over 20,000 online course video and audio files from public access. The move was in response to a letter from the DOJ that found their lack of captions a violation of the ADA. This outcome was
unfortunate, but it shows how seriously captioning is taken by the government and how much of a liability un-captioned, public-facing videos can be for institutions.

**Miami University**

In January of 2014, Miami University, Ohio was sued for not providing speech-to-text software and Braille alternatives to a blind student, and eventually agreed to settle the case in 2015. As a result, the consent decree required that the university not only provide these alternatives for blind and low vision students but also “requires Miami to provide [all] qualified individuals with disabilities with an equal opportunity to participate in and benefit from Miami’s services, programs, and activities, as Miami increasingly relies on web-based, digital, and emerging technologies.” In other words, the rest of the campus, including its course videos and other web-based course materials, had to be made accessible, too. The decree lists timely access to closed captions and audio description as requirements for making video accessible.
Conclusion

As technology continues to advance, video accessibility will become more and more important – both in terms of legal requirements and the needs of our world. Captions have become much more ubiquitous with the growth of online video. Laws continue to close the accessibility gap by requiring captioning in more places where video is found, but people are also starting to appreciate the universal benefits of transcription and captioning. As video continues to dominate the web and other public spaces, captioning will become less of an afterthought and more of a necessity in an increasingly accessible world.

About 3Play Media

3Play Media provides cost-effective, premium quality captioning, transcription, and audio description to more than 2,000 customers in higher education, enterprise, entertainment, media, and government. 3Play Media simplifies the process of making videos accessible through flexible API's, integrations with video players and platforms, simple plugins, and a user-friendly online account system. 3Play Media is based in Boston, MA and has been operating since 2007.
3Play Media's Captioning Process

At 3Play Media, we use a 3-step process which combines ASR and human editing. Our advanced technology enables us to offer competitive rates (including volume-based discounts), and our multi-step quality assurance measures ensure we deliver premium quality captions, subtitles, and transcripts far more efficiently than traditional methods. Our average measured accuracy is 99.6%, and we guarantee over 99%, even in cases of poor audio quality, multiple speakers, difficult content, and accents. Our staff of more than 1,500 transcriptionists gives us the flexibility to assign complex or technical content to transcriptionists with a discipline-specific expertise or a familiarity with a certain accent, enabling us to process a broad range of complex content to a consistently high quality.

Contact

3Play Media
34 Farnsworth St, 4th Floor
Boston, MA 02210

Website – www.3playmedia.com
Email – info@3playmedia.com
Phone – (617) 764-5189
ii http://www.3playmedia.com/services-features/tools/integrations/
iii http://www.hearingloss.org/content/basic-facts-about-hearing-loss
vi http://www.3playmedia.com/customers/case-studies/mit-opencourseware/
viii http://www.3playmedia.com/resources/white-papers/wp-seo/
x http://www.3playmedia.com/customers/case-studies/university-wisconsin/
xi https://www.w3.org/TR/2008/REC-WCAG20-20081211/#media-equiv-captions
xii https://www.fcc.gov/consumers/guides/captioning-internet-video-programming
xiv http://www.3playmedia.com/2016/03/15/fedex-ground-sued-by-eeoc-nad-for-ada-violation/
xvii http://www.3playmedia.com/2016/10/05/uc-berkeley-responds-to-doj-pressure-over-web-accessibility/
xviii https://www.ada.gov/miami_university_cd.html