

KELLY MAHONEY: My name is Kelly Mahoney, and I'm a member of the marketing team here at 3Play Media. Just a quick note on my physical appearance for anyone who can't see me. I'm a young woman with long, dark brown hair, and I'm wearing a green dress today. And as a content marketing associate at 3Play, I create content on anything accessibility and the cutting edge of captioning using platforms like our blog, social media, and doing presentations like this.

So speaking of presentations, let's talk about today's agenda. Today, for most of our time together, I'll be discussing the details of captioning, including live captioning and [AUDIO OUT] asking yourself, what is the difference? That's why you're here. So after that, I'll go into a little bit about 3Play Media, who we are, and what we do. And then with the time left over, we'll dive into that Q&A session.

So to start off, what are captions? We should probably start by defining the basics. The most important things that you need to know about captions are that they are always time-synchronized. They are the text that can be read while watching a video, and they're usually indicated by a capital-CC icon.

Captions are provided as an accommodation for deaf and hard-of-hearing viewers, and they were originated as the result of an FCC mandate in the 1980s that originally required captions on all broadcast media. But the use of captioning and the requirements [AUDIO OUT] expanded to include online video and internet applications. We'll get into that a little later.

But because captions are intended for viewers who cannot hear audio, they should also include relevant non-speech elements, like sound effects or speaker identifications, that are critical for a viewer to understand who is speaking or what sounds are happening in the environment. So for example, if, onscreen, a character is unlocking a door, and you can see their keys jingling, you wouldn't have to describe that the keys are jingling. But if that noise is taking place offscreen, you would want to convey that information to your viewers.

Now, it's important to distinguish between captions versus subtitles versus transcripts, because these terms are often used interchangeably but actually [AUDIO OUT] different processes. So captions, we just went over. These assume that the viewer cannot hear the audio. They are time-synchronized, and they include non-speech elements.

Subtitles, on the other hand, assume that the viewer can hear but cannot understand the audio. They are also time-synchronized, but the primary purpose of subtitles is to translate spoken audio. And finally, transcripts are a plain-text version of a piece of content's audio. These [AUDIO OUT] audio-only content, and they're literally just a written version of anything that is spoken in a piece of content.

Now, we make this distinction because in the United States, the difference is important between captions and subtitles. But in other countries, those words may be used interchangeably. In the UK, for example, captions and subtitles mean the same thing. So if you're joining us from somewhere else, it may mean something different to you. But this is what we are intending when we describe captions versus subtitles.

Out of those three, the most common is, by far, closed captioning. And there's actually a variety of ways that you can get closed captioning. You can do it yourself. You could use an automatic speech-recognition software, or you could work with a captioning vendor. So [AUDIO OUT] to manually transcribe and create captions yourself.

And the reason I say if you have the time is because this is definitely the most labor-intensive method of captioning, particularly when you factor in the amount of time that it takes to add those non-speech elements and to synchronize the time codes so that everything matches up. DIY captioning could take an inexperienced captioner up to five to six times longer than the video's length to complete. So this method could also potentially be very costly to scale if you have a lot of video content that you're looking to have captioned.

So the second method of captioning entails the use of automatic speech-recognition software to create a baseline, first-draft transcript. This software is great for a first draft, like we say, because it'll give you a rough idea of what is being spoken using automatic speech-recognition software. But ASR engines are usually not sufficient on their own, and they do not provide a high-enough level of accuracy to be fully considered completely accessible. So this method often requires additional manual editing.

If your organization does choose to go this route, it is totally possible. And we like to talk about YouTube's automatic caption generator to create those captions. You can upload a transcript, and YouTube will automatically align the transcript to the audio as well as break it up into appropriately-timed caption frames. And it does a pretty decent job of doing this. Certain factors will aid in an easier first-round transcription, which I'll get into a little later when we talk about live captioning. But once again, you would probably still have to go back and edit that script for accuracy, add those non-speech elements, and just make sure that there were no weird substitutions.

Finally, if you don't feel like worrying about any of this, you could always outsource the work to a captioning vendor [AUDIO OUT] media. No matter who you work with, you should always be looking for a minimum guarantee of 99% accuracy because that is the industry standard for accessible captions.

All right, so captioning quality. There's a lot that can affect accuracy rates. The industry standard is 99%, like I said. And although that's pretty close to 100% accuracy, there's still a 1% chance of error. So practically, what this means is, in a 10-minute video file that contains 1,500 words, there's going to be about 15 errors total.

And while it's nice to strive for perfection, a truly 100%-accurate captioning file is not really attainable, even when captioning is performed by a human. And in certain cases, there are different ways that you should write your captions, depending on what the content includes. So if your video is scripted content, you should ensure that captions are written verbatim.

For example, if you've ever turned on a TV show or a movie, and you see um's and uh's written into the captions, that's because they're intentional, and they're a part of the spoken dialogue. On the other hand, for lectures or live events, a clean read is preferable, meaning that those filler words will be eliminated so that viewers can more easily read the captions that are in front of them. In either case, there are general guidelines for making sure that captions are easy to read, whether you're writing them verbatim or clean read.

So each caption frame should have around one to three lines with about 32 characters per line. And although they should be time-coded to go along with the media, they should appear on screen for a minimum of one second to make sure that viewers have time to read. Best to use a non-serif font. Those are easier to read. And captions are typically placed in the lower-center portion of the screen. But they should be moved if they're in the way of other important text or visual elements.

For silences or long pauses in dialogue, you want the captions to disappear so that viewers don't mistakenly believe anything is still occurring when nothing is happening. And if you want more information or you're curious about what you do or don't need to caption, we have some links in the chat about the Described and Captioned Media Program, the DCMP, as well as the FCC guidelines on what should be captioned, as well as the Web Content Accessibility Guidelines. But we're going to talk about the Web Content Accessibility Guidelines more in-depth towards the end of this presentation as well. So stick around.

All right, so once you have your captions written, you have to publish them. There's a lot of different ways you could do this, just like with writing the captions. The most common way to publish captions is as a sidecar file, which is essentially a file that stores the caption data and is associated with your video. So if you've ever uploaded a caption file alongside a video onto YouTube, you've already done this. You already know what you're doing. And this format still gives users control of being able to turn on or off the captions when they're viewing the end product of your video. So regardless of providing captions, the users are still in control.

Another way to do this is to encode captions directly onto the video. These, again, still give users the ability to toggle on and off. But these kinds of videos, with encoded captions, are most often found on kiosks or offline videos that don't always have internet.

On the other hand, open captions are burned into a video and cannot be turned off. So the user doesn't have a choice whether they see these captions or not. And the most common example of these would be on social media. So platforms like Instagram and Twitter don't support sidecar files or any sort of additional file upload with videos. And so burning open captions into your video is a great way to overcome this barrier and still make sure that your social-media videos are accessible.

And finally, integrations are like an automatic publishing process for captions. So if you use a captioning vendor, integrations are essentially a pre-set workflow that they create for you that streamlines your captioning process into the existing video-publishing process that you're already working in.

So next is live captions. What are live captions? They're actually much different from closed captioning. Closed captioning is used for recorded content, whereas live captioning, as the name suggests, would be used for live events or other live streams happening in real time. So for example, webinars like this, a virtual meeting, fitness classes, conferences, online learning environments, et cetera. Live captions ensure the accessibility of live events to deaf and hard-of-hearing individuals as well as work to make your content more engaging for everyone.

The two primary ways to create live captions involve the use of an automatic software similar to the ASR that we talked about with closed captioning, but it's a little bit more advanced to keep up with the live element, or using a human captioner. In either case, depending on what platform you use, there may be a slight delay in caption appearance, which we call latency. And this is just the time that it takes for the ASR software to process the words that it's hearing or for the human captioner to actually type the words that they're hearing. So in either case, there's just going to be a little bit of time for things to process.

But when discussing live captioning, these are some key terms to know. It can all get very confusing. But LPC, or Live Professional Captioning, specifically implies the use of a human. So live captioning, as I just went over, can be created in a variety of ways. But live professional captioning implies that there is a professional or a human expert involved in the process somewhere along the way.

Another method of live captioning is called CART, which is Communication Access Real-Time Translation. This also involves human stenographers that mostly work remotely. And they are the ones doing the captioning and connecting to the events. Finally, we have ASR. Again, this is Automatic Speech Recognition. This is used for live automatic captioning solutions. So there is no human involvement in this process at all. It would just be relying on the automatic solution.

And you're never going to believe this, but just like with closed captioning, quality is important for live-captioning accuracy. So in closed captioning, we talk a lot about the 99% accuracy rate and the industry standard that I went over. But since live captioning is happening in real time, the accuracy rate can be, and often actually is, a little bit lower. Nonetheless, you should still be striving for the range of 80% to 90% accuracy, if not higher.

Like I said earlier, there are a lot of factors that can affect captioning accuracy or the likelihood that softwares are going to correctly pick up what is being said. Some of those include filler words, like um's and ah's. Homophones, or words that sound like other words, can cause substitution errors and distort the meaning of sentences. And the external environment or any type of background noise happening can also have an effect. And finally, whether or not humans are involved in the transcription or if you're solely relying on automatic technology-- that has a huge difference.

And something to think about when you're weighing automatic solutions versus professional solutions is to think about the fact that, typically, humans tend to be more accurate but commit more omission errors, which means they skip certain words or don't transcribe things because they have to keep up with the pace of the event.

By comparison, ASR solutions will not miss words, but they can be less accurate. So if it hears a word that it doesn't understand, it may write something down, but it may not necessarily be the correct word. And [AUDIO OUT] also does not include speaker identifications or non-speech elements. So these are the things that you have to weigh when you're thinking about automatic solutions versus professionals.

And another [AUDIO OUT] element of live caption quality is the latency. Like I said, there's pretty much always going to be some sort of latency. But the average you should strive for is around three to five seconds. And this can also depend on the quality of streaming equipment and the general connection between all the devices that you're using.

Currently, there are no legislative guidelines or a governing body for live captioning, but certain states have implemented guidelines on captioning for instances of live court reporting or live events, things like that. But there is no one centralized piece of legislation on live captioning yet. We hope that there will be soon.

So I want to expand on those factors that can affect live captioning and give you some best practices to understand how to make these work for you, whether you're using an automatic solution or a professional solution. First and foremost, a strong network connection is never a bad idea. We all practically live online now, so I'm sure [AUDIO OUT] but if you [AUDIO OUT] signal during an event, we recommend using a wired internet connection via an ethernet cable. So this will help to avoid your Wi-Fi signal going down, because you won't ever have to worry about that.

The second thing is good-quality audio. If you're going to be hosting a live event, we recommend investing in a microphone or a headset, because the built-in microphones in computers can pick up a lot of surrounding noise or distort the input, making the speaker sound far away or like they're in an echo chamber.

But if you don't plan to invest in a new microphone, our next recommendation is to monitor your surroundings. Do your best to work in a space with little to no background noise, because a computer is not as good as humans at filtering out the noise. So it will pick up everything, and it will translate everything. Do your best to avoid echoey rooms. Or if you even need to, you could hang blankets around the room to dampen the sound.

And last but not least, having a single speaker at a time is probably one of the most helpful things, because like any of us, a computer would be very confused by hearing too much dialogue all at once. So try to only have one person speaking at a time, using clear speech and pronunciation when possible.

All right, so then after you have your live captions, once again, you need to figure out a way to put them somewhere. So when it comes to broadcasting options, popularly-used video players and meeting platforms already support this functionality natively. But if, by chance, your platform does not, a lot of captioning vendors, like 3Play, will offer integrations so that you can seamlessly incorporate a captioning workflow into the production process that already exists for your recorded video content as well as live events.

So with live captions, is that good enough? Can your event accessibility stop there? Not entirely. The difference is between automatic solutions and professional solutions, again. So from our tests and research, automatic captions tend to range from 80% to 90% accuracy, which is good. That's acceptable.

And in our 2020 State of ASR Report, we analyzed ASR software from various providers, like IBM Watson, Google, and Amazon. And we found that accuracy rates are improving across the board. However, they still can't compare to having a human captioner. There's still not an independently-sufficient solution that you should rely on for post-production captioning, because they are not the most accurate.

If you would like to view that 2020 State of ASR Report, we actually just put a link in the chat. And this is the report that helped us to conclude that humans are crucial for providing high-quality, accurate captions, especially if you want to reap the benefits that are offered by captions.

What do you think I'm going to talk about next? I'm going to talk about the benefits of captioning. So there are innumerable reasons that you should be captioning. But the biggest reason is for accessibility, which you've already heard me talk about a bunch. More than 20% of Americans have some form of hearing loss, and there are over 300 million deaf and hard-of-hearing individuals worldwide. So adding captions helps make your content accessible to them and engaging to everyone beyond even that particular audience.

A 2019 Verizon study actually found that 92% of users are watching videos on mobile devices with the sound off due to environmental factors. So for example, someone forgot their headphones in a quiet space, like the library or on public transportation, or you want to multitask at work. People prefer to watch video without sound. And if you're not using captions, they may be scrolling right past your content. So again, having captions available on social media is very, very valuable.

And video accessibility, at large, also offers tremendous benefits for improving SEO, user experience, brand awareness, engagement, because it makes your content searchable and easier to comprehend. Search engines cannot watch videos in the same way that humans do. And so when you incorporate a caption file, it can read and index that file to better understand what is taking place in the video so that your content can be more appropriately [AUDIO OUT].

So again, that improves your Search Engine Optimization, your SEO. And you also have the opportunity to expand your audience, because more people are finding what they're looking for. And specifically within learning environments, captions and the transcripts derived from captions can be a focus tool. A lot of students like to use captions and transcripts as study guides. So this can boost classroom comprehension as well.

All right, so I would say one of the biggest benefits of accessible video is that you get to avoid legal trouble. So I want to dive into the most pertinent accessibility laws. I'll just try to give you a brief overview without getting too, too technical. So the first major accessibility law in the United States was the Rehabilitation Act of 1973. And there are two sections that specifically impact video accessibility.

Section 504 is a broad anti-discrimination law that requires equal access for individuals with disabilities. And this applies to federal and federally-funded programs. Section 508 [AUDIO OUT] requires information technology to be made accessible. And the most recent refresh of Section 508 actually references the Web Content Accessibility Guidelines 2.0. So this makes the Rehabilitation Act unique because it specifically references requirements for closed captioning and audio description. Once again, I'll get a little more in-depth about the Web Content Accessibility Guidelines on the next slide. So if you have any questions about that, give me one slide.

The next major accessibility law in the US was the Americans with Disabilities Act. So again, two sections specifically impact video accessibility. Title II of the ADA applies to public entities, and Title III applies to places [AUDIO OUT]. This includes private organizations that offer or provide a public accommodation. So for example, a doctor's office, a library, a hotel, a restaurant, and many more. And the context of public accommodation has actually been tried in many lawsuits and many legal contexts in regards to how this impacts internet-only businesses.

And in several cases, Title III has been extended to include online spaces. So the most notorious example was when Netflix was sued in regards to their lack of closed captioning and audio description. And the outcome was that they had to provide these services because the content that was offered on their platform was considered to be a public accommodation. Any member of the public should be able to access that content.

Next, the [AUDIO OUT] Accessibility Act, or CVAA [AUDIO OUT] captioning requirements, the CVAA online video [AUDIO OUT] on television. Any online video captioned when it goes online, including video clips and trailers. As for audio description, the CVAA refers to the rules set by the FCC in [INAUDIBLE] which include the requirement of audio description by most major broadcast networks.

So again, I won't get too, too technical about that. But zooming out a bit, one thing to know about standards for web accessibility is the Web Content Accessibility Guidelines. These are not technically a written law that everyone must follow. These are pointed to, again, in the Rehabilitation Act as the international standard for web accessibility. So now let's get into that.

Web Content Accessibility Guidelines-- we lovingly refer to them as WCAG because it's much quicker. Once again, this aims to make the web accessible for everyone, but specifically users with disabilities. And the criteria by which they measure this is whether the content is universally perceivable, operable, understandable, and robust. And there are three levels that WCAG defines their compliance, with varying accessible features.

So on the screen, you can see some examples of the accessible features included in each level. But as a general summary, level A is the easiest to maintain. Level AA is what most people are aiming for. It's where most people end up. And this is the mid-level of standards. And then level AAA is the most comprehensive, the highest accessibility standard. And this is often not as attainable for organizations at great scale. But it is a good thing to strive for. Obviously, if you strive to be compliant with level AAA, you will likely be compliant with the other levels of WCAG.

So in terms of captioning, specifically, as you can see on here, just as an example, level-A compliance requires captions on prerecorded video, while level-AA compliance additionally requires captions on live video. So that's just an example of the distinction of, as you get further up, there will be more accessible feature requirements.

Most laws and lawsuits regarding web accessibility reference WCAG 2.0 compliance. So for now, that's what's legally required. However, the World Wide Web Consortium, the organization responsible for writing the WCAG guidelines, is constantly improving and constantly making them better. So a version 2.1 and 2.2 do exist. 2.2 is currently still a working draft. It was recently released in May of 2021, I believe. So it's not a finalized, published version, but they are also great resources for web accessibility. But 2.0 is what is required.

All right, now I'm going to tell you a little bit about 3Play Media. So at our core, we like to think that we combine the best of both worlds. Our process combines cutting-edge AI and machine-learning technology with professional human editors in a strategic and careful way to provide the highest-quality captions possible.

So first, a file will go through ASR to produce a first-draft, rough transcript, like I was talking about earlier. Then one of our editors will correct the transcript file. And finally, during another round of quality assurance, a QA manager will conduct a final review of the transcript and the captions. And this combination allows us to ensure that we meet our guaranteed 99% accuracy rate, which-- again, that's the industry standard. That's what you should be looking for.

We do more than just provide captions. We are a full-service accessibility partner offering a range of services to help you create compliant and engaging videos from closed captioning to live captioning, with both offering professionally-created and automatically-generated options, to subtitling, translation, audio description, or anything in between, we've got you covered.

Our goal is to provide a future-proof solution that makes captioning into something that is easy, flexible, and scalable for you. [AUDIO OUT] their services at any time. So if you come to us only needing one thing, but later on realize you'd like to add something else, you can absolutely do that, and vice versa. We have support management, support and account management, to help you with anything or discuss your account strategy, and of course, answer questions that you may have along the way.

More than anything, one of our biggest offerings is flexibility. We work across industries, and we understand that every organization is different. Nobody needs the same things. And that's why [AUDIO OUT] turnarounds, formats, and more. So anything you need regarding video accessibility, seriously, we've got you.

But we do also offer a lot of free resources, if you're not currently looking for services. If you want to know more, you can head to our website. You'll find weekly blogs, free ebooks, checklists, and research studies. We also have tons of monthly webinars, which I'll share more about on the next slide [AUDIO OUT] We have a podcast, *Allied*, where you can hear from accessibility professionals across industries on a variety of topics, like the accessibility of gaming, architecture, theater arts, and so much more.

To learn more about any of these resources, we're going to send a link to our website in the chat. [AUDIO OUT]. You'll have that to visit. So we have a webinar dedicated to podcasting as well as audio description. We have our Live Professional Captioning Launch Party, which we're very excited about. And then we also have our event, ACCESS Boston. So if you haven't signed up for ACCESS, we would love to see you there. You, once again, get to hear from an amazing panel of accessibility professionals. It's a full-day event with workshops. So we would love to have you learn more.

And with that all being said, I'm going to open the floor for Q&A. Let me see. Yes. OK, so somebody asks, can you send us a file for this presentation? That's an easy question. Yes, we can. We absolutely will do that. You'll get an email with the webinar recording as well as the slide deck. So just keep an eye out for that in the next couple days.

Somebody also asked, can you clarify how live professional captioning and CART are different? Yes, I can. So CART, Communication Access Real-Time Technology, is one method of live professional captioning. So CART is basically just an example of how you could complete live professional captioning. And then LPC just refers to any sort of live-captioning workflow that involves the use of a human being. And if you'd like to reach out to us, please do with any other questions. Thank you all for joining us. I hope everyone has a wonderful day.