**SAMANTHA SAULD:**      Thanks for joining this webinar, entitled Quick Start to Captioning. I'm Samantha from 3Play Media, and I'll be presenting today. So let's begin. Today we'll cover the following topics. The first question we'll cover is what are captions, followed by how do you create captions, where do you publish captions, why should you caption, and who is 3Play Media. Finally, we'll finish off with a Q&A at the end.

So what are captions? Captions are time-synchronized texts that can be read while watching a video and are usually noted with a CC icon. Captions originated as an FCC mandate in the 1980s, but the use has expanded to online video and internet applications. Captions assume the viewer can't hear, so they include relevant sound effects, speaker identifications, and other non-speech elements to make it easier for the viewer to understand who is speaking. An example of this additional information would be if someone is opening the door and you can visually see their keys jingling. You wouldn't need to caption that. But if the keys are jingling offscreen, you would include that non-speech element.

Now it's important to distinguish between captions, subtitles, and transcripts, as they all mean something different. Captions assume the viewer can't hear the audio. They are time-synchronized and they include relevant sound effects. You can spot if a video has captions when you see a CC icon. Subtitles, on the other hand, assume the viewer can hear but can't understand the audio. Their purpose is to translate the audio. Like captions, they are also time-synchronized.

Transcripts are a plain text version of the audio. It's not time-synchronized and it's good for audio-only content. In the US, the distinction between captions and subtitles is important, but in other countries, like in Europe, these terms are used synonymously.

So next we'll talk about, how do you create captions? One way to create captions is to do it yourself. The best way is to upload your video to YouTube for auto-captioning. Although accuracy is a big issue when using automatic

captions, YouTube does allow you to go back and edit the script for accuracy. Another way, and if you have the time, you can manually transcribe the video yourself. But like I said, you'll need plenty of time, as this method can take five to six times longer than the length of the video. This doesn't include the time it takes to add non-speech elements. This method could also be a very costly method at scale.

Now, in terms of accuracy, like I mentioned before, the industry standard is a 99% accuracy rate. Going below this rate can cause a lot of inaccuracies. It's important to note, the accuracy percentage is based on word for word, meaning each word has that percentage chance of being wrong. In other words, an 85% accuracy rate means each word in a sentence has an 85% chance of being incorrect.

When you use ASR technology, the accuracy rates are pretty abysmal. A lot of ASR errors make sense acoustically but not linguistically. Here, I will show you an example of a transcript captured by ASR. Listen closely to the audio and compare with the words on screen and see if you catch the errors. You can type any errors you notice in the Chat window.

[AUDIO PLAYBACK]

- --challenging aspects of choosing a career is simply determining where our interests lie. Now one common characteristic we saw in the majority of people we interviewed was a powerful connection with a childhood interest.

[MUSIC PLAYING]

- For me, part of the reason why I work here is when I was five years old, growing up in Boston, I went to the New England Aquarium. And I picked up the horseshoe crab and I

touched a horseshoe crab. And I still remember that and I still-- I love those types of engaging experiences that really register with you and stick with you. As a child, my grandfather was a forester and my childhood playground was 3,600 acres.

[END PLAYBACK]

**SAMANTHA SAULD:** Well, I see some people say speaker identification, punctuation, [? the length ?] [INAUDIBLE]. So yeah. [INAUDIBLE] is the lack of punctuation. In this transcript, there are a few periods of incorrect capitalizations. Another issue is that hesitation words are not removed, which spill over to other words and make it inaccurate. Speaker changes and speaker IDs are also not captured.

Now a lot of these errors are harder to catch when you are listening to the transcript, but if you are reading it with the errors, it clearly makes no sense. So in this example, when the speaker says New England Aquarium, ASR picked it up as new wing of the Koran. And when the speaker said a forester, the transcript read four story. A human wouldn't make these errors.

So now, where do you publish captions? Captions are supported on most devices where you can publish video. One limitation, though, is on social media video. While a lot of platforms have more recently built in caption functionality, some platforms, like Twitter, still don't allow it. Originally, captions were mandated for broadcast TV, but in recent years, this has changed and now they are published across all devices and platforms.

Sorry, guys. Can you hear me? Let me know if you guys are having any trouble with the audio. OK, I'll continue. Please let me know if you guys are having a hard time hearing me. So as I was saying, originally, captions were mandated for broadcast TV, but in recent years, this has changed and now they are published across all devices and platforms. Most players and platforms have caption compatibility, although some are more advanced than others. Many also [INAUDIBLE] control for customizing their captions,

such as adjusting font size or font type.

Now, there are many ways to publish your captions. The most common is through a sidecar file, which is basically a file that stores the captions so they can be associated with the corresponding video. So when you upload your caption file on YouTube, you are uploading it through a sidecar file. These types of captions give the user control to turn the captions on or off.

Another way is to encode captions onto the video. For example, these are found in kiosks or offline videos and can be also turned off or on. Open captions are burned into a video and cannot be turned off or on. For social media videos on Instagram or Twitter, adding captions as a sidecar file is not possible. Open caption encoding is the one way to overcome this barrier and keep your videos engaging. Lastly, integrations are simply a publishing process for captions. It's a preset workflow between your captioning process and video publishing process to really make everything more streamlined.

So what do caption formats look like? Caption formats vary depending on the platform and player. The two most common types are presented in the slide. To the left, we have an SCC file, and on the right, we have an SRT file. Both of these are the same exact file. Both are presenting the first three caption frames. But as you can see, they are very different.

SCC files use hex codes, so they are more complex and harder to read if you're not a computer. SRT files, on the other hand, are more readable. You can easily tell the number of the caption frame, the time codes, and text contained in each frame. SRT files read more like a script, including the words, and are easier to create, so it's recommended if you want to DIY your captions. Whether you have an SCC, SRT, Web BTT, or other types of caption files, you can always use a caption format converter to change the format for the appropriate video player you're using.

So next, we'll talk about why you should caption. There are many reasons for why you should caption. The biggest is accessibility. There are 48 million Americans with hearing loss, which is about 20% of the US population. Captions help make your content accessible to them.

Another reason is for better comprehension. In a study by the Office of Communication in the UK, they found that 80% of users were not deaf or hard of hearing. Captions help people understand difficult content and accents. They are also useful for people learning English as a second language. Captions give people flexibility, allowing them to view videos in sound-sensitive environments like on the train or at the library. Captions also help with video search. MIT surveyed their students and found that 97% of students said interactive transcripts enhanced their learning experience. Interactive transcripts allow users to search and jump to relevant spots in the video. We'll talk more about this later.

Next, captions help increase SEO. Google can't watch a video, so they don't know what the content of your video is about beyond the title and description. Captions help search bots comprehend the video, thus helping your video to rank higher in search engine result pages. In a study by Discovery Digital Networks, they found adding captions to YouTube videos led to a 7.3 increase in views.

Captions also help with translation, making a video accessible on a global scale. With captions, you can easily translate the native language of the caption file into other languages. Captions and transcripts are also reusable. You can create lots of derivative content with infographics, white papers, blogs, case studies, and course materials. The University of Wisconsin found that 50% of students were repurposing their transcripts as study guides.

Captions also help in the social media realm. Social media platforms like Facebook auto-play videos without sound, so adding captions has become increasingly important in order to keep viewers engaged. And of course, the legal requirements. In many cases, captions are required by law. We'll explore more of this in a second.

So we conducted a nationwide study with Oregon State University. We surveyed students to see how and why they use captions. The results prove that captions truly help students learn. 98.6% of students found captions helpful. In addition, 75% of all students who use captions, not just those who are deaf or hard of hearing, use captions as a learning aid. And the number

one reason students used captions was to help them focus on video content.

Now, there are three key laws that relate to captioning-- the Rehabilitation Act of 1973, the Americans with Disabilities Act of 1990, and the 21st Century Communications and Video Accessibility Act. Under the Rehabilitation Act, there are two sections that mention captioning, Section 504 and Section 508, which was recently refreshed. Section 504 is an anti-discrimination law that requires equal access for individuals with disabilities. This applies to federal and federally-funded programs like colleges, airports, and statehouses.

Section 508 was introduced in 1998 to require federal communications and information technology to be accessible. It applies to federal programs but often applies to federally-funded programs through state and organization laws. Closed captioning requirements are written directly in Section 508 and are often applied in Section 504. Many states have their own mini 508 laws as well. The Section 508 refresh was recently refreshed in January 2018 and now references WCAG 2.0 double-layer guidelines as accessibility standards to meet.

Of course it's crucial to know what standards should be met to avoid legal action. Web Content Accessibility Guidelines, or WCAG, is the international set of guidelines making digital content accessible for all users, specifically users with disabilities. It outlines best practices for making web content universally perceivable, operable, understandable, and robust.

It has three levels of compliance. Level A, which states that captions are provided for all pre-recorded audio content. Level AA, which states that, in addition to level A compliance, captions are provided for all live audio content. And level AAA, which states that, in addition to levels A and AA compliance, sign language interpretation is provided for all pre-recorded audio content. Most laws and lawsuits mention WCAG 2.0 compliance. So for now, that's what is legally required. Only if a law explicitly states that web developers have to adapt to the newest WCAG version do you need to make your content WCAG 2.1 compliant. The World Wide Web Consortium does suggest that any new websites should be created following WCAG 2.1

guidelines, since they are the most inclusive and mobile-friendly.

The Americans with Disabilities Act of 1990, or the ADA, also mentions captioning. Of the five titles in the act, Title II and Title III apply to captioning. Title II applies to public entities and Title III applies to public accommodations. There are precedents that bypass lawsuits. Title III has been extended to the private sector and is now being tested against online businesses. The key cases to know regarding the ADA are the *National Association of the Deaf,* NAD, *v. Netflix,* and *NAD v. Harvard* and *MIT.*

In the case between *NAD v. Netflix,* Netflix was sued by the NAD in 2012 for failing to provide closed captions for most of its Watch Instantly movies and television shows streamed on the internet. It was during this case that Title III of the ADA was broadened to include internet-only businesses as part of a public accommodation. Before, it was really only applied to physical structures such as wheelchair ramps. The court ruled in favor of the NAD, and Netflix settled, agreeing to caption 100% of its streaming content. This set a profound precedent, leading to similar lawsuits against FedEx for failing to caption training videos, and Hulu and Amazon, who also settled with the NAD and agreed to caption their streaming content.

In the education realm, MIT and Harvard were both sued by the NAD for providing inaccessible video content that was either not captioned or was inaccurately and unintelligibly captioned using auto-captions. This was the first time that accuracy had been considered in legal ramifications for closed captioning. The NAD argued that educational online videos should be constituted as a public accommodation. After several years of motions and hearings, the lawsuit against Harvard is now moving forward in federal court.

On March 28, 2019, Federal Judge Katherine A. Robertson denied a large part of Harvard's second request to dismiss the cases. Robertson declared that Harvard's website will not be excluded from Title III of the ADA and Section 504. However, she did agree that the university could not be held responsible for captions on third-party content, under the Communication Decency Act. Litigation is still ongoing, but the outcome will surely have huge implications for higher education.

So the last key law [? is ?] 21st Century Communications and Video Accessibility Act of 2010. This states that all video previously aired on television needs to have captions when published online. So an episode of *Game of Thrones* that aired on HBO needs to have captions when published online for later viewing. Similarly, clips and montages need captions.

Now a little bit about who we are at 3Play. We are a video accessibility company specializing in transcription, captioning, subtitling, and audio description. We spun out of MIT in 2007 and are currently based in Boston. We have over 2,500 customers, spanning higher education, media, government, e-commerce, fitness, associations, and enterprise. And our goal is really just to make the whole captioning and video accessibility process much easier. We also have a number of turnaround and workflow options to choose from.

Like I mentioned, our goal is to make the whole captioning and video accessibility process easier. With our easy-to-use online account system, you can seamlessly upload videos for captioning directly from your computer via links, FTP, APIs, or through our video platform integration. We also offer a number of turnaround options and over 50 caption formats for you to download. We offer tools that allow you to go back and edit your captions, upload cheat sheets for editors, move captions so that they don't obstruct important information on the video, and even allow you to upload your own caption and transcript files to have them translated into other languages or time-coded by our system.

We integrate with most leading video platforms, like Brightcove, Kaltura, YouTube, Vimeo, and Mediasite. Our integrations actually allow you to select the files you want for captioning directly from your video platform. All you have to do is tag 3Play Media. The file will then come to us, we'll caption it, and then automatically post it back to your video. You can also upload from the cloud, which allows you to upload videos for captioning from your Dropbox, Google Drive, or Box accounts.

The 3Play plugin was designed to make your publishing process simpler and

more flexible. It's a compilation of all our features. Basically, all you have to do now is check a box with the feature you want to add. Our features include caption embed, which allows you to post captions over YouTube videos you don't own or on video players that don't support captions. LMS compatibility allows you to add captions, interactive transcripts, and audio descriptions to learning management systems like Blackboard.

Interactive transcripts allow you to search key terms within a video and jump directly to that point in the video. SEO embed helps search engine bots read your video for search engine ranking. And audio description embed allows you to publish audio descriptions on video players at the platforms that don't support it. Since the 3Play plugin it's an iframe embed, it works seamlessly with the major video platforms and players.

So we use a multi-step process that uses a combination of technology and human editing. So when you first submit a file, it will go through ASR, or Automatic Speech Recognition. This gives us a very rough draft, but it makes it easier and more efficient for our editors. Once it has gone through ASR, it's released into the marketplace, where one of our more than 1,500 editors will self-select your video and start editing. Finally, we have one more QA review process, where another human editor looks back on the file and video to ensure we meet a minimum 99% accuracy rate. All our work is done in the US, and we have very strict standards to make sure that consistency is achieved.

On this slide, we have three of our editors. Like I said, we have over 1,500 editors who go through a rigorous certification process well before they touch a file. Our editors also come from very diverse backgrounds and areas of expertise, which is very valuable to us as we get a lot of different type of content. As you can see, we have one editor who raises sheep, another who refers to himself as a digital nomad, and another editor who is a mother and wanting to raise extra income for her son.

Now I've mentioned interactive transcripts briefly, but this is another innovative tool we offer. An interactive transcript is a time-synchronized transcript plugin that highlights words as they're spoken and allows you to

search, click, and jump directly to any point in the video. You can even have multiple videos in a playlist and search within them. This feature is called Playlist Search.

This GIF is a demo from MIT's Infinite History site. This is an archive of video interviews from alumni, as you can see, and they are using YouTube videos. As MIT has done, we allow you to customize your plugin so that color options can fit your branding. This tool is extremely easy to implement and really transforms your user experience.

So we're at the end of the presentation and we can now begin Q&A. So the first question is, how does an integration work? Integrations link disparate systems or platforms to make it easier to share information and build workflows between the two. Our integrations are engineered to make the captioning process a whole lot easier. As I mentioned, we integrate with most leading video platforms. Our integrations allow you to select the files you want captioned directly from your video platform or cloud storage. Integrations save you a ton of time by streamlining the whole captioning process.

The next question is, does color contrast matter with captioning? The DCMP, or the Described and Captioned Media Program, outlines best practices for caption quality formatting. Captions should be readable and not obscure any visual content. They should be on a background that you can clearly read. The typical color for captions is white on a black background.

The next question is, how much does a plugin cost? Our plugins are free. You just have to have a 3Play Media account. At the moment, the plugins only work if you captain with us, but they are free and included in the service.

The next question is, do we need both captions and the transcript? Captions and transcripts are two different things. Transcripts are not time-synchronized, so they're preferable for audio-only content, but if you provided the transcript, it doesn't necessarily take the place of captions, so it's sometimes possible that you will need both. And you can see from the research that I mentioned that there are definitely benefits for each of these

different types of tools.

The next question is, can we upload transcripts and your service will time-code them? Yes. We have a service called Transcript Alignment, where you can upload an existing transcript and we'll time-code it to create a caption file.

The next question is, how do you ensure accurate captions on more complex or challenging content? Good question. So our process includes two rounds of human editing, so that no matter what the content is, we're giving back at least a 99% accuracy file. Our editors choose the file to work on from the marketplace, which allows them to choose content that they know really well and are competent in. Our editors come from many different backgrounds, and we're able to produce that same quality file, whether it's a high-level math, any sort of STEM work, or no matter what the content is. Any medical content, we can certainly still provide a high accuracy rate for. We also allow users to upload cheat sheets, which are basically a glossary where you can upload specific terminology or any names that you think the editor may not be super familiar with.

The next question is, we are doing [? synchronous ?] live sessions online. Do you provide real-time services? I may have missed this in the presentation. At this time, we don't provide live captioning, but it is on our roadmap. OK.

And then we have time for one more question. The next question is, are there quality standards for how captions should look? So, with accuracy, the FCC states that captions must match the spoken words in the audio to the fullest extent. This includes preserving any slang or accents in the content and adding the non-speech elements. For live captioning, some leniency does apply. Captions must also be synchronized. They must align with the audio track. And each caption frame should be presented at a readable speed, between three to seven seconds on the screen. Completeness is also important. Captions must run from the beginning to the end of the program and not drop off. Captions must also be placed so that they do not block other important visual content.

And we have one more quick question. So how much is the service? We start at $2.50 per minute, but we have bulk discounts depending on the hours of video you need. That's all we have for today. Thanks everyone for joining.