

3Play Webinars | Intro to Live Captioning for Broadcast

JENA WALLACE: And without further ado, I just want to welcome Josh Summers, who is our senior manager of training, development, and technology for 3Play Media Canada. Thanks so much for presenting today, Josh.

JOSH SUMMERS: Yeah, you bet. Thanks, Jena. Yeah, very quick intro to myself. I am Josh Summers, as Jena said, senior manager at 3Play Media Canada. One of the things that I do here is I train live broadcast captioners. Prior to that, I was 3Play Canada's live voice captioning manager. I also work with our broadcast customers on live captioning workflows and solutions. I have been a broadcast captioner and I've worked in closed captioning over the last 20 years, give or take.

OK, we'll get into the slides here. So the agenda for today, pretty straightforward. We'll talk about the basics of live captioning for broadcast. We will talk a little bit about who 3Play Media are and what we do. And then there'll be some time at the end for some questions and answers.

So, yeah, let's define the basics of what live captioning is. And we'll start with a kind of really, really basic look at closed captions. Specifically, fundamentally, time-synchronized text that can be viewed alongside a video and typically denoted with the Capital CC icon in media players and on TV remotes and things like that. They are an accommodation for deaf and hard of hearing viewers in the US mandated by the Federal Communications Commission in the 1980s-- similar time frame in Canada also, if not somewhat behind that. It was actually in the late 1970s that the Canadian Association of the Deaf petitioned the CRTC to compel live broadcasters in Canada to start captioning their content.

And then also worth pointing out here that closed captions are not just about the spoken word. There necessarily must be non-speech elements that are captured in this way. And we're talking about things like speaker identifiers and relevant sound effects as well, all of which help to paint a fuller picture of what is happening in the broadcast.

So then looking at live closed captions-- specifically by definition, they are produced in real time. The gamut of programming-- so if you're watching broadcast television in both the US and Canada and obviously other parts of the world, typically, if it's a live program, it will carry closed captions. So we're talking about sports events, news, live events, political programs, current affairs. The whole spectrum of programming will typically be live closed captions. Generally speaking, vendors are using highly-trained, live professional captioners, human beings. And generally speaking, using one of two disciplines-- voice writing or voice captioning methodology or stenography.

And then we are just spotlighting an element I think most viewers of live closed captions will have observed themselves. And that is that typically there's a slight latency in captions versus the soundtrack. And that's because of the processing time. Across the chain, it starts with a closed captioning of themselves. So in the human brain, interpreting the soundtrack, deciding how to write, what to write. There is a little bit of latency in captioning software. And then there is processing time in the transmission of caption data also, all of which contributes to typically speaking a three to seven-second caption lag in the broadcast space.

Let's go to the next slide. So I think we need to go back one just real quick. I just want to touch on this. We're going to talk about specifically the nuances of broadcast television and why it's different from other types of live closed captioning that you might have observed if you, like most people you're on lots of corporate town hall or internal meeting type events.

So some key terms that we want to flesh out that you should know about in this space-- voice writing, stenography, and ASR-- these are all methods of closed caption production. Voice writing is a transcription method where a trained captioner will dictate or re-speak, as it's often known, speech into specialized voice recognition software. Stenography-- similarly, trained captioners very quickly typing what they hear into a specialized steno machine, less commonly so into a keyboard.

And then automated speech recognition is AI. So we're leveraging AI to somewhat intelligently interpret the soundtrack and transcribe what is heard. Generally speaking, not recommended on the ASR side for broadcast captioning, just given that the quality is often not where we would want it to be for compliance reasons. But there are use cases for ASR, depending on the content type, there are parameters that it's well-suited to. And if you do deploy it, it should be deployed by experts who understand those parameters and can manage it effectively.

Caption quality is key. Yes, that is very true. This is something that in broadcast captioning circles is always front of mind. We think about accuracy and are concerned with accuracy very much so in the US. 95% to 98% and above is the typical accuracy rate that we see. It's a little bit different in Canada. 98% is the kind of minimum benchmark. And generally, we're seeing above that. And that's just because of differences in the way the closed captions are assessed or that the accuracy is assessed.

As I said before, quality also includes nonverbal elements-- so things like speaker IDs, descriptions of sound, both of which help particularly deaf and hard of hearing viewers take away more from a program. So you can think about in a sports game, applause that is unseen from an audience. And that would be a description of sound that we want to describe-- but anything that kind of adds necessary flavor.

Vocabulary is super important. We want our captioners to be well-prepared so that they can write correct spellings of people's names and things like that. And then obviously, we think about continuity of service, minimizing downtime in closed captioning, making sure that captions are in place at the right time and providing an end-to-end service. And then very much front of mind is compliance in this space as well.

So both in the US and Canada, there is regulation with real teeth. The US-- the Federal Communications Commission has standards around live captioning. There's some leniency that is sort of allowed in that space, which takes into account the various challenges that live caption vendors face. There are other legal precedents in the US that still compel live programming to be captioned at a high accuracy rate. But there is no technical threshold there in the US. And in Canada, again, it's a little bit different.

Not only are most broadcasters required to caption 100% of their programs within a broadcast day, there are also minimum accuracy benchmark standards that must be adhered to. And then just at the bottom of the slide there, it almost goes without saying that high accuracy is possible in live closed captioning, but for the benefit of this audience, it very much is possible.

The example that we're giving here is the US House of Representatives, which provides real-time closed captioning of televised proceedings. And they are mandating a 98.6% level of accuracy in captions. I think it's fair to say that a well-trained, well-equipped captioner can write well over 98% accuracy sustained and at speeds of 200 plus words per minute.

So some best practices, then, in live broadcast captioning-- lots of them. We'll start with prep. So prep materials-- this is crucial to successful broadcast captioning-- any captioning, in fact. We are generally looking for materials from broadcasters that include word lists, key terms, proper name spellings in advance of the broadcast so that our captioners can take that material. And they can make sure that they are able to correctly spell certain names correctly come broadcast time.

Without that information, we can still do a good job. We can write around the spellings of names if they're unknown. But obviously, that creates gaps in the captions, which is an accessibility concern. So the more prep material, the better. We do ask-- I think most lenders do this-- are asking their captioners to do research of their own to try and uncover some of these spellings. So we're doubling up the effort there.

Having a strong network connection-- depending on how you're receiving captions, delivering captions, any sort of interference in internet signal, cable, antenna signal can disrupt the caption delivery. So you might see where you have issues, characters being dropped from captions, placement inconsistency, and that type of thing-- even color changes in captions as well.

So when you're talking to and working with a vendor, just try and make sure that you have a strong network connection. If you're using captioning hardware-- so things like closed captioning coders and connected telephone lines, again, you want to make sure that equipment is well-maintained and upgraded if necessary, if it's aging.

Clear speech from programming as well, super important. So making sure that you use basic things like using professional grade microphones, asking your on-air talent to enunciate as best as they possibly can. It just helps the captioners interpret the information that bit more easily. And they can write a bit more accurately.

Loading background noise. Obviously, this is content-dependent. This can be difficult to avoid. But anything that makes it difficult for the captioners to hear the intended dialogue, that's going to provide a challenge. Single speakers-- ideally, particularly in studio settings, if we can avoid crosstalk, that's always preferential. It's very difficult for captioners to parse through crosstalk and reorder it in a logical way. So, again, trying to have your on-air talent and other speakers speak one at a time is the ideal whenever that's feasible.

Use highly-trained professional captioners for sure. And again, when you're talking to vendors, ask them whether they use trained professional captioners. And if so, are they using a proven methodology like voice writing or stenography? Ask them about quality as well and their approach to that, whether they measure it. It's always good to know how mindful vendors are of that.

And then as we've just been talking about, being mindful of quality guidelines as well. As I said, the FCC and the CRTC in the broadcast space have regulation with real teeth. There are other precedents out there as well that you need to be compliant with.

So how are live captions delivered to broadcasters? Fundamentally, and in one key way, and you can think of this in two streams-- we use encoders. An encoder is a device that is used to embed live captions into a video stream. And the encoder is either a piece of hardware or it's a cloud-based virtual encoder.

On the hardware side, professional-grade captioning software integrates with encoders very, very easily. We're doing this in a few different ways via Telco or Telnet connections. So we're either delivering captions over the internet basically with an IP address and a port number. We can integrate with telephone lines using fax modems as well. And broadcasters can also register their encoder with the manufacturer who can provide us with third-party software and a sort of simple gateway access code.

And then on the virtual side, this is a growing technology. More broadcasters pivoting towards online streaming and tapping into cloud-based encoders. The workflow is essentially the same. That CC stream is being embedded into the video stream but in the cloud. And then the resulting stream can be sent to one or more destinations.

Accessibility laws-- for anybody that isn't familiar with the term A11Y, it is shorthand for accessibility. The 11 represents the number of characters between the letters A and Y. It's a term that is becoming sort of quite widely, commonly-used in accessibility circles. But these are laws that we'll look at now that are specific to broadcast that is going into a little bit more detail there. We'll look at the US and Canada specifically.

So starting with the US, then. The Federal Communications Commission has very strict guidelines, then, around closed captioning television programs and live broadcasts. They look specifically at accuracy, timing, placement, and completeness. And then the CVAA looks specifically or particularly at programming that has aired on US television with closed captions originally. So we've got sort of two sides of the coin there.

In Canada, we have the Canadian Radio Television and Telecommunications Commission. Again, they have created closed captioning standards to ensure consistency and reliability around caption quality in the Canadian Broadcasting system. That is federal regulation. And then just underneath that, we have-- we're spotlighting some provincial legislation here, which does exist in other provinces. But in Ontario, the AODA was kind of broadly created to set accessibility standards so that nobody with a disability is prevented from fully participating in all aspects of society because of a disability.

What's interesting about the AODA is that, yes, it calls for accessible video. It often references another set of legislation known as the WCAG. And it enforces upon most Ontarian companies to provide closed captioning for videos that they have online.

OK. So a little bit about 3Play Media and 3Play Media Canada and what do we do. So these are our services-- recorded, closed captioning and transcription at 99% accuracy guaranteed that is compliant, reliable. And we have many, many flexible workflows. Live captioning, as we've been speaking about, that is highly accurate, driven by highly-trained professional human being captioners.

It's reliable and easy to schedule powered by 3Play Media technology. We offer subtitles and translation, live translation in over 100 languages, and subtitles and translation of recorded content in over 40 languages and then, finally, high-quality audio description, or described video if you're in Canada, that is competitively-priced. And our offering includes both synthetic voice and professional voice artists as well as extended AD.

So our solutions are future-proof. You can upgrade your services at any time. And that's of real consideration. I think in the live broadcast captioning space where, as we were talking about, if you are looking to put a video online that has aired on US or Canadian television with captions, to be compliant, you're going to want to consider upgrading your captions to verbatim standards.

Our solutions are scalable. We have 99.9% deadline compliance on the recorded captioning side. Specifically, we can process large quantities of files. We offer dedicated account managers to review goals with you and also to help you keep up-to-date with accessibility news. And we can offer flexible solutions as well. So we're accommodating many, many different workflows, file formats, as well as turnaround times that match your needs.

So just wanting to spotlight a few free resources here, then. We have our resource portal for blogs, white papers, checklists, and research studies, webinars that are taught by accessibility experts, and our video accessibility course where you can learn and test your knowledge on video accessibility.

Couple of upcoming free events to bring to your attention. On November 29, the Impact of Accessibility and Gaming Fireside Chat with AbleGamers, which is a fantastic non-profit organization that works with disabled people to open up access to gaming-- and then on December 8, the Digital Accessibility Legal Update with Lainey Feingold, who is a disability rights lawyer, public speaker, and author. So a couple of events there that I think you want to check out if you can.

And that's the end of the Intro to Live Broadcast Captioning session. I think we can open the floor to some questions. I think I saw a few coming into the chat there.

JENA WALLACE: Yeah, yeah. We have a few really great questions. So we'll try and get through as many as we can. So a couple of the questions are kind of similar. So I think we can probably combine the answer. So how is accuracy quantified? Is that 95% to 98% accuracy a government standard? Or is that just a target? Can you expand on that?

JOSH SUMMERS: Yeah, I can definitely speak to the Canadian side. 98% accuracy is-- it's a CRTC mandate. So, yes, effectively, it's federal legislation. And that uses what's known as the NER model of accuracy. It is sort of-- not to get into the weeds too much-- but it looks at errors in captions in a slightly more nuanced way than I think is the case in the US.

95% to 98% in the US, I'm not 100% on. I would have to defer to my US colleagues there. I don't think that there is any kind of necessarily mandate for a benchmark there. But certainly in Canada, there is. And we have to submit regular reports. All caption vendors do in Canada to the CRTC.

JENA WALLACE: Great. Yeah. Couple questions around broadcasting and how we're defining broadcasting. Because I think there's a few attendees looking to transition to different forms of broadcast. So they're wondering what we mean by broadcast.

JOSH SUMMERS: Yeah. Great question. I suppose traditionally, we would think of broadcast as being sort of linear television networks and channels, so cable and satellite. But with a lot of content sort of transitioning to online spaces, it may be the same networks-- traditional networks that are moving that content across. And so, yeah, we should consider that broadcast as well. The captioning workflows, whether it's broadcast-based or online in terms of delivery, are typically identical. Yeah, there's not really a distinction to be made there.

I think anything else is typically in corporate environments where you have users, attendees using Zoom and other video platforms for meetings and that kind of thing.

JENA WALLACE: Great. So a couple of questions around prep materials that we talked about. What is an acceptable amount of time to provide prep materials for professional captioners to review before a broadcast? And when would this prep material not be used?

JOSH SUMMERS: OK. I'll take the first part, then. An acceptable amount of time-- the more time the better, clearly. Very often, we don't receive any material. And even within a 30-minute window before the beginning of a broadcast, particularly if it's something like live news where the content is evolving, that is sufficient for our captioners to be able to prep at least a chunk of the content that they're captioning. It's not ideal.

I think day of is fine. But if we can do that up to 24 hours in advance, that gives our captioners plenty of time to prepare for at least most of a broadcast. And then, sorry, I forgot the second part of the question there.

JENA WALLACE: Are there any cases in which those prep materials would not be used?

JOSH SUMMERS: Would not be used? It depends. Broadcasters often have information that is useful to captioners that they may not be aware of-- so scripts for news anchors and reporters and even rundowns of programs. Sometimes, that information is helpful. We can glean technical terminology and names from-- but it's typically that kind of material that we're looking for.

We want to be able to spell, fundamentally, people's names correctly and technical terminology correctly. So there aren't many use cases where we wouldn't use print material, I don't think.

JENA WALLACE: All right. We're short on time. But I think we can try and squeeze in a couple more questions. What is a good resource to go to for best practices in captions considering grammar, punctuation, spacing, timing, et cetera?

JOSH SUMMERS: The CRTC and FCC guidelines for sure, which we could provide. The 3Play Media Canada and 3Play Media websites as well, I'm certain, have plenty of best practices there as well as me, if anybody wants to reach out directly.

JENA WALLACE: Yeah, you're definitely a great resource, Josh. One final question kind of touching on that delay within captioning-- can you speak a little more to that latency and delay that happens?

JOSH SUMMERS: Yeah. So one thing I didn't touch on is that, typically speaking in the broadcast space, our captioners receive live-- live, live, audio. So it's in advance of the TV broadcast itself, which helps to offset some of the latency that we were talking about in the processing chain, which is a good thing, obviously. It is, as I said, it's typical though to still see a delay in captions of somewhere between three and seven seconds in broadcast. If it's a non-broadcast platform, video calling platform, typically, it's a little bit shorter than that.

Obviously, there's less equipment to go through the chain there. But, yes, it's pretty standard to see that.

JENA WALLACE: Great. Well, thank you so much, Josh. For a fantastic discussion. Thank you, Kelly, for the wonderful ASL interpreting. And thank you to everyone for joining us for today's presentation and asking some really, really great questions.

I just want to direct you to the chat right now. Please take a moment to share your feedback on today's session via the survey that's going to pop up in the chat. Thanks again for joining us. And we hope everyone has a great day.

JOSH

Thanks. Bye.

SUMMERS: