

3Play Webinars | Real-Time Captioning in Higher Education

KELLY MAHONEY: Thank you, everyone, for joining us for today's session. This session is titled The 3Play Way, Real-Time Captioning in Higher Education. On the screen, we have our website and social media handle, which are 3playmedia.com and @3playmedia respectively, and that is using the numerical 3.

We also have the hashtag a11y, or ally. That's the shorthand for accessibility. And it's commonly used by advocates and professionals in this space. So if you feel so compelled to share this session on social media, or LinkedIn, or whatever that may be, definitely use that hashtag to make sure you're reaching the community.

So we can go ahead and hit the next slide. I will introduce myself before we proceed any further. My name is Kelly Mahoney. I am 3Play's Partner Marketing Specialist. I use she/her pronouns. I'm a young white woman with long reddish brown hair and wearing a black top.

My email address is also listed on the screen, which is kellym@3playmedia.com. You can feel free to reach out to me if you just feel like saying hi or if you have any questions after today's presentation. Don't hesitate to reach out.

All right, we'll first start off with an overview of Real-Time Captioning 101. This is just for everyone to get on the same page about terminology and concepts, et cetera. Then, we'll talk about 3Play's solution, how we fit into the landscape that we just described.

Next up is real-time in real-life. We'll discuss a use case for real-time captioning from an actual university that we have helped. And last, we'll describe how we are partners in accessibility, how we deliver more than just captions.

Like I said, any remaining time will be for Q&A. And that is the last that you have to hear from me. So with that, I'm happy to welcome today's speaker, Erik Ducker. Erik, go ahead and take it away.

ERIK DUCKER: Thank you, Kelly. My name is Erik Ducker. I am the Senior Director of Product Marketing at 3Play Media. I am a white male in my early 30s, and I go by he and him pronouns. I have brown hair and brown eyes as well.

Very excited to be here. I'm going to probably go off video. My internet here is not great, so I'm going to protect my bandwidth. But I'm going to go off video during the presentation, but that is to make sure that we have a smooth experience throughout.

So let's get started. We're going to dive right into Real-Time Captioning 101. And what we're talking about here is we're going to start with some definitions here around real-time captioning. And what we need to start with are some basic understanding of what are the definitions here.

So CART is probably a likely term that we're all familiar with in live captioning scenarios, especially in the edu use cases. But largely, we want to center our understanding around it as really 1-to-1 accommodations, typically delivered via a second screen, of being able to translate audio that you are experiencing into real-time text. 3Play Media focuses on CART accommodations through the terminology, really, real-time captioning. So these are considered solutions to CART experiences.

And there are two general ways in which we consider delivering real-time captioning. One is live professional captioning, which is relying on experienced voice writers, like we're doing today in this webinar, and then, live automatic captioning, which can be generated in real-time using Automatic Speech Recognition ASR technology. There are other mechanisms that we come across in the market for student accommodations, such as meaning-for-meaning as an alternative, which in some use cases, might make sense for your students, but is ultimately less accepted as being the standard for accommodations as defined by law. And so when in doubt, we tend to recommend that customers seek out real-time captioning, text-for-text transcription services for students for any type of accommodation.

At the end of the day, what we're trying to do is, for accommodations, our goals are to provide the most accuracy given your constraints. So before we talk about how we do this, let's talk about accuracy and what we mean by that. So when we think about measuring for accuracy, we're thinking about two specific ways to measure. One is Word Error Rate, or WER, and that is effectively the transcription of word for word. How many words did you get correct in the transcription from the audio content?

Most likely, this is what most companies refer to as their accuracy rate. Because in a lot of use cases, word error rate is all you need to focus on in terms of accuracy. But in accessibility, word error rate is not enough for supporting an accurate experience for student accommodations. And so this is where the other accuracy measure called Formatted Error Rate, FER, comes into account. And what this is tracking are formatting errors, like capitalization, punctuation, and other grammatical considerations, as well as non-speech elements, like speaker labels that are being used today in our live caption output today, or sound effects that are relevant to the understanding of the content.

So why are these accuracy measures so important? Because at the end of the day, errors are opportunities for problems of comprehension. So if you have formatting errors, punctuation can dramatically change the meaning of a specific sentence. So "let's eat grandma" and "let's eat, grandma" are two very different meanings with a simple problem of not having that comma.

So if you are deaf or hard of hearing, and all you have to go on is the context of the text in front of you, punctuation is incredibly important for you. In addition, there are words that are just confusing to hear. So, "I can't attend the meeting." "I can attend the meeting." Very similar words, but a challenge, nonetheless, potentially, for either a speech recognition system, but it's really important that we get those right.

In addition, there is obviously complex vocabulary. So it could be proper nouns, like Ehrhardt, could easily be translated to air, or Bowen can mean could be translated to bone very quickly. So these errors are obviously illustrative of what could possibly go wrong, but ultimately, there is real impact in these errors, and they add up very quickly. So just at 85% accuracy in word error rate or word error rate plus formatting error rate, you can have almost one in seven words incorrect, or one in seven opportunities of punctuation and words incorrect, and that can dramatically reduce your ability to comprehend.

So why do we do this? Why do we care so much about this accuracy for student accommodations, or any accommodation in the university-- in university-- in the university industry? One, we want to make sure that we're providing inclusive learning environments. We want to make sure that everyone, regardless of their abilities, have equal access to information. And ultimately, in some cases, there's going to be benefits to all learners.

So if captions are turned on for a virtual classroom for one student, they're turned on-- they have the option for all students to access those captions, and that can provide a better experience, a better learning experience for everyone involved. And at the bottom line, there's also a question of legal compliance. At the end of the day, we are support-- we have to support our students on an equalized basis and make sure that we are providing accommodations to those who need them.

But why we do this doesn't always match up to what are we up against. What are the challenges that we face in the accessibility offices at universities? How can we manage all of these expectations of adhering to the legal compliance and improving student outcomes when we're up against budget constraints? So whether you're doing a one-time event, or student accommodations, or stadium captioning, you may have budget constraints. And so it's important that the solutions that you pick align to your budget and your ability to deliver that experience consistently to your students and faculty, and anyone who's involved in your audience.

The other piece that we commonly hear is the urgency of these accommodation requests. If it's a student accommodation, we might get that request the day before the semester starts, and now, we're scrambling to try to find all of the captioners lined up to make this a great experience for the student. Maybe there's a colleague that isn't aware of accessibility accommodations and they're running their first virtual event. And the day of the event, they're like, oh, crap, I need accommodations for this particular individual, and I'm going to be calling the accessibility office. That type of urgency creates challenges for you, and we want to make sure that we're solving those problems.

This ultimately leads to limited resources. You guys are doing heroic work to make sure that the accommodations are served across the campus, whether it's at the department level, or at an entire campus. The limited resources that you face are real challenges, and it's important that we're finding ways to help you save time, because you're likely stretched very thin.

And then, regardless of who you are or how many people you have, it's sometimes just a complex task. Scheduling captioners, scheduling captioning events, knowing what's coming up next can be a real challenge. And so I've seen everything from really, really, really large Excel spreadsheets, to student accommodation platforms, to whiteboards for managing all of these requests. And so this is what we'll be talking a little bit about today with 3Play Media and the solutions that we're providing to solve a lot of these challenges.

So before we talk about the solutions, what challenges, what use cases are we really talking about? And I've sprinkled them in so far today. But first and foremost, we're talking about classroom accommodations, real-time accommodations for students, and accommodations for on-demand video that's present in those courses.

Second, in-stadium captioning-- the sports side of the university also has requirements to support accommodations for those who need them, and this is just one of the many use cases in which live captioning is a requirement for in-stadium experiences. And we'll be talking about that as well.

The final use case is commencements, and really, general campus-wide events where you're typically streaming to an external audience. So it could be a music concert, or it could be-- or it could be a talk from an external professor. These events, we find, are typically streamed out to your Kaltura, or your Vimeo, or a Zoom, and these events often require accommodations, both for your known students who need accommodation, but in general, you're going to run into that with your external audience.

So what is 3Play Media's live captioning solution? We're a platform of accessibility services, stretching from live captioning, all the way through localization, including recorded captions and audio description services. We have a one-platform approach so that your management of all of your media for accessibility is accessible to you in a visible UI, and through APIs and connections to third-party platforms.

Our platform and our services are really focused on providing accessibility compliant with all of the US guidelines and US laws and case law. In addition to the services that we provide for transcription and audio description and localization, we also provide user experience tools as well, like our access player, and our second screen experience that is being utilized today. The live platform and-- sorry-- the 3Play Media platform allows you to have unparalleled visibility into everything that is happening around your campus, and we're going to dive deeper into different aspects of the platform and how your institution can benefit from using this platform to organize all of your accessibility needs.

But for right now, we're going to just focus in on live captioning and the services and the solutions in which we can support your needs. So with 3Play Media, as mentioned, we support two primary mechanisms in which we caption content, live automatic captioning, which is produced by a world-leading automatic speech recognition system, and live professional captioning, which is produced by 3Play-trained voice writers and contractors. Each can be ordered separately. However, live professional captioning, in some instances that we will talk about, have automatic failover capabilities. So in the unlikely scenario that live captioner loses connection, we can failover to automatic captions and make sure that your event is still receiving captions.

So what are the options? Why LAC versus LPC? Besides the accuracy, which we'll talk about briefly, but LAC, L-A-C, or Live Automatic Captioning, biggest reason why you might be considering this is it's a lower cost option.

But with that lower cost tends to come with lower accuracy and a limited ability to capture audio for that content. A phone call, we can't really do ASR on a phone call. The data isn't being transferred to 3Play, and thus, we can't use that as an audio capture, as an example. So the use cases that we tend to see with auto captioning are for low-visibility events or high frequency with a low budget and meetings that don't require professional captioning.

Live professional captioning, on the other hand, obviously, it's going to have the higher accuracy and quality, but also provides the flexibility for really any audio capture capability. For example, our captioner today is just accessing the Zoom directly as, basically, a participant, just like you. On the other side, admittedly, it is a higher cost. And so you do have to put that into consideration when you're thinking about your overall budget and your ability to provide that consistent experience from event to event so that your audience can have those expectations set.

Use cases for this are really going to be centered around student accommodations, where meaning and accuracy is highly important for the student, high-visibility events where you're going to have lots of people or very, very important stakeholders attending that event. And typically, when you're running into complex content, similar to classroom content, this is really an opportunity to leverage professional captioners who have robust experience captioning that type of content for students. So that's a little bit about LAC and live professional captioning. Let's talk about the workflows within delivering those captions.

So what we think about when we think about captioning, we think about LAC and automatic captioning-- sorry-- professional captioning and automatic captioning as really just the tip point, the tip of the iceberg for the discussion. We also need to understand, how are we going to hear the audio, and how are we going to deliver those captions? And based on your answers along the way, you're going to be pigeonholed into different solutions.

So when we think about professional captions, we can really listen to any audio source to generate professional captions, but it will limit our ability on where we can deliver. At the very minimum, we can always deliver a second screen experience where you'll have a live text transcript being filled out with any professional captioning service. However, based on how we're collecting an audio source, we may not be able to deliver those captions into a specific player, for example. But we have solutions that we'll talk a little bit more about on how we can make that more flexible for you with professional captions.

On automatic captions, we're much more restricted. Specifically with 3Play Media, it requires us to have an RTMP stream, which is a very basic live streaming protocol that delivers that audio to the 3Play Media platform. That can be through Zoom or a basic software or hardware streaming encoder. Then, we can deliver those captions, most likely through second screen, back to Zoom, or through our embedded 608 solution that we'll talk about in a second. So those are just some of the questions that we ask and what you should be thinking about when you're deciding what is going to work for my solution, for my scenario.

So we're going to dive, really, just into one particular use case. We think that this is a really unique solution that's perfect for large school and commencements. We call it virtual captioning coding.

And the benefits here are the captions are-- you don't have to think about it. You send us your RTMP stream. We send you a link. We send you the RTMP to whatever video player that you want, whether it's Kaltura, Brightcove, Vimeo, and we're going to deliver those captions in near real-time to the visual content in the content.

So it's going to feel like almost you're watching an on-demand video, but it's actually going to be live. In addition, you get the benefits of being able to use our auto caption failover, and as I mentioned, being able to send that stream anywhere. And this is really a great use, a great solution for your school events or commencements where you're likely having an external audience that you're streaming out to.

So that's a lot about the streaming and the-- sorry, the captioning side of the workflow, but what about the admin side? How do I schedule? How do I make sure that I know that I'm going to have caption coverage? And that's where 3Play Media's solution comes in.

3Play Media offers a live platform for scheduling your events in minutes. So we provide an intuitive UI to create events in our platform that get matched to a captioner in our marketplace, or you can schedule live automatic captions as well. Within that scheduling event UI, you're able to provide custom event instructions, things like speaker names, word lists that are both used by professional captioners and our ASR technology to better improve the accuracy of your captions.

Once you're done submitting your event, you're really done. You just show up at your event, and you're ready to go. We'll help finish-- we'll make sure that the captioner is all ready, and then you can just focus on streaming that event and the technical aspects of that.

After the event, whether it's student accommodations or an external event, we'll have a full transcript ready for you. That transcript can be delivered directly to a student, or it can be downloaded from the UI. In addition to our UI, we also support an API that can be integrated with your workflows today. And so you can automate some of your workflows through our API.

And we're really excited about an upcoming integration in the next few weeks with one of those platforms that you might be using today for student accommodations. We can't share it right now, but we're really excited about the opportunity it might make in terms of saving you time around scheduling all of your caption events for all of your courses. So stay tuned for that in the next few weeks. I'm sure you'll hear more about that from 3Play Media.

So in addition to the live workflows, we also have generalized integrations for all of our post-production or recorded content that you need captioned today. And we work with all of the platforms that you work with on a day-to-day basis. Kaltura, Panopto, Vimeo, Blackboard, YouTube, and Zoom, just to name a few, are all platforms that you can automate much of your accessibility workflow with 3Play Media.

And ultimately, we are a platform that's going to scale with you. Whether your scale is volume of content, or numbers, or types of services, we're here for you to build your accessibility practice around 3Play Media. We handle well over one million files every year for accessibility, and we're always ready to handle more. We have built out robust processes that allow us to scale this incredible rate without compromising our quality. And so it's really important to understand, we're not just a live captioner platform. We are a full digital accessibility platform for your institution.

And part of that scale is really making sure that we're adhering or accommodating your billing needs. At every institution, there's different policies around billing, POs, and payment, and so we make it really flexible within our platform to have full visibility, whether it means I need to see how my entire institution is spending on accessibility services, or I need to send a budget, set a specific budget for a department, or, hey, there's a single project that has a very specific PO attached to it. All of those billing needs are capable within the 3Play Media platform. And that's the power of this enterprise platform for education institutions.

And at the end of the day, behind the scenes, we have unparalleled support throughout the organization. We have our world-class customer-facing support team backed by our robust team of operations, engineering and product, and account management to help you no matter what you have in store, whether it's a technical issue you run into, or you're looking to seek advice on your next use case and guidance on what solution might be ready for you. That support is end-to-end, from the moment that you make contact with 3Play, all the way through all of your events and beyond.

So to wrap up this section, I wanted to just comment on one particular customer who's found great success in working with 3Play Media, and that is North Idaho College, who uses 3Play live professional captioning at their annual accessibility symposium, which is an external event. And I'm just going to read the quote, because I think that Jeremy sums it up the best.

"Live professional captioning gave me a lot of peace of mind for Accessibility Camp Symposium. I definitely feel this product was worthwhile and helpful. It saved my sanity, relieved a lot of headaches, and it helped carry the event's accessibility. It's been a great experience working with 3Play Media and a breath of fresh air to be able to put your trust in a vendor and know that we're going to get quality live captions. Not only does 3Play care about accessibility, they care about me, and that translates clearly when utilizing their services like live professional captioning."

And I think we highlight this one specifically because it summarizes that we're partners in accessibility. We're really trying to make it possible for you and your small team, with your limited budgets and the urgent requests that you run into, that we're there for you. And so I want to just leave everyone with these three kind of themes that we really discussed today.

One, consistent challenges that we face-- I've talked to dozens of universities in the last couple months, and we understand there's consistent challenges that you face. Accommodations are not easy, and it's likely stretching your team thin. And we're trying to find solutions that allow you to solve for those problems.

We also know that, in live captioning, not all events are equal. You might need auto captions in some events. You might need professional captions in some events. But we have solutions that match for your live captioning.

And ultimately, 3Play Media is your partner. Our solutions are aligned to solve your problems, and we're continuing to invest in solving more challenges you face, and we're excited to learn more about those challenges as we work together. With that, I'm going to hand it back to Kelly to wrap us up and open up to Q&A.

**KELLY
MAHONEY:**

Thank you so much. Yeah, so a strong note to end us up on. I'm just going to tie a little bow on some things that we did not directly address in this presentation, but I think are important for your consideration.

First and foremost, 3Play Media services are legally compliant out of the box. We comply with the FCC, ADA, and WCAG requirements, among others. Our services are always performed by trained professional captioners to make sure that you're always getting the highest quality possible. Our support team has been rated five out of five stars by our customers, so we're there to help you in the event that you do run into any type of issues.

You have access at any time at no additional cost to all of the files from your live recordings. And there's often easy and discounted upgrade paths if you desire to order more services aside from what you originally needed. Finally, you can customize all of those additional services or your live captioning experience. We give you the flexibility to upload speaker names, speaker labels, curated event instructions, or other word lists that would be helpful for us to prepare and better serve you on event day.

And next up, I'll just say that the proof is in the pictures. We wanted to show you a sampling of some of the 1,500 plus colleges and universities that have trusted us to help them simplify their accessibility and accommodation workflows. I'll spare you reading them all out individually, but once again, this slide deck is going to be shared at the same time as the recording, so don't fret. You'll have access to see this deck again.

With that, we've got some time left over for Q&A. So we've gotten a couple questions. We can go ahead and dive into those. Let me just pull them up, make sure I address them properly.

One of the first ones that we received, Erik, I think you may be able to speak to this. Someone says, Pam says, I've spent many hours calculating accuracy rates, so they're familiar with the sort of WER/FER distinction, but formerly didn't have a way to capture formatting error rates other than qualitative comments. Can you share a little bit about how this may be calculated, whether 3Play specific, or generally within the industry?

ERIK DUCKER: Yeah. So 3Play Media, every year, manages a state of the ASR report. And so what we're doing there is we're compiling about six to eight of the top ASR engines in the market and we're running them through a control set of content. I think it's about 10 hours of content, or I forget if it's 10 or 100 hours of content. And we're also doing a-- we're comparing that against a proof set of human-edited content, so that human-edited content.

So we have both a control of the full edited version, and then, we can compare those to the ASR output. And so we're doing a calculation based on, really, the number of opportunities to get something right, and then measuring the accuracy of that. For example, 3Play Media, we really bias towards not over-- we want to eliminate false positives in our ASR, and so we really focus on finding high quality or high confidence in terms of our ASR output instead of best guesses. And so we do-- similar to you, Pam, we do kind of a big-- it takes hours and hours to do this, and we do it once a year to make sure that there's a state of the industry report. And as mentioned in the chat, Kelly has created a link to that state of ASR report to learn more.

KELLY Yeah. Feel free to browse that report. Lots of good information in there, not only about measuring accuracy, but
MAHONEY: it gives you good context for the industry generally. Sort of keeping in terms of something that 3Play knows a little bit more about, someone asks, is the virtual caption encoding device a cloud service?

ERIK DUCKER: Yes. It is a 100% cloud service, and it's fully integrated into the 3Play scheduling system. So all you need to do is schedule a live event as if you would schedule any live event with 3Play Media. It's just an option of configuration, and we will handle all of the cloud service needs for spinning that service up.

KELLY And then, a good follow-up to that, could you maybe explain a little bit more about the differences between
MAHONEY: virtual caption encoding and something like live automatic captions?

ERIK DUCKER: Yeah, so live virtual-- sorry, the virtual captioning coding is more about how we encode captions into the stream. So CEA-608 is the standard delivery of caption data into web streaming and broadcast, and that virtual encoding solution is agnostic to whether it's using automatic captions or professional captioning. So it's really just a process in which how do we deliver captions versus how captions are generated from machines or humans.

KELLY And speaking of how questions are generated, we have a really, really good one from Elizabeth. She is wondering
MAHONEY: how automatic captioning or AI captioning, she says, is still ADA compliant. She's always heard that automated captions do not rise to the level of compliance, and 3Play is very well-equipped to elaborate on that. So, Erik, if you want to take a crack at it, go ahead. Otherwise, I've got some things to say.

ERIK DUCKER: Yeah. So I think in the general case, it's not ADA compliant. And specifically, most of the case law has pointed towards WCAG standard 2.1 AA in terms of remediation, and those standards point to providing closed captioning that includes non-speech elements. And ASR technology, to your point, does not typically cover non-speech elements, like speaker labels, and they don't do it accurately if they do. And they also don't provide non-speech elements.

And so when you look at it from a pure interpretation of what's out there legally, it's definitely not ADA compliant. Now, in practice, that's different than theory. In practice, we have found that automatic speech recognition can be acceptable in some use cases in the live ecosystem. We have found much bigger pushback specifically in the prerecorded space for ASR being acceptable for compliance. So you're not wrong. [LAUGHS]

KELLY MAHONEY: Yeah, you encapsulated it perfectly. And we have a question more technically about how the actual captions are delivered. Someone asks, does 3Play's tool put the captions into video recordings when they're made available, such as captioning Zoom recordings?

ERIK DUCKER: Oh. So it's really dependent on the platform. So 3Play Media does not necessarily have an influence over that. I would have to-- my understanding with Zoom is that they have their own recorded captioning for-- they have their own recorded captioning service.

And so I think any recorded Zoom calls will go through that transcription service. But for example, if you use our captioning service and virtual captioning-- virtual caption encoding service with Brightcove or Kaltura, they're going to convert that caption track into a WebVTT file in their recorded content. So it kind of depends on the platform, at the end of the day.

KELLY MAHONEY: That makes sense. An easy question, I feel like, for you to answer-- do we have captioning in other languages besides English?

ERIK DUCKER: Yes. Can the-- sorry. So for on demand, yes, we support English and Spanish in a robust manner, and then, we work through our partners for non-English or Spanish language captioning. In the live space, we can support English, Spanish, Portuguese, and French live captioning, depending on your needs.

KELLY MAHONEY: Great. Thank you. Also in the chat, I saw a question about 99% accuracy and whether that is a WCAG standard. I don't know off the top of my head with 100% certainty whether it is a WCAG standard, but Casey on the back end did share a blog or two about caption quality so that we can provide more information for you that way. It's definitely an industry standard, if not specifically a WCAG standard.

Next in the Q&A window, Meg is asking for a little bit more elaboration on the audio description service. You make a great point that I did not audio describe the slide with all the different schools, and that's a note that I will take with me to keep in mind. So that would have been inaccessible to anyone who is blind or low vision.

Typically, the accommodation there would be I do better about making sure that I audio describe what is on the screen. I can also describe the audio description service. But Erik, I'll let you take a first approach at it again.

ERIK DUCKER: Yeah, so the 3Play Media audio description service is very similar to our other services. We'll generate a transcript using our 99% proprietary system of human editing, and then, we will send that to an audio description job where we have a marketplace of audio description scriptwriters who can script for either a standard audio description, which means that you're trying to fit within the exact time length of the video, or an extended audio description, which is being able to-- where you have the ability to extend beyond the time of the video.

So that service is going-- that scriptwriter is going to write scripts based on our 3Play standards, which are a combination of audio description standards published across the US. And then we use either voice-synthesized speech to record that audio track over the scripts, but we also have voice artist description audio tracks available as well as necessary. In the education space, we have largely found voice-synthesized speech to be sufficient for our customers.

KELLY

Thank you. Yeah, and that's the detailed information about AD. On a zoomed out level, it basically is just accommodating blind and low vision users and describing what you see visually verbally.

MAHONEY:

So I want to say thank you again for pointing that out. It's always great to be mindful of how we can improve our accessibility. Thank you, everyone, for being here. We appreciate your attendance. Thanks again for joining us. And that's all we have for you.