

### 3Play Webinars | WBNR 12-12-2019 ASR

ELISA EDELBERG: Thanks again for joining this webinar, "What to Know about Automatic Speech Recognition." My name is Elisa Edelberg from 3Play Media, and I'll be presenting today alongside my colleague Roger Zimmerman.

So just a little bit about myself so you know who is presenting here this afternoon. Like I said, my name is Elisa. I'm a content marketing manager here at 3Play. I'm super passionate about web accessibility and making the world a little bit more accessible for everyone. And outside of work, I love dogs, and I love all things crafting. And I'm going to hand it over to Roger to introduce himself and get us started.

#### **ROGER ZIMMERMAN:**

Great. Thanks, Elisa. My name is Roger Zimmerman. I run the research and development here at 3Play Media. Research and development at 3Play is basically the application of machine learning and data science to our services, which include captioning and audio description. And so that's very relevant to this presentation, where we're going to talk about one of the hottest areas in machine learning, which is automatic speech recognition. And you'll get a little feel for how that is integral to our process here at 3Play and why this report is so important.

So we're very interested in keeping up with the latest technologies. I don't like to call it artificial intelligence. I think that term is overused-- but machine learning and data science are truly pushing this field, and we need to stay on top of it if we're going to do our job here. And my hobby is playing guitar-- sort of badly, but I enjoy it. And my dog doesn't seem to mind. So that's what I do.

So here's the agenda. I'll just read it through. We're going to talk about the annual state of the automatic speech recognition report that we published. We're going to say what the research goal was, how we did it, and what the results were, sort of in basic scientific paper style, but it's not going to have the kind of depth and rigorousness of a peer-reviewed paper. That's a caveat I want to put out front.

So that's kind of the theory of what we are examining here. And then Elisa will take it over and get into more of the practice-- what these findings mean for you, examples of automatic speech recognition in captioning, typical automatic speech recognition errors, and then kind of wrap it up with key takeaways. And

as Elisa mentioned, we will have collected your questions and answered those and have some more time for that at the end of the presentation.

OK, so this is what we are doing. We've done this now approximately three years in a row. There was a little bit of gap before the research three years ago. This is the first year we've actually reported on it. We want to know the current state of automatic speech recognition, because automatic speech recognition is the first step in our process of producing captions for accessibility and engagement, and all the good reasons to have captions.

Automatic speech recognition in our process is then presented in front of our expert transcriptionists, who we actually call editors, because what they are doing is they are editing the speech recognition output in order to create a perfect transcript-- or near enough to perfect as the eye can measure. And so basically, the better speech recognition gets, the less editing work our transcriptionists need to do, and therefore the more they can focus on the higher level tasks that are required to create a high-quality caption. And we'll be discussing what some of those tasks are as this webinar proceeds.

But what we're basically trying to do is look at as many speech recognition technologies out there as we possibly can and evaluate them in terms of their impact on our process. So we're aware of reports of automatic speech recognition approaching human accuracy. And to be fair, there have been many improvements in automatic speech recognition, and we leverage those improvements in our process. But we need to be very careful about when we talk about accuracy, we need to ask, for what task is the accuracy being measured?

And this research is all about applying speech recognition technology to our task, our task of captioning a very wide range of very varied data for a lot of environments and a lot of situations. And this differs quite a bit from some of the other tasks that are out there. So that's what we are doing now. That's what we will continue to do in the future. We are going to publish this report annually. As I said, we've been doing the work approximately annually. But we are going to publish it so that the community can be aware of the progress in these technologies.

So this is the task, and this is the universe of APIs, basically, that we examined this year. We tested six of the most popular ASR technologies across content from e-commerce, higher education, fitness, media and entertainment, and enterprise industries. The speech recognition APIs we tested were Speechmatics, which is the one that we currently use; IBM Watson; Google, which is the one used in YouTube live captioning; Microsoft; Temi, which is the one used by Rev or Rev.ai; and Trint. As it turns out, Trint also uses the Speechmatics technology, although there are some differences between how Trint uses it from how 3Play Media uses it, and we'll discuss that. So it's a very wide range of very popular technologies, but the key to this project is how we tested them.

So let's go to the next slide, Elisa. So this is where the test and our sort of objectivity in doing this research overlaps with our self-interest. 3Play wanted to test these engines on our own data. So we collected real content that is reflective of the most common type and volume that we receive at 3Play Media. So we wanted to get a clear picture of where we stand.

We chose to test content that spanned across the industries and the conditions which best represent our broad customer base, and the content type that we receive on a regular basis. Let me flesh this out a little bit here. This content is very distinct from a lot of what you hear about being tested for speech recognition-- or reports you hear about speech recognition accuracy.

They're distinct from academic reports. There's a lot of work going on in the academic community about speech recognition, but they are using these kind of benchmark databases which are very useful for academic research, where you sort of want to keep the data constant, but become kind of a mountain that everyone is trying to climb. And you can get very skewed perception of accuracy when you focus on these benchmark databases.

Rev recently published a report on speech recognition accuracy using 20 different podcast files. They gave the links to those in the report. And podcasts is one of the domains that we are interested in. So we definitely included some podcast-like data in this test. But it is not at all the universe that we're interested

in.

And of course, we test it on a lot of data. You will hear reports about people who test it on one or two files. And obviously, those are anecdotal kinds of reports and not really relevant to our interest here.

And there's one aspect of speech recognition which is very well known. A lot of people are using Siri and Alexa and those kind of voice assistant types of technologies, and are getting extremely good success with those. But this is very distinct from those tasks as well. Those tasks, the speaker is known, the queries are very constrained, the kind of things you can ask Siri or Alexa are very limited.

And most importantly, the words themselves when you're talking to Siri and Alexa are not important. It is not important to transcribe the words that are being spoken. What is important is that Siri, Alexa, and the other voice assistants get your meaning about searching the web or turning on a particular song from your music player. So it's a very different task. So it's not reasonable, it's not possible, really, to extrapolate from experience with Siri and Alexa and those types of technologies to this transcription task where getting every word and more, as you will hear, is required for the captioning task.

All right. So let's go on to what the data actually looked like. We collected a total of 423 files from our data set, comprising more than 8 million words and almost 100 hours of data. So this is a very substantial database. They were collected across a number of different industries-- and I'll just read the number of files from those industries so you get a feel for this. Education, 178 files. E-learning 85. Entertainment and media, 62. Corporate, 37. Online video, 34. Market research, 9. Faith, 7. Fitness, 6. Government, 5. Again, totaling 423 files.

A large variety of industries, and this doesn't quite capture all the variety. So I want to dive into a few of these cases to just give you a feel for what's going on here. So in the education and e-learning bucket, we're talking about high school curricula, college, live presentations in a MOOC setting, over many different subject areas-- some very technical, some artistic. You get the idea-- a wide range of content. Usually one speaker presenting. Sometimes there's Q&A in

those kind of presentations, but a lot of variation within that content.

Entertainment and media-- there's feature films, there's live content, there's television, there's sports. Again, you get the idea-- a large variety of content there. In corporate, that encapsulates things like internal training, advertising content, instructional content for customers, and the like. There's a lot of variety there.

And in the online video category, we have content that's published on YouTube, Vimeo, Wistia, et cetera-- kind of the whole array of online video that's involved here, which can involve different kinds of audio characteristics, different kinds of presentation styles, et cetera. Interesting that YouTube, obviously, was mentioned, does use Google for its live captions. So just a point to note.

OK, and then how did we measure on this large database? Well, there's a pretty standard measurement, which is called word error rate. And we measured word error rate for all these engines on this data set. And we found out, as you can see, sort of stealing the punch line here-- Speechmatics had the lowest overall error rate, an accuracy of 90.9%. Google and Trint were within 1% of that, but it's always important when you're looking at statistics to do a statistical test to see if there's significant differences. We use the very well known in the speech community McNemar's test, which judges, essentially, whether the word error rate differences between two systems are statistically significant. And we found that that was true in this case.

So that was for basic error rate, for raw word error rate. But then we looked at what we call formatting error rate. And formatting error rate, Elisa will get into with a lot more detail later. But just as a first-level understanding there, that includes notions such as punctuation, capitalization, numeric formatting, speaker labels, all of the things-- editorial notes. The examples will follow, but all of those things that are required for excellent captions.

So generally speaking, you will only hear about word error rates in the academic community. In the captioning task, we need to be focusing on formatting error rates. And so we did analyze those. And again, Speechmatics was the most accurate engine there with 77.9% accuracy. And we'll dive a little bit into the

details there now.

So here is a breakdown of now word error rates-- not formatting error rates. But this is kind of what the speech community judges as the quality of a speech recognition engine on a particular task. And I'll just read the first column, which is the percent word error rate, which is Speechmatics 13.06, IBM Watson 26.77, Google 14.01, Microsoft 14.90, Trint 13.17, and Temi 22.90 on this task.

Note that error rates do not equal percent correct, which is what is in the second column here-- 100 minus percent correct, because speech recognition engines can insert words that are not there. Yes, I will address that. I did misstate the number of words involved-- 800,000 words, not 8 million words. Sorry about that. It was 800,000 words-- still statistically significant amounts of data.

But anyway, so the difference between, you might expect Speechmatics to have a, roughly, 9% word error rate, but in fact you have to count insertions as errors as well. So there were approximately 4% insertions for Speechmatics, and you can see that in the second-to-last column. So those were the numbers we got on this task. And yes, that sort of speaks for itself.

So let's go to the next slide, which is the formatting error rate. So here you see greatly increased error rates. And again, we're taking into account all of the formatting aspects of the task that we need to here. Now, it's worth noting here-- so you see more divergence between Trint and Speechmatics plus 3Play here. So we replaced the first row with Speechmatics plus 3Play. Again, I'll read these for everybody in terms of word error rate or formatting error rate.

Speechmatics plus 3Play 25.37, IBM Watson 39.90, Google 27.76, Microsoft 32.73, Trint 26.15, Temi 33.72. So the point about why is there more divergence now in formatting error rate between Speechmatics plus 3Play and Trint than there was in word error rate-- that has to do with a technology we use internally that we call our mappings technology. And basically what we do at 3Play is we learn from our transcriptionists' corrections to speech recognition.

Most of those corrections involve formatting types of changes. And then we can apply those. So really the engine here is the Speechmatics off-the-shelf API plus this mappings technology. And as you can see, we can get a gain in the

formatting error rate relative to those that do not use that technology.

So that was the picture. We feel very confident-- sorry, next slide. We are very confident based on these results that at this point, we are using the best available technology for our problem. And that is always the qualification. I want to emphasize when you're evaluating a speech recognition technology, you have a particular problem to solve. You need to choose the technology that is best for that problem. And again, our self-interest here is entirely aligned with an objective assessment of this. So we're confident that we're doing the right thing.

On the other hand, ASR has improved of late, and it still continues to improve. So we need to keep our eye on the entire landscape here. As such, we will be continuing to do this assessment continually each year and making sure that these results stay the same-- or if they don't, making a move there technologically.

And then I guess it's, obviously, when you look at those formatting error rates and you think about the fact that even with the best technology here, we are not even getting 80% accuracy on average. You will understand that speech recognition by itself is not sufficient for solving this problem at this point-- at least the problem that 3Play has to solve, which involves this large variety of data. So I think I have made my point, and I will hand it over to Elisa now for some more practical stuff.

**ELISA EDELBERG:**

Thank you, Roger. So now that everyone kind of has an idea of what the testing consisted of and what the results from the testing looked like, it's really important-- we want to make sure that you know and can walk away with what this actually means for you. So the biggest takeaway is that, while technology continues to improve and we've seen that over the years, there's still a really significant leap to real true accuracy from even the best speech recognition engines. So at this point, humans really are a crucial part of creating accurate captions.

So you might be wondering why is this the case and kind of how so. So some of the things that we're going to cover in the next half of this presentation are looking at examples of common ASR errors and kind of why those happen. We'll

look a little bit more at word errors versus formatting errors. We'll talk about function words and why ASR technology typically fails in some of those words and the kind of implication that that has. And then we'll look at what 85% accuracy really means. And we'll touch on the ASR captioning lawsuit.

So I want to play this example for everyone. It's a transcript generated by automatic speech recognition, and I'm hoping that everyone is up to participating a little bit this afternoon. So hopefully you can take a listen, and then feel free to enter in the chat window any errors that you notice. And I'll collect those and touch on them aloud after the clip.

[VIDEO PLAYBACK]

- One of the most 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 a horseshoe crab, and I touched a horseshoe crab. And I still remember that, and I'm 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 of trees and wildlife that he had introduced to me. So my entire childhood was around wildlife and in wild places. It just clicked.

- When I was a kid, all the cousins would use my grandparents' driveway.

[END PLAYBACK]

**ELISA EDELBERG:**

Great. So I'm going to read some of the comments coming through. A lot of people are pointing out the lack of punctuation, wrong words. The "New England Aquarium" was completely inaccurate. It was transcribed as "the new wing of the Koran." No indication of speaker changes. "Four-story" became "forester." Yep, exactly.

Someone else pointed out capitalization, punctuation, speaker ID, no indication of music. So these are all-- really long run-on sentences, no breaks, no capitalization. Yeah, everyone did a really great job of noticing the many errors. And kind of the point that I want to make here by showing this example is that these are the type of errors that a human who has the ability to think reasonably and use context clues would not make.

So these are things that completely change the meaning, made it really difficult to follow. And again, ASR technology can't make these rationalizations. So that's just a really clear example of why humans are still important. And then Roger touched on this earlier, but when it comes to captioning accuracy, it's really important to consider both formatted error rate and word error rate.

So I've listed some examples on the screen that I will read aloud. But these are just sort of common causes of ASR errors. So when we're thinking about word errors, a lot of common causes include multiple speakers, maybe overlapping speech, background noise, poor audio quality, false starts, and then acoustic errors, which is what you heard or what you saw in the examples. So "New

England Aquarium" sounds similar to "new wing of the Koran," but it doesn't actually make any sense. And then function words, which we'll get into in just a moment.

And then as far as formatting errors, again, you saw a lot of these examples. But we have speaker labels, punctuation, grammar, numbers, relevant non-speech elements-- so for example, a door slamming or keys jingling-- and no inaudible tags.

So for me, I think that the best way to really understand the importance of punctuation and other formatting errors is to see some examples. So I have a couple of silly examples that I kind of wanted to share with everyone, but they really do sort of hit home as to why punctuation is important. So there's an image on the screen. It says, "Let's eat Grandma!" And then there's a little cartoon of an old woman saying, "What?" And then, below that, is the sentence "Let's eat, comma, Grandma."

And it says, "punctuation saves lives." And obviously this is a little bit tongue in cheek, but the point is punctuation really changes the meaning of the sentence. And if we think about this a little bit more seriously, if you're thinking about students in a classroom that are trying to use captions to learn and follow along, or maybe it's a presentation, someone's relying on the captions for understanding, to access the topic and the content, this can really have a big implication on whether or not they are able to follow along.

And then I have another example kind of similarly-- it's the cover of a magazine, and it says, "Rachael Ray finds inspiration in cooking her family and her dog." There is no punctuation, and this implies that Rachael Ray likes to cook up her family and her dog rather than the fact that she enjoys all of these three things separately. So again, this just kind of puts it in perspective of what formatting and what punctuation really does for content.

And then we've mentioned function words a couple of times. And this is another area where ASR usually fails. The example that I have on the screen is "I can't attend the meeting" versus "I can attend the meeting." This is really common, especially with background noise or if a speaker is not emphasizing or

enunciating. It's really easy for an ASR to miss these. And again, a human can kind of use context clues to understand what's going on.

If I say, I can't attend the meeting, can we reschedule? It indicates that I was implying "cannot." But these sentences really mean 100% opposite of each other. So again, they seem like small errors, it's just one tiny apostrophe, but the meaning is really completely changed.

And then I have an example on the screen of complex vocabulary. And these are several examples from the actual text that we did. They tend to focus on names and complex vocab, which really require human knowledge and expertise. So in each case-- and I'll go through these in just a minute-- the truth is on the left and the ASR is on the right. And there are a lot of errors, again, both incorrect words as well as formatting. So I'll kind of go through them side by side.

First, it says, "Picked up really well by Ehrhardt." And then the ASR transcribed that as "Picked up really well by air, quick passing front bone slaps at home Virginia one loyal nothing." And that was just one really long run-on fragment, really, no punctuation. And then what was actually said in the following sentence was "quick pass in front. Bowen slaps it home. Virginia one, Loyola nothing." So again, the ASR really kind of butchered the meaning and the content here.

And then the next example, it says, "This week, you will focus on identifying who primarily experiences precarity, who makes up the growing precariat." And the ASR transcribed that as "this week. You will focus on identifying who primarily experiences precariously who makes up the growing prokaryotes." So those are just some really good real-life examples.

And then it's important to note that what we're looking at here-- at 85% accuracy, one of seven words are incorrect. And while this may not seem like it makes a very big difference in a smaller sentence, when you have two sentences, a paragraph, and multiple paragraphs and it becomes longer form, this really adds up very quickly. So even 85% accuracy really is not sufficient in terms of providing an equal experience. And the longer the content, the more inaccurate it gets.

So I wanted to just kind of touch on this Harvard MIT lawsuit. Obviously, you may know there were some updates to this quite recently. But I would like to point out this example because it does indicate the importance of accuracy for captions. So this was a quote from that suit. "Much of Harvard's online content is either not captioned or is inaccurately or unintelligibly captioned, making it inaccessible for individuals who are deaf or hard of hearing."

And the important thing to note is that those inaccurate captions were from ASR technology. So it really speaks to the importance of having accurate captions and the fact that ASR alone is just not at that point yet. And then Roger spoke to this a lot in the first half, but I just wanted to bring it up again because it really is important, and really important when looking through our report and the data that we collected. It kind of compares why Siri may seem so good but captioning may be so bad.

And some things just to reiterate to consider-- when you're talking to Siri or Alexa, it's a single speaker, really constrained tasks. They can ask for clarification. Did you catch my drift? And the speaker is usually very close to the assistant.

And then with captioning, a lot of the times there are multiple speakers. Tasks are more open-ended-- they can really be on any topic. It can be broader discussions. There's usually background noise, possibly poor audio, lost frequencies, and most of us don't speak perfectly-- and then the last item, changing audio conditions. So these are just some things to consider when comparing the different applications of ASR.

And then a couple of key takeaways-- so the application of AI in captioning is really still in the works. There's still a lot of improvement that needs to be made when it comes to formatting error. None of the solutions that we tested are fully sufficient. Some of the best ASR systems can achieve accuracy rates in the 80s and low 90s if all conditions align perfectly.

Even the best automatic captions, as we mentioned, have a lot of room for improved accuracy. And there are still some really fundamental advances in machine learning in order to replicate professional human editors. And lastly,

3Play plans to continue to monitor the landscape for improvements in these technologies, and as Roger said, make sure that we're constantly using the best of the best in our process as well.

So before we move on to Q&A, I just want to point out a couple of items for next steps. So the full ASR report is available for download, and I will send the link out in the chat window. But I encourage everyone, if you're looking to learn more, to certainly take a look at the full report. And I'm sending that in the chat window right now.

And then I also wanted to share some exciting news. We have a video accessibility certification course coming out very shortly next month-- right in new year. And the video certification is online, it's free, and it really just offers a lot of content to help you become more of an expert and be more well-informed to make decisions and be an advocate for accessibility at your organization.

If you have found our webinars and white papers helpful, this is kind of a more in-depth and more well-rounded resource that you may enjoy. And then it also offers the opportunity to network and meet other accessibility enthusiasts along the way. So I will also send the link for that.

You can sign up immediately to make sure-- just want to make sure that I am sending this out to everyone. You can sign up immediately at this link-- [3playmedia.com/certification](https://3playmedia.com/certification) to make sure that you're staying up to date as it becomes released. And then with that, we are just about ready to move into Q&A, so keep the questions coming in.

Great. So someone is asking how formatting error rate factors into legal standards. So that's a really great question. I think the thing to consider here is some of the examples that I showed. So a lot of the legal requirements, they don't really specify accuracy. A lot of them say things like providing equal access. And it's really important to consider the overall accuracy and quality of the captioning and transcription.

So it's hard to say exactly. They don't really specify about formatting error rates specifically, but kind of big picture some of the examples that I showed and that we talked about, you can tell how those formatting error rates really do have a

big impact on the accuracy of the captioning and on being able to provide that equal experience.

**ROGER ZIMMERMAN:** --addressed this one already.

**ELISA EDELBERG:** So someone else is asking if this makes us ADA compliant. Again, the laws don't specify exact numbers, but our captioning is guaranteed 99% accurate, which is along the standards of the industry.

**ROGER ZIMMERMAN:** Yeah. So by this, the automated captioning obviously no, but in reference to 3Play's full solution, Elisa's answer is on point there.

OK, so then there were a bunch of technical questions on ASR. Let me see if I can get through these pretty quickly here, because a lot of them are related.

"Is it possible to test engines that don't have APIs such as Otter AI?" The answer to that is it is difficult to do so. You have to buy the software. You have to install it on a machine. And really it's not relevant to us. If there was a compelling case that one of those engines was very superior, we might think about it, but, you know, we're a workflow solution here, so we need APIs.

And then somebody asked specifically about Amazon Transcribe, which does have an API. And we do intend to test Amazon Transcribe and publish about it next year. Amazon has some restrictions on publishing their results, which require that you make the data available, so we're going to need to do some careful data curating when we do that. But we should be able to include the next year and get some useful numbers out of that.

Another very interesting question is how did these results relate to our new live automated captioning service. And the answer is live is harder. All the results presented here are in batch mode. And batch mode, you can do a lot more computation on the audio and do a better job. So generally speaking, we've seen that the live automated accuracy is about 20% relative worse. So if you're talking just word now, if you were at 15% word error rate, the live solution would probably be about 20% word error rate.

So what does this mean? Well, this means that if you want to have live

automated captions that are of use, you better choose the content very carefully. It better be very high-quality audio, preferably one speaker, preferably scripted. A webinar type situation might be a good basis for that if you want highly accurate.

And I guess I should say the numbers presented in this presentation were averages over a large set of data. There is definitely a subset of the data where the batch recognitions can get into the 90s, sometime into the mid-90s, in terms of accuracy. And so for that type of content, live audio captioning may be a viable solution in order to get something up there for accessibility enhancement. But we would not want to say at all that live automated captioning would make things compliant, if you will.

And then there was a specific question about graduate level-- someone said, "I have graduate-level medical education content with scientific and Latin terms, and ASR has been challenging. Any tips?" Well, I'm not surprised it's been challenging. ASR is challenged by content, certainly by out-of-vocabulary words. My tips would be if you're going to stay with ASR, you should get the absolute best possible audio. You should try to script these presentations of the content so that the speaker doesn't have hesitations and restarts-- spontaneous events.

And there are capabilities in many of these engines, and 3Play will be adding these next year, actually, to add word lists prior to running the speech recognition. So I don't know how they'll do on the Latin terms. They will probably do better on the scientific terms. So that could help. My biggest tip is if you want really good captions, you should use a human solution for those-- technology-assisted human would be best, from 3Play.

Another technical question-- "do any systems learn as the conversation continues?" Yes, I think most of these technologies now are doing what's called runtime adaptation. They're adapting to the audio characteristics. They're adapting to the language characteristics as the kind of events progress. So you may see improvement in accuracy at the margins as events progress. But all that is baked into these results.

And then another question was talking about, again, the distinction between

video on demand and live video, where two-hour turnaround time is not acceptable sometimes in either of those situations. And that is kind of the minimum turnaround time that 3Play and other caption vendors commit to.

And so that's a situation where having something using automatic speech recognition can get you something quickly that will be less accurate. So there is a valid trade-off consideration there, which is that human captioning does take some time. And if you need the captions right away, human solutions are probably not going to suffice.

So we've had customers use kind of a hybrid approach, where they take the speech recognition output quickly, or even do live speech recognition. And then they upgrade to the full transcription capability, perhaps with a two-hour turnaround, perhaps with more. And then you can have the best of both worlds over the long term.

I think that was the list of questions I saw. Were there any that got added? Yes, I addressed that. There was somebody that-- right. Yeah, that is an interesting question about our process. "Given a transcriptionist an ASR transcript, how much time do they end up spending to transcribe?" And yeah, that's a very relevant question. 3Play is obviously premised on the idea that editing speech recognition transcripts is faster than, essentially, typing from scratch.

And the answer to that is typically, human transcription from scratch goes in the range of four to six times real time depending on content. So a one-hour video would take four to six hours to transcribe, whereas editing a speech recognition draft can typically take two to three times real time-- so about half that.

I think more important than the time versus real time, however, is the focus of the transcriptionist. A couple things-- one is they don't have to worry about precise timing, because the speech recognition effectively helps them with the precise timing. So to the extent that you need synchronized captions, that's a big help.

And then the other thing is just focus in terms of content. Typing from scratch involves getting all of the elements-- word accuracy and formatting accuracy-- correct simultaneously, whereas if you have a speech recognition draft, you can

focus much more on the missing formatting elements and kind of spot-checking the word errors. So we believe it enables a more accurate end product.

**ELISA EDELBERG:**

Awesome. Yeah, I think that that is all the questions that we've gotten in today. So I want to thank everyone for joining us this afternoon. Our contact information is on the slide if you have any additional questions.