

Captioning Trade-Offs in Fast-Paced Media: User Insights on Speed, Accuracy, and Delay

[00:00:00.17] KELLY MAHONEY: Let's get started with today's presentation. So first, before anything, I want to thank everyone for joining us for today's session, Captioning Tradeoffs in Fast-paced Media, User Insights on Speed, Accuracy and Delay. My name is Kelly and I'll be moderating today's session. I use she/her pronouns and I'm on the marketing team here at 3Play Media.

[00:00:21.24] Just a brief self-description for you. I'm a young, white woman with long, reddish brown hair, and I'm wearing a green collared shirt today. With all of that taken care of I'm happy to welcome today's speaker, Deborah Fels. Thank you so much for being here with us today, Deborah. And with that, I'll pass it over to you for what I'm sure will be a great presentation today.

[00:00:41.69] DEBORAH FELS: All right. Thank you, Kelly. Much appreciated. That was a lot of information about yourself.

[00:00:47.02] [CHUCKLES]

[00:00:48.23] I appreciate it. And I go by pronouns she/her, and I am a professor in the Ted Rogers School of Management at Toronto Metropolitan University. And I shortened the title just a little bit, although it seems long. But we're going to look today at some of the research that we have been doing over the past five years or so, maybe a little bit longer, on what the impact of the speed, accuracy, delay tradeoffs are in captioning for the fast-paced media on different users. So we'll figure out who those are.

[00:01:35.23] I'm also OK, if you have some burning questions while I'm going through this with you, to stop and answer them. Because sometimes at the end of the talk, you forget the questions or they don't really-- they're not that important anymore. So I'm happy to do that. OK, so let's get going here. All right.

[00:02:02.28] So I'm going to tell you about three different pieces of research that we have carried out, along with some final thoughts. And the first one is going to be on subjective mental workload of captioners, who caption live, fast-paced sports live. And just maybe for a quick show of hands or show of something, how many people in this group are captioners-- specifically live captioners, but captioners? A show of something-- got a couple.

[00:02:44.71] KELLY MAHONEY: Yeah, we've got some hands raised. I see about three or four.

[00:02:47.83] DEBORAH FELS: Yeah. OK, good. That's very good. All right. So this particular study was the first ever of its kind from a research point of view. But I think that most captioners would be understanding of the results. And they would say, well, that's kind of obvious.

[00:03:10.18] All right. So the first thing that I wanted to go through with you-- and I wanted you to type them in the chat about, what do you think the issues are with live captioning? And it can be as a viewer or as a captioner.

[00:03:29.42] KELLY MAHONEY: So I'll read out what I'm seeing in the chat here. We've got latency, accuracy multiple times, fast talkers. Sometimes things get complicated and garbled. Background noise, technical jargon.

[00:03:44.45] Accents sometimes can change the interpretation of these things. Someone calls out scrolling is problematic as well. Lots of-- lots of engagement. So this is a great question to start off with.

[00:03:57.71] DEBORAH FELS: All right. Yes, very good. OK, so here's my list so far. There we are. All right. So a lot of people said it's fast.

[00:04:11.60] So what does that mean, fast? So oftentimes the speakers are going more than 220 words per minute, which is very high. And so what does that mean, though, for captions? It means it's hard for captioners to keep up without errors. And there's a whole bunch of different kinds of errors, but there's errors.

[00:04:37.57] It's also a high workload for captioners. So captioners get tired, and as they get tired, they make more mistakes. That's just the nature of human beings.

[00:04:48.52] And then on the other side, when words are going by really fast as captions, it's hard to read. And reading is not the same as talking. So it takes time to read and it takes time to type. And so we end-- as we saw in the chat, things going fast is was identified as a problem by many people. And that's true.

[00:05:18.78] OK, another one is, as a result of things going fast and the live nature, is that errors just happen and they don't get corrected. So there's different kinds of errors that happen, but a big one is where the end of the sentences are left off. And so that means that things are incomplete or things are not quite right. And I think there was misspellings of player names and things like that.

[00:05:51.97] So misspellings, missing words, the ends of sentences cut off, and those won't ever get corrected. So there's no chance for that. So you have to put up with the mistakes.

[00:06:04.99] And then a number of people said delay. Absolutely. So what happens with delay when you're trying to watch a sports game? Well, the play continues and the captions and the play don't match. So now it's confusing.

[00:06:26.57] And I was looking up what the average delay time is in the live environment, and it's about five seconds-- kind of the average. So in a fast-paced sport like basketball or hockey, lots happens in five seconds. And if the caption is for a play that happened five seconds ago, it's kind of confusing. So what ends up happening then? For the viewers on our chat, what do you do?

[00:07:18.23] Rewind? Yeah. So-- what Megan and Cheryl said, is they just ignore the captions. They turn off the captions. And that's a really common result of all these things that are going on.

[00:07:42.05] Rewind if you can. But then if you are watching it with other people, rewinding makes it hard. So it just makes it really hard to watch the game. So there's a competition between reading the captions and watching the game.

[00:08:03.24] And so that's frustrating as a viewer. It's frustrating as a captioner because they know this is happening. And so overall, it makes it for a difficult situation and maybe not as fun. OK.

[00:08:22.50] So I'm going to go through a couple of things that we have done in order to try and examine some of these problems and find some potential solutions. So the first one we did was the mental workload assessment for live captioners. So what we were trying to look at is, can we measure how much workload live captioners experience?

[00:08:50.54] So we had 17 captioners. Most of them had more than eight years of experience as live captioners. We used the NASA TLX mental workload assessment tool, which was initially launched in 1980 at NASA, and so it's been around for a long time. It's a very common tool that's used to measure mental workload.

[00:09:16.07] And we did this during the COVID restrictions. So it was an online-- which has limitations. And we did some post-survey interviews with some of the captioners-- not all of them. Let's see what we found-- so hang on.

[00:09:32.21] Before we go there, this is what the NASA TLX looks like-- the questionnaire looks like. There are six main factors-- mental demand, physical demand, time or temporal demand, performance, effort, and frustration. So a participant would rate how they feel about those things-- the workload that those things are imposing on them for doing the live captioning task.

[00:10:07.37] The other component to a mental workload measure in its full version is that people will also rate the importance between the pairs. So, is mental demand more important to you than physical demand, for instance. So that's a set of extra questions that are done in the full version, which we did for this very first study. If you're interested in the technique and the measures, I put on the URL so you can go and look up how to do that.

[00:10:48.54] OK, so what did we find? We have a chart with the mean and standard deviation for the scores. So we take the pairing scores, which is called a weighting, by the actual score that they scored for the live task. And we multiply them and add them and multiply them and divide and we come up with a rating. So for each of those factors, we plotted those.

[00:11:22.61] And what I want you to take a look at here is the rating for performance. And the question is, how successful were you in accomplishing what you were asked to do? So for captioners, they are asked to caption the live speech that's going at more than 220 words a minute as accurately as possible-- the standard being verbatim or 100%.

[00:11:58.56] That's a lot of pressure on captioners to do that. And they scored that part of their mental workload as the worst. So the lower score on the scale is perfect and the higher is failure. So they were scoring themselves as failing their performance-- as failing at their performance more than anything else. So their opinion of their performance was contributing the most to their feeling of workload. All right? And I'm thinking that some of you may be nodding your heads or agreeing with this, that your concern about your performance gives you the most stress.

[00:13:04.55] And probably not surprising either is that the physical demand-- the actual act of typing or respeaking-- is the least. It contributes the least to stress or mental workload. Again, that probably is not so surprising to you, but we actually got a chance to measure it. And I think it probably follows what most people who are live captioners think.

[00:13:38.42] OK, so what does this all mean then? All right. So first of all, after you multiply and divide and add and do what's necessary, you come up with a score out of 100. And this is the subjective mental workload score.

[00:13:54.50] And for our study, it was-- 67.7 out of 100 was the mean with a standard deviation of about 23. The range was from about 31 out of 100 to very high 99. And as we just looked at, the performance factor is the thing that contributes most to this score, and physical demand-- typing-- the least.

[00:14:23.66] So how does this compare? The problem with subjective mental workload is trying to compare with other jobs. It's difficult because there's many factors that influence subjective mental workload and how you would rate yourself.

[00:14:39.08] But if we look at the other literature in subjective mental workload, the workload of a captioner is similar to workload ratings in the medical domain. So people like surgeons have this kind of levels of workload. So that's considered pretty high. This is a high score.

[00:15:04.68] So live captioners, we can conclude they experience pretty high mental workload in their task of live captioning. Anybody surprised, of our captioners who are on here? Does that make sense to you?

[00:15:34.65] Yeah, right. So I think that having these data helps in arguing about your job and some of the things that need to be done in order to reduce that. Because high mental workload causes performance issues. And I think everybody probably knows that. And you're going, oh, that's why.

[00:16:08.22] So, Mike, that was a good question. Hang on one sec and we'll answer that. So high subjective mental workload or high mental workload affects your stress levels, burnout rates, and performance. So as you fatigue and as that mental workload increases over time, you will perform worse. And that causes people stress.

[00:16:42.00] All right. So Mike Wald asked whether the captioners were using typing rather than ASR. So of the group, the majority were stenographers, and there were a couple people who were respeakers. So in this study, nobody used automatic speech recognition. This is for the

conventional captioning that is done still today by mostly stenographers, or people who are using the stenography keyboard.

[00:17:14.79] But you're asking a good question. Perhaps there's a place for automatic speech recognition to reduce some of this high workload. And I'm going to show you a study in just a second-- not quite yet. But there are some other solutions and other technologies that maybe could have an impact on reducing this workload.

[00:17:41.24] KELLY MAHONEY: Deborah, I'm sorry to interrupt you, but while you were answering an attendee question, we got another one here asking for a little bit of clarification. You mentioned that there's a delay between the action in televised sports and the captions populating that's about five seconds. Are you able to provide a citation for that-- either now, or maybe we could add it to the slides later for reference. We just got a Q&A in the chat, so wanted to shout that out.

[00:18:09.03] DEBORAH FELS: Yep, absolutely. So I've been trying to find this actually, as much as I can, and I haven't seen enough data to actually give me a hard number on that. But there's a fair number of people that are citing this as a typical delay. I'm going to be measuring it in another study that's just underway. So I don't have those numbers yet. So.

[00:18:46.76] Sometimes the broadcasters will delay a transmission to either-- it's often to allow for captioning, but usually it's for censorship-- that delay in the broadcast. So I know, Mike, you are-- so that will see-- so usually that delay is not necessarily for live captioning, but for the censorship part in the live stuff. And captioning will-- it's not enough.

[00:19:24.35] So there's a comment from Casey. We often cite a general three- to seven-second delay. Yes. So the regulations in Canada-- the Canadian regulation is six seconds. But it doesn't matter because that's a long delay in a fast-paced sport. like basketball or hockey. Sorry for those of us in Europe-- I didn't mention soccer, but it is not fast enough compared to basketball or ice hockey. And basketball is actually the fastest--

[00:20:02.61] [CHUCKLES]

[00:20:04.08] Yes, sorry-- the fastest that we worked with, followed by ice hockey. And for the golfers, yes-- tennis players. Anyway, so there's a lot of play that can happen within a three-second, seven-second, six-second-- it doesn't really matter. So that makes it difficult.

[00:20:27.85] And a number of people that we worked with said they just turn off the captions, mostly because of the delay and the mistakes. So they don't even watch it. Well, as a broadcaster and as a regulator, if people are turning off the accessibility. elements, then the point of having them maybe needs to be reconsidered. I'm going to just leave you with that because we're going to talk about in a bit. All right.

[00:20:59.07] So in 2016, we did a study with viewers-- hard of hearing and deaf viewers-- and we asked them what they wanted. That doesn't mean that's what they need, but it's what they

wanted. And as probably not unexpected, there was a difference between hard of hearing viewers and deaf viewers.

[00:21:31.17] Although the top four things stayed the same, the order of priority was different. So people who are hard of hearing thought that caption speed was the most important, followed by not blocking the images. Having verbatim accuracy was third, and then the delay factor was fourth. For deaf viewers, the delay factor was the highest priority. Blocking images next, similar. Verbatim accuracy next, and caption speed was the fourth priority of about 15 things that they could rate.

[00:22:11.64] So the problem is-- and it's a very well-known problem-- is that there is a speed-accuracy tradeoff. So as you go faster, the accuracy decreases. And that's been in the literature since at least the 1960s-- related to all kinds of different things, not just captioning. But it applies to captioning too.

[00:22:42.94] So we have added delay as another factor that has to be traded off in this equation of, as you go faster, it gets less accurate. But when it's delayed then, especially in the fast-paced sports, then it compromises the accuracy. And if you make it go faster, then the accuracy is going to be compromised even more.

[00:23:10.86] So it's a complicated mix of things that have to be considered for captioning, and it has caused a lot of problems. And these are longstanding problems that have been around for a long time. And it's further complicated by the politics of captioning.

[00:23:36.07] So Ben mentioned something about where the captioning should be. But that's true in general. And for post-production captioning, generally it's below. But when it's in sports, the bottom section of the screen will-- they often put things like the score of other games or the score of that game. And it'll shift between the top part and the bottom part. So often you'll see the captioning shift so it doesn't block.

[00:24:18.14] If you want to put the captioning separate from the TV screen, that's maybe another option. And we've tried that in the past, especially with the smart TVs. You can change the screen. So the problem with the live environment is, they put text on the screen in lots of different places-- player stats, et cetera.

[00:24:48.10] So this is what people said they needed. So what we did-- and this is one of the technology solutions-- is we looked at, well, all right. If people are turning off the captions and the delay is causing problems with keeping up with the game or making it a confusing situation, or having to do all these workarounds to try and get the captions and the game play more synchronous. So we looked at a study where we said, all right, we're just going to take away some of the captions.

[00:25:27.62] And so what does that mean? In the traditional captioning or conventional caption, everything that's said is captioned-- or should be captioned. All right. So in live sports-- and we focused on live sports. We didn't work on news or weather or talk shows. But I'd like to.

[00:25:55.06] So conventionally, there's two different kinds of speaking that happen in a live sporting event. There's the play-by-play announcer that says what's going on in the game, what each player is doing, where the ball is or the puck or whatever is, and who has it and what they're doing. So that's the play-by-play announcing.

[00:26:17.99] And then there's the commentary, which is in between, when nothing's going on or there's a replay or there's a break of some sort. And then the color commentators-- which is their name-- they talk about other things, like who's on the injured list or what other teams are playing at the same time, or give their opinion about the coaching or whatever. So that's the commentary. And the color commentary is something that you don't see. There's no evidence of that in the visuals.

[00:27:03.85] So we split those two things. And we had a study which compared the conventional captioning, which captions everything that's said, and we had a version of the same sport clips which had only captioning for the color commentary-- the non-gameplay announcing. So there was a lot of captioning that was missing-- a lot of captions missing. And we wanted to look at what people were looking at and for how long. We wanted to know what their opinion of those two different styles of captioning-- and we wanted to try and understand what they took away from those clips, or their comprehension.

[00:27:54.50] So I'm not going to talk too much about the comprehension, because that's a whole separate thing. We've come up with a different method for measuring comprehension. In most of the other kinds of studies which are trying to look at captioning and the benefits and drawbacks of captioning, often comprehension is measured with some kind of test, like a multiple choice test or a focus group.

[00:28:20.30] Or people are being asked to talk about facts. And that's not how we watch TV. That's not the kinds of things that we remember generally. We're not asked to do that after we watch TV generally.

[00:28:39.17] But we are usually watching television-- especially live sports is a fairly social thing, where you're watching it with other people, either away from your house or in your house. And so what you're doing while you're watching the game is, you're talking to those other people in a social conversation. So we developed a methodology to simulate that conversation style in order to try and understand people's comprehension rather than doing tests. And if you're interested in that, I'm happy to talk to you about that after I've also given you some extra references regarding that particular methodology.

[00:29:27.45] OK, so our play-by-play version was then verbatim, as much as live captioning is verbatim. It was very fast and it had lots of errors, typical of a live captioned hockey game. And even with the very best captioners who know a lot about hockey, which is Canada-- hockey is important in Canada.

[00:29:53.64] And so our best captioners are assigned to caption hockey. And there's still lots of errors, lots of dropped sentences. And I'm not saying that that's bad captioning, it's just, they're doing the best that they can, and it's still going so fast that they just can't keep up, more or less.

[00:30:19.13] And then we had a version that was commentary only, which had fewer errors because the captioner didn't have to go so fast. It was much slower because often there was just no captions. And because there were no captions, there was no reading task except for the text that was on the screen normally anyway, which is like-- the logo, the advertising along the boards, and the score of the game and that text-based information that's often appearing on the screen that competes with captions.

[00:31:03.68] All right. So we used the eye tracking system, and there are advantages and disadvantages to eye tracking. But it was really interesting to see what happened. So I'm just going to give you a little bit of the data, show you a little bit of what came up. And I'm just going to show you two. One is for the deaf viewers for hockey-- got to have hockey. And I'll show you the hard of hearing ones.

[00:31:38.91] This is a heat map which essentially shows you where people's eyes were-- where they were looking at. So there's a question about PbP and CO-- Play by Play and Commentary Only are those acronyms. Play by Play is the conventional style of captioning where the play-by-play announcing is captioned as well as the commentary, and the Commentary Only is no play-by-play captioning, only the color commentary.

[00:32:17.53] So the color's intensity-- yes. Thank you, Casey. So the colors are the intensity or how often and how much time people are spending in any particular area on the screen. So the red color is where there's more intensity. So people are spending more time there and they're there for longer time.

[00:32:51.13] So the gameplay happens in the center of the screen. And the cameras around the arena or around the field are oriented and they change depending on where the players are so that they keep the play in the center of the screen. And then the captioning usually appears at the bottom-- sometimes at the top, sometimes at the bottom-- most of the time at the bottom.

[00:33:23.29] So in the Play by Play version, the people are watching the screen. Their eyes are in the center of the screen, just not as much as in reading the caption. In the Commentary Only, there's more intensity, more often looking at the center of the screen compared to the Play by Play version, which means they're watching the hockey game instead of reading the captions.

[00:33:52.27] But they're also reading the captions. So their eyes are now more equally distributed between the gameplay and the captions. That's for the deaf viewers.

[00:34:04.83] For the hard of hearing viewers in our study, it's a little bit different. In the Play by Play and the Commentary Only, it's about the same-- the distribution of how much time and frequency and how long people are watching the center of the screen, which is the game, and how much time and intensity or how long they're staying on the captions. And it's about the same between the Play by Play version and the caption only version for hard of hearing viewers.

[00:34:40.41] So there's an interesting difference. And this is for hockey. Basketball, it's a little bit different, but not much. And just for our subjective comments-- no. Cat, CO is color commentary only-- not caption only. Commentary Only.

[00:35:09.07] So Mike asked a question about people who have-- thank you. Final answer-- asked a question about people who are hearing about the commentators. So the play-by-play commentators being fast-- probably, but I don't really care. I don't care about hearing people. Sorry, I do care about them, but I'm working on captions. I'm not so worried about it.

[00:35:46.60] So Mike, you're asking about how commentators should be trained. That's a good point. They should talk less, maybe not as fast. That's a good question.

[00:36:00.65] And I can't really comment on that because I'm not that familiar with how commentators are trained. But I was trying to look at what literacy level is that they speak at, and it's about a grade 11 level. And I did find it takes about 20 years for play-by-play and color commentaries to actually get trained. So it's a very long training period for people that are doing professional sports. I can't really say much about college level or university level sports or the other levels. Olympics are quite a beast by themselves.

[00:36:55.43] OK, so when we asked them what they liked better, 65% of the people that we talked to of those 27 people said they didn't have a preference. But of those that had a preference, 33% were deaf, 18% were hard of hearing. And the majority said they preferred the Commentary Only style of captioning.

[00:37:28.16] OK, so in terms of the kinds of comments that they made in the interviews and the comprehension, was-- as I think we've seen in our comments, it's stressful. It's too much. There's too much work for the eyes. But on the positive side, that they knew everything that was going on, particularly the players' names.

[00:37:57.10] Players' names and numbers are important to know. And in the play-by-play style or the more conventional captioning, that's there. In the Commentary Only, that was missing. So that was a concern. But the overwhelming comment was that they could actually watch the game and not worry about missing something.

[00:38:30.76] Because this was new for everybody-- they'd never seen captioning this way-- they actually thought there was more delay. But once they realized that there was no play-by-play captioning, they went, oh, it's not delayed. So once that was realized, I think they understood.

[00:38:51.15] So for the comprehension, which was measured by simulating that informal social conversation that people would have while watching the game, or even a little bit after where they'd do a debrief or what they liked and who did what. And they didn't agree with the coaching or they didn't agree with the refereeing or they agreed with it, or those kind of very social things that happen when people are watching a game together. So for comprehension in general, the people who watched the play-by-play had lower scores. So they remembered less or they could talk about less.

[00:39:37.58] And the people who watched the Commentary Only had higher scores. And the kinds of topics that we addressed in this conversation came from the gameplay itself, as well as the color commentary comments. So people who were watching-- everybody got to watch the Commentary Only. Everybody got to watch both. It was just in a different order.

[00:40:06.31] So from the Commentary Only version, people were able to remember more and be able to talk about more. Not remember because we weren't asking, remembering kind of questions or recall kind of questions. We were asking-- talk about their opinion of the coaching or refereeing or whatever. And so they were able to integrate more about what was going on in the game than those who were watching the Play by Play, which had that play-by-play described. So that was interesting.

[00:40:49.40] Well, all right. So what does this mean? From this research-- there is some evidence that verbatim captions maybe are not necessary for live fast-paced sports. Right? So we have evidence from people's opinions as well as some eye tracking evidence. I don't mean a lot of evidence, but there's some evidence maybe we should pursue this a little bit more.

[00:41:21.80] We also know-- and I don't think this is surprising for anybody in this group-- is that hard of hearing needs and preferences are different from deaf viewer preferences and needs. So this is also interesting because, how do you deal with that? As a broadcaster, you're not going to have two different kinds of captioning. As a captioner, how are you going to make two different kinds of captioning at the same time? That doesn't make sense.

[00:41:58.92] So that's an interesting question. And Mike, I hope-- or-- yeah. I hope I'm getting to the ASR part, or the Automatic Speech Recognition. But the last thing is that it's possible to measure comprehension using a conversation method rather than a multiple choice test or a focus group kind of thing. All right.

[00:42:23.67] So the question is, then, can we look at some of these issues with automatic speech recognition and AI? So the question is, can speech-to-text and AI help? And I think for some of you, a collective--

[00:42:52.26] [CHUCKLING]

[00:42:53.97] Yes, right. OK. So naturally-- yeah. OK. Our next piece of this project is that we built a captioning tool that combined AI captioning or automatic speech recognition with some AI or artificial intelligence, with the captioner's duty as the supervisor of the AI. Our belief is, the captioner has to stay in the equation.

[00:43:32.82] They are the quality control. And people are good at that. AI, not so good at that. And we need the captioner to be in the quality control domain.

[00:43:48.20] Maybe we're not so good as typists. And we just saw lots of evidence that we're not so good at that as humans. And it makes us upset or it causes us stress when we don't perform as well as we want to and as well as the regulations say we should.

[00:44:05.96] So we built a tool called PAVOCAT, and I'll show you the interface in just a sec. And we used a model of supervisory control, which has been around since about the 1950s, and began with the automatic pilots for airplanes. So we built this tool-- and I'll show you in a sec. And we used it with 10 novice captioners and 11 experienced captioners-- novices with less than a year experience.

[00:44:45.52] Now with three segments of less known content-- and we used Olympic snowboarding. Snowboarding is fast. There's a lot of technical terminology and some very colorful announcing with the snowboarding genre.

[00:45:10.48] So we wanted everybody to be on a more even playing field with respect to their knowledge level of the sport. When we worked with the hockey and basketball group, lots of people knew lots about hockey and lots about basketball. And so they were bringing that knowledge to bear on how they looked at the captions and what they understood about those particular types of games.

[00:45:36.94] So this one, we took away that knowledge. Most people don't know much about snowboarding, and that includes the captioners. And the captioning for the Olympic snowboarding was atrocious, I have to say. And I'm sorry. It was CTV that did the captioning in Canada, and I didn't watch the American captioning and I didn't watch any of the British. But it's a very unfamiliar genre sport, so it's hard.

[00:46:09.54] So we measured a bunch of different things, including subjective mental workload, which we measured before. But we measured some other things like trust and satisfaction. Ah, there's more messages. Sorry. OK. I won't worry about that, Dennis. Can't even read it.

[00:46:33.51] So the PAVOCAT interface then looks like the AI or the speech-to-text is doing the majority of the work, converting the speech to text. And it's actually pretty good at that, minus the speaker identification. But we could talk about that as a separate thing.

[00:46:56.76] It's also fairly reasonable at finding errors of its own. So the tool then offered the errors to the captioners. I think this is an error. That's what the system said. And it gave the error's listing and suggestions on how to modify it.

[00:47:23.83] So if the captioner agreed-- in our test, it was a mouse. But we can imagine this being on a tablet where they select the correction. So if they agree that it should be corrected, they can select that. If they don't agree, they can just ignore that.

[00:47:43.03] The other thing that this tool gives you is the actual display-- the video display. When we looked at captioner workspaces, we often had multiple screens, TVs, all over their workspace. And their attention had to be divided amongst all these screens.

[00:48:04.18] And they weren't given the visuals. They were only given the audio to do the captioning with. And so they had to use their own TV to watch the game and to see the captions. So they were seeing the delayed captions and their own captions.

[00:48:21.63] So it's very confusing and another stressor that's making this a very difficult task-- tiring and difficult. So we combined things so that they could get a visual from the broadcaster themselves rather than watching it on their own TV. All right, everybody understand that?

[00:48:49.13] On one screen were all the speech-to-text-- and the one part of the screen, the central part where people's eyes were mostly looking. On this side, there was the suggestions and

the error. Possible errors were marked on the main caption screen. And then at the bottom area was where they got the feed from the-- visuals from the broadcast all in one place. OK.

[00:49:18.59] So I'm an academic. I got to throw a little bit of theory at you. Just a little bit. This is a model of the human as supervisor. And what I wanted to point out here is that there's a number of things that affect your ability to be a good supervisor. And that's workload, engagement, and complexity.

[00:49:45.01] Plus, the other kinds of things that affect your satisfaction with being a supervisor is your trust level in the automation-- how much you trust it and how much attention you can give to it. What other things do you have to look at or listen to? So this is a model of the different pieces that combine together to be a supervisor of an automated system.

[00:50:14.32] So whether that's an automatic pilot or robots in manufacturing or as an AI supervisor as a captioner-- so that's the model that we use to try and look at how people moved from being typist to supervisor with PAVOCAT. And as prototype tools go, they're not perfect. Again, we had to do this virtually. And that technology always interferes and gives us usability problems and interrupts or interferes with the smooth functioning of things.

[00:50:57.99] But let's just take a look at some of the things that we found. So first is the mental workload, which you've already seen. And so if you remember, in this study, we had three different sessions over three days. So the captioners were asked to caption three 10-minute clips over the three-day period. So the first-- with PAVOCAT together-- as friends, we hoped.

[00:51:29.40] So the first session, the mental workload was 61. If you remember the first study that we did with the captioners who were stenographers as live captioners, they were in the 67 neighborhood. So it's in the neighborhood. Session two was lower, and session three was lower again.

[00:51:52.53] So the mental workload is going down. Less pressure, less stress. All right. It doesn't mean it's going to be a lot less. And in other kinds of application areas like automatic pilot or robot supervisor, the mental workload actually doesn't go down that much. But people are happier with their jobs. So satisfaction goes up.

[00:52:26.61] All right. So in interviews, we asked them what they thought about this. Novices, people who are new to captioning, were more willing and trustworthy of the AI system than the experts. Again, not really a surprising result. We think that the workload increase or the levels of workload were affected by their attitude towards AI. And that can come from different kinds of places.

[00:53:00.02] We also didn't let people take over from the AI. That was intentional. The next study where they will be able to take over-- and I want to look at how often that happens. But that negative attitude and trust of AI-- and in the captioning world at this point in time, this common sentiment is captioners are going to be replaced by AI. And so it's an area of considerable concern.

[00:53:34.73] And people were coming in with that kind of attitude. The people who were more experienced were coming in with that, I don't trust AI. I don't like it. It's going to take over my job. OK.

[00:53:52.77] So what did we find about this? Well, first of all, nothing's perfect. AI systems are not perfect. They don't do speaker identification. In sports, maybe not as important as in other things like talk shows.

[00:54:08.70] And captioners are not perfect either. So we have an imperfect system that we're trying to make better. Captioners' jobs are changing, and that's going to be the reality. And as a stenographer or as a respeaker, that could be a concern. But I would propose that captioners' jobs are not going to be eliminated. They're going to be changing.

[00:54:33.14] And they're going to become supervisors of captioner systems, which, from a human point of view-- humans are better at being supervisors. They're better at finding mistakes and making decisions. AIs are not so good at that. So they need help and the captioners are going to provide that help.

[00:54:54.26] I think job satisfaction, workload, stress levels, will go down as a supervisor. So I suggest that AI and captioners work together, and that we humans do what they do best and AIs do what they do best. And maybe some of these very longstanding issues can get more resolved. I'm not going to say they're going to get solved, but we'll work towards that.

[00:55:22.85] The other thing that happens as we have AI and captioners as supervisor is, now we can look at customization options. Because an AI can do different things. You just have to ask it. I know, we're going to stop in a sec. And then this is my last slide.

[00:55:45.62] So there's some possibilities there with respect to customization, not just with deaf or hard of hearing needs, but also maybe low literacy levels or literacy-- changing it to be a tool that second language learners can use and improve on and change. So they need lower literacy at the beginning and higher literacy later on.

[00:56:11.57] And finally, the big thing that I'm hoping will cause some reaction is that the politics have to change, and that the equity versus equality question needs to be reconsidered. It should not be equal to what hearing people get. It needs to be equitable.

[00:56:37.05] Because what hearing people get, it's too fast. And hearing people have the same problems, but we don't worry about them. It's too fast. There's too many mistakes and it's too delayed to be useful. So we need to rethink that politic. And I'm just going to put that out there.

[00:56:57.06] OK, there's a number of acknowledgments. I will just leave that up. Thank you, 3Play Media. And additional information-- we've published a lot of this. So you can find lots of details. The final report for all of the research that I just talked about is available in English and French. And I believe that these slides are going to be made available so that people can go and look these things up if you'd like.

[00:57:25.11] KELLY MAHONEY: Yes.

[00:57:25.37] DEBORAH FELS: OK, done. Thank you.

[00:57:27.44] KELLY MAHONEY: Perfect. That's exactly what I was just going to say. Unfortunately, we're out of time to ask or answer your questions live. But as Deborah said, we will make these slides available. So feel free to reach out to us or reach out to her with any continuing questions we have.

[00:57:41.89] Thank you so much, Deborah, for joining us and giving us such an insightful presentation. It was wonderful. And thank you to our audience for being so engaged. I'm so glad to see that you enjoyed yourselves as well. Thanks again.

[00:57:55.66] DEBORAH FELS: Thank you.