Captions and Beyond: Building XR software for all users – Evan Tice Accessibility VR Meetup – June 17 2021

THOMAS LOGAN: Thank you so much. So I am very thrilled to see so many of you all here. My name is Thomas Logan. I'm the organizer of a monthly Meetup, Accessibility Virtual Reality, and we typically have been meeting in Mozilla Hubs. This is our first time for our event, hosting inside of the AltSpace world. I'm a huge fan of AltSpace. This is the social VR platform I actually use the most. And we are very excited, obviously, about the captioning feature that is going to be discussed tonight. It's something very exciting. And best in class for things -- for other VR properties to learn from.

We want to thank, tonight, our sponsors from Equal Entry. And we also want to thank Joly MacFie from the Internet Society of New York, who does our livestreams on YouTube, and makes sure that we have recordings of all of our great presentations. And I want to thank Mirabai Knight from White Coat Captioning, for the captioning for today's presentation. And one other thing I would like to do is just take a moment to have Lorelle come down from Educators in VR and introduce herself. My very first experience meeting and presenting in AltSpace VR was at Lorelle's conference, Educators in VR. So I'm really thrilled that she's here tonight with us, and I want to let her have a chance to tell you about what her event and her organization does.

LORELLE VANFOSSEN: I appreciate that. Did not expect it. And it's lovely to see all our Educators in VR members out there. Come on! Give me some hearts! You know I love that! Educators in VR was founded here in AltSpace. Almost three years ago. And has now grown to be almost 4,000 members, who are educators, researchers, students, trainers, businesses, et cetera, that are determined -- we're evangelizing and determined to integrate XR technology into education at every level. And we have training programs that we offer regularly, and very, very active Discord, we...

(no audio)

LORELLE VANFOSSEN: Okay. Sorry about that. I have terrible internet. So the fact that I'm here is always a miracle. So we just finished a great discussion on Discord's Stage, which is the equivalent of the new Clubhouse live chat, which needs some serious accessibility features on both Discord and Clubhouse. On -- what can you not teach in VR. Which was a fascinating discussion. And we put it on YouTube, which luckily has automagical captions, so you can follow through on that. But we're just really glad to support everything that y'all are doing. And we have been working with educators... Educators in VR has been working with AltSpace and other platforms, from day one. To encourage accessibility. And I am so excited about this topic. Because captions are great. Then we need to take it up. Because what's next is making sure we have the visually impaired getting here. That's my dream. But thanks for the opportunity. Thank you!

THOMAS LOGAN: Yeah, thank you so much. And do you want to let everyone know that is joining us on the YouTube stream -- we do monitor the YouTube stream for your comments and questions for our presenter tonight. So please feel free. There's a slight time delay, obviously, on YouTube, but we will be monitoring those and bringing them into the world. And without further ado, I would like to hand it over to our presenter tonight. Evan Tice, from Microsoft, and the AltSpace Team, and this great captioning feature. So take it away, Evan.

EVAN TICE: Hi, folks. I'm Evan Tice. Thrilled to be here today. And talking to you about captions in AltSpace. And beyond. Just to give you a brief overview, I'm going to give you an introduction of myself. And my background. And how I came to work on captions. I'm gonna introduce myself two ways. We'll get to that.

We're gonna talk about captions in AltSpace. And I'm gonna do this both from a feature overview, talking about what we built, and then I'm gonna teach you all the code and show you how easy it is to make caption software for yourself. I'm only half joking. We won't go too deeply into code. At the end, I'm happy to take questions and discuss captions and accessibility in general. One thing I will note: I am not a Microsoft spokesperson. And I'm super thankful for Microsoft's support in giving me this talk. But I won't be announcing anything today. So... Just a heads up on that.

And with that, I will introduce myself. So... I learned to program on a Radio Shack TRS 80, which was very old even for the time. I went to high school in Montana and studied computer science in Dartmouth, graduating in 2009. I finished college in the middle of the recession and was very, very happy to get a job at Microsoft. I started my career fighting hackers, working on Windows. I worked a bit on bolstering security of the Windows Heap. If anyone knows what that is, send me some emojis or hearts right now. Seeing none... Yes. I'm a super nerd. The Windows Heap is the thing that gives applications memory and it's a frequent target of hackers. I also worked a bit on a dynamic defect detection tool, which is a nerdy way of saying I worked on a program that finds security and reliability bugs in other programs.

I'll pause and mention here that mixed reality was not an obvious career move for me. But I eventually decided that I wanted to build things. And not break things. And breaking things is a lot of what security engineering is all about. And I'll save my story about joining the mixed reality team for the end of this talk. It's a good way to conclude. But I made the leap to a secret project in 2014, which turned out to be HoloLens. I worked on persistence for HoloLens. So if you put an object on the wall, say, Netflix, on the wall, or... If you put a model, an engineering model, on a table, and you come back the next day, that model or that Netflix screen should be in the same place, in your house or in your office, in spite of any changes in lighting, maybe you've moved a small amount of furniture around -- that was my contribution to HoloLens v1. And right after that... Oh, goodness. I forgot to

advance the slides. Right after that, I worked on articulated hands for HoloLens v2. And this is where my journey into accessibility in XR really began. So can I get a show of emojis from anyone who's used a HoloLens V1? A handful of you. That's great.

The input and interaction model for HoloLens v1 was state of the art at the time. It was called gaze-gesture-voice. It basically meant you pointed your head at the thing you wanted to click at and you made this awkward clicking gesture on the screen. And in HoloLens V2, this was my feature -- we added articulated hand support. As the technology evolved over time, we could do individual joint tracking, and you could do much more natural interactions. You could reach out and grab an object and manipulate it with your hands. In this slide, the user is actually interacting from afar. You can do that as well. But it was a much more natural progression.

After that, and actually I should mention that all throughout this time, I was on loan to the AltSpace team, occasionally. And periodically. I'm responsible for most of the Unity upgrades that the AltSpace team has done. Apologies, content creators. I know you hate those. And on one of those times when I was on loan to the AltSpace team, doing the Unity 2019 port, I did captions as a bit of a side project. We'll talk more about captions in just a second. I joined the team full-time in November of 2020.

And that's me and my journey to AltSpace. In a nutshell. But I want to introduce myself another way. And talk to you a bit about how I experience XR. I'm admittedly not particularly neurodiverse. I wear glasses. And they drive me freaking crazy. I'm wearing contacts now. My boss actually messaged me on Teams. He's like... I've never seen you without glasses! And I'm like... Well, I'm giving this talk and I want to be able to wave my hands around, so it would be nice if I wore one of these fancy headsets I have through work. But I don't want to scratch my glasses. And aside from my vision issues, I probably experience XR in a way that isn't particularly... Remarkable.

Why am I telling you all this? And why am I giving this talk about accessibility in XR? The obvious answer is I'm obviously part of the team that built the captions implementation in AltSpace VR. I'm super proud of that. And we're gonna get into the nitty-gritty of how this

feature works. I'm sure I'll hear your feedback on how it could be better. But more than that, I'm gonna soapbox for a second.

And joke's on you, Thomas and company. Thank you so much for inviting me. But it's very likely I will learn more for this talk -- from this talk -- from the discussion that follows this talk -- than you learn from me. I'm really thrilled to be here. I tried... I titled this talk "captions and beyond", because we are gonna talk about captions. We're gonna talk about some cool cognitive services that make apps more immersive and accessible. And I have a handful of small ideas on that topic. On the "beyond" topic in particular.

But I'm really looking forward to the conversation at the end of this talk. This audience is likely more experienced and well versed in accessibility, and is likely to help me learn things. And I think that will make me a better engineer. And I think that will make the products I work on more accessible. So thanks again, Meryl and Thomas, for having me.

Let's see. One other thing, before we dig into captions. I went to the Chi Conference in Montreal, in 2018. And I think it's the thing that lit the interest in accessibility -- set it off for me. I'm just gonna play this video. It's a short video. But it was personally inspiring for me. And I will say, before I saw this talk, and some of the others that were presented at Chi this year, I hadn't really thought about pushing the boundaries of accessibility in XR. And I want to credit the authors of this paper and this research for their inspiration.

I think they set the bar high. I think we have a long way to go. And we'll talk more about that throughout this talk. If I can play this video.

PC: We created a novel VR experience by using haptic and audio feedback to enable people with visual impairments to experience the virtual world. Our novel haptic controller simulates the interaction of a white cane to help people who are blind understand and navigate a virtual space using their existing orientation and mobility skills.

PC: I found the domes and the traffic light. I found the pole with the traffic light button.

EVAN TICE: All right. That's it! So... Chi 2018. I said a few slides ago, and I'll say again: I am not the accessibility expert. I happen to build this really cool feature. And I'm happy to talk about it, and I'm super psyched to learn from you and see where we go beyond that. So let's talk briefly about AltSpace. Particularly for those of you not in AltSpace, out on the livestream. You're missing out. Come join us, if you would like. AltSpace is a social VR application, primarily designed for events. Events can mean everything from a virtual Burning Man to an LGBTQ Meetup. Educators in VR. Or a Holoportation concert. Aside, Halley Greg is giving a concert in AltSpace tomorrow, at 7:00 Pacific. Attend if you can. Holoportation is really cool. And if you can't attend, we also had it at our Ignite conference in March, holograms of James Cameron along with a giant squid. You can find videos of that on the internet. AltSpace is a great piece of technology to bring people together, even when we're social distancing. Thanks, COVID. We shipped captions preview. We shipped it during the pandemic. And I'm really proud of this. Because we can bring folks together across language barriers. We can allow folks who have difficulty speaking and hearing to participate in AltSpace. And we did so at a time that... These interactions were sorely lacking in our personal lives, for most of us. Because of pandemic and lockdown.

Yeah. So I'm proud of this feature that we built. Let's talk a bit about it. So we have live captions, powered by Microsoft Azure. My speech is being sent to Azure and being sent back to all of you as text. And we'll talk about how specifically that works. Captions in AltSpace can be enabled in select events. You'll notice, if you ask a question later on in this event, or if you were talking prior to the event, that in order for you to speak in this space, either via audio or text input, you must accept the captions consent prompt. Because this space is caption-enabled. And you can stay muted if you want. Captions are lazily initialized. That means if none of you had shown up for this event today, and I was in here, and talking alone to myself, the captions feature wouldn't turn on. It's waiting around for an audience member who wants to see captions. Before we actually light up the captions feature. That's just a basic cost mitigation.

Captions are currently translated into 8 languages. English, German, Spanish, French, Italian, Japanese, Korean, and Portuguese. We present captions in two different ways. There's the presenter captions, that you're looking at right now. Those can be viewed from

afar. They have my name on them. They're our initial take at a more traditional speaker-style or presentation-style closed captioning.

And then there's also the speech bubble, the social captions feature. That you'll have when we finish up this event. And everyone is mingling around, and talking amongst themselves. There's a speech bubble that's attached to the player's head. And you can read it. For some, it's easier to read than others. Admittedly so. But that's our social caption viewing mode. And there's rudimentary text input on PC. We'll talk more about this in a second. I have a slide or two on it. And there's also the option to speak and view captions in different languages. That's the feature as a whole. And I'm realizing as I'm looking at my slide now that I actually have in the picture... I have a picture of the social captions. Those are those speech bubbles I was talking about.

All right. Rudimentary text input. I have to press enter to enable this. And I get a little box that shows up. Into which I've typed "hello captions". And when I hit enter or click send... On the audience member side, we have a perspective of a Spanish caption viewer in the audience. And where, as I typed "hello captions", they see "hola subtitulos". This feature is admittedly limited. It's only available on 2D PC right now. And due to a weird quirk I won't delve into too much, it still requires you to unmute your microphone in order to type.

The other captions feature I want to talk about is our settings. I mentioned earlier that you can view and speak in different languages. That's not the default. You have to flip that little slider. That says "view captions in another language". And when that appears, you get a second language selection. In this case, set to German. You can also adjust the size of the caption text box. All right. Here's an architecture diagram. And Lorelle told me my slides are impossible to see. So I'm going to describe this for you.

Client one sends audio data to the Microsoft Speech Service in Azure. And the Speech Service sends back captioned text. We're gonna deep dive into the Speech SDK in just a moment. But it's the area of this diagram in green. After the client gets the caption text back, it sends it to our backend. Where it is sent out to other players. That's the current implementation. And let's deep dive into that green caption or Speech SDK part.

So the cognitive services platform is a comprehensive set of Microsoft technologies and services that's aimed to accelerate incorporation of speech into applications. As well as to amplify, like, the impact of those applications. This is a set of technologies that wasn't necessarily developed with virtual reality or augmented reality, XR, in mind. But I think it has a lot more to offer us than we're using it for today. Again, the title of my talk: Captions and Beyond. So this software stack was -- is typically used for scenarios like call centers.

Call centers really like transcription and translation. Also, more recently, voice assistants. No one wants to touch an elevator during COVID. Wouldn't it be great to say: Take me to the fifth floor? We use two of the core capabilities of the Azure Speech SDK. We use speech-to-text and speech translation. And we'll dig into each of those in a moment. Yes, I promise, I will teach you to code. It won't be too painful. The platform enables us, as applications developers, to do a lot more.

I am really impressed, as I was researching my talk and the capabilities of the Azure Speech Service, particularly in the areas of custom keyword creation, you can create a keyword like: Hey Cortana, that lights up the speech service and starts listening. And they also support things like custom commands. That's like... Run the vacuum or open the menu. That sort of stuff. I'm super impressed with this speech technology. And I'd like to demo you some of the features we're not using in AltSpace.

PC: Oh, well, that's quite a change from California to Utah!

EVAN TICE: No, don't play! Don't play yet!

PC: Heavy snow.

EVAN TICE: I couldn't get the other video to start when I hit the slide, but this one did. This is the voice gallery within Speech Studio. You can play around with this at speech.Microsoft.com without writing any code. One of the things that particularly struck me is how the affect of voices can be adjusted. These things sound more human-like over time. I'm gonna play two clips for you. One conversational, and one is the voice of a newscaster. So this is... Text-to-speech. In two different styles. First, conversational, and then a newscaster.

PC: Oh, well, that's quite a change from California to Utah. Heavy snow and strong winds hammered parts of the central US on Thursday.

EVAN TICE: And those are just the built-in ones. If you're so inclined, you can build your own custom or neural voices, and make them even more expressive and emotive. So lots of untapped potential in the Speech SDK. All right. Who's ready to learn to code?

(popping noises)

Actually, one or two more slides. There is... Several processes at play in speech translation. One is capturing the conversion of the speech into text. And the other is the translation of that text into the desired language. In the caption and conversion stage -- so this is the part where the application captures audio, sends it to Azure, we run automatic speech recognition in Azure, it performs an initial conversion, that initial conversion is just words without context, and that initial conversion, that automatic speech recognition, is refined using TrueText. TrueText uses text matching patterns. Microsoft research has a few papers on TrueText. But suffice it to say that the transcript is refined over time. This TrueText stage is where we might correct or disambiguate between two words. Think: H-E-R-E, here, versus H-E-A-R, hear. Can you hear me. The TrueText is where we apply our learnings about language, and further refine that recognition.

In the translation stage, so now we've got text, it's been refined, the text is routed to another machine learning model, that's been trained with up to 60 languages, and we get

partial and final translations. We'll talk more about these partial translations in a moment. If an application is using it, the text-to-speech system can also convert the translated text back into speech audio, like we demoed a few slides ago. And obviously, AltSpace isn't doing this today.

All right. Now comes the code. So I'm gonna give an example of transcription without translation. We'll add translation in a few slides. And my goal in... The code is less important here. The annotations I put on the code are more important. My goal here is to convince you all that adding captions is really relatively straightforward. So the first thing you do is you take in some sort of speech configuration. This includes the input language setting. The language that the speaker is speaking. In my case, United States English. As well as a credential from whomever is paying the bill.

And then two lines of code initializes a speech recognizer from an audio source. An audio source could be a microphone, a sound file, some other source. And then we wait around for a single phrase to be recognized. And in this demo, we print it out to the console. We'll talk about continuously recognizing a phrase and multiple phrases in just a moment. But first, let's define a phrase. I gave that example -- can you hear me. I carry it over here as well. You can think of a phrase similar to a sentence, but in reality, it might be multiple sentences. If you're noticing, sometimes my captions overflow, off the screen. That's generally because AltSpace itself doesn't really understand sentences. It understands phrases from the Speech SDK. As the phrase builds up, the TrueText technology refines the estimate that it's getting. Using its understanding of human language.

So in this example here, we get four updates before the phrase is finalized. Can. Can you. Can you here, spelled incorrectly, and can you hear me. And on the AltSpace side, when we're getting this text back, to save bandwidth, and because some of these phrases can get quite lengthy, because we're doing this in multiple languages, we try to only send the difference between the previous phrase and the current phrase. So that's what that right column is about. Okay. More code.

If we want to speak more than one phrase, we can wire up some callbacks. Our previous demo wasn't that useful, because it just could only recognize a single phrase. So as the

user speaks, the recognizing callback fires repeatedly. And when the phrase is done, the recognized callback fires. And there's also callbacks for the session being canceled or stopped that can also fire. Okay. Translation doesn't change much. We create a translation configuration with credentials and settings. Much like we did just for simple English transcription.

And we specify a source language. For nerds in the audience, this is the BCP47. Which is a standard for identifying human languages. You can look it up on Wikipedia. But the important thing to note is that: For the speaker, we include a region. A US English speaker might get better text recognition out of a US English model, versus the UK English model. And that's why we do that for the speaker.

And then for translation languages, we specify 8 in AltSpace. I named them off earlier. And for these, we simply denote the language. The text a reader in Britain is looking at, and the text a reader in the United States is looking at -- that's gonna be the same for both of them. And if you've played around with your settings, and you turned on that view captions in another language, you'll notice that there are many, many, many more options. Differentiating different types of English and Spanish and what-not. For the speaker. And not for the viewer. And that's why that is.

So yeah. We create an audio source and a recognizer, the same way that we did before. And the events that we looked at, those recognizing and recognized events, the canceled events, they fire just like they did for pure English translation. The only difference, for the nerds in the audience -- we get a dictionary back, as opposed to just a string. So we have this mapping between the language code and the actual text in that language.

All right. It's gonna be hard to see. I apologize. I downloaded this sample from the Microsoft Speech SDK samples on GitHub. This is very, very similar to the code I just showed you. And the code that we have in AltSpace. But I want you to pay attention to something, if you can see it, and if not, I will blow it up for you. Pay attention to the phrase changing over time. Here we go.

PC: The captain has turned on the seatbelt sign in preparation for our descent into the Seattle area.

EVAN TICE: I'll play it again. Did you catch the change?

PC: The captain has turned on the seatbelt sign in preparation for our descent into the Seattle area.

EVAN TICE: Okay. So about halfway through... That recognition... It ended a sentence, and it started a new one. The captain has turned on the seatbelt. Period. Sign in... Is the first part of that second... That second sentence. This is all one phrase, by the way. So it's a little weird. Maybe it didn't have confidence that the phrase was actually ended. But, as I kept speaking, and as it got more context, and ran through these models that have been trained, it figured it out. The captain has turned on the seatbelt sign in preparation for our descent into the Seattle area. I'm not looking in VR to see if it captioned it as I said it correctly. But you get the idea. I'll play it one more time.

PC: The captain has turned on the seatbelt sign in preparation for our descent into the Seattle area.

EVAN TICE: Yeah. I would encourage everyone to -- who does have some engineering skills, or interests, or wants to learn: Take what I've taught you today. You're now coding experts. Go download the Cognitive Services Speech SDK, and you too can experience captions. That's just about it.

THOMAS LOGAN: Evan, quick question on that last part. Thomas here. So for the continuous recognition -- is that a different API call? Or is that just the default that it does that kind of...

EVAN TICE: Can you repeat the question once more?

THOMAS LOGAN: For the continuous recognition, or where it does the recognition kind of... After it's processed more of the text, is that on by default? Or are those different API calls for, like...

EVAN TICE: Yeah, it's a different API call. There is the... I forget what it was called. But it was the one I showed originally. That's just like: Capture a single phrase. But it's a related API call, where you can wire up the recognizer and the recognizing. It's a sibling to that API, if you will.

THOMAS LOGAN: Okay. Thanks!

EVAN TICE: Yeah. Okay. I have some ideas about what could make XR more accessible. From readability to support for more languages, to improved text input for those unable to speak. But I want to leave you all with a thought. And I want to spend the portion of our Q and A talking about the future. Before we do that, I want to look back to the past.

When I joined the HoloLens team in 2004, we were just a code name project. It was actually interesting. It felt like... For those of you who have seen the movie The Matrix, it felt very much like the Take the Red Pill or Take the Blue Pill and see how far the rabbit hole goes. That quote from the Matrix? They didn't tell me I would be working on HoloLens. I had an idea the project was gonna be awesome. And around that time, when I was interviewing with this team, and trying to decide: Do I really want to leave security? The hiring manager described what we were doing in a way that really stuck with me. He talked about when the mouse became widely available in the '80s, and even into the '90s, we didn't quite know how to write software for it.

Here's Microsoft Word 6.0. I had a blast installing this today, in a VM. It was kind of a pain. To get it running again. And I couldn't find later versions of Word that had the menus

within menus within menus. But I always remember using software as a kid. We had the mouse. We didn't really know how to build UI for it. Clicking -- misclicking out of a menu was a pain. Misclicking out of a nested menu was an even bigger pain. And we had the mouse, and we knew it was going to be amazing, but it took a while. And... Similarly, for HoloLens, we had this paradigm, this gaze-gesture-voice paradigm. It felt so amazing at the time, but it was nothing compared to articulated hands.

And I think about that often. That it took many years to design intuitive UI. You can say what you want about the ribbon in Microsoft Word, but I find it much more accessible and easy to understand and approachable than searching for something in a menu. And similarly, I think we've improved on gaze-gesture-voice, with what we did with articulated hands in HoloLens.

And now, with accessibility in VR, I think we're really at the Word... At the Microsoft Word 6.0 stage and the gaze-gesture-voice stage today. I'm really looking forward to hearing your ideas for the future and having a discussion about what we build next. But yeah. My big takeaway is: We're at the -- we're the vanguard of something new and exciting here. I want to acknowledge some folks, before I open up for questions. Meryl and Thomas, thank you so much for having me. Joly, for your help with the stream. Lorelle, for jumping in. And helping moderate at the last minute. I want to really thank the team that built the captions prototype for a hackathon that ultimately evolved into what we have in AltSpace today. That's a project for -- or that's probably a talk for another day.

The AltSpace team -- I see a lot of you in the audience right now. Thank you for attending, and thank you so much for helping me troubleshoot random caption and projection issues, the last few days. I'm really excited about what we build next. I love working with you. You inspire me every single day. I want to thank the Azure Speech Team. For your help on building captions, as well as on this talk. And last, but certainly not least, I want to thank each of you, for attending.

And with that, I'm super happy to open it up for a discussion. And I'm gonna put my headset on properly. Because it's starting to hurt. And look at all of your beautiful faces.

THOMAS LOGAN: Great. And hello, everyone. This is Thomas again, from a11yVR. We're gonna take a couple questions from the YouTube stream first. Because we do have people that aren't with us here in the world. We're gonna ask their questions. And we'll take your questions here. And then if you are here on streaming, please continue typing questions. We're excited to get a lot of questions. And thank you so much to Evan tonight. I really appreciate that you were showing concrete code examples. That we could try out. So first comment for you. This is just more of a comment. But Makoto Ueki-san may be in here. He was using the Japanese translation feature tonight.

For your presentation. I'm in the morning in Tokyo. And the translation was working very well. In Japanese. Makoto, are you in the room? If so... Can you give your comment directly? I'm not seeing...

EVAN TICE: Oh, I would love to call on someone who is speaking one of our eight languages. That I don't speak. And have everyone be able to see that. That would be cool.

THOMAS LOGAN: Okay. Well... Makoto, if you are here, give us an emoji, and we'll pull you up to show that off. But I thought that was a great comment. And Deb Meyers, on the YouTube stream. How is the Speech SDK for people who have speech impairments, such as stuttering, thick accents, et cetera?

EVAN TICE: I would encourage you to try it and let me know. My guess is improving over time. I think we as an industry have realized that when we train machine learning algorithms, if we only train them with people who look like us or sound like us, they're only going to work well for people who look and sound like us.

So try it. Let us know. If it's not great, send us the feedback. And it gets better, over time.

THOMAS LOGAN: Cool. And I'm trying to flex some recent knowledge, Evan and Lorelle. On using the host panel here in AltSpace tonight. But I've just turned on the "raise hand" feature in my host tools. And if you would like to ask a question or a comment, if you would do the hand raise, and then I will call on you. So next person I'm calling on is Lorelle.

LORELLE VANFOSSEN: Hello! I have a bunch of questions. But I'll keep them specific and scattered. First of all, you talked about some of the training. One of the problems that we have from the very beginning is that AltSpace comes out as Old Spice and variations thereof. Is working being done in the training process for those things? Better for brand names and... Old Spice? Ha-ha!

EVAN TICE: That's a really hilarious bug. One thing I'll say... So we don't send AltSpace data for the purposes of training right now. We would consider doing that. But we would probably have to allow users to opt in. So I'm not surprised to hear that the name of our product maybe isn't transcribed correctly. I will look into that. Someone should shoot me a bug or a feedback. I think it's altvr.com/support. And I will look into that. Because that's funny.

LORELLE VANFOSSEN: We've sent it in. And then... Really quick, I want to talk... Could you talk about the two-way translation? Because I think that's gonna be the biggest, biggest game changer. Is that I can speak in my language and it's translated into theirs. And then they can speak in their language and it's translated into mine.

EVAN TICE: It's actually 8-way translation. At any given moment, so... And you can test this out right now, by just going into the settings panel and changing your language preference... The words I'm speaking in English are being sent to all of you in 8 different languages. The two-way feature I think you're talking about is the ability to view captions in a different language. That's just a choice we made in the UI. Because I think it would be really hard to, perhaps, see your speech in more than two languages. Or to see it in more than one language. But yeah.

I envision folks being able to use that. Maybe they're trying to learn a new language. Maybe they're proficient in a language. Or maybe they're not proficient in a language. Maybe they're talking in English, but they're more comfortable speaking in German. And they want to see how their speech is being translated. It's probably not the most common use case. I'll admit. During this talk, I've kept my captions on English the entire time. But... There are those that you suspect will get value out of it. Did I answer your question?

LORELLE VANFOSSEN: Yes, you did. And I see in AltSpace when it's ready to grow up and come and join us with the captions... Because I cannot wait for that moment... Of having that choice to -- of a language set upon installation. So that people who come into AltSpace, they land in the info zone or the campfire, wherever else, can immediately have answers to their questions, and they can immediately connect with people that are around them. In their language. Not just... Don't speak Italian! We can't help! The problems we've had supporting that. I love this. Thank you.

EVAN TICE: I hear you!

THOMAS LOGAN: All right. Thank you so much. And I'm gonna take a couple more questions from YouTube. But for those of you in the room with us here in AltSpace, please use the raise hand feature, and we'll be calling on you next, after we handle a few more on YouTube. So Evan, we have a question from Wendy Daniels. Do you know what the word error rate is for your project?

EVAN TICE: I don't. I have no idea. I will say... I spend a lot of time developing this feature. Reading the Gettysburg Address over and over again. And tuning how captions appear and when we break and when we don't. When we break up phrases and when we don't. I found that with certain types of text, it seems to work better than others. Lorelle just mentioned the name of our product is not always properly transcribed. I've noticed with other words... You know, in mixed reality, we're sort of a niche. We're not, as I said, the primary use case for translation.

And some of our technical jargon and phrases don't translate as well as the Gettysburg Address. But I have no idea about the error rate.

THOMAS LOGAN: Okay. Thank you. Next question from Jocelyn Gonzalez. How has AltSpace managed to add all these deep speech models without creating lag? Doesn't it make the app significantly larger?

EVAN TICE: It does not. I think... Don't quote me on this. Actually, I probably shouldn't say that. I was gonna say... I think the fonts themselves, for the languages that we display, particularly the Korean language, and some of the other Asian languages, those files are very, very, very large. But the actual models don't live on your device. They live in Azure. So we don't -- we need the binaries that know how to interact with the Azure service. And we need the fonts. The character atlases. But that's the bulk of the file size. And all of the compute happens in the cloud.

THOMAS LOGAN: Great! Thank you. All right. Last question we'll take from YouTube, and then we'll come back to YouTube and we'll do some more questions here in AltSpace. But last one from YouTube for right now. From Simaspace. You have (audio drop) why not?

EVAN TICE: I lost... Your audio cut out. At the end of the question.

THOMAS LOGAN: I'm sorry. Do you have any Deaf people on your team? If not, why not?

EVAN TICE: We, on my immediate team, I do not have any deaf folks. We do have colleagues within Microsoft and our broader organization that are deaf. And why not? I would love to have more deaf folks on our team. And... Yeah. Careers.Microsoft.com. Tag AltSpace VR.

THOMAS LOGAN: (laughing) Cool. I'm gonna go back now... Chris Turner? I'm gonna turn on your on-air and megaphone. I think that'll put your microphone on. If you would like to ask a question.

CHRIS TURNER: Thank you very much! Yeah, thanks. And it's great that you guys are doing this. My hat is off to you. And I like the idea around "beyond". Because when we think about the many different types of disabilities, from vision... Hearing... Motor dexterity, cognition, mental health, speech -- there's lots of opportunities. A few that I was wondering about is: You know, maybe if someone had a reading disability or something, the ability to click on a button and actually speak a message, or speak, maybe, a menu, or an event menu, or something like that. And I have many other suggestions and ideas around different disabilities. I was wondering if there's a place where we could submit some of our ideas. So that they could get to you guys for review.

EVAN TICE: Please do submit a ticket on altvr.com. Not all vr. Altvr. Yeah. I'll check tomorrow. I keep telling people to put things in there. And I'll collect any feedback that's submitted.

CHRIS TURNER: Great. I took a class recently from Hector Mento. He's on the accessibility team at Microsoft. And it really opened my eyes. The class is called Digital Accessibility for the Modern Workplace. It's on LinkedIn Learning. And there's really a lot of things we need to think about, when we're creating content, whether in the workplace or outside of the workplace. And how that content might be interpreted by those that have different preferences or different needs. I think we all need to build towards that. So that we're inclusive.

EVAN TICE: Yes, absolutely.

CHRIS TURNER: Thank you again.

EVAN TICE: Thank you!

THOMAS LOGAN: Thank you, Chris. Next, we've got Kurt. VRDude18. I'm going to give you the megaphone.

KURT: Hello. Hopefully you can hear me.

THOMAS LOGAN: Yes, we can.

KURT: Yeah. So are you gonna have text-to-speech in AltSpace in the future? Where people can come in through 2D and whatever and do translation? Where people can speak...

EVAN TICE: I would love to have text-to-speech. I can't make any product announcements today.

KURT: Okay. The other thing was... I would like to see in-world, where if we put up pop-up messages in our worlds, that the pop-up message can be translated to that local language for a person coming in. Who has a setting for a different language. That would be nice.

EVAN TICE: Great feedback.

KURT: Yeah, okay.

EVAN TICE: Please do submit a ticket. All of these ideas. I can't take notes and wear a headset at the same time.

KURT: No problem.

THOMAS LOGAN: Now we're gonna call on Makoto Ueki, who will be speaking Japanese and using the language feature here live in the room. Let me turn on Makoto. Let's see. I need Lorelle or Evan's help with this. I'm not sure. It doesn't look like I'm able to enable that.

EVAN TICE: There we go. Why can't I do it? Oh, because they're not in the room.

LORELLE VANFOSSEN: Yes, they just left, sorry.

THOMAS LOGAN: All right. Well, Lorelle, thank you! Next question.

LORELLE VANFOSSEN: Yeah. We're all clicking it. When you had your videos... The videos were not being captioned in AltSpace. So it wasn't working. And... Is there... Because it's tied to our microphones, right? How do we take that next step, other than to caption all of our videos manually, or bring in... Somehow caption in the web projector, whatever? How do we pick up those sounds that are not truly mic-based? For captions?

EVAN TICE: Oh, brilliant question. That is an absolutely awesome question. I mean, we clearly have the technology to send audio off to be captioned. Just in a different place than the web projector stack. Send me that as a bug as well.

LORELLE VANFOSSEN: You got it!

THOMAS LOGAN: I'm excited to see that. I'll plus one that, if I can plus one that. All right. Makoto-san. Here we go. I think you're live!

MAKOTO UEKI: Hello, everyone. Can you hear? Good morning. Now I'm in Tokyo, Japan, now. And I speak Japanese. My Japanese is translated into English on your screen. Whether it has been translated into your own language?

EVAN TICE: I see your text. Your Japanese text was translated into English. You just sent goosebumps down my back. Because I've never done that with a Japanese speaker before. Wow! This is what this is all about, folks.

MAKOTO UEKI: And now you speak English to me, but it has been translated into Japanese. I see subtitles translated into Japanese on the screen. Thank you very much for each interesting story.

EVAN TICE: Well, thank you!

MAKOTO UEKI: Cheers!

THOMAS LOGAN: That was a very good demo. Hopefully that's captured on the stream as well. Because it was a very good translation. Okay. I'm gonna take another question here from YouTube. And everyone here in the room. If you would like to add a comment or question, please click the "hand raise" feature. We have a question, Evan, from Rebecca Evans. What are AltSpace's plans for localizing web-based tools and also in-app menu items?

EVAN TICE: Great question. I can't make product announcements today. I would encourage you to keep those accessibility bugs flowing. I hear you.

THOMAS LOGAN: All right! So now we're gonna go back into AltSpace. KiliChicChick. I'm giving you the megaphone.

>> Hello! Thank you. Can you hear me?

THOMAS LOGAN: Yes.

EVAN TICE: Yes, hello.

>> I have goosebumps from that. Thank you, Makoto, for doing that. That was a good experience. This is more like a feature request. When you are a presenter, and you're speaking, I can see the text. But there's no gray box behind it. So it's kind of difficult for me to read it, if it's not behind a solid background. So if that could be added on a list of to-dos, that would be much appreciated.

EVAN TICE: Please send it in. I've heard this one from Meryl and others before. The more feedback we get, the better. Once again, I hear you.

>> Thank you!

EVAN TICE: I picked this darker room, rather than the outdoor space. With that in mind.

>> (laughing)

THOMAS LOGAN: All right. Lorelle?

LORELLE VANFOSSEN: Sorry. I'm crying right now. I don't know about the rest of you. I have been fighting with web browsers to make translation and captioning and what-not be accessible. For decades! And to actually see this, when it was first announced, like... Five, six years ago, that AltSpace had that, that's when I chose AltSpace to be where I wanted to be.

And to just witness what I just did, with the translation... Thank you! (sobbing) So much! From my heart and everyone else in the world! So thank you very much!

THOMAS LOGAN: Yeah. And I completely agree. It's like... It's such an exciting feature. And it's something that we've never seen. And it's really exciting to see it in this environment. And... Yeah. That was awesome. Echo that. Thank you! FleetAdmiralGraham9753? Should be unmuted now.

>> Yeah. I was wondering if you could talk about some of the technical challenges? For me, I've used this with some people before, who were very comfortable with two or three languages. And they found... They've told me that it's been a little bit hard for them to use sometimes. Because in order for it to properly translate -- let's say from French to English -- they have to stay locked into speaking French. And they're more used to... When it comes to talking to someone who is monolingual, kind of... Jumping between the two languages. Is that possible to allow for? Or is there some technical challenge that makes it so you have to stick to one language and speak it very fluently and clearly for it to be translated?

EVAN TICE: I am not an Azure or even much of a machine learning expert. But I can guarantee you that we've heard this feedback before. And I'm sure the folks in the Speech SDK Team have heard this as well. I would imagine that training a larger model, that is capable of doing... Cross-language... Would be more difficult. Require more compute. But if we've learned anything, and if the spirit of my talk is -- really, like, we need to keep pushing boundaries. I'm with you. I would love to live in a world where -- I mean, you saw the slide: We have to tell the Speech SDK today what language we're speaking.

Who knows what speech recognition technology a year, five, ten years from now will do? We're not privy to any of the Speech SDK Team's plans. But I'm sure they've heard that feedback as well. And you mentioned technical challenges. Oh, sorry.

THOMAS LOGAN: Go ahead. Go ahead.

EVAN TICE: I can comment on that. Going into this, I knew nothing about font atlases. So that was an interesting learning journey. Doing those in Unity. It was challenging. I don't think I'm an expert, by any stretch of the imagination. But we pulled it off. The other thing that was challenging, strangely, was bandwidth. And you don't think about it for text. But these large phrases, over time, in 8 languages, it can build up. And if you play around with this feature when you're not megaphoned, one of the things you'll notice is that when you walk further and further away from a user, and their speech bubble suddenly disappears, or eventually disappears, and you stop hearing them, you'll also stop seeing their captions. And in fact, around the time we stop sending the audio, we also stop sending the caption data. That was challenging as well.

But kind of a fun problem.

THOMAS LOGAN: Cool. And Evan, I'm gonna take a moment just to steal a quick question about... I was curious for, like, the SDK calls that you were showing -- is there any kind of plugin or adapter for, like, Zoom? Or if someone wanted to get similar functionality, how hard is it to use the SDK? Is there any kind of helper functions? Or would you, if you wanted to, for example, use this with Zoom, need to write your own code?

EVAN TICE: I have never written a Zoom plugin. Does anyone know how extensible Zoom is? Or what their SDK looks like? We can leave that as a rhetorical question. I'll tilt my headset back up and go back to the cognitive services website I have open in another tab. And I will... In just a moment, I can read off the language projections that they have. But it's more than just C# and Unity. Let's see. Azure Speech SDK. Maybe while I search for that, we should take another... Oh, here we go! Um... C#, C++, Go, JavaScript, Objective-C, Swift, and Python. That's from docs.Microsoft.com. But they have language projections and engine projections for quite a few things. I don't know about Zoom specifically, but getting started in Unity C# was not hard.

THOMAS LOGAN: Thank you! We'll be posting links to resources you highlighted in your talk on our meetup page, Meetup.com/a11yvr, and we'll be doing recaps on our blog at Equal Entry. I'll read a few more things here from YouTube. We have from Patrick Donnelly: An appreciation. I love how the live translation gives an appreciation for syntactic

differences between languages. And a question. It might be another secret one. But from Leon: Any plans for Chinese translation?

EVAN TICE: I would love to do Chinese translation. Can't make any product announcements today. Wish I could. I'm gonna give a shout out back to Patrick. Patrick is actually my husband. And studies speech and hearing and the reading brain. So I appreciate that feedback as well. I don't know that he's actually seen this feature ever before. So... Thanks for joining, Patrick.

THOMAS LOGAN: That's cool! All right. From Joly MacFie, our Internet Society of New York streamer. He says: Captions was a beta feature. Is it now live?

EVAN TICE: We're still calling it a beta. Oh, sorry, your audio cut out. I didn't mean to step on you. We're still calling it a beta. You might have noticed that it's appearing in more events than it used to, and that's a good thing. And I hope that continues. For those of you that are on Mac, you've probably realized that captions are not working. I would be hesitant to remove the beta label until we're... You know, at parity, across all of our platforms.

THOMAS LOGAN: Cool. And then we have a comment, which I believe they can send to your support link. That switching captions from large to small and vice versa did nothing for someone. Someone had that feedback. Is there any extra guidance on how they submit feedback?

EVAN TICE: Please file a ticket on altvr.com/support. I will personally make sure that our feedback gets collated. I'll probably look tomorrow morning.

THOMAS LOGAN: Cool.

EVAN TICE: I don't want to call her out. But somewhere in the audience one of my coworkers, who is also passionate about accessibility, is hanging out. Actually, I see a bunch of coworkers. Thank you all for coming.

THOMAS LOGAN: Oh, that's awesome!

EVAN TICE: Yeah.

THOMAS LOGAN: Yay! All right. I'm gonna take a few more questions here. On YouTube. And we're getting close to the end here. For our stream. But... Obviously as long as anyone is available to hang out in the space afterwards, that's fine. We have a question still on the YouTube stream, from Miles de Bastian. I heard AltSpace was going to support hand tracking, such as Facebook, Oculus Quest. Any comment? Because it would allow Deaf people like myself to converse in our native Sign Language.

EVAN TICE: I have also heard that folks are using VR chat to sign. So I understand the ask. I wish I could agree to implement all of these features. And roll them out tomorrow. Party line is I can't talk about... Anything... In development at the moment. But I do hear you. I would love to be able to do it.

THOMAS LOGAN: Awesome. All right. And then from Rebecca Evans, she said: Thank you for your response. This feature is incredible. I was lucky to have it enabled, by special request, for an AltSpace event, and everyone was amazed. Wonderful way to practice learning languages.

EVAN TICE: Yeah. Actually... So thank you for that. The view captions in another language -- we actually thought learning a language might be a primary scenario for that feature. So I'm glad you're using AltSpace to learn. I'm guessing you're probably using that "view in a second language" feature. Thanks for trying it out, and thanks for the feedback.

THOMAS LOGAN: Great. All right, everyone. Well, I really appreciate it. We want to thank you all for coming to the event. We're going to be now sort of just... Letting you know that, again, this is Thomas Logan, from the Accessibility Virtual Reality Meetup. We do meet once a month. Usually meet in Mozilla Hubs. But we've met tonight in AltSpace VR. We've met in Engage VR. Sometimes we change the locations. We really appreciate such an awesome set of attendees tonight, both in the room and on YouTube. And thanks to Evan Tice, with his excellent presentation today. So we are, as we've always said, going to make sure that the recording of this presentation and the different assets that were referenced -- those will be available to you. This presentation will be something you can come back to. And thank you for all the emojis. Evan, we really appreciate your presentation tonight. I want to end with saying: Thank you again to the Internet Society of New York, Joly MacFie, who always makes sure that we get our events streamed on YouTube. And recorded and available for anyone to view in the past. I want to thank -- I'm very excited tonight that we have Barbie Greenwater from LotuSIGN. If you are watching our YouTube stream, we've started some of our events at a11yVR and accessibility New York City having Sign Language interpretation at our events, upon request. So we're very excited to have that on the Zoom call. I didn't really say this at the beginning of our call, but we are always in VR wanting to learn best practices and make more improvements.

So we want to have the Sign Language interpretation. Also in the virtual space. And we want to have our human captioner, Mirabai Knight, from White Coat Captioning, who if you're watching the livestream, we'll be having her captions provided. We are always working in these spaces to advance it. We are an open book, in the sense that if you send us feedback, we'll communicate with people like Evan at AltSpace on your behalf. Or people at Mozilla Hubs. We are always trying to make sure that our events are as inclusive as possible. And that's another reason that we appreciate being able to have the YouTube stream as well, for people who may not have a headset, or cannot join, for example, in AltSpace at the time.

Our next event will be coming up in July. So please check out our website, meetup.com/a11yvr. And I'm looking forward to hanging out with you all in this space. And getting to know you all better here in AltSpace. And thank you so much for coming tonight! And thank you again to Evan.

EVAN TICE: Thank you! Yeah. I'm happy to let people talk again. Let's see.

THOMAS LOGAN: All right.