**Globally Speaking**

**Podcast 90 Transcript**

**Talking Transcription — Part 3**

**Featuring Sam Liang of Otter.ai**

Michael I’m Michael Stevens, and this week on Globally Speaking, we are continuing our series, number three of three, on transcription. If you haven’t checked out the first two episodes of the series, listen to this one and then go back and listen to the other two. I think you’ll really appreciate it. We started with a look into how transcription is traditionally done, what the human being actually does for this type of work. In our last episode, we talked to the founder of a company who’s looking to transform work as we know it by distributing work around the globe so people can live in flexible areas, and the first type of work he’s looking to distribute is AI with human transcription services.

This week, we follow the trend, and in our third interview, we talk to a technologist who believes that in many cases, the machine is going to be able to do the work for us. I think you’re going to be really impressed with how far advanced this technology is, the AI, and I think you’re going to want to go and sign up for this service after you hear what’s happening in this space.

So, let’s let our guest introduce himself.

Sam My name is Sam. I’m Founder and CEO of Otter.ai.

Michael In a nutshell, what does Otter.ai do?

Sam Otter.ai is a product that helps people keep track of their conversations, their meetings. You can either use it individually for yourself or use it for a team in a team meeting scenario. Think about enterprises, small and medium businesses or large businesses. Everybody has tons of meetings these days, either in-person meetings in conference rooms or you just have a stand-up, quickly, every day, or you have phone calls, you have video conferences. So, there are a lot of studies that show that people in enterprises spend 30%, even 40%, of their time in all kinds of meetings.

We look at this space and say, hey, you know, a few years ago when we looked at this, where is the data? People spend 30% or 40% of their time talking or listening. You can take notes or you can remember part of it, but most of the data actually just disappeared in the air.

Michael Yeah, right now it just goes off after the meeting, it’s not capturable unless you’re working at like Blackwater or some company that has a…

Sam Bridgewater.

Michael Bridgewater, Bridgewater Capital, who has the concept of recording every meeting they have and providing most of the notes publicly to the members of their company. But most companies, it’s just entirely, it’s gone, except for what a certain person takes notes at.

Sam Right.

Renato So, the solution that you found for this is essentially you have incorporated text-to-speech technology and you are able to capture conversations and create an automatic transcription of that conversation in text format, right?

Sam Yes, we actually created the speech recognition technologies ourselves. We’re not using a third-party API. We actually created everything by ourselves. This is why you can see very high accuracy with very high performance in terms of a real-time transcription. Right now, we’re having this call and running Otter on my laptop where I’m using Zoom, so while we’re talking, all this data is recorded by my laptop, but it’s sent to the Otter server in the cloud. The Otter server processes the audio immediately and sends back the transcript immediately. So, within a second, you can actually see the transcript right away.

Michael And a moment ago, when you shared your screen with us while we were just chatting, and I could see my text come up, it was a really interesting…like, I could feel the difference in my brain. For those who are listening, I was looking at the Zoom thing, seeing the words that I was saying automatically be populated on Sam’s application there.

Sam Right.

Renato By the way, if you are listening to this on a smart phone, just go to the Google Play or App Store or whatever store you have, download the Otter app and play with it. You’re going to be amazed. Today I just spent eight hours in a meeting with 18 people. I put the app on and it was capturing every conversation with a degree of accuracy that shocked me. What is the underlying technology behind this? Because we know that voice interface is something that is growing with these Alexas and Google Homes and Apple Siri. How did you come up with this technology and how does it fit in this world where voice has become the main user interface?

Sam We are actually relatively young; we started only three years ago. At that time, whenever people talked about a voice AI, people thought about Siri, Alexa you mentioned, Google Home, though it was not necessarily in that device. We are actually working on a totally different problem. When you use Alexa, you use a hot word like “Hey Alexa” to wake up the device. Then you ask…

Renato You just woke up a ton of Alexas. [Laughter]

Sam Then you ask a short question like, “What’s the weather tomorrow?” Then Alexa, which is a robot, will answer that question. So, the whole interaction is a chatbot interaction, you know; a human being interacts with a robot. It’s useful, however, we see—just look at, you know, a hundred people—how many times does each person talk to Alexa? Probably no more than three times on average every day. But, look at the average person: how many hours do they talk to other people? Either they’re speaking or they’re listening to other people speaking.

So, you’re going to be at least two to three hours every day. So, for myself, one of my own problems is…I used to work in Google; I led Google’s map location service 10 years ago. We actually built the blue dot system, Google mobile map, on your iPhone and Android. Then I quit Google in 2010 to start doing startups.

I have a lot of meetings with venture capitalists. I have a lot of meetings with customers, a lot of meetings internally with our team. Now, I have this pain point that I just can’t remember the things people told me. I can’t remember all the things I promise other people as well. So, I thought about it, you know: okay we spend so much time talking and I can’t remember this. Is there any tool I can use?

I looked around and I just can’t find such a tool. On the other hand, I just feel like, you know, life is short, you know. You meet a lot of people who are interesting, but you talk to them and then they go away, you know? I want to remember that conversation. A lot of them have, you know, a lot of value, either technical or insightful conversations or, you know, with my family, it’s more sentimental.

I wish I could play back whatever my mother told me. So, several motivations that motivated us to do this. So, three years ago we started to work on this, we built a team, and we looked at a number of technologies: the Google API, you know, IBM, Microsoft, but none of them are good enough. The reason is that the original speech recognition—think about Siri or Alexa—they are all trained to support the chatbot model. And Google, when you can use a voice to search using Google, so their system is mostly trained to optimize a voice-based search.

The big difference between that and the conversation we are talking about is that the way people speak is very different when you talk to a robot versus talking to another person. When you are engaged in a conversation, people talk much faster; they talk casually. They don’t speak very clearly or even grammatically correctly, right? People interrupt each other a lot. So, there is a lot of noise; usually either in the conference room with the air conditioning running or you meet somebody in Starbucks or a restaurant. A lot of noise.

Different people have different accents. You know, I came from China so I still have my special accent. So, we have to build a model from the ground up to handle all of these problems, which actually [are] not [what] the traditional speech recognition system can handle.

Michael This this is absolutely fascinating because a few episodes ago, we did an interview with a traditional transcriptionist, and one of the reasons in [the] case she was making was environments where there’s lots of background noise, multiple accents, even speaking the same language, and there is someone speaking another language in that moment. And it’s fascinating to now speak with a technologist who’s thinking of those same exact issues and trying to address them.

Sam Right. So, we’ve been working on this really hard for the last three years, which is actually a short time compared to the work other companies have been working on this problem. But, you know, we’ve been working hard, work smart and on top of that technology, we built the Otter product which is a full-fledged standalone product everybody can use.

It’s designed with a consumer-friendly user interface. It’s designed for non-technical people. Anybody can download it from the App Store, Google Play, or you can run it on your laptop. I just go to [the] Otter.ai website.

Michael And Sam, I was going to ask you a question. The purpose of Otter.ai, and you mentioned it’s where conversations lived, kind of took me back to when Evernote started as a company and they would say it’s your second brain. Right? Like, it’s this thing where you can store information. And so then it made me think, you know, Evernote never quite got the collaboration piece. It’s good for organizing, but outside of that, it always seemed very far.

When you just described your service, I imagined the episode of “Black Mirror,” or multiple episodes, where they have the lenses that recorded all of life, and you could go back and skim through different scenes in your life to hear what people were talking about in the room. Instead of having a lens that records things, it would almost be like a user could have a microphone on them, record their day and then have multiple AIs who create tasks from things you say or commitments you make, or working in that background.

Is that a longer-term, futuristic version? Is that what you would see as the ultimate benefit, or is there something else there as well?

Sam It’s absolutely a part of it and an important part of it. Right now, we’re focusing more on the meeting use cases, whether it’s a corporate meeting, it’s a stand-up project meeting, or, you know, interviews like this, or podcasters; you know, a lot of podcasts are interviews. Even for today, the value is more than just the recording. It’s because once you have the transcript, you can do a lot of natural language processing on it to understand the topics people are discussing, automatically extract the action items for project meetings.

So that’s why we actually see a lot of project managers using Otter. I have friends who work at Google, Facebook, Amazon; they told me they’ve seen people using Otter there, although it’s not officially [laughter] endorsed by their IT system. They don’t even ask about security and all that.

You mentioned Evernote. I like what they are trying to pursue, they say remember everything. I really like it. But, so far, you know, they require you to manually enter the notes.

Renato This is the part that fascinates me, is that I’m a lazy consultant, and it’s very hard for me to be talking and presenting and discussing with my clients during a consulting engagement and taking notes at the same time. And this is a faithful compilation of everything that is being discussed, and as we’re discussing, I can see the transcription going on right here. It has excellent punctuation, it has full stops, it has paragraphs; the text looks totally natural.

And what I loved today is that at the end of the meeting when I saved and I closed the meeting, there is also automatic tags of the topics that were most discussed or the words that were most used. I don’t understand exactly how that tagging happened, but there was an automatic tagging process. How do you do that?

Sam [Laughs] That’s what I meant by natural language processing, is actually analyze the text, look at the semantics, look at the way people talk about it, look at the frequency they use it. Right now, it’s still relatively rudimentary, but this is just beginning. It will get more and more intelligent in terms of understanding the higher-level concepts that people are talking about.

Michael How much of these are sort of a rules-based system where someone’s gone in and said, ‘if this frequency of words, we’ll put it into this category,’ or ‘if these words [are] together, this topic will come up’. How much of it is that, and how much are you being able to use neural networks to be able to do that processing?

Sam Right now, the result you saw, it was all done by machine. It’s actually not done by a human being. Although, in the training process, we do have a human involved to look at the classification of words, look at the association of different words. Are you talking about a medical problem, are you talking about AI, are you talking about sports? Right? These people use different words. It’s a big research area still. It’s definitely far from solved yet, but just like self-driving cars, it’s not fully realized yet, but maybe we can still get some partial results by using Tesla’s autopilot, right? It’s not level five yet, but even at level two, it is still useful. You know, what we are doing is actually similar.

Again, this is a technically very, very challenging problem. Even human beings, you know, have trouble understanding each other, right? How can humans understand what they’re talking about?

Michael Absolutely. And, for a number of our listeners, a natural question is going to be—it’s in your frequently asked questions—what languages do you support? And right now, it’s currently English, but what do you see the future looking like? What are some of the unique challenges when you decide to go beyond English for this type of technology?

Sam We get a lot of questions like this. We get people who ask us when are you going to support Spanish, when are you going to support French, Mandarin, Japanese? We actually have a lot of large enterprises in Japan who ask about this. This is definitely on our road map, although we have a small team, right now, so we cannot do everything at this point.

We’d love to expand the scope and work on other languages, and on top of that, translation; like, translate the transcript from one language to the other language, also incorporate speech recognition—exactly doing speech-to-text—and reversely, you want to also do text-to-speech. Once you have the transcript to Spanish, for example, you want to have a human, natural human voice, say those words after translation.

Michael There’s a company out of Boston called Vocal ID who’s doing a lot of interesting work on the text-to-speech. They started through accessibility, for people who are in wheelchairs and using a monitor to type and communicate with folks. And the founder was sitting at a conference giving a speech and there were two people in wheelchairs, one was like an 18-year-old female and the other was like a 40-year-old male, and they were talking back and forth and had the same computer voice talking to each other, and she thought, ‘what would it mean for us to start reflecting the human being in the voice as they interacted?’ And so they’ve done a lot of interesting work. I’m not sure if you know them but they’re one of my favorites out there, Vocal ID.

Sam I don’t know about this specific one, but I will look for them. I’ll send you a video actually; we recently were actually invited by USDA, United States Department of Agriculture, to speak about using innovation to enhance productivity and also help with accessibility. We actually have a lot of users who are deaf or hard-of-hearing. They use Otter to help them understand people’s conversations.

We’re actually recently working with a number of universities like UCLA, Western Kentucky University, because every university like this has a few hundred people, hundred students, who actually have learning disabilities. Either they have a hard-of-hearing problem, or they have other, dyslexia or ADHD.

The school actually spend a lot of money to hire people to take notes for them. When they discovered Otter, they actually contacted us and we’re working with them to provide the service to them. We actually already know there are hundreds of students in like Stamford, Berkley, UCLA, that are using Otter.

Another part is, actually because you talk about localization and globalization, for international students whose native language is not English, they found Otter really helpful for them, even if and before we do translation, with the transcript, it makes it a lot easier for them to understand the professors and the presentations because, you know, they can listen and read at the same time. They can usually read better than listen.

Michael I experienced that first-hand in grad school. We had a large Korean student population and they had every class for, I believe it was the last five or six years previous to mine, all the major courses transcribed. You bought those notes and you had them to work with. But yes, I benefited from folks who were using that as a tool, and now Otter.ai makes that available. You didn’t have to know the secret handshake anymore; you can just use Otter.ai, which is great.

Sam Yeah. Another part is, we see this as not just an individual note-taking service, although podcasters can use it, students can use it to take notes, but we also see this as a collaboration system. While we are having this group call, actually, I sent you a URL which, if you click on it, you can actually see the transcript on your laptop. During the meeting, you can actually highlight certain interesting things to help you remember the important points. We will allow people to add comments later, as well.

So, think about, you know, Slack is a text-based messaging system. But Otter can help you communicate better with a voice. You know, in Slack you have channels for each function like the product channel, engineering channel, marketing channel. But in Otter, we allow you to create groups for each function. You can have all the product people in one group so that all the meetings or conversations related to products is shared in that Otter group with everybody.

Renato I’m thinking here of your work at Google, of the Google Glass period when everybody at Google was wearing Google Glasses and then it didn’t take off. Are you the kind of person that goes around and records every single conversation that you have? You mentioned recording your mother’s voice and so on. Is that something that you do under the guise of testing the product?

Sam Not yet, but I do want to be transparent. I actually already tell everybody that, most likely when I talk with them, I’m going to use Otter. For myself, they already know because I work on Otter, if they don’t want to be recorded, they actually tell me, “I want to opt out.”

Renato Okay. [Laughter]

Sam Internally, in our own team, we actually use Otter for every single meeting as a backup. You may never look at it again, but when you need to look for information, it’s available.

Renato And it’s searchable. It’s very user-friendly.

Sam Yes. I don’t know if you’ve seen, we actually were selected by TechCrunch Disrupt San Francisco to use Otter for the entire conference. So, all the speeches were available. You can search for anything: you can search for a speaker’s name, you can search for a specific topic. Suppose you record all your own meetings, and months later you try to recall, “hey, what did Sam say about Alexa?” Then you can quickly search through your notes in Otter; you can find it quickly.

Because another thing we did was a voice print. You can label the paragraph with the name of the speaker and our system will create a voice print profile for that speaker. Then, later on, when Otter hears the same speaker, it can label it for you automatically. That’s also maybe another dimension there, so you know who said what; you can search based on the speaker name as well.

Michael I was nervous for Renato when we started recording our podcasts because I knew there would be things that were recorded that he was saying. Now, and the fact that he could have it in text on a regular basis that people could search, Renato, do you realize how many more things you’re going to be held accountable for?

Renato Well, yeah. That’s terrible!

Michael It’s dangerous!

Renato All my empty promises. [Laughter] But, one question. I live—and I’m sure you do too because, as you said, you come from China—I live in a multilingual family. In a regular conversation at home, we speak Portuguese, French, English; that’s the everyday languages that we speak at home. Occasionally we have some Spanish and some Italian. You’re not going into that space yet. What is the barrier? What prevents AI and voice recognition technologies to work like the human being is capable of doing, of switching languages and still maintaining the same conversation and the same topic? What is the technology challenge for something like this?

Sam It’s just a matter of time, right now. If we have, you know, five times more engineers as we have right now, we can tackle those problems. I did my PhD at Stamford; I’m a technologist myself. The problem you just mentioned is actually not that hard. We are able to tell what language you’re talking even in the middle of a sentence. We can tell you switch from one language to another language. We can run the audio through multiple models: one English, one French, one Portuguese, Spanish, then we can get the result from each model and merge them and show the transcript.

Michael I hear a lot of people saying, you know, code switching is a challenge because sometimes we’re saying things, there’s a meaning beneath, and how do these things get captured? You seem very optimistic that this could be solved, which is good to hear.

Sam I’m very optimistic. I think the AI is so powerful, it is learning from tremendous amounts of data. I don’t know when it will happen, but before long, the machine actually can understand the text better than any single human being. The reason is that the machine is actually taking advantage of the global data available to it. You know, any single person, you just have a limited number of brain cells. But the machine potentially is almost unlimited. So, the limited number of brain cells cannot compete with [an] unlimited number of machines.

Michael Well, I also like the fact that, at least the use cases you’ve described so far, they’re still in service of people’s conversations. Do you think it’s going to be too long before we’re just having conversations with AIs that have been developed in this space?

Sam There is a movie called ‘Her’…yes, it will happen, but right now we’re not focusing on that space. I think we’re trying to solve the problem to listen and understand human-to-human conversations because that’s the main use case today. People are talking to each other. But what you describe can be achieved as well. You know, again, as we learn more and more about how human beings think, then they can emulate how a human being will answer a question.

Michael Yeah.

Renato Sam, what is the underlying technology here? How do you train this? Because here we have people using three different accents and the transcription is showing up perfect. Do you use multiple engines? What is the underlying technology for Otter?

Sam Yes, in short, we are using machine learning. That’s the high level: machine learning and neural networks with deep learning. This is why, actually, a few years ago, no matter who you looked at, whether it was Google, Microsoft, IBM, nobody had this type of accuracy. People, you know, scientists came out with some really nice models and real nice more powerful methodologies to train the neural network. So, for us, we actually have millions of hours of audio data we collected from all kinds of sources. We use that to train the model.

Among these millions of hours of audio, there’re actually all kinds of accents. There’s a lot of work to do; obviously, not all the data is useable, we actually crawl tons of data from the internet. We have algorithms to clean the data, filter out the bad data. Then we tune the model. There is constant enhancement of the model to make it more accurate.

And also, we allow users to correct. You know, the transcripts are very accurate today, but there are still errors. We do allow people to correct the words in [the] Otter user interface. So, when you correct it, then the machine will actually take advantage of that correction and try to pick out why did it do it wrong the first time and look at your correction, and if a lot of people correct the same words, you know, we will incorporate that into the model as well. So, longer-term, this becomes a self-improving system. It will get better and better. Six months later, you use Otter again, you will see it will work better.

Another new feature we recently released is called Custom Vocabulary. There are always words, new words people are inventing that then the dictionaries don’t have yet: company names, new acronyms and new jargon. So, if you stay still, nobody can understand those, but, you know, a machine can constantly crawl new data.

And also for words we never heard, we allow users to add that into the dictionary. That’s the Custom Vocabulary. This is a premium feature; once you get the premium plan you can do this. Once you add those, and also people’s names, you know, right? There’s a lot of international people who don’t have traditional English names, and when you pronounce them, how do you transcribe them correctly? So, we allow people to add those names into Otter as well so that their names can be transcribed as well.

Renato This is fascinating. I’m just wondering, you know, everybody, most people have seen that Facebook funny video of the voice-activated elevator where people don’t understand the Scottish accent. I wonder if Otter can understand Scottish and Welsh and these very strong [laughs] regional accents. It would be an interesting thing to train.

Sam Scotland, Ireland, UK, Australia, South Africa, you know, even [in] the United States, people have different accents, Canada, Texas, and also, again, a lot of international people, you know, whose native language is not English, including myself.

Renato One of the things that I found fascinating—and this is something again I told our listeners in the beginning of this conversation, download Otter—but it’s that you can have basically 10 hours of transcription for free per month, so some people don’t need more than that. And then you have a professional plan which is essentially very affordable: it’s like $10 a month or $9.99. So, it’s something that is really life-changing. I can always imagine this miserable job that—for me, it was always miserable—you go to a meeting and you have to choose somebody to take the minutes, and I said, “Oh my God! If I have to take the minutes, I cannot pay attention to what people are talking and I cannot interfere and intervene.” But with a tool like this, everybody has the time to participate and to engage, and even you can go back and look at your notes and say, “Ah, you didn’t say that. You said this.” [Laughs]

Sam That’s part of the benefit of using Otter, is to improve communication, improve engagement, right? Especially when you’re meeting with somebody in person, it’s actually rude, you know, if you keep typing on your laptop; you know, you’re not having eye contact. But, you know, with Otter, you know that information will be available so you can focus more on the moment talking with each other.

Of course, Otter itself will be further enhanced if you also identify important things. It’s not available yet, but we’re working on new technologies to actually automatically…I’m talking about certain things that usually people consider important: you know, numbers, money, you know, for a project meeting, usually action items are important. So, we’re building a new algorithm to identify those and highlight those automatically.

So, this is why also we allow people to manually highlight the transcripts so that the machine can also learn from the way people highlight the transcript; to learn, you know, what kind of a sentence is important for people.

Renato So, would you say, Sam, that this is the doom of the human transcriber?

Sam [Laughs] Well, yeah, people will say, “Is AI going to replace people’s…replace jobs,” right? I think yes and no, you know, this is the usual answer. When you replace the repetitive jobs, then people can learn new skills and do more interesting things.

Michael And now Otter provides a way that you can do it almost simultaneously, all the time. It’s fabulous.

Sam Right. Also, you know, when you are driving, you are having dinner, you’re talking with someone, it’s hard to take notes at those moments. When you’re debating with someone, it’s hard to take notes because, you know, you have to think.

Renato Is there anything that we didn’t ask you that you would like to share with us, or something that you expected us to cover and we didn’t?

Sam I want to stress the meeting use case. We see Otter will be part of the collaboration system. You think about Zoom, think about Slack, think about other storage software, the Box or StarBox. Otter actually can work really well with those systems. Zoom already licensed Otter as the exclusive transcription system for Zoom. So, if you are a Zoom user, you can actually use Otter inside Zoom. But Otter can also work with other video conferencing systems like Google Hangout, Skype, WebEx. During those meetings, you can turn on Otter and transcribe those meetings.

Because you are podcasters, actually, we already see some podcasters publish the podcasts using Otter because not only you can listen to it, you can also see the transcript, which is a big help, and because the transcript is searchable, the listeners, the users, don’t have to listen to everything. They can jump around and select the part they’re most interested in.

Renato It would be awful for us. We want them to listen [laughter] to every single word that we say.

Michael It’s all good!

Renato Well, Sam, one of the things that we learned from our listeners is that I have a very good voice to get babies to sleep. So, we serve multiple purposes.

Sam I see! Another actual benefit for podcasters is once you’ve published the transcript along with the audio, Google actually will index your transcript so that users, when they search for certain topics, make your podcast more discoverable.

Michael That’s awesome.

Renato Absolutely. Thank you for your time.

**End of conversation**