Rodney: Calvin, thank you so much for stopping by to tell me about your app. Before we jump into that, why don't you tell my audience, a little bit about your background and how you got involved?

Calvin: Absolutely, so for me, growing up second generation Chinese American it was always very difficult to learn Mandarin. If you could imagine. So for Me: besides having to study very diligently, I took the adventure to go study in Beijing. And I studied at Tsinghua University, I was very happy. But when I came back there. I knew there was something missing. Because that type of immersion really accelerate my learning a language, but when I came back, I was like while I'm learning augmented reality for computer science. Why couldn't I at least see a see whatever I'm hearing, and then also hear it in a different language or see it in different language, so I could learn a vocabulary that's of interest, permanent and relevant to what I'm learning? So I came back and I built the best team I could, we found the best technologist for advanced speech recognition, machine translation, and we built a system that: everything a speaker says or instructor when they use the software will be subtitled on screen in two languages. We get a transcript afterwards, captions and subtitles for the videos, abstract and analysis for afterwards, and that is what we are here to discuss today.

Rodney: Wow I mean that sounds pretty incredible. I wish I had a tool like that when I was studying German. So it's an artificial intelligence platform that I assume it's hosted on a web, or is this something that can be actually downloaded onto your mobile device.

Calvin: So most of the computation is done on the cloud level, just so that the processing is much more robust. We're slowly integrating our GPU cloud, because we partnered up with NVIDIA, and it's something that we're looking to make a switch to. But right now it's really for enterprise software for higher education or corporate learning settings. But afterwards we are slowly transitioning. We have a team of interns from Temple university that's helping us create a mobile app. So that when you're attending your next conference you can download the app, and if you're very far back or you're not even in the same room, you can see what the speaker saying, the video, their slide deck, and the captions along. And then you could choose a target language, so that whether you're studying German or Spanish or Korean, you can see it live in your language that you want.

Rodney: That's pretty cool. I know we're looking at different audience response systems here on campus. One of them is Glisser, which I'm fond of, but I don't think it has that sort of captioning ability, so if I can compare to it's different than say Google translate app, which actually I've never tried it for very long period of time. But it's something on your phone. And you're traveling abroad, you could have a two-way conversation with somebody and in another language. Can yours do that as well or is it, primarily designed for presentations like in a class?

Calvin: Right now, we're more presentation oriented. The speech-to-text application for consumers is down the road. But Google Translate and you know Bing or any of the other big providers, it's very generic, and the sense that it's like a dictionary. And it doesn't have personalization for your voice or your intentions are anything of that nature. So we are working very hard with our natural language processing and artificial intelligence to understand each one of our speakers very well. So right now, we really aim it for academia, meetings industry, and multinationals, and governments. So we do apply it more for right now just speech to text and translations, but later on we'll have a more advancements to come with wearable technologies.

Rodney: The whole idea of using artificial intelligence now is amazing to me. I mean this is you know, Google said they're becoming an AI first company, and it seems like possibilities are endless. How do you build something or do you take advantage of Google tools? Or a how do you how do you get started with building an A.I. application?

Calvin: Well you have to have very sound fundamentals in computer science, which I never had growing up with an international business and marketing background for my undergraduate studies. I found the best talent I could around the world that study engineering to a very high degree. So for instance, our Chief Technology Officer, Nagendra Goel. He earned his PhD in Johns Hopkins in speech processing, so he spent 25 years in industry and in academia to study how to make speech recognition the best it possibly can. And actually, we are recently stacked our engine up against Google's for recognition, where they performed on 80's on average. We performed Upper 90's, but it took let's say a lot of dedication to build something internally that we can create the privacy for those who want data protection and that we can secure it well.

Rodney: So what stage would you say your company is in now? You're still considered a startup? Do you have paying customers? How's that going?

Calvin: Right now we finished with R&D. We have a product that we have been testing at locations. And right now we have a partnership with ACT (American College Test), NVIDIA and we'll be sponsoring the African Trade Investment Global Summit in D.C. So we have a lot of things in the pipeline that we're working on a lot of fun partnerships that we growing and expanding in each one of the markets that we are in. For instance, we are partnered up with Meeting Professionals International (MPI) they're a large meetings group, one of the largest ones in the world. And for their academy of corporate learning, they've taken us on tour to a lot of different events in Vegas or wherever they choose to go.

Rodney: That's pretty exciting. It seems like there's a big need for something like this. You know, any company worth international companies they have offices everywhere. I can imagine this is really something that could go far. How do you see it being used in the education environment? Could you give us a scenario or two? How this might be used?

Calvin: Absolutely, I can you a real reason why we do this; the concrete reason is one out of five U.S. citizens are deaf or hard of hearing. That means those students always grown up had to sit in the front rows. They always had to read the lips of the professor or the teacher. Can you imagine paying a lot for an event and not even able to hear? How painful is it? And actually so many students unfortunately never speak up about their needs. Even though there's a whole office of disability services. They never mentioned that oh we need accommodations. Then if they never notified their teachers, they always get ridiculed, because they're like why do you have a headset on? Are you listening to music? So I've heard too many bad stories. This is one thing that really is pertinent and useful for them to have a visual tool that they could see. Then the language learning component, when you do rapid language learning it is very efficient when you learn information that's something that is to your liking. And so you're taking all these classes, now you can take them in another language, as well simultaneously. And then the third aspect really is the neuroscience behind it is when you see and hear one thing using two parts of your brain to process, you're statistically significant to retain that information so much more.

Rodney: Interesting.

Calvin: So we're not the first to create this real-time captioning was done back in the 2000s by IBM. They helped a handful of schools. And at that time, it was say anywhere from 15 million to 100 million dollars that they were generating from these schools that they were helping for disability purposes specifically. But given that they're IBM, they didn't produce over a billion dollars, so the company said no just sell it. It still lives on to this day, but that type of research really proved that actually using 10 years of data to prove that with using real-time captioning, transcription, and then an interactive transcript for videos, so that you could do keyword search or anything of the like and captions for the videos, students improved improved their grades by a whole letter grade. So it was really effective, it was just sad that they had to can the project, because of financials.

Rodney: Is that the company that became Nuance?

Calvin: So they they sold ViaVoice, their speech recognition engine.

Rodney: I see.

Calvin: And now they have it in the consortium, where they still use it.

Rodney: I see, but it's amazing how voice recognition like that just spread and like you said there services that has been around for a while, where you can translate the audio from a video into captions. But this is real time. What's the latency?

Calvin: So it all depends on how much we soup it up, because if we equip it with more computational power, it's faster. But on a standard, it should be less than a quarter of a second.

Rodney: Wow!

Calvin: It should be instantaneous.

Rodney: Because, wow, this all fascinates me how you train something; how you train AI. This is kind of the situation, I remember hearing about Google. And how the reason it got so good at text translation was because they didn't just go through rule sets, and in parsing all the text, but they would compare two different text in two different languages as a basis. And that has became the basis of their translations. Is that sort of the way this works with?

Calvin: Yeah, so neural networks is one of the bigger fields now, in the sense that AI is just taking lots of data and lots of computations like algorithms to figure out and solve common trends, that they see within that data set. So that when you dig it down another layer, when you add layers and layers and layers of layers of different types of data, then it can pull and retrieve that information faster or in a more efficient manner. But yeah that is it. In essence, how the software gets better over time that it collects more use case, it collects more case study, so to speak of a speaker profile, vocabularies. And it knows how you like to say things. How often you use certain words. And it understands both grammar and context. That's the hard part, because it's very easy just to load up lots of dictionaries, but it's heavy on the size filing. But for it to be a smart system, it really has to understand deeper about linguistics. And that's why there's a whole field that spun off called computational linguistics. Of which, one of our other partners, he has his master's in computational linguistics, so he heads the translation system. So, that is in efforts, I think everyone in the industry is really moving toward the deep learning route of you know multiplying layers of information. And the more we see it being used, the more it'll get better over time.

Rodney: It would seem to me that at least in the beginning stages do you have to have human beings that are looking at these translations? Like to see how accurate they are? So is that still something that's ongoing that you still have to have a lot of? Is it labor intensive to validate these systems?

Calvin: It can go both ways, you can automate it with a systems like blue scores. Or you a lot of times we do like to do it internally to see where we stack up against the industry, and how we compare? But it's always nice to feed the computer back with supervised learning, so that it can do more unsupervised learning over time. But yeah it depends. That's why you pick things that you like to translate about. So we pick TED Talks or anything that's of intellectual value out or just of interest.

Rodeny: Like my Podcasts!

Calvin: Yeah!

Rodney: You know right now my podcasts are an audio podcast. Occasionally, I'll do a video. But I've seen other folks that do podcasts that even though there's no visual interest, they put it up on YouTube. So, in fact, I was thinking doing that just as another platform people could find my material. So wouldn't it be neat to have my audio podcast on YouTube with with the captions running across the bottom in a language that you could pick is that something that's a doable for somebody with my means.

Calvin: Absolutely, actually let's take your use case Rodney. Being a professional podcaster what the advantage of using our software as services is: right now we're captioning and transcribing, this podcasts actually, you can then upload this transcript to your podcast, so that it has greater SEO or SMO, so that you can look for keywords or anything of the like, word verbatim text. And then afterwards, if you do choose to upload to YouTube, you can find a greater performance than just using the default captioning that they have built in. Let's just be honest Alphabet spends a lot of their time doing a lot of different things. So, I will say Google is a great innovator, but at core, they are in terms of their revenue dollars, they are an ad tech company more so than anything else nowadays.

Rodney: Yeah, for those of you obviously since you are listening on audio, you don't know what I'm looking. Why don't you explain what your screen is showing here?

Calvin: Sure, so we have a transcript at the top of our software, which has been recording the whole time. Then we have another transcript for German. But then as an overlay, we have augmented reality, there is a transparent or translucent text box that can be overlayed for PowerPoints or PDFs in both languages at the same time. I see it happening at the bottom of your screen now.

Rodney: So that's pretty impressive! And the latency is very quick. It's almost like hearing an echo, when I'm reading what I'm just saying, it's kind of weird, but that's pretty impressive.

Calvin: That is lots of years of research and development. And we're finally glad that we can use this to help people because of our core mission. We are a social mission driven company, so for every one contract we sign, we donate a system to a primary or secondary school of your choice in your honor, so that we can help accelerate students learning. And allow them to learn how they learn faster. At the end of the day, this is actually very basic. It's almost like the primary colors of learning: visual; auditory; and kinesthetic learners. It is just giving more routes for them to obtain the content. That's why I've been such a big proponent of Universal Design for Learning, because the way we deliver the content is also as important as how the students can express it and be tested on it. So it is very important to keep all users in an inclusive environment. Given that you know they always say, we're the melting pot in the U.S., and this year in the Open Doors report statistics show that we had increased international students. There are over a million this year of out-of-country students. And as we have more and more students overseas ourselves and internally in the U.S., we have to keep in mind all those users all the ESL learners, and then also even anyone with any type of different birth circumstances that they can't choose, the different health circumstances, that they still have equal access to equity and learning. And it's not so much the access to the information. It's about making it useful for them, so that they can turn it into wisdom. And apply it where it needs to be applied.

Rodney: Sure, I mean I can certainly see how this could be a wonderful adjunct to the way we teach. We have a lot of foreign students; English was not their first language. I'm sure it makes it very hard for them. They love to have lecture capture. We use two different systems on campus for lecture capture, and you can see how especially before the exams it gets a lot of traffic. And they watch these lectures. You know we are trying to get away from lectures, like many schools and flipping classroom. But still there is a lot of lecture material going on, and to have this translated into their native language would certainly be a no brainer. Are you working yet with some the major lecture capture providers or do they have their own way doing it?

Calvin: So for the MOOcs like edX and Coursera, they've been using a volunteer service. So because I want to learn let's say costs accounting, I would watch the video and transcribe it in Chinese or caption in Chinese. It's all done by a volunteer basis. But doesn't go to say that we are not willing to go help them with more tool sets, so that the students can do it in a more efficient manner. So our post production tool sets are going to expand out, so that we can help more and more providers who offer free education. We have some world class education online. But again, some people may have access to it, and some may not in different ways.

Rodney: Now the way, it's being deployed, the speaker, the lecturer, or the presenter would be connected using your software at the podium. How long is it going to be for a student, who can just hold up his phone, smartphone, and just capture the audio that way and see the translated text? Audio quality is certainly I imagine it's an issue?

Calvin: So it's doable. Given everything is very speaker or keynote speaker oriented, so it makes the audio visual team facilitate that process so much more when it is done on the presenter side. If a student does transcribe it, it depends on how clean the audio is; the better it is, the more accurate it'll clearly be. So it's something that we want to put in the hands of the students, and we're slowly working on. But there has to be some sort of a good capturing device, so like an audio input device.

Rodney: Right right. The wheels are spinning here, and thinking about a student watching a recorded video in their lecture capture. Why couldn't they input headphone jack out from their computer, where they are watching it, into the microphone input of their smartphone, and have that translated on-the-fly? Seems like that could be doable.

Calvin: Yeah, we also do a lot of video captioning. So we actually have a loop back feature, where a student watching a lecture, they could run it through and they could see the captions live or we can do it on the post production side, and have it cleaned and edited. But it is something that anyone with our software could actually do. You can watch, any video you want and it can be done on a live translation and captioning style.

Rodney: What would you say is the accuracy these days of on-the-fly type without being cleaned up in someway?

Calvin: So speech recognition has always been ahead of the curve in terms of the industry standards. It should be upper 90s, depending on the speaker, the acoustic settings. High 90's like 97s we've seen or 99s. And then translation has always been a the one catching up so to speak. It is hard to not just translate the words, but the meaning behind it and the reasoning behind it. We've seen 80s is about the standard or the average. That is what a lot of researchers have been developing over time. We're very keen on making it more accurate, whether it be statistical machine translation or rule based machine or neural network. Any type of hybrid is what we play around with, so that it can be a more useful for all parties.

Rodney: Great! Well listen, I want to be respectful of your time. And I have a challenge for you. Maybe you can make this happen? It can be a first for me as well. Maybe this podcast? You could arrange to have it translated, actually you have already translated into German, haven't you? So maybe we could make this a public and do what I was talking about. And put this on YouTube with the simultaneous translation. Maybe you can help me make that happen?

Calvin: Challenge accepted! Okay, very good Calvin thanks for stopping by, appreciate it. Danke Schoen!