EPISODE 1025

[INTRODUCTION]

[0:00:00.3] JM: Descript is a software product for editing podcasts and video. Descript is a deceptively powerful tool and its software architecture includes novel usage of transcription APIs, text-to-speech, speech-to-text and other domain-specific machine learning applications.

Some of the most popular podcasts and YouTube channels use Descript as their editing tool, because it provides a set of features that is not found in other editing tools, such as Adobe Premiere, or a digital audio workstation. Descript is an example of the downstream impact of machine learning tools becoming more accessible. Even though Descript only has a small team of machine learning engineers, these engineers are extremely productive due to the combination of APIs, cloud computing and frameworks like Tensorflow.

Descript was founded by Andrew Mason, who also founded Groupon and Detour. Andrew joins the show to describe the technology behind Descript, as well as the story for how it was built. It's a remarkable story, actually. There's some creative entrepreneurship. There's numerous takeaways for both engineers and business founders. Honestly, my favorite part is the fact that Groupon, which is Andrew's first extremely successful company was born out of a product that was not working. The same thing happened with Descript. Andrew had been working on a company called Detour. For many years, he was trying to make it work and eventually through having his back against the wall with the product not working as well and as popularly as he had planned, he found adjacent problems. That turned into Descript, which is an amazing product.

I really like this story as an example of how innovation actually works in practice, because it can be very, very messy.

[SPONSOR MESSAGE]

[0:02:10.3] JM: DigitalOcean makes infrastructure simple. I continue to use DigitalOcean, because of the low friction and attention to user experience. DigitalOcean has kept the experience simple and I can spin up a server in less than a minute and get high-quality performance for a low price.

For an application that needs to scale, DigitalOcean has CPU-optimized droplets, memory-optimized droplets, managed databases, managed Kubernetes and many more products. DigitalOcean has the flexibility to choose the right instance for the right workload and you can mix and match different configurations of CPU and RAM.

If you get stuck, DigitalOcean has thousands of high-quality tutorials, responsive Q&A forums and a customer team who treats customers respectfully. DigitalOcean lets developers focus on what they are building.

Visit do.co/sedaily and receive $100 in credit over 60 days. That $100 can be put towards hosting, or infrastructure and that includes managed databases, a managed Kubernetes service and more. If you want to get started with Kubernetes, DigitalOcean is a great place to go. You can use your $100 to start building your distributed system and you can get that $100 in credit for free at do.co/sedaily.

Thank you to DigitalOcean for being a sponsor of Software Engineering Daily.

[INTERVIEW]

[0:03:46.8] JM: Andrew Mason, welcome to Software Engineering Daily.

[0:03:48.7] AM: Thanks for having me. Glad to be here.

[0:03:50.5] JM: It's 2020. We've had podcasts for more than 15 years. Describe the state of podcast tooling.

[0:03:58.2] AM: Well, you have to go a little bit further back, because the truth is that podcast tooling is mostly just music production tooling. Back in the 90s, you started to see tools like Pro Tools taking non-linear audio editing off of tape machines and onto the computer. Nothing has changed too much since then. I mean, everything's gotten better. There's many different options. If you're making a podcast, you're using timeline-based editors with waveforms as an abstraction, such as Pro Tools, or Adobe Audition, or maybe GarageBand, or Audacity. There's a high learning curve, powerful flexible tools, but not really within reach of the normal person.

The way Descript came along is that we were working on a audio tour platform called Detour and we were basically making these long-form podcasts, and just saw what a tedious workflow it was for these audio producers. This was at about the same time that speech-to-text was reaching an inflection point, where it was actually accurate enough to be usable. We thought, wouldn't it be cool if you could just automatically transcribe audio and then build a audio editing tool, or a podcast editing tool that was really designed with narrative-based content in mind and not just trying to repurpose these tools that were designed with music in mind and allow people to edit audio by editing text, the same way they would in a word processor? That's what we're doing with Descript.

[0:05:41.8] JM: I can think of a few main types of podcasts. You have the two-person interview format, which we're engaged in right now. You also have the this American lifestyle podcast with some emotional narrative that is going concurrently with music. What are the other important podcaster use cases that you want to accommodate in building a podcast editing tool?

[0:06:09.8] AM: Well, the main thing that we are trying to accommodate is the editing process. If you're just doing something that's totally unscripted, where you're just letting the tape roll for two hours and you don't feel the need to edit any of that, you want it to come out exactly as it is, then the tools that are already out there do that thing pretty well. I mean, you can just use tools that ship on your computer for that.

If you want to craft your content in any way, if you want the ability to edit and make things sound better, that's the space that we're playing in. Like you said, that applies to both unscripted content, like this conversation that we're having, as well as scripted content, like at this American Life episode, or so all the above.

[0:07:01.3] JM: As you said, Descript can take audio files and then transcribe them and then allow you to edit them as a text document. This requires at least two technologies that I can think of to machine learning technologies. You've got speech-to-text and you also have the audio text alignment. How good is the quality of those algorithms today?

[0:07:26.0] AM: Very good. If we were to transcribe this conversation that we're having right now, there'd be very few errors. The word error rate might be 2% or definitely less than 5%. In fact, you would probably be more distracted by punctuation errors, than the actual word accuracy.

As the quality of the recording starts to diverge from this broadcast quality, or people with heavy accents, then the quality can still – word error rate can still be as high as 30%. Increasingly, for just your general podcast, the transcription is incredibly accurate.

The forced alignment process that we use is doing a phonetic mapping of the text to the audio and giving us timestamps at the beginning and end of every phoneme. That's incredibly accurate as well; accurate enough that further improving the accuracy of those word boundaries hasn't really been a major area of focus for us.

[0:08:38.3] JM: You do also have this facet of the system that if I want a transcription that has a higher degree of reliability, I can key it off to a white glove human editor. What are the circumstances where I would want to do that? What are the circumstances where a human editor is going to catch things, or edit things that are more fine-grained than what the automatic transcription can?

[0:09:05.1] AM: I think there's two categories. One is like I said when you have a strong accent, or when the audio quality is lower. If you just have an iPhone that you have on a table in a crowded restaurant and you're recording a conversation, the transcription accuracy gets low enough that somebody's going to need to go through and clean it up for it to really be useful.

The other situation is just like you work at an organization that has a budget for human transcription. Even 2% error rate is something that you don't need to bother with, so you'd rather have it transcribed. We see both of those.

[0:09:43.7] JM: Descript is a desktop application. Can you give me the brief overview of the stack of technologies that go into it?

[0:09:51.6] AM: Sure. Back in maybe 2015 or so when we started working on Descript and we were incubating it inside of Detour, this audio tour company that I mentioned, it was a native Mac application and written first in Objective-C and then in Swift. Then once we spun it out to be a standalone company and launched, surprise-surprise, some of the first feedback we got was from users requesting a Windows version of the app.

This is I think 2018. In that brief period of time, it felt like the landscape had really changed and we were on the cusp of the point where we're building this in a full web technology stack was viable. We've obviously seen this transition happen to different categories of tools, starting with CRM and documents and spreadsheets and now design tools, like Figma.

We were very cautious about doing it for an audio/video production tool, because of the requirements of processor, intensive audio/video editing, the size of the files that we're dealing with. There were just a lot of things about it that made us nervous. At that point, we looked at what was out there and we felt the benefits of using a web technology stack outweighed the costs.

Descript is built in React. The desktop application runs in Electron. Then we have parts of our media engine that are native and it runs offline. Just have being able to run it in Electron has allowed us to make this possible in a way that we couldn't have in purely a browser, at least not yet.

[0:11:44.2] JM: There are some elements of the machine learning stack that can be in the cloud. There are some elements that can be, or that need to be local. What are the models, or the machine learning systems that you need to keep locally and what are the ones that are fine to be in the cloud?

[0:12:02.6] AM: Well, we don't have our own transcription engine. That just seems like the thing that there's a bunch of really big companies with really smart people where it's a core strategic focus to be good at speech-to-text. Rather than try to compete with them as a small startup, we would just continuously survey the field of options out there and use whatever we thought was the most accurate and the best experience for our customers.

That as a result is all in the cloud. We have our own alignment process and we do that both in the cloud and locally, depending. In addition to the initial alignment when we've transcribed a file, whenever you're correcting a transcript, will actually re-align the seconds around that correction to make sure that the audio and text remains aligned, and it's important for us to have that work offline, so that as much of the experience can work offline as possible. We do that locally. A lot of the new overdub text-to-speech stuff that we're doing is in the cloud.

[0:13:14.3] JM: Are there any particularly annoying bugs that come to mind that you've solved, that might illustrate the difficulties of working with audio?

[0:13:26.4] AM: I think part of it is we have a design paradigm that the document that encourages people to use the tool creatively and push its limits. As soon as we launched, we saw people creating these documents that were – or that are timelines that are five or 10 hours long with incredibly large assets and thousands of edits and crossfades and so on. We just got pushed to its limits very quickly and it took us some time to just catch up with the way in which people were using it and make sure that that all worked with audio and video. There's a long list of things like that.

[0:14:12.4] JM: With such an open-ended application, what's your process for testing it?

[0:14:17.7] AM: We do a lot of internal testing and then we have some manual QA. We have some automated QA as well, but we haven't found that to be quite as useful. It's just the automated QA that we do, because the manual QA because it is so open-ended and often our customers are using it in ways that are difficult to replicate with automatic test cases. We also have a pretty active beta user community who get early access to early versions of the app and are really good about logging reproducible bug reports for us.

[SPONSOR MESSAGE]

[INTERVIEW CONTINUED]

[0:16:03.0] JM: In 2018, you acquired Lyrebird which has this technology that allows you to train an audio model and create new audio. I remember when Lyrebird first launched with their demos and it was scary, but also exciting; is incredibly useful potentially, but it's also potentially dangerous. Why did Lyrebird decide to sell their product, instead of trying to productize it themselves?

[0:16:31.4] AM: The Lyrebird team is a team of AI researchers that worked in Yoshua Bengio’s lab up in University of Montreal, who are just incredibly smart PhD researchers and are doing, I think the best work I've seen of anyone out there in making easy to reproduce voice model of yourself. We had had our eyes on that space really since the beginning of Descript. The stuff that we know how to do is build product and product engineering, build a business. We don't know anything really about the deep learning stuff that those guys were doing.

When we started talking to each other, we realized that we had a similar vision and a skill set that really complemented each other, and so we decided to join forces. It’s really been, of all the acquisitions I've done in my career, perhaps the most hand-in-glove fit between two teams.

[0:17:32.9] JM: Describe how the Lyrebird technology is used in Descript.

[0:17:38.3] AM: Increasingly, it'll just be scattered throughout. We have a few things that Lyrebird is built that are in production now, like AI speaker detection, that's just automatically labeling speakers in a mono file that you transcribe. The big thing that we'll be launching is overdub and that'll be launching soon. It's in beta right now. That'll let you upload about 10 minutes of your voice. From that, we can give you a text-to-speech model that you can use to generate audio.

[0:18:10.6] JM: The text-to-speech application, this is one of these ideas that could be potentially dangerous, but I know you have a limitation on it, where you can only use it on your own voice. How do you enforce that?

[0:18:26.0] AM: It's pretty simple. Lyrebird used a similar mechanism when they were an independent company and successfully trained 300,000 or 400,000 voices without any cases of fraud. All you do is you give people a script that they have to read and then you validate the transcript of what they read against the original script. Unless, you can fool someone into reading a 10-minute script, you can't really fake it.

[0:18:56.8] JM: Describe some of the other subtle uses of machine learning in Descript.

[0:19:02.0] AM: One of the other ones that we're rolling out now is disfluency detection already in the app, or 'um' and 'uh' detection. If you say 'um' or 'uh', we have a one-click button you can use to zap all of those. We're doing something similar now that'll catch contextual filler words like, 'like' or 'you know'. It'll also catch stutters and false starts, so that you can easily clean up your audio.

We're looking to offer more tools around that, around mixing and leveling your audio, noise reduction and room tone detection. There's really an endless list of the types of things that we could we be adding that's right for tools that are enabled by machine learning.

[0:19:55.2] JM: Something like AI powered compressor or EQ. People have subjective opinions on what a compressor, or an EQ should be on a podcast. If you listen to Joe Rogan, it's going to be mixed a little bit differently than NPR. Is that important? Is it hard to find the right granularity of which to give people? Is it something like the Instagram filters, where you just want to give people some same defaults that make things sound nice that they can sample?

[0:20:31.1] AM: Yeah. The cool thing about Descript, I mean, we have compressor effects and EQ effects just built into the app. When we think about applying machine learning, it's not a destructive process. It's really just tweaking parameters that are in the app. Then while it's true that there's some subjectivity and taste involved in how someone might mix something, I think that's true for – the vast majority of people just want something that sounds professional and good and the subtleties of different compressors are really of interest to them. If we can do something that gets it to a general place that things sound good and then the user can continue to tweak from there, we think that's a pretty good option.

[0:21:19.0] JM: How does a domain of video editing compare to that of audio editing?
[0:21:25.0] AM: That's an interesting question. My experience is mostly in audio. I went to school in music technology. I worked in a recording studio for a couple years. Then at Detour, we were working very closely with radio producers and got a first-hand view at their workflow and what worked well about it and what didn't.

Like most people, I've done some video editing. Before joining Descript, I really had to get myself up to speed with the workflows and the way people work. I think when you look at the video editing tools and the audio editing tools, there's a lot of similarities, but there are some differences. Some of those differences are for example, the way that you set the in-and-out point for a range selection, appear to largely be path dependency and just convention that one tool started one way in another – a different way and everybody got used to it that's in that field.

Others are legitimately different. For video editors, tend to apply effects on the clip level primarily, while audio editors tend to apply them on the track level. That's for legitimate use case reasons and the nature of what kinds of effects you're putting on video versus audio. We've been looking at that and working through the differences and tried to come up with something that works for both.

I will say that I think the primary dividing point between for media editing tools should be whether it's narrative or music. That to me seems a more natural division than audio or video. The fact that they're divided that way is more just circumstances of history. If you were building something from scratch today, I think you would probably divide things more along the lines of music versus narrative.

[0:23:27.8] JM: Given that you studied music production, music technology and now you're working on audio editing tool, I did a show with Splice and one with a company that was similar to Splice, these online music collaboration platforms. I've always found it curious that music, there has not really been a github-like experience for music. There has not been mass collaboration on music. Do you have any perspective for why that is?

[0:24:01.4] AM: Having people tried to do it, I thought Splice and Gobbler –

[0:24:04.3] JM: Exactly. Exactly. They tried to do it. Splice tried to do it and they basically pivoted to being a cloud sample library, which they're having a lot of success with. People just didn't want to collaborate.

[0:24:15.7] AM: I mean, I don't know if that's the problem, or the problem is that it was I never tried to use it, so I'm seriously just speculating here. The other way to think of it is that it was bolted on non-first-class feature for the software platforms. I mean, just to take an example, like if you look at the way Adobe has added collaborative tooling to their suite of products, it feels like illustrator. If you compare the way it works with illustrator to the way that it works, like a tool, like Figma has treated collaboration, Figma has reimagined everything from the ground up with collaboration, collaborative workflows as a first-class citizen. It's worked wonders for them and really given them a major superpower.

I think a lot of these music editing tools, whether it's Pro Tools, or Ableton Live, or Logic, or any of them, have not made the decision to rethink what they're doing as a collaborative tool. All of these timeline editors are so feature-rich, or they've all been around for 10 or 20 years just adding feature upon feature upon feature. If you're making music, if you're in a creative workflow, you want to have access to that stuff a lot of the time. The idea of switching to a startup that's building a collaborative timeline editor, it's a big trade-off, right? It's going to take a long time before somebody's starting from scratch on a music production tool can reach the minimum feature set that is going to allow someone to say, "Yeah, I'll use this instead of Ableton Live and I don't feel in any way restricted in terms of what I can do creatively."

[0:26:11.6] JM: Collaboration is a first-class citizen in Descript. What are the implications of that? I think about the complexities of Google Docs. You have two people editing concurrently. You can have these kinds of merge conflicts that occur when somebody goes offline for a little bit and somebody else has edited the same piece of text that they have been editing offline. You have a merged conflict. Do those kinds of issues emerge in the collaboration experience with Descript?

[0:26:41.1] AM: Yeah, they sure do. There's some types of those conflicts that there's not really a good single solution to and we'll just let the user know that it's happened and save both in their version history and make it easy for them to pull it back. We've had to work through that same set of issues that the other collaborative editing tools – live collaborative editing tools have had to work through.

[0:27:01.9] JM: Is it just enough of an edge case that it very rarely happens and causes an annoyance?

[0:27:08.6] AM: Yeah, I think that's right. It's a near-term annoyance. The main thing for us is making sure there's never a situation with data loss. Because we have full version history and the way everything – and because Descript works offline, we're saving everything to your hard drive before we even try to push it to the cloud, so the chances of anything happening there are very slim.

[0:27:31.9] JM: When I think about the business of podcasting, we have five shows per week, we have four ads per show and I record them all as host-read ads, which is what most of the market wants these days. You want the host to read the ads, so it's influencer marketing. One application that people talk about sometimes is like, oh, if you wanted to get dynamically inserted ads where you could have this real-time bidding, or some advertiser bids on let's say I'm a listener, I've suddenly tuned in to an episode about JavaScript. An advertiser bids on that listen in real-time with a JavaScript error monitoring tool ad, but I have not recorded that ad. You could have this real-time process where all of a sudden, the advertiser has purchased that ad and they dynamically create an ad-based off of a voice model that has been trained from my voice. That's a potential application. I don't know if it's something that people will want, but it seems like something that's it's almost an inevitability. Do you have any perspective on that potential application?

[0:28:42.7] AM: Not really. I agree with you on that it will be possible. I agree with you that I'm not sure if it's something people will want.

[0:28:51.5] JM: Right. Do you think you as a consumer, thinking like a consumer who has probably consumed a lot of podcasts, would you know the difference?

[0:29:00.3] AM: I think we'll reach a point where people will not know a difference. If you just assume that people won't know a difference and then think about what a world looks like where this is happening and whether it changes the value of these host-read podcasts, I mean, you can't look at that in isolation of all the other ways in which synthetic voice will be appearing in culture and how that will change people's perception of voice. There's just too many variables in play to even be able to speculate for me.

[0:29:37.1] JM: Yeah, because in that – I guess in that world, you can have a ghost – entirely ghost-written podcast episode. Somebody could totally script an entire episode of me talking to somebody else and maybe that's something people want. We just have no idea at this point. We're too far from that.

[0:29:53.2] AM: I mean, there's probably visionary people out there that will answer your question more affirmatively, but I certainly don't feel I have any idea.

[0:30:03.0] JM: In terms of synthetic voice technology though, how good is it today? If I tried to write an entire ghost-written podcast episode, would it sound like me?

[0:30:15.2] AM: The use case we've been focused on for now is pickups and editorial corrections, in part just because we think it's more interesting; something that's augmenting organic audio, rather than being a complete substitute for it. If we can allow you to correct a couple of words in something that you've recorded and give you the same editorial flexibility that you have when you edit text, that seems like an incredibly useful thing.

It's also I think the longer a string of text-to-speech is, the more likely you are going to notice a glitch in the matrix. There are use cases for longer form text-to-speech in its current form. It is useful for stuff and I think we'll start seeing that once we come out of beta. I don't think in the next year, you're going to be listening to audiobooks of your favorite celebrity's synthetic voice, or something like that.

[0:31:19.3] JM: What's the hardest engineering problem you've encountered building Descript so far?

[0:31:24.4] AM: I would have to defer to the engineers on that one. I'm thinking about trying to answer and then thinking about my engineers listening to be trying to answer that and then just wanting to punch me. I'm just going to –

[0:31:39.6] JM: Let's take a different angle. I've listened to a lot of podcast interviews with you. You're one of these people who I think is you serve a pretty valuable touch point of inspiration, in the same way that these other founders who have beaten their head against the wall with product that might not have a perfect market, like your whole thing with The Point and then that transitioning into Groupon and then with Detour and that transitioning into Descript. I think there is probably some nugget of wisdom, or nuggets of wisdom that you can offer about how to evolve an idea that is not working perfectly into something that is more honed and has a better market. Do you have any lessons that you can share with the audience?

[0:32:26.3] AM: In both cases, I've been through two sets of companies that had a pivot in them. The first was The Point, which was this collective action platform that pivoted into Groupon. Groupon was a very narrow application of the broader technology, or the broader idea, which was this idea of a collective action tipping point get a critical mass of people to agree to do something and only once it's achieved does it go into action. Groupon is that applied to group purchasing. In the early days, Groupon was you would only get the deal if 20 people joined or something like that.

That was a situation where our backs were up against the wall at risk of losing our funding and we had to come up with something and we were just frantically experimenting with all the different use cases that we had imagined. With Detour, it was this audio tour platform. From the moment we started thinking about Descript, I was immediately thinking of as the potential for a separate business and we treated it as such internally, where it had its own team. Honestly, it's like, I probably wouldn't recommend doing something like that in a normal startup, especially if it's your first startup. It's this whole other thing to manage.

I don't know. It's this weird combination of being stubbornly persistent about sticking with your idea, but also keeping your ears open and looking for opportunities, being open to the idea that the thing is not the thing that you thought it was. It's this other thing over here. It's really a wonderful part of the process is working through the wreckage of your idea to find some gem that's in there. It's often said that the obvious ideas are taken, so you have to get yourself into a little bit of a mess and look at what your assets are, to have those constraints to figure something out.

The one thing I've done consistently is just recklessly jumped into things and got myself into a mess that I have to work my way out of. If you're thinking about doing a startup, just do a startup and get started and don't overthink it. Just be open to the idea that maybe things will come out differently than you expect it, although not badly.

[SPONSOR MESSAGE]

[0:35:08.3] JM: When I'm building a new product, G2i is the company that I call on to help me find a developer who can build the first version of my product. G2i is a hiring platform run by engineers that matches you with React, React Native, GraphQL and mobile engineers who you can trust.

Whether you are a new company building your first product like me, or an established company that wants additional engineering help, G2i has the talent that you need to accomplish your goals. Go to softwareengineeringdaily.com/g2i to learn more about what G2i has to offer.

We've also done several shows with the people who run G2i, Gabe Greenberg and the rest of his team. These are engineers who know about the React ecosystem, about the mobile ecosystem, about GraphQL, React Native. They know their stuff and they run a great organization.

In my personal experience, G2i has linked me up with experienced engineers that can fit my budget and the G2i staff are friendly and easy to work with. They know how product development works. They can help you find the perfect engineer for your stack and you can go to softwareengineeringdaily.com/g2i to learn more about G2i. Thank you to G2i for being a great supporter of Software Engineering Daily, both as listeners and also as people who have contributed code that have helped me out in my projects.

If you want to get some additional help for your engineering projects, go to softwareengineeringdaily.com/g2i.

[INTERVIEW CONTINUED]

[0:36:57.3] JM: That process of incubating Descript, were you thinking of it like a hedge? Were you thinking of it like a Hail Mary? Were you thinking of it like some combination of the two? It sounds there's almost like you were setting up the company with a sense of cognitive dissonance, but you had a sense that that was the only way to actually do this.

[0:37:19.6] AM: It was a little bit of a hedge. It was a little bit of well, we can do both these things and it was a little bit of just self-indulgent, I'm personally really interested in this, both because I'm interested in audio and media production, but also because I just love building tools. Yeah, I think it was all those things.

[0:37:40.4] JM: I've worked in a recording studio. Whenever I see those recording studios, they have all this gear inside of them. Are you a believer that we need analog gear? Or do you think everything can be turned into software in the audio production process?

[0:37:57.8] AM: The recording studio I worked at was Electrical Audio Studios in Chicago, which is a famously analog recording studio. Everything at the time was done on 2-inch tape. It depends on what you're doing, I guess. I really appreciate the constraints that that puts on the creative process. I'm a firm believer in all aspects of things that making things easier doesn't necessarily make them better. That's really about one's creative process.

Beyond that, in terms of the sonic qualities of it, the reason I never became a recording engineer and didn't stick with that is because I just have terrible taste in that thing. I wouldn't be a good person to ask. I think people spend way too much time fixated on gear as a way of procrastinating on their actual craft.

[0:38:58.6] JM: As a musician, is there a part of you that wants to quit this whole business thing and just write music all the time?

[0:39:04.6] AM: Not at all. I mean, I'm an amateur musician. I'm not very good. I love to play music, but I don't really have any desire to write music.

[0:39:15.0] JM: How big is the podcast tooling market?

[0:39:17.8] AM: We'll find out. There's a lot of podcasts out there and it's growing quickly. It's big enough for us and a bunch of other people and a bunch of investors to be excited about it. We think of ourselves as part of a larger new media space than podcasting specifically. In the new media world, content tends to blur the boundaries between podcasting or audio and text. In video, you're distributing across all three for any given piece of content. When you think of it that way, when you think of the size of video, then it really starts to get quite exciting.

[0:39:57.4] JM: Is there a way to make podcasts more social?

[0:40:01.8] AM: I don't even know what you mean.

[0:40:03.9] JM: There is this common meme around the fact that podcasts can't be shared. Maybe you just don't – you don't even explore the world of podcast marketing at all. You're just more in the podcast tooling space. People say that there's a problem being able to share a podcast is the idea, because they're all in this fractured RSS ecosystem and there's not a one canonical way to listen to a podcast, like a YouTube video.

[0:40:28.7] AM: I'm sure that could be better. Yeah, I don't have that – I don't have anything to say about it.

[0:40:33.5] JM: Okay. Fair enough. Do you have a sense for why the podcast advertising market is so underdeveloped?

[0:40:39.9] AM: No.

[0:40:40.9] JM: It's very hard to get podcasters to change their workflows. I watched the Descript commercial that you have and you have this old guy who looks somebody with old, dusty podcast recording tools. How do you get somebody like that to change their workflow?

[0:40:58.5] AM: I think it's hard to get people to switch from Pro Tools to Audition, or one timeline editor to another timeline editor, because all of these tools are so mature, they've all aped each other's features. At this point if you talk to someone and you ask them what they're using, if they're using Pro Tools, or Logic, or Audition, it's almost definitely because the reason they're using it is because it's what they learned on, because all the tools are equally good. Descript is really the first editing tool since editing tools made the jump to the computer, that's a complete reimagining of the interaction paradigm, where you're editing as a document, instead of editing purely as a timeline. It's one of the changes that is happening, because the technology exists for it to happen. When YouTube came around, it wasn't just that it was a brilliant idea, it's that technology only at that moment had reached a point where such a platform was even achievable.

Words are just a better abstraction of narrative audio, narrative media than waveforms. They're more expressive, they're faster and easier to work with. Once people start using it, it's just such an obvious improvement that in terms of speed and simplicity. The fact that you can remain in your editorial brain and not need to continuously switch back into your technical or engineering brain as you're making these waveform edits to your timeline, that those are the reasons that we tend to see people switch.

On top of that, podcasting is going through a golden age. Most people who are podcasting now are not that crusty old audio engineer. They're print journalists, or just whoever that don't know timeline editors from Adam, right? They're new to it and they might be outsourcing their editing, if they're doing any editing at all because of the technical complexity and learning curve. Descript is now making editing possible for them when it just simply wasn't within reach before.

[0:43:23.7] JM: For people who have not worked in a digital audio workstation, or who have not edited a podcast, can you hone in on that distinction between the document-based editing style, versus the timeline-based editing style? What are you describing there?

[0:43:41.5] AM: Yeah. When I say a timeline style, you're imagining like GarageBand, or iMovie, where you have on the Y-axis you have tracks of your different media and the X-axis is time. You're just seeing, your tracks, or different speakers, or music, or effects when it gets into video titles, or B-roll that you might be overlaying on top of your main video track.

Then the document is just exactly what it sounds like. When I say, in Descript you're editing a script, you're editing something that looks like you're working in Google Docs, but when you're editing the words, you're also editing the underlying media.

[0:44:18.5] JM: When you're thinking about the market size for podcast editing, or video editing, is there a specific company, or set of products that you benchmark against to know, or even just to reverse engineer how much you need to charge for this when you're trying to calculate, like what is the market? How much do I need to charge in order to have a good business here? Are there analogs? I'm just wondering how you're thinking about the market sizing.

[0:44:47.7] AM: There's a little bit of that. There's a lot of looking at the zeitgeist and just seeing that we're in a world where there's really no difference anymore between the consumer and the creator and the prosumer. Everyone is a content creator. The tools that are out there are difficult to use. If you believe that you can build something that can be a solution for the creators out there, then it's a process of combining existing markets, creating ones that don't exist into something new and that's happening all the time with startups that are creating new spaces.

[0:45:31.9] JM: Are there any machine learning applications that you really wish you could have in Descript, but the models are just not quite good enough yet?

[0:45:43.1] AM: Don't know yet. There's a backlog of things that we’re excited to start exploring, but haven't yet and we'll see when we get there.

[0:45:51.7] JM: Are there any particular business adjacencies that you're considering exploring, like hosting, or the recording process itself?

[0:46:00.7] AM: Right now, we're focused on building a great editing tool. That's really the main thing.

[0:46:06.4] JM: As we get to the end of the conversation, could you just give me a description for how the management of the company is structured? How do you organize the product teams the testing process and just the overall management of the company?

[0:46:19.5] AM: It's still a small team. We're maybe 22, 23 people. I have a VP of engineering and the engineering and AI research teams report in to him. Still don't have any other product managers other than me, although that'll change soon. We have a couple of designers who report in to me. The team here is senior by comparison to some of the teams I've worked on in the past. The engineering team has high consumer judgment and high business judgment and that's why we haven't had the need for dedicated product managers.

So far, we just have a lot of very creative, high-judgment engineers who are able to work directly with design to take a fairly high-level goal and work things out. I think the team here has really enjoyed that. A lot of the team has been together for going on five years since the Detour days. It's all the same team. At this point, we're growing. We're hiring now after we did our Series A. It has a strong core that's really figured out how to work well together.

[0:47:35.2] JM: If you aren't building Descript, what company would you be working on?

[0:47:39.2] AM: I don't know that I would be doing a company. I have a couple random side projects that I'd love to work on at some point. Yeah, I'm not sure.

[0:47:51.4] JM: Can you share anything? I just want to know what the next reckless decision is going to be.

[0:47:55.6] AM: Oh, it's too self-indulgent. I'll share at some other time.

[0:48:00.5] JM: Fair enough. Andrew, thanks for come on the show. It's been great talking to you.

[0:48:03.2] AM: My pleasure. Thanks for having me.

[END OF INTERVIEW]

[0:48:14.2] JM: As a company grows, the software infrastructure becomes a large, complex distributed system. Without standardized applications or security policies, it can become difficult to oversee all the vulnerabilities that might exist across all of your physical machines, virtual machines, containers and cloud services.

ExtraHop is a cloud-native security company that detects threats across your hybrid infrastructure. ExtraHop has vulnerability detection running up and down your networking stack from L2 to L7. It helps you spot, investigate and respond to anomalous behavior using more than 100 machine learning models.

At extrahop.com/cloud, you can learn about how ExtraHop delivers cloud-native network detection and response. ExtraHop will help you find misconfigurations and blind spots in your infrastructure and stay in compliance.

Understand your identity and access management payloads to look for credential harvesting and brute-force attacks and automate the security settings of your cloud provider integrations. Visit extrahop.com/cloud to find out how ExtraHop can help you secure your enterprise.

Thank you to ExtraHop for being a sponsor of Software Engineering Daily. If you want to check out ExtraHop and support the show, go to extrahop.com/cloud.

[END]