**Interview with
Thad McIlroy**

**For podcast release
Monday, January 17, 2022**

KENNEALLY: High-level text-to-speech technology has made it possible for Siri and Alexa to answer questions and follow commands for a decade. Now, the latest in TTS may mean TTFN – ta-ta for now – for traditionally produced audiobooks.

Welcome to Copyright Clearance Center's podcast series. I'm Christopher Kenneally with Velocity of Content. Audiobooks are an increasingly important piece of the revenue pie for publishers. In 2020, audiobook sales topped $1.3 billion, 12% jump over 2019. With few exceptions, human narrators – authors themselves as well as actors and other artists – are heard in such recordings. AI-enabled automated audiobook creation, however, lies just beyond the horizon, says electronic publishing analyst Thad McIlroy. The shift from analog to digital voices, he reports, would lower production costs and lead to greater choice in titles, as well as mean a lot less work for Hollywood actors between pictures. What stands in the way of that vision, though, is a contractual requirement of the leading audiobook platform.

Thad McIlroy joins me now from his Vancouver office. Welcome back to the program, Thad.

McILROY: Thank you, Chris. Great to be back with you.

KENNEALLY: Well, it's a fascinating subject that you have highlighted for us in a recent article for *Publishers Weekly*. And I want to go over with you the main points, because it really points to an interesting direction for the industry. We have to start, I think, by telling people a little bit more about TTS, text-to-speech, which in recent years has moved away from clunky voice bots to voices that are nearly indistinguishable from a human's.

McILROY: Yeah, it's been fantastic. We're brought up on Siri, which always sounded suboptimal, let's say. And now, when you listen to some of the samples – I think we'll get some links for people up by the end of the podcast – hearing is believing, and some of the latest voices are indistinguishable from human – maybe not quite that. That's for anyone to judge. I can be fooled by some of them.

KENNEALLY: It is, I think, certainly approaching the point where it will be hard for most to tell the difference. And it would seem that with this advancement in the technology, we

are getting to the point where we can move beyond the use cases that we're familiar with to others that ultimately could include book-length content.

McILROY: Yes, that's the quest at this point. We can do a short snippet no problem. And even when you listen to some of the samples, you think, OK, that was a 20-second clip. How is that going to sound over an hour's length? That's perhaps where we see the separation of what's not quite there. We're waiting just to be able to sit through two hours, four hours, whatever the length of a longer book, and still be delighted by the sound of that voice.

KENNEALLY: And we're probably going to get there, because high-quality TTS is, as you wrote, a holy grail for Google especially, but for all the other big tech players – Amazon, Apple, Facebook, IBM, even Microsoft. So it seems to me with those names behind it that we're going to get there soon.

McILROY: I'm with you on that. Google – we see that they can do whatever they want. One of the experts that I spoke to said Amazon's got so much skin in the game already, and Google has got the deep expertise and are doing so much work in this area that even without some of the smaller players, when you consider what Amazon and Google are up to, you have to figure it's just about there.

KENNEALLY: And we have to think as well that this has gotten the attention of publishers, because while producing audiobooks has been an important piece of sales for them in recent years, probably helping to save their bacon, as it were, over a number of years recently, it's still true that producing audiobooks is a big investment. It's an expensive project. So the appeal that automating audiobooks would have is certainly easy to understand.

McILROY: It is. I've had a couple of experiences on behalf of clients of producing audiobooks, and I was surprised at how much effort it was. It sounds so effortless when you listen to the final product. But, gosh, we went into the studio, in and out of the studio, and what took the most time was the correction process. There just were the inevitable errors. You'll hear it in a podcast like this. There's stumbles and stops and starts and those kind of things that have to be corrected in the final recording, and that takes a long, long time. It's worth mentioning as a sidebar to that, of course, those kind of errors creep in in automated audiobook production as well, which remains one of the flies in the ointment.

KENNEALLY: Right. You know, that's why we appreciate our producer, Jeremy Brieske. He helps this all sound so smooth at the end of it. But in between time is when I can't quite read my script or sort of start to stutter and stumble. And it is, as you say, expensive and important, because people's expectations are pretty high. They've become accustomed to

a high level of production. And the costs there are pretty significant. You looked into it to come up with some numbers there.

McILROY: Well, the big cost up front is the level of talent that you hire. Amazon has this Audible Audiobook Creation Exchange, it's called, its ACX, where you can get, let's say, semiprofessional talent, and get that for $100, $200 an hour. But when you start thinking about the level of a Hollywood actor – I put a base figure of about $1,000 there, and I think I'm really lowballing in terms of some of the talent that's being used these days by the big publishers.

KENNEALLY: And of course, for publishers who are looking to save some money and are watching the bottom line, it sounds harsh, but getting rid of the talent could mean a big change in that bottom line.

McILROY: Yes, it could. And it's funny – when the article was published, that's the one big criticism that people were throwing at me in the article and saying, Thad, you're recommending that we do this – that we get rid of the talent. That wasn't a very close reading on their part. I'm trying to make a distinction, where for certain books, it is simply not economical to bring in talent at that kind of price level and to go into the studio and to have the kind of production values that are in a podcast like this. So the choice becomes an either/or. It's a binary choice. Are we going to get an audiobook for this backlist title or not? Because if we have to go through traditional production methods, it's simply not going to be financially possible.

KENNEALLY: That's the part that addresses the point of greater choice. You're right. If we can use this technology, we'll just have so many more books to choose from. But what kind of books do you think are best when it comes to possibly automated narration?

McILROY: That's been one of the controversies. I was surprised that some of the vendors that I spoke to said that they could do fiction. You think of the difference between nonfiction and fiction, where nonfiction, you're trying to get a steady cadence in the voice. There's some emphasis and change of emphasis thereof. But compared to a stage play, let's call it, for a fiction title, that takes a lot of nuance in the voices. As one of the vendors was saying, just in a short piece, you're going to have to have anger, delight, laughter, passion, and that's just within one book. So that's a big, big challenge for the voices. So from my point of view, for publishers listening to this, I would say stick with nonfiction at this point. You're more likely to get a successful result.

KENNEALLY: Well, we get some of that even in this podcast, Thad. We have some delight, some passion. Never anger, at least not so far. (laughter) So I understand just what you

mean about the modulation in tone – the modulation in emotional tone that is so important. So it's much easier to capture with nonfiction.

And it's remarkable enough that this goes beyond historical narrative. I understand that last fall, Google worked on a scientific article.

McILROY: Yeah, that was something to see. They claimed to be creating the enabling technology, and it took some looking around to find where they're actually putting it to use. And indeed, I found an instance where they used multiple technologies on a scientific journal article. But what they first were able to do was convert the text into digital text, and then they were able to narrate that digital text. And the quality was surprisingly good.

KENNEALLY: There is, though, this matter of a contractual obligation that a submitted audiobook be narrated by a human. At least that's the obligation for ACX and its audio submission requirements – the Audible platform that you were telling us about. That's a significant hurdle.

McILROY: It is. It's a big hurdle, and all the vendors recognize it. The way I think of it is that hurdle will go away as soon as Amazon makes a larger commitment itself into using this automated technology. Ultimately, every expert that I spoke to felt that Amazon's going to come along and endorse it 12 months, 24 months from now. But in the meantime, yeah, the vendors are stuck using alternate distribution channels.

KENNEALLY: Well, they're stuck on Audible, but not so much on many of the other platforms. Apple Books, Google Play, and others do allow for AI-generated voices right now.

McILROY: Yeah, they do. I don't have figures on market share, but my sense is that Audible controls, let's say, about 50% of the market, so you can still get substantial sales through the other players, through the other distribution. Particularly, you can get into the library and educational markets, and those are substantial.

KENNEALLY: And it really comes down to hearing is believing in this particular case. So you've done some listening. You can share with us your own sense of just how close we are to being comfortable with this technology and perhaps highlight for us a couple that really stand out for you.

McILROY: Well, the best examples are going to Google's AI voices and Amazon's AI voices so you can hear from the master players. But some of the vendors have customized those voices and done some custom voices, where they're using actual clones of human actors

and creating a speech simulation, let's say. Those – there's DeepZen. There's Speechki. If you go to those two sites, you'll hear some interesting samples.

Speechki took my article from *Publishers Weekly* and converted that into an automated voice. It created a little (laughter) – my own mini-audiobook from the article. So I got to hear something that I'd written converted into this automated voice. On the one hand, it was amazing. On the other hand, there were errors throughout, and it left me aware that this is not yet an automated process. We've still got a ways to go.

KENNEALLY: Well, this is not an automated process here, our podcast interviews. We can assure the listeners that there were two human beings involved, and we're very grateful that one of them, Thad McIlroy, joined us today for the program to tell us about the coming AI revolution in audiobooks. Thad McIlroy, thanks so much for joining me.

McILROY: Thank you, Chris.

KENNEALLY: Our producer is Jeremy Brieske of Burst Marketing. You can subscribe to the program wherever you go for podcasts, and please do follow us on Twitter and Facebook.

I'm Christopher Kenneally. Thanks for listening. Join us again soon for another Velocity of Content podcast from CCC.

END OF FILE