



Interview with Bill Wolfsthal

**For podcast release
Monday, March 21, 2022**

KENNEALLY: Audiobooks may add millions to their bottom lines each year, but publishers have yet to realize the digital category's full revenue potential. So what keeps publishers from growing the audiobook market? If I said Amazon, would you be surprised?

Welcome to Velocity of Content. I'm Christopher Kenneally for CCC. Audiobooks are booming. The US-based Audio Publishers Association reported audiobook revenue rose in 2020 to \$1.3 billion, a 12% hike, and the ninth straight year of double-digit growth for the format. Yet only a fraction of published titles are commercially available as audiobooks. While some books cannot reasonably become audiobooks – heavily illustrated children's storybooks as well as art books and photography books – most fiction and many nonfiction titles likely can.

Bill Wolfsthal, who advises book publishers and publishing technology providers on strategies to grow sales and revenue, says the opportunity presented by machine-generated narration is too good to pass up. He joins me now from his New York City office. Welcome to the program, Bill.

WOLFSTHAL: Thanks for inviting me on.

KENNEALLY: We're glad to have you. This is a part of a response, if you will, to a program we did in January. At that time, Thad McIlroy told me about the several factors holding back an AI-driven explosion in audiobooks. The shift from analog to digital voices, Thad noted, would lower production costs and lead to greater choice in titles. What stands in the way of the machines, though, is a contractual requirement of the leading self-publishing audiobook platform, ACX, which is part of Amazon's Audible service. According to submission requirements, your submitted audiobook must be narrated by a human. So, Bill Wolfsthal, is Amazon the chief reason why more books aren't available in audio at the moment?

WOLFSTHAL: As much as I'd like to blame all of the world's curses on Amazon, it's just not true. It is disappointing that they won't allow audiobooks recorded with synthetic narration on their platform, but I think they're just failing to keep up with the times. Two or three years ago, if you listened to synthetic narration, it sounded like Siri or Stephen Hawking – very robotic. But in the last two or three years, great improvements have been made, and listening to an audiobook with synthetic narration is really an enjoyable



experience. I believe that Audible and Amazon are going to change that policy once they realize that it was put in place to protect their consumers, and their consumers don't want protection. They want choice.

KENNEALLY: Interesting point. Thad McIlroy I think would agree with you, as he also expects that Amazon will eventually come around to supporting this option, the synthetic narration. So what at the moment are the other obstacles to the growth of usage of synthetic narration in audiobook production?

WOLFSTHAL: Like everything else in the world, the major obstacle is money. If you are a book publisher, you are working on a P&L, a profit-and-loss statement. Some publishers have actual spreadsheets for each book that they plan ahead of time on what they plan to spend and what they hope to make, and they make choices based on that. Others just do it on the back of an envelope, trying to figure out how best to publish.

But the question is if you're going to spend money to create an audiobook, will you get a return on that investment? And up until now, even though you could create audiobooks for \$2,000 or \$3,000, most good, high-quality audiobooks with a first-rate narrator cost between \$5,000 and \$10,000 to create. If you're spending that kind of money as a publisher, you have to sell enough books to get that money back. And up until now, it's been impossible. So creating books with synthetic narration for much less – for \$500 or \$1,000 – means for publishers, they're taking much less of a risk and have a greater chance of getting their money back.

KENNEALLY: You've been telling us, Bill Wolfsthal, that the synthetic narration now sounds much better, much more human. So I need to ask you about how this all works. You consult with Speechki, an audiobook recording platform that uses synthetic voice. Can you tell us how Speechki creates audiobooks?

WOLFSTHAL: Yes. So Speechki has multiple ways of creating their audiobooks, but it all starts with a voice. We can license a voice from someone like a Google or an Amazon and then use that to create an audiobook, or we can create our own voices, which requires a narrator being in a recording booth for 10 to 50 hours recording certain words, phrases, and sentences that can then be put together to create any audiobook. But whether you start with a licensed voice or a voice that you're creating from scratch, what's happened in the last few years is artificial intelligence. Artificial intelligence can make decisions that make artificial voices more human-sounding. Even though you have a sentence with three words in it, the artificial intelligence can tell the three words how to sound within that sentence, and it can be much more human-sounding than the very stilted narration or voicing you get from a Siri or Alexa, which is most people's experience with synthetic narration.



KENNEALLY: So it's about the pauses. It's about the breaths, if you will, that we all take when we speak. It's the way we give emphasis to words – all of those things, AI can sense in a text and then have the voice replicate.

WOLFSTHAL: That's correct. What Speechki does is it takes a manuscript, a PDF, or a Word document or an EPUB, puts it through our software, and the AI catches most of the corrections that need to be made to make it sound good. We then give it to a human proof listener who catches the rest, and we have a very powerful editing platform that allows you to change the emphasis if it's wrong. For example, the word REcord and reCORD is spelled exactly the same, and it all depends on context. Many times, our artificial intelligence gets it right in the proper context, but once in a while, it'll get it wrong. And all you have to do is make a click in our software to correct it. When there's a human narrator, you have to put that human narrator back in a booth to re-record that sentence or paragraph. So not only is the creation process shortened and simplified and less expensive, but so is the editing and correction process.

KENNEALLY: I understand that in the Speechki case, there are something like 300 distinct synthetic voices for audiobook publishing across some 77 dialects and languages. So do they all go by names? Is one called Bill? Is another called Chris? And do you have any personal favorites?

WOLFSTHAL: Each one gets assigned a name, and they're appropriate for the language they're in. It's just an easier way to keep track of them. Of course, you have to have male and female voices, because if you have a woman author, you want it recorded by a female-sounding voice. You want the voice to sound as much like the originator, the writer, as possible, so that it has the opportunity to be not just a pleasant, but an authentic listening experience.

KENNEALLY: We're talking about the listening experience, but let's talk about the business experience behind audiobooks, because they certainly have become an important part of the mix of revenue for many trade publishers. But university presses and other kinds of publishers have different goals, and text-to-speech technology, like Speechki's and others, could make academic and scholarly books available in audio. That would mean a real opportunity, because it would make books accessible to students and scholars that might have some visual impairments and learning disabilities.

WOLFSTHAL: Right now, our best estimate is that only between 5% and 10% of books in print are available in audio. That's a disappointment both because it could be new revenue for publishers who need it and in terms of accessibility, because there are so many book lovers, so many readers, so many students that have a visual impairment and can't read, or



have a learning disability like dyslexia that makes reading difficult for them. Audiobooks are hugely popular in that audience.

As books have moved from print and ebook into audio, books have also found new audiences. There are people who listen to podcasts who never go into a bookstore or buy a hardcover book anymore. They could be audiobook listeners even though they're not book buyers. That's a chance for publishers to expand the marketplace of what they do.

KENNEALLY: Bill, it's no surprise, but the reporting on AI narration for audiobooks has created a bit of a stir, because it's always included responses from the narrators, the actors who do this work. They've expressed concerns about their livelihood. They think they might be ridden out of town by these machines. Should they be worried?

WOLFSTHAL: I've been in therapy long enough to know not to tell anybody not to worry. People are going to worry whether they should be worried or not. My perspective is that audiobook fans who want to hear a great experience – a novel read by a first-rate narrator – are not going to change and suddenly want to hear synthetic narration. But there are millions and millions of books published where it's not a choice between human narration and synthetic narration. It's a choice between synthetic narration and having no book available at all in audio.

So in my world, in my hope, every narrator that's now working should have four times as much work next year as they have this year, because there should be more audiobooks, and there'd still be millions and millions of books that should be recorded with synthetic narration, so that instead of only 5% or 10% of books published being available in audio, it's 90% or 95%.

KENNEALLY: All right. Bill Wolfsthal, put Speechki and the boom in audiobooks into the context of business development for publishers. Based on your experience, are audiobooks with synthetic narration the next mass-market paperback – in other words, a format and a technology that put publishers in the black and kept them there for years?

WOLFSTHAL: It's a really good question and one that speaks to the heart of what I've done my entire career. I started in a part of publishing called special sales, which was the part of the sales department that sold books to people who weren't bookstores. So I put fishing and hunting books into sporting goods stores. I put books on how to use woodworking tools into hardware stores. I helped put gift books – cookbooks, books of humor – into gift shops, so that publishers had different revenue streams than simply counting on their bookstores for their money to come in. To me, that's what's happened over the last 10 years, first with ebooks. Ebooks were a brand-new, out of left field revenue source for publishers, and it's one that everybody was worried about, and in the end has only



expanded the amount of revenue that publishers can reach. The same is true with audiobooks.

Book publishers have learned over the last 10 years because of ebooks that they're not just book publishers. They own content. And the question is, what can you do with this content to monetize it? It can be a first-rate audiobook with human narration that should be up for awards and win awards, and it can be an academic or scholarly or nonfiction book with synthetic narration that simply gives people an option on how they're going to get this information.

A large number of people that listen to nonfiction audiobooks don't listen to it at regular speed. They speed up the audio to 1.2 or 1.5 speed, so they're getting the content as quickly as possible. If they're interested in cryptocurrency, theoretically they can listen to four books on cryptocurrency a day by just speeding it up and going through it as quickly as possible. So having an audiobook experience is not simply I'm listening to a novel and having it performed well. It's also access to information.

KENNEALLY: Bill Wolfsthal, at any speed, it's been a pleasure speaking with you. Thanks for joining me today.

WOLFSTHAL: Thanks for having me on.

KENNEALLY: That's all for now. Our producer is Jeremy Brieske of Burst Marketing. I'm Christopher Kenneally for Velocity of Content from CCC.

END OF FILE