

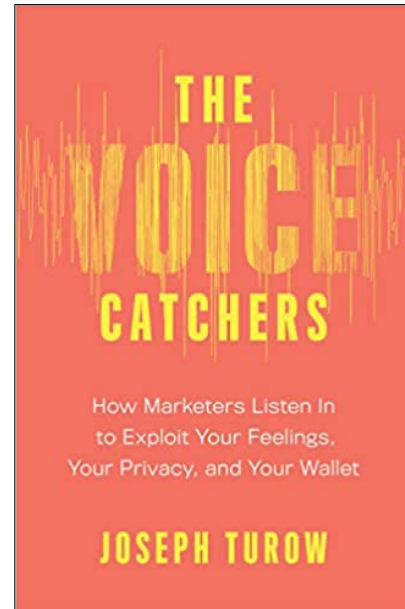
Joseph Turow, **The Voice Catchers: How Marketers Listen In to Exploit Your Feelings, Your Privacy, and Your Wallet**, New Haven, CT: Yale University Press, 2021, 344 pp., \$16.70 (hardcover).

Reviewed by
Edward B. Kang
University of Southern California

Your voice is unique. No one else has it. And because your voice belongs to no one else; it's extraordinarily valuable, not only to you, but also to a new sector of society that is designed to exploit it: the voice intelligence industry. (p. 1)

In ***The Voice Catchers: How Marketers Listen In to Exploit Your Feelings, Your Privacy, and Your Wallet***, author Joseph Turow presents an extensive overview of what he calls the *voice intelligence industry*—a group of technology companies that develop voice-related technologies such as voice assistants like Apple's Siri and Amazon's Alexa, voice-controlled smart home devices, voice analysis and emotion prediction technologies, and voice biometric authentication/identification systems. Acknowledging from the onset that this is a still-developing industry in what insiders call the "scale-building period" (p. 10), Turow employs a range of creative methods to navigate the difficulties of examining such a nascent business, one of which includes his telling of a fictional narrative based on scenarios described in various patent documents granted to "voice intelligence" companies (p. 101). In fact, much of the book is primarily concerned with *imagining potential future* dangers that the "voice intelligence industry" might introduce to society. Considering the undeniably increasing interest in voice-based technologies, this book is both timely and effective in expressing an urgent warning toward a rapidly developing "voice intelligence" industry, as well as in providing relatively concrete policy suggestions that can veer the industry in a direction that can mitigate potentially negative consequences (pp. 263–267).

The first five chapters are primarily concerned with documenting some of the most relevant corners of the "voice intelligence" industry—voice assistants, smart homes, and voice analysis and profiling. Drawing from a variety of sources that range from patent documents, press releases, company websites, and personal interviews conducted by the author himself, the chapters examine how the principal players use and conceive of voice-based technologies. In this sense, the direct conversations with industry insiders are perhaps one of the most salient strengths of the book. By speaking directly with leading figures and practitioners, Turow is particularly successful in painting for readers a seemingly more "authentic" picture of how "voice intelligence" systems are currently being conceived and developed in tandem with existing technologies, especially in the realm of marketing. There is an impressive coterie of industry insiders, ranging from company CEOs, project managers, researchers, and patent attorneys, brought together by Turow, who nuance this conversation. One patent attorney named Tim Cowan, for instance, provides invaluable insight with regard to identifying the multiple and often conflicting agendas that are baked into a patent application (p. 100), while several marketers



Turow interviewed express that "voice" might be a way for "data scientists to use artificial intelligence to achieve unprecedented and near-immediate insights into shoppers' identities and inclinations" (p. 23). This book is thus undoubtedly informative for gaining a broad understanding of the inside operations of the "voice intelligence" industry, but this extended focus on the *business* of "voice intelligence" also often deters opportunities to engage with some of the more fundamental questions that underlie these voice-based systems.

A lingering question throughout the text, for instance, is how Turow himself is understanding and defining "voice," and, as an extension, *if* and *how* Turow is challenging the "voice intelligence" industry's limited conception of "voice." To provide some context, numerous scholars have grappled with the ontological complexity of voice, understanding it primarily as a vacillating sociocultural phenomenon as opposed to a fixed biological or sound object (e.g., Dolar, 2006; Eidsheim, 2015, 2018; Stoever, 2016; Weidman, 2015). For example, Eidsheim (2018) has provocatively and convincingly argued that the source of voice is the listener and not the speaker, emphasizing that the meanings inferred from voice are generated at the point of reception and not at the point of projection. Similarly, Weidman (2015) points to the voice as not only a sonic and material phenomenon "embedded in social relations that shape how voices are produced, felt, and heard" (p. 241), but also as a powerful *metaphor* for identity and agency. How would these understandings of voice complicate the "voice intelligence" industry as it is described by Turow, considering that it primarily positions voice as a minable site of sonic—"tones, speed, emphases, pauses, and much more" (p. 25)—or physiological (pp. 78–79) data that can be analyzed to make deductions about the speakers? Engaging with these more fundamental questions around voice can lead to opportunities for rich critical analysis regarding the "voice intelligence" industry's dependence on these more limiting and modular conceptions of voice. Although Turow's mobilization of industry hype around the analytic potentials of voice might prove to be an effective way to warn readers of the future dangers that development in "voice intelligence" might engender in society, one is also led to wonder how realistic some of these dangers are. In fact, according to one of the patent attorneys interviewed by Turow,

Sometimes [companies] will file a patent on a device [they] are never going to make, just so no one else can . . . Pushing it through to a manufacturing floor takes ten more years. So you can have your patent, but a product that is going to hit an end-user is nowhere in sight. (p. 99)

Taking this into consideration, it would have been illuminating if the book expanded its focus toward not only expressing alarm with what "voice intelligence" *can* do, but also with critically interrogating what "voice intelligence" realistically *cannot*. As Leonardo Cardoso (2019) points out in his analysis of a different acoustic identification technology called ShotSpotter, designed specifically to identify and locate gunshots, the "stunning agenda," of novel sonic technologies described in patents, press releases, and websites such as those examined by Turow (p. 81), are often not so spectacular and sometimes even completely unusable when deployed in practice, most often due to the highly variable nature of sound. While the author does occasionally point to the questionable efficacy of "voice intelligence" systems, such as when he documents his interview with computer scientist Rita Singh and her hesitation toward drawing conclusions around emotion and personality based on voice analysis (p. 82), his overarching position tends to focus on the anxieties around "voice intelligence," which unfortunately also comes at the expense of readily *accepting* the technologies behind these anxieties as probable realities. Not only that, but the anxieties that Turow expresses around "voice intelligence" are also often tied to a simple "surveillance is bad" (for lack of a clearer phrase) argument. While surveillance is certainly not "not

bad," it would be more effective if additional context was provided as to *how* the added surveillance from "voice intelligence" exacerbates existing inequalities or introduces new forms of mass monitoring and control to society. Put differently, what is different, and more specifically, what is differently *bad* about the kind of surveillance that "voice intelligence" introduces?

Although I have certainly pointed out opportunities for Turow to have engaged in deeper critical analysis, I believe this is in part a reflection of the nascency of the field, both in the tech world, as well as in academia. As one of the first books on the "voice intelligence" industry written by a social scientist/humanist, *The Voice Catchers* proves to be an impressive and well-researched text that will provide a worthy introduction to the field, especially as it relates to marketing technologies. It is written accessibly and not embedded in a trove of academic references, which will deem it suitable not only for undergraduate- and graduate-level classes, but also for general readers who are interested in learning more about the world of voice-controlled technologies.

References

Cardoso, L. (2019). Translations and translation gaps: The gunshot acoustic surveillance experiment in Brazil. *Sound Studies*, 5(1), 52–71. <https://doi.org/10.1080/20551940.2018.1564495>

Dolar, M. (2006). *A voice and nothing more*. Cambridge, MA: MIT press.

Eidsheim, N. S. (2015). Race and the aesthetics of vocal timbre. In O. Bloechl, M. Lowe, & J. Kallberg (Eds.), *Rethinking difference in music scholarship* (pp. 338–365). Cambridge, MA: Cambridge University Press.

Eidsheim, N. S. (2018). *The race of sound: Listening, timbre, and vocality in African American music*. Durham, NC: Duke University Press.

Stoever, J. L. (2016). *The sonic color line: Race and the cultural politics of listening* (Vol. 17). New York: New York University Press.

Weidman, A. (2015). Voice. In D. Novak & M. Sakakeeny (Eds.), *Keywords in sound* (pp. 232–246). Durham, NC: Duke University Press.