Is Machine Listening Listening?

Today, one can find references to machine listening across the human sciences, engineering, computer science, music, and other fields. One also finds it increasingly in the press. Generally, the term refers to the process of converting sound into data and then subjecting those data to a machine learning process. Applications include everything from digital assistants (Siri, Alexa, Cortana), music identification (Shazam), speech transcription (Otter), automatic signal processing (Izotope, Landr), and music recommendation (Spotify) to more esoteric hopes and fantasies for the technology: speaker identification/voiceprinting, music composition, voice synthesis for people who do not speak, voice synthesis for deception (so-called deepfakes). At the extreme end of these fantasies is a sort of aural phrenology: classifications of speakers along lines of mood, personality, body type, race, gender, sexual preference, truthfulness, health status, and any number of other vectors. Is this actually possible now? Absolutely not. Will some promoter say it is possible? Absolutely. Deception about the technology is part of the technology.[1]

In this context, what does it mean "to listen" and to say that machines listen? In a recent essay, Domenic Napolitano and Renato Grieco attempt to answer this question by arguing that machine listening is not the same thing as a human-centered definition of listening. They conclude "that the distinction between human listening and machine listening becomes blurred, and the co-determination of the two emerges."[2] Drawing on the media archaeological tradition, notably Wolfgang Ernst's concept of implicit sonicity,[3] they argue for moving beyond an anthropocentric model of listening to understand machine listening. Specifically they claim that machine listening systems operate according to their own logics, which are not the same as the logics of human hearing. Their argument is worth entertaining for a theory of sound, especially if it is placed alongside Indigenous and ecologically-based challenges to anthropocentric approaches to listening.

While Napolitano and Grieco attempt a machinic definition of machine listening, there is another political question to be asked: what do the researchers who build machine listening systems think they do? What do the corporations and states who deploy them think they are doing? Do their users treat their listening machines as listeners?

---

[1] Burrell, "How the Machine 'Thinks'"; Gray and Suri, *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*; Natale, *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*.

[2] Napolitano and Grieco, "The Folded Space of Machine Listening," 186.

[3] Ernst, *Sonic Time Machines*.

With Mehak Sawhney and Andy Stuhl, I have begun researching this question. There are at least three major subfields in machine learning that deal with machine listening. Music Information Retrieval deals with extracting data from music and is used in music recommendation (such as Spotify, though most recommendation systems do not primarily use audio data to make their suggestions), music identification (such as Shazam); and music production (automatic arranging, mastering, composition, orchestration, editing). Auditory Event and Scene Analysis deals with identifying sounds, such as remote monitoring of industrial equipment for malfunctions, or triangulating sounds—most famously in the case of "shot spotter," which aims to triangulate the location of gunshots in American cities. Natural Language Processing deals with speech recognition, for instance in digital assistants, voice identification, and voice analysis. To our surprise, although these fields share some basic assumptions about listening—drawing from a core of concepts in psychoacoustics and information theory—they are not broadly in conversation with one another. Further, there is no general agreement within those fields on *what listening is* and *whether machine learning systems listen*. For instance, Music Information Retrieval converts sound recordings into multidimensional representations—the closest human analog would be a spectrogram. Is extracting features from a spectrogram qualitatively related to listening? MIR researchers don't agree—some call it listening, some do not. Most of these debates seem to be grounded in the very anthropocentric definition of listening that Napolitano and Grieco hope to supersede.

Even if there was consensus in computer science, we would also have to re-examine the operative definitions of listening in communication studies. Our field also lacks a consensus around the term. Some writers cling to romantic definitions of listening as openness to alterity; others connect it to a process of self-assertion and connectedness to the world; and still others cling to an ableist concept of listening that ties it to biological understandings of hearing.[4]

My own approach has been to understand listening as a set of techniques that can be developed, repeated, and transformed over time.[5] These techniques of listening have no necessary political valence: they can be used in contexts of care or community,

---

[4] James, "Affective Resonance"; Moten, *In the Break: The Aesthetics of the Black Radical Tradition*; Robinson, *Hungry Listening: Resonant Theory for Indigenous Sound Studies*; Brooks, *Liner Notes for the Revolution: The Intellectual Life of Black Feminist Sound*; Friedner and Helmreich, "Sound Studies Meets Deaf Studies"; Friedner, Michele Ilana, *Becoming Normal: Cochlear Implants and Sensory Infrastructures in India*.

[5] Sterne, *The Audible Past: Cultural Origins of Sound Reproduction*; Sterne, *MP3: The Meaning of a Format*; Sterne, *Diminished Faculties: A Political Phenomenology of Impairment*.

or they can be techniques of extraction or subjugation. They can be attached to human bodies or delegated to machinery (or machinery can delegate back to bodies). In this framework, machine listening extends and permutates processes that are already in play in culture: it is not a "break" with human listening, both because human listening is already technological, and because it is not a radical departure from prior modes of engaging with sound.

We should therefore study machine listening as one more scene in which the form, function, and meaning of listening is debated, and where listening weaves yet another tangle of humans, technologies, and institutions. In our moment, machine listening is intimately connected to corporate attempts to enclose more and more domains of human interaction, and to state surveillance and authoritarian projects in many parts of the world. Therefore, any theory of the *listening* in machine listening needs to also be a theory of power. As with other forms of machine learning, the most basic use cases in the illustrations above—recommending music, locating gunshots, executing verbal commands—are laden with cultural values, structural inequalities, and institutional agendas. Just as there is no non-technological form of listening, machine listening never "just" listens.

Bibliography

Brooks, Daphne. *Liner Notes for the Revolution: The Intellectual Life of Black Feminist Sound*. Cambridge, MA: Harvard University Press, 2021.

Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms , How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3, no. 1 (June 1, 2016): 1–12.

Ernst, Wolfgang. *Sonic Time Machines*. Amsterdam: University of Amsterdam Press, 2016. https://www.aup.nl/en/book/9789089649492/sonic-time-machines.

Friedner, Michele, and Stefan Helmreich. "Sound Studies Meets Deaf Studies." *The Senses and Society* 7, no. 1 (March 2012): 72–86. https://doi.org/10.2752/174589312X13173255802120.

Friedner, Michele Ilana. *Becoming Normal: Cochlear Implants and Sensory Infrastructures in India*. Minneapolis: University of Minnesota Press, 2022.

Gray, Mary, and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*. New York: Houghton Mifflin, 2019.

James, Robin. "Affective Resonance: On the Uses and Abuses of Music In and For Philosophy." *PhaenEx* 7, no. 2 (December 16, 2012): 59–95. https://doi.org/10.22329/p.v7i2.3814.

Moten, Fred. *In the Break: The Aesthetics of the Black Radical Tradition*. Minneapolis: University of Minnesota Press, 2003.

Napolitano, Domenico, and Renato Grieco. "The Folded Space of Machine Listening." *Sound Effects: An Interdisciplinary Journal of Sound and Sound Experience* 10, no. 1 (2021): 174–89.

Natale, Simone. *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*. New York: Oxford University Press, 2021.

Robinson, Dylan. *Hungry Listening: Resonant Theory for Indigenous Sound Studies*. Minneapolis: University of Minnesota Press, 2020.

Sterne, Jonathan. *Diminished Faculties: A Political Phenomenology of Impairment*. Durham: Duke University Press, 2021.

———. *MP3: The Meaning of a Format*. Durham: Duke University Press, 2012.

———. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham: Duke University Press, 2003.