ECHO

ORPHEUS INSTITUUT

a journal of music, thought and technology

Title: Synthesizing the Signal from the Noise: Reflections on Mary Had a Little Lamb (2019)

Subtitle:

Author(s): Kazuhiro Jo, Nobuhiro Masuda

Issue: #7 Speculative Synthesis

Publication date:

Review status:

Article DOI:

Intro

Since Edison's invention of the phonograph, countless debates have revolved around the problems posed by acoustic reproduction technologies. Yet when compared to recent AI-based models of sound synthesis, these inquiries—premised on sender–receiver transmission models—risk reaching an impasse. To address this, the paper turns to Gilbert Simondon's theory of communication to examine its applicability through the author's artistic work *Mary Had a Little Lamb* (2019). By foregrounding the multiple processes through which sound is synthesized and rendered sensible, this study seeks to clarify both the limitations and the possibilities inherent in the concepts of noise and signal.

1. Backgrounds

In a 1906 essay titled "The Menace of Mechanical Music," the American composer John Philip Sousa —known for "The Stars and Stripes Forever"—famously described the sound from the then-proliferating phonograph, one of the earliest recording/reproduction technologies, as "canned music" (Sousa, 1906). Here, the analogy of canned music to recordings is employed to contrast the act of listening to recordings, despite the availability of live performances, akin to consuming canned fish instead of freshly caught fish. This comparison is deliberated against the backdrop of Sousa's concerns regarding the potential abandonment of active music-making (performing/catching) by individuals facilitated by the mechanized production of canned music, leading them to assume the role of passive consumers (listening/consuming) (Lessig 2008).

About a century later, Japanese composer Masahiro Miwa, advocate of "reverse simulation music," argued that the "music" perceived through the playback of recorded materials should be discussed—or even denied—as something fundamentally different from music in its traditional sense (Miwa, 2007). To distinguish works experienced through speakers or headphones, he coined the term *rokugaku* ("recorded music") and proposed that this be distinguished from the Western musical tradition grounded in the performer's body and the musical score (Miwa, 2008). While this concept may seem extreme, it breaks with the understanding that a "work" receives the "soul" (the performer's interpretation) anew each time via the medium of music sheets, redirecting attention toward expressions specific to technologies of reproduction.

Jonathan Sterne, a pioneer of sound studies, also draws on early debates on recording and reproduction technologies to argue that "face-to-face" or "live" acoustic events are social practices that are fundamentally different from the acoustic production that accompanies technical reproduction (Sterne 2003) and criticizes the distinction between the original and its copy. According to Sterne, the idea of the original as an object of reproduction is impossible without the reproduction process, and

recording and reproduction mask the differences between technologies, each with its specific characteristics. Following this line, one might say that the value of freshly caught fish is obscured without canned fish, and the value of live performance cannot be established without *rokugaku*.

Despite differences in historical context and aims, these three positions share a communication model in which music (sound) is transmitted from a sender (performer or technology) to a receiver (listener). Framed through Shannon's notions of "signal/noise" (Shannon 1948), their arguments can be summarized as follows.

For Sousa, the understanding that canning diminishes the freshness of caught fish treats the difference between listening to live performance and to recorded media as "noise." Miwa's proposal to see live performance and recording as distinct forms of expression might seem to escape the signal/noise opposition at first glance. Yet ontologically divorcing the two raises the question of how to position the various noises involved in the recording/reproduction technology and process in relation to the recording as signal. Sterne's assertion that the concept of the original cannot exist without the process of reproduction can be read to suggest that the authenticity of the signal (or, originality) is clarified only through the noises that accompany reproduction.

In these frameworks, what happens if we add sounds synthesized by AI, which has made remarkable progress in recent years? Beginning with WaveNet (Oord et al., 2016), AI techniques that synthesize audio waveforms directly through deep learning developed into systems such as Jukebox (Dhariwal et al., 2020), and—by October 2025—prominent examples include Suno (2023) and Udio (2024). These suggest a future in which virtually any sound (music) can be synthesized insofar as it can be represented as digital data inside the computer.

At first glance (or hearing), AI-generated sounds that can be appreciated through loudspeakers or headphones without the intervention of the human body seem to be the ultimate form of canned music, or *rokugaku*, lacking the human soul (interpretation by the performer). One might also take the human-provided prompt test as the "work," with AI acting as a performer, but this seems merely an attempt to forcibly preserve the Western tradition of the music. The fact that different sounds (music) are "synthesized" each time in accordance with a text (prompt) is qualitatively different from "reproduction" via recording and may be better understood as a purification of only the required signals from among innumerable bits of information. As different sounds and interpretations are produced without bodily performance, they verge on "copies without an original" or "signals without noise."

The foregoing suggests that AI-based sound synthesis marks the critical point of the sender–receiver transmission model assumed by reproduction theories exemplified by Sousa, Miwa, and Sterne. The technological lineage of sound (music) synthesis and its current culmination call for a different model of communication that cannot be fully described by the Shannon–Weaver model. The next section turns to Gilbert Simondon, whose works have been reappraised in recent media studies and the philosophy of technology, to argue that his perspective, especially developed around 1970 and later collected under the title *Communication et Information* (2015), offers effective insights into this problem.

2. Simondon's Theory of Communication

For Simondon, "communication" possesses a scope that markedly departs from the term's ordinary meaning. Communication cannot be restricted to exchanges that occur between living beings (Simondon, 2015, p. 70). His discussion extends beyond signaling and transmission in the sense of language and message, to other animals and microorganisms, including microscopic organizations such as intracellular processes and DNA, and at times even to the inorganic. Yet this is not a mere anthropomorphism of physical or chemical reactions: as interactions of action/reaction recognized

within atoms or cells are situated as a primordial stage of communication for organisms and machines responding to environments and stimuli.

Simondon organizes communication into an evolutionist model of three stages. Simplified, at the most primordial (1) biological level, one finds stimulus–response for the organism, including DNA-level translation/replication, protein composition, and enzymatic reactions. At the (2) ethological level, cries of animals through to human language mediate struggle, mating, and parent–child affection via signals grounded in instinctual motivations. Finally, at the (3) psychological level, significations are related within individual memory and cognition, certain knowledge is accumulated, and collective, social, and cultural systems are formed (Simondon, 2015).

Why did Simondon develop such a distinctive theory of communication at this particular time? One primary reason, I suggest, was to have been his rebuttal and opposition to the Shannon-Weaver communication model, which had been formulated somewhat ahead of his time.

Rooted in a biological philosophy of technology, Simondon's argument acknowledges communication not only in organs, living beings, and ecosystems but also in elements, individuals, and ensembles of machines. While this resonates with cybernetics to some degree, what is emphasized in his argument is the *gnosie* in French, or perceptual discrimination, of unicellular organisms. "Communication cannot take place without gnosis in addition to information, and gnosis, at its most primarly level, is not far away from "good" or "bad" depending on the tendencies, needs, and motivations of living beings." (Simondon, 2015, 75, hereafter trans. by author). Simondon further positions this gnosis, or primarily perceptions, as a condition for what Jakob von Uexküll set as the basis for the *Umwelt*: the "functional loop," i.e., the organism's perceptual–behavioral circuitry. As with the tick that approaches mammals by detecting butyric acid even without vision to suck their blood, the "functional loop" indicates the organism's basic perceptual–behavioral mechanism toward its milieu; its principle, as is often noted, foreshadowed the "feedback" mechanism, somewhat preceding cybernetics in the 1930s (Uexküll, 2010; Pasquinelli, 2016, 2017).

From this perspective, Simondon's following remark becomes salient:

"Gnosis introduces a possibility of error in communication that exists at the most elementary level (here it's not a matter of 'noise in the channels', nor 'receiver noise' nor 'transmitter noise'), and which increases with the complexity of the work to be accomplished for gnosis between the reception of information and the action or reaction. (Simondon, 2015, p. 76)"

Interestingly, Simondon adopts Uexküll's stimulus-response model (feedback) of organisms, yet sharply distinguishes the "possibility of error" arising within it from issues of noise in receivers, transmitters, or channels. It is clear that the latter "noise" noted in parentheses refers to Shannon and Weaver's communication model. In contrast, the 'possibility of error' he refers to is imbued with a more active meaning than mere transmission failure between sender and receiver.

What exactly does this mean? The subsequent part, "Examples of Acoustic Communication," is instructive. Explaining the first of the three levels as communication mediated by vibration and sound, Simondon notes that such energy transfer requires, first and foremost, a "milieu": gas for birds' voices, liquid for fish and aquatic mammals, and solid for tools and machines. Here, Simondon takes communication precisely as the phenomenon in which a signal as "figure" emerges from the background noise as "ground," saturating that milieu.

This points to a decisive difference from Shannon–Weaver, which grounds communication in the transmission of specific signals. As a human example, Simondon cites the then-recently discovered "cocktail party effect." From the murmur of a crowd, one can recognize one's own name; "Living

beings are capable of extracting from random background noise a signal carrying information whose intensity is lower than that of the ambient noise or background noise." (Simondon, 2015, p. 91). Here, noise does not indicate errors along a communication path; rather, it is the milieu—the background/ground—from which the signals that an individual organism prefers come to emerge. Put differently, the "possibility of error" refers to the process by which signals are generated or individuated from noise as a metastable state (Simondon, 2016; 2020).

Such a communication model is meaningful not only for rethinking conventional music–reproduction media theories that have treated noise as something to be eliminated vis-à-vis signal, but also for critically interrogating AI-based sound synthesis realized at the extreme of such tendencies. The next section explores this through the work *Mary Had a Little Lamb* (2019).

3. Considerations: *Mary Had a Little Lamb*

Mary Had a Little Lamb (2019) is a collaboration between the present author (Jo) and artist Paul DeMarinis (Jo & DeMarinis, 2023). The work combines electromagnetic induction with the deposition of ink onto paper, widely used in various printing technologies. Electromagnetic induction is the phenomenon by which a changing magnetic field induces current, discovered by Faraday and Henry in the 1830s. By combining these technologies, it becomes possible to synthesize sound from printed matter.

In the work, the phrase "Mary had a little lamb" synthesized by the AI waveform-synthesis technology "WaveNet" is converted into binary digits (0 and 1) on the computer. This binary data is rendered as black–white stripes akin to a barcode, arranged in a circle, and then printed onto paper with a high-resolution laser printer. When this print (whose black regions form fine raised ink ridges) is placed on a turntable and a permanent magnet is pressed onto the ridges while the record spins, the minute impacts and vibrations between the magnet and the ridges induce an electrical signal in the coil connected to the cartridge via electromagnetic induction; from the connected speakers a voice reading "Mary had a little lamb" is heard.

This production process might seem to forcibly reel a digital, generative-AI voice back into the analog and rebuild it as sound synthesis via a humble electromagnet. However, our concern is not an analog/digital opposition but synthesis as effected by the technology. How can the process by which sound is synthesized be made sensible without fixing it as a single or dominant process? Simondon's communication model, examined above, is productive for a critical reconsideration of this question.

As noted, Simondon treats communication as something that occurs irrespective of living or nonliving systems. Applied to the work's process of voice synthesis, we can organize it into two phases: synthesis by the electromagnet (nonliving), and synthesis through human audition (living).

On the one hand, the audible sound here is an aggregate of collisions between ink and magnet and the oscillations of magnet and coil; it is therefore difficult to claim that the technology "faithfully reproduces" the original string of digits. At the same time, the phenomenon arising from the printed black–white stripes is enabled by an environment = milieu of magnetic fields produced by the magnet and the coil. Although variables such as hand jitter while holding the magnet and the magnet–coil distance intervene in that environment, these noises are assimilated into the ground or background necessary for the emergence of voice as the signal–figure. Here, noise is positioned not as jamming for telecommunication but as nourishment that renders the electromagnetic field–as-milieu—an element necessary for the emergence of signal-as-figure, i.e., the voice.

On the other hand, to consider the synthesis of voice in listening, an observation by Thomas Edison, inventor of the phonograph, is illuminating:

"Another phenomenon I have noticed is that if two simple but different sentences are put on the machine, and a person who had never heard of such an apparatus is brought in and told to listen, he will not, even after a dozen repetitions, be able to say what it is, but if the first sentence is told him & then reproduced he generally says why that's perfect. The second sentence is reproduced when he generally reads it or part of it the first time and the whole second time, if simple. The same thing has been noticed on the telephone, and I think it lacks confidence or has some obscure effect of the mind on the hearing apparatus. They do not expect or imagine that a machine can talk; hence, they cannot understand the words. "(Edison, 1878)

Phonograph playback of cylinder grooves (or the telephone preceding it) undoubtedly yielded voices far noisier than today's standards. Edison's anecdote reports that listeners, once told in advance what the message was, could pick it out from noise that falls short of fully realizing meaning or melody. While Edison attributes this to a lack of expectation or imagination that "machines can talk," the inverse implies that human animals, through perceptual discrimination that prefers particular messages, can generate signals out of noise. "It is not the most physically powerful acoustic signal that is perceived, but the one that best corresponds to the motivation; the distinction between noise and signals is not predetermined physically in the nature of the stimuli." (Simondon, 2015, p. 92). Edison's anecdote from the dawn of the phonograph can thus be reinterpreted as illustrating Simondon's principle of communication more than the transmission models of the signal or speech, later assumed by Shannon–Weaver or Saussure. (Tkaczyk 2023)

The aim here is not to set the two models against each other or argue for the superiority of one over the other. Nevertheless, in *Mary Had a Little Lamb* as well, once the words and their meaning are recognized, listeners cannot escape their interpretation. In this effect, an AI-synthesised phrase traverses the history of audio technology to generate a signal within the background noise of electromagnets once more—yet, needless to say, this is not "reproduction" in the conventional sense. Because this work achieves a dual phase shift in sound synthesis, from physical phenomena to acoustic phenomena, and from perceptual identification to semantic content, as the preceding discussion has demonstrated. By multiplying and rendering perceptible the processes of sound synthesis as such, it liberates us from the communication model that has long haunted arguments of sound reproduction. In this sense, the significance of this work lies in its embodiment of the"possibility of error".

4. Conclusion

Assuming that debates around technologies of sound reproduction have reached a certain limit in the face of AI-based sound synthesis models, this paper has re-examined media-historical discussions of sound through the lens of signal/noise. After confirming the need to part ways with the Shannon–Weaver model—which seeks to minimize noise's intervention in order to transmit signals—we turned to Simondon's original arguments to show how voice or music comes to be synthesized from background noise as a milieu that harbors boundlessly the potential to become a signal.

To be sure, we do not claim that this model offers a definitive schema encompassing not only acoustic reproduction but all future AI-based generation. Rather, combined with *Mary had a Little Lamb*, it may remain a singular case. Yet while AI-based synthesis appears to prune away noisy parts to purify the signal, the work presented here shows, rather, that noise itself can serve as an effective milieu for rendering the signal salient. Gathering and making sense of such possibilities of error from within the historically noise-saturated background of audio technologies carries significance at a moment when technology seeks to render everything synthetic.

References

Dhariwal, Prafulla, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. 2020. “Jukebox: A Generative Model for Music.” *arXiv preprint* arXiv:2005.00341.

Edison, Thomas A. 1878. “Letter from Thomas Alva Edison to Alfred Marshall Mayer, February 11, 1878.” *Edison Papers Digital Edition*. Accessed October 24, 2025. https://edisondigital.rutgers.edu/document/X095AC

Jo, Kazuhiro, and Paul DeMarinis. 2023. “Producing Sounds from the Past of Media: *Mary Had a Little Lamb* (2019) and *We Were Away a Year Ago* (2023).” *The Journal of Media Art Study and Theory* 4 (2): 21–33.

Lessig, Lawrence. 2008. *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. London: Bloomsbury Academic.

Miwa, Masahiro. 2007. “Reverse-Simulation Music.” In *Cyber Arts 2007, Prix Ars Electronica*.

———. 2008. “What Is Reverse Simulation Music.” In *SITE ZERO/ZERO SITE No.2 Information Ecology Theory, Media for Living*, edited by Dominic Chen, 86–96. SITE ZERO, Media Design Institute. (In Japanese)

Oord, Aaron van den, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. “Wavenet: A Generative Model for Raw Audio.” *arXiv preprint* arXiv:1609.03499. https://doi.org/10.48550/arXiv.1609.03499.

Pasquinelli, Matteo. 2017. “The Automaton of the Anthropocene: On Caricatures of Machines, Animals and Humans.”*The South Atlantic Quarterly 116:2, doi 10.1215/00382876-3829423*.

———. 2016 “Abnormal Encephalization in the Age of Machine Learning,” *e-flux journal* #75 https://www.e-flux.com/journal/75/67133/abnormal-encephalization-in-the-age-of-machine-learning

Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical Journal* 27 (3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Simondon, Gilbert. 2015. *Communication et information: Cours et conférences*. Paris: Presses Universitaires de France.

———. 2016 *On the Mode of Existence of Technical Objects*, Trans. Cecile Malaspina and John Rogove, Minneapolis: Univocal Publishing.———.

———. 2020. *Individuation in Light of Notions of Form and Information* Trans. Taylor Adkins, Minneapolis: University of Minnesota Press.

Sousa, John Philip. 1906. “The Menace of Mechanical Music.” *Appleton’s Magazine* 8 (3): 278–284.

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

Suno. 2023. “About.” Accessed May 11, 2024. https://suno.com/about.

Tkaczyk, Viktoria. 2023. *Thinking with Sound: A New Program in the Sciences and Humanities around 1900*, The University of Chicago Press.

Udio. 2024. “About.” Accessed May 11, 2024. https://www.udio.com/.

Uexküll, Jakob von. 2010. *A Foray into the Worlds of Animals and Humans: With A Theory of Meaning*. Translated by Joseph D. O’Neil. Minneapolis: University of Minnesota Press.

Acknowledgement

This work was supported in part by JSPS KAKENHI Grant Number JP23K25288, JP23K25276.