Phantasmagoria: Sound Synthesis After the Turing Test

Vincent Lostanlen

▶ **To cite this version:**

Vincent Lostanlen. Phantasmagoria: Sound Synthesis After the Turing Test. Speculative Sound Synthesis Symposium, Sep 2024, Graz, Austria. hal-04650754v2

HAL Id: hal-04650754

https://hal.science/hal-04650754v2

Submitted on 22 Jul 2024 (v2), last revised 24 Aug 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Phantasmagoria: Sound Synthesis After the Turing Test

Vincent Lostanlen[1]

[1] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

vincent.lostanlen@cnrs.fr

Abstract. Sound synthesis with computers is often described as a Turing test or "imitation game". In this context, a passing test is regarded by some as evidence of machine intelligence and by others as damage to human musicianship. Yet, both sides agree to judge synthesizers on a perceptual scale from fake to real. My article rejects this premise and borrows from philosopher Clément Rosset's *L'Objet singulier* (1979) and *Fantasmagories* (2006) to affirm (1) the reality of all music, (2) the infidelity of all audio data, and (3) the impossibility of strictly repeating sensations. Compared to analog tape manipulation, deep generative models are neither more nor less unfaithful. In both cases, what is at stake is not to deny reality via illusion but to cultivate imagination as "function of the unreal" (Bachelard); i.e., a precise aesthetic grip on reality. Meanwhile, i insist that digital music machines are real objects within real human societies: their performance on imitation games should not exonerate us from studying their social and ecological impacts.

Keywords: Clément Rosset, computational creativity, critical performance practice, generative AI, Turing test.

1 Introduction

The opposition between "real" and "fake" data plays a structuring role in the scientific literature on generative machine learning. In the case of music, a digital audio signal is presented as "real" if and only if it can be traced to a form of life. Conversely, "fake" signals are those which are synthesized procedurally by the machine, either at random or from a non-audio input such as a text "prompt". Under this prevailing definition, which i will refer to as *anthropocentric*, machine musicianship is a kind of illusion; that is, a duplication of known perceptions while bypassing the creative process.

The appeal behind the anthropocentric paradigm is that, if enacted fairly, it holds all musicians to an equal status: anyone taking the trouble to play a song has a legitimate claim to "real" music. In particular, the paradigm is purposefully indifferent to the conditions of music production. For example, all guitar chords in a commercial music catalog are assigned the same degree of reality, regardless of whether the guitar in question is acoustic or electric; recorded with one or several microphones; played through a distortion pedal; sampled from another record; or even simulated by means of a Karplus-Strong algorithm. Such ability to encompass diverse sound synthesis technologies, without preference for allegedly "natural" or "live" settings, is clearly a strength.

However, now in the age of deep neural networks for audio synthesis, it is high time we call into question the premise that machine-made music is "fake" until proven otherwise. Indeed, the embarrassing implication of this premise is that a "fake" song may *become* "real" if released online, adopted by human listeners, and eventually included as training model for some next-generation model. Under such a premise, the dividing line between "real" and "fake" is not material but temporal: it gradually recedes in the direction of the machines as new synthesis technologies gain adoption. The anthropocentric paradigm essentially equates the real with the *déjà vu*—or, in this instance, the *déjà entendu*. By doing so, it runs the risk of negating the reality of music in the present time.

In this article, i attempt to subvert the anthropocentric paradigm "from within"; i.e., without erasing the importance of human perception in music technology. More precisely, i borrow the concept of "phantasmagoria" from philosopher Clément Rosset. Historically, the phantasmagoria is an entertainment device which projects light patterns on a surface so as to make the scene look like a magical or paradoxical phenomenon, such as the apparition of a ghost. By design, the phantasmagoria makes no promise of fidelity. Quite the opposite: it deliberately produces unfaithful perceptions in a way that puts the spectator in a state of disbelief, which gradually extends to the whole of visual perception. Spectators feel like they are in a dream while remaining aware that the feeling will dissipate after the séance is over. The magic lantern only projects as far as the curtain goes, and for a fee.

I claim that sound synthesis is a kind of phantasmagoria, on the basis of Clément Rosset's two books *L'Objet singulier* (1979) and *Fantasmagories* (2006). This is for three reasons. First, electronic amplification is an integral part of musical reality, as made evident by its congruence with a non-electronic kind of sound reproduction: the echo. Second, the advent of synthetic voices and instruments is not only a source of "deep fakes" but, more fundamentally, leads to disbelief on the faithfulness of *any* sound played through a loudspeaker. Third, it is clear that our disbelief would soon dissipate should the sound synthesis machines in question lack energy supply or maintenance effort in the future.

Together, the three elements above set the stage for a phantasmagorical conception of sound synthesis, one whose logical backbone is no longer real versus fake but real versus imaginary. More so than a philosophical disputation on the concept of the real, the reference to Rosset yields precise consequences in AI ethics, artistic practice, and music ecology. Ethically speaking, the phantasmagoria urges stakeholders in generative AI to stop advertising for "high-fidelity" models and to never conceal the presence of the machine from the audience. Methodologically speaking, it encourages artists to not underestimate the willingness and ability of listeners to imagine that a sound is machine-generated even so it isn't. Lastly, ecologically speaking, it reiterates the urgent need of "making infrastructures audible" (Devine 2021) by situating sound-making machines within the material and energetic flows of industrialized societies.

2 When the past counterfeits the present

In a famous 2014 article, Goodfellow *et al.* offer a colorful analogy to describe the learning objective of their generative model, termed "generative adversarial net" (GAN):

> In the proposed adversarial nets framework, the generative model is pitted against an adversary: a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles. (Goodfellow 2014)

What Goodfellow *et al.* omit to mention is that a collection of such "genuine articles" must be available *before* training both the generator and discriminator in a minimax game. Carrying on with the counterfeiture analogy, the authors assume that "the police" has some kind of special agreement with the central bank in order to receive a comprehensive and up-to-date supply of genuine currency. For academic datasets such as MNIST and CIFAR-10 (i.e., those imitated by Goodfellow *et al.* in 2014), the assumption is perfectly reasonable. But now in the year 2024, it is not unconceivable that scraping a contemporary music collection for "real data" would occasionally return GAN samples. Facetiously said, the counterfeiters of yesterday have gotten a job in the police and are now taking bribes the counterfeiters of today. Real and fake currency coexist lawfully to the delight of the mob.

Among the 100,000 songs which are uploaded to Spotify every day, how many are supposedly "counterfeit"? It is difficult to say; but in recent months, some Spotify users have reported receiving aggressive recommendations for what is seemingly AI-generated music[1]. Therefore, the hypothesis of "fake" songs soon becoming "real", under the anthropocentric paradigm which i presented earlier, has an empirical foundation.

Having postulated the historicity of the real, the machine learning community begins to recognize that this postulate is putting the whole profession at risk. A group of computer security researchers has proven that web-scale crawls of images and text are susceptible to malicious "poisoning", i.e., the introduction of adversarial noise in training samples so as to trigger targeted mistakes (Carlini 2023). While the adversarial examples of Goodfellow *et al.* are deemed "fake" and fool humans, the poisoned examples of Carlini *et al.* are deemed "real" and fool machines. Together, these two effects erode the validity of the anthropocentric paradigm, since they demonstrate that neither machines nor humans are perennial guarantors of conformity to the real. The erosion is strikingly apparent in the recent work of Shumailov *et al.*, who have trained generative models on their own output, recursively, causing them to "misperceive the underlying learning task":

---

[1] https://community.spotify.com/t5/Content-Questions/Release-Radar-this-week-was-almost-all-AI-generated-music/td-p/5630466

> It is now clear that large language models (LLMs) are here to stay, and will bring about drastic change in the whole ecosystem of online text and images. In this paper we consider what the future might hold. What will happen to GPT-{n} once LLMs contribute much of the language found online? We find that use of model-generated content in training causes irreversible defects in the resulting models, where tails of the original content distribution disappear. We refer to this effect as Model Collapse and show that it can occur in Variational Autoencoders, Gaussian Mixture Models and LLMs. […] We demonstrate that it has to be taken seriously if we are to sustain the benefits of training from large-scale data scraped from the web. (Shumailov 2023)

In their article, Shumailov *et al.* note that "preserving the ability of LLMs to model low-probability events is essential to the fairness of their predictions: such events are often relevant to marginalised groups". Yet, they show empirically that "only early signs of model collapse can be detected", thus casting doubt on the technical feasibility of filtering out any "model-generated content" during dataset curation. Instead, the authors recommend

> […] community-wide coordination to ensure that different parties involved in LLM creation and deployment share the information needed to resolve questions of provenance. Otherwise, it may become increasingly difficult to train newer versions of LLMs without access to data that was crawled from the Internet prior to the mass adoption of the technology, or direct access to data generated by humans at scale. (Shumailov 2023)

The text above is remarkably lucid. More so than wishing that model collapse could be solved by *future* advances in LLM technology, the authors scrutinize the *past* (and present) use of such technology in our society. As a result, they suggest that the Big Tech rush towards massive LLM deployment might soon turn into a Pyrrhic victory—one that inflicts such a devastating toll on the victor that it is tantamount to defeat (Gill 2019).

I wish to go in the same direction as Shumailov *et al.* and retain a lesson for sound synthesis at large, whether based on machine learning or not. The lesson is that anthropocentric categories of real and fake have betrayed the dualists who believed in them—a rare case of *self-poisoning* in computer science. But before we search for an antidote, it is worthwhile to return to a founding text of the field: Alan Turing's *Computing Machinery and Intelligence* (1950)[2].

3 Turing, then and now: who's testing whom?

> I propose to consider the question, 'Can machines think?' […] I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. […]

[2] I write in American English but preserve the British English spelling in Turing's text.

> The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. […] The interrogator is allowed to put questions to A and B […]. It is A's object in the game to try and cause C to make the wrong identification. […] The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers.
>
> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? (Turing 1950)

It may come as a surprise to some that the original exposition of Turing's imitation game makes no mention of adjectives such as real, actual, or genuine. Markedly different is the coverage of "the musical Turing test" among technologically enthusiastic media. Consider, for instance, *Singularity Hub* (emphasis is mine):

> Did you ever want to hear another Beatles album, or jam with Miles Davis? Of course, these things are impossible—but could we create a similar experience that people would genuinely value? Even, to the untrained eye, something indistinguishable from **the real thing**? (Hornigold 2018)

Or, on a "Turing test for sound", *MIT News* (again, emphases are mine):

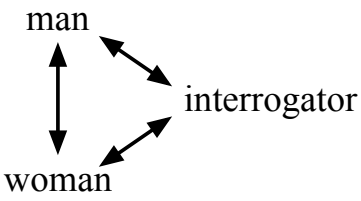
> To test how **realistic** the **fake** sounds were, the team conducted an online study in which subjects saw two videos of collisions — one with the **actual** recorded sound, and one with the algorithm's — and were asked which one was **real**. The result: subjects picked the **fake** sound over the **real** one twice as often as a baseline algorithm. (Conner-Simons 2018)

It has already been noted that these modern-day tests do not involve two people interacting with one machine, as initially proposed by Turing; but instead, one person being exposed to pre-recorded media. Here i concur with philosopher Daniel Dennett:

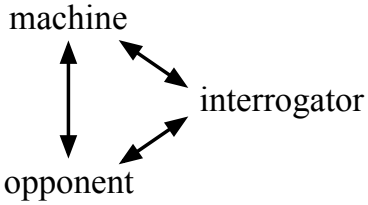
> there is a common misapplication of the sort of testing exhibited by the Turing test that often leads to drastic overestimation of the powers of actually existing computer systems. (Dennett 1998)

To be specific, while Turing aimed for an intersubjective test, current-day tests are intrasubjective: they purely rely on silent deliberation. This is for lack of an opponent playing the role of "the third player (B)" to "help the interrogator" via human–human dialogue (Wiggins 2021). The removal of opponent may be justified by practical considerations: while the 1950 version requires to recruit two humans who consent to play the game simultaneously, the current-day version may be implemented for individual people browsing the web at any time of day. Furthermore, the removal of dialogue offers the opportunity to replay the same data to different people and averaging out differences in judgment, thus treating subjects as a statistical cohort.

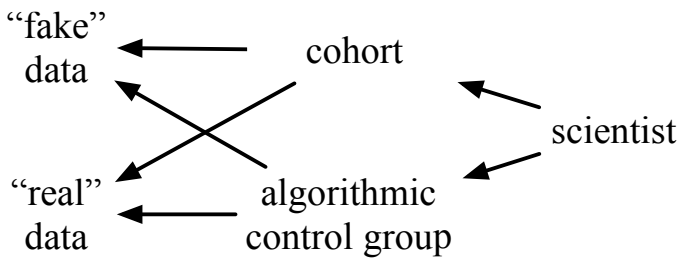
The diagrams below depict the variants of the Turing test in its original version (1950) as well as a typical current-day "misapplication"—borrowing the term from Dennett.



(i) The imitation game (Turing 1950): man-versus-woman variant.



(ii) The imitation game (Turing 1950): human-versus-machine variant.



(iii) A typical misapplication of "the Turing test", without opponent nor dialogue.

A previous publication (Ariza 2009) has argued that the difference between the two versions warrants a refinement in terminology: i.e., Musical Directive Toy Test versus Musical Output Toy Test. I believe that this refinement is helpful, but i immediately want to add that Ariza's categories refer to the *object* of the test; i.e., what is being tested. In this article, i wish to draw attention on the *subject* of the test: i.e., who is being tested. My point is that Turing's person C is now being pressed to give an answer to the question: "real or fake?". As artist Hito Steyerl has noted, the interrogator is subjected to scientific interrogation (Steyerl 2017)[3]. Or, much like in the digital modeling of analog signal processors: "the software passes the test when the user fails it" (Sterne 2020).

With Web Audio technologies, the same "real" and "fake" stimuli may be served to a large population of humans, therefore yielding scientific problems: for example, does musical training correlates with ability to avoid "picking the fake sound over the real one"? Although Turing did conjecture about the winning rate of the "average interrogator", he seemed uninterested about demographic variations around the average. This is unlike current-day methods for the evaluation of musical metacreation, which carefully distinguish expert versus non-expert ratings (Agres *et al.* 2016). Moreover, much in the same way that A and B are opponents for Turing, the human cohort has its own machine

[3] Before Steyerl, philosopher Sadie Plant had described how "Turing was subjected to his own test" in her cyberfeminist *Zeros and Ones* (Plant 1995, p. 100).

opponent: namely, the "baseline algorithms" mentioned by Conner-Simons (2018), effectively playing the role of a control group.

Since the 1980's and the massification of digital media, the historical question of Turing, "can machines think?" is becoming less and less pressing. The concept of thought has lost the level of scientific prestige it had in the time of Turing. In its place, another concept is becoming essential to machine learning research: the concept of the real. In the next section, i shall contend that, as machine learning researchers who operate on audio, we must make efforts to stand clear of essentialist categories such as "real data" and "fake data". More precisely, my position is that the substance of sound synthesis is not fake but *double*. To uphold this position, I shall discuss two essays by a French philosopher Clément Rosset: *L'Objet singulier* (1979) and *Fantasmagories* (2006)[4].

4 Listening to Clément Rosset's singular objects

In *L'Objet singulier*, Rosset defines the real as "a non-closed set of non-identifiable objects" (Rosset 2006, p. 22). By non-identifiable, Rosset does not mean that these objects would all be ordinary and interchangeable. On the contrary, they are all extraordinary and peculiar:

> An identification consists in bringing an unknown term back to a known term; such operation is impossible in the case of the real, which is unique to be and thus, so to speak, uniquely unique: "a unilateral being whose mirror complement does not exist", said of the universe physicist Ernst Mach. Thus, the real is foreign to any characterization: and that is precisely its proper character to have no assignable characteristics. […] The most direct relation of consciousness to the real is thus a relation of pure and simple ignorance. (Rosset 1979, p. 23)

Despite this relation of ignorance, real objects are in no way hidden to the senses. This is unlike Platonic realism, which only ascribes reality to universal ideas and dismisses particular things of the sensible world. While Plato distinguished works of art in terms of varying degrees of faithfulness in the representation of nature *(mimesis),* Rosset rejects the criterion of faithfulness in aesthetics. His anti-Platonism is radical in the case of music:

> Music did not wait until becoming "concrete" or "stochastic", to celebrate, with Xenakis, a seemingly late reunion with an art of noise from which it has never truly divorced, before revealing to whoever wished to listen its veritable function: to not imitate, to talk about nothing, to be elsewhere, out of a plot where, without exception, the other forms of human creativity find themselves more or less engaged. (Rosset 1979, p. 60)

---

[4] To my knowledge, neither *L'Objet singulier* nor *Fantasmagories* have received official translations to English. For lack of a better option, I have taken on to translate Rosset myself. I should disclaim that I am not a professional translator. I can only urge my reader—*a fortiori* of the philosopher kind—to refer back to the original text if possible.

It is helpful to contrast Rosset's philosophy of music from that of Cornelius Castoriadis, who analyzed the power of music in terms of creation of affects and desire:

> […] music neither 'expresses' nor 'represents' affects known otherwise, it creates them. […] And there is a desire, perhaps close to the desire of the state of nirvana (Schopenhauer, Wagner …): the desire for it to last forever […]. (Castoriadis 1990)

Compare with Rosset, for whom this power is essentially destructive:

> The musical object fascinates because it situates itself outside of the realm of the desirable, realizing this paradoxical condition, often coveted but never granted outside of the musical space, to be "unlike any other", that is to be perfectly *other* […]. No thoughts, no feelings, no so-called affective reactions other than this unique musical "affect" which consists in liquidating all affects. (Rosset 1979, p. 62)

Moreover, Castoriadis emphasized the meditative element in music listening, calling it an "abolition of the world" (Castoriadis 2007) which causes a "complete absorption into something else that oneself" (Castoriadis 1990). In this respect, Rosset does concede that musical listening "listens for something and hardly cares about the rest" (*ibid.*, p. 63): it is for that reason that "humans and animals listen to Orpheus with stupefaction and awe." (*ibid.*, p. 62). However, Rosset and Castoriadis disagree on the effect of listening:

> Faced with music, the listener is always taken aback, taken by surprise. Indeed, the musical effect is, before anything else, an "effect of real", and indeed the real is the only thing in the world to which one does never completely get accustomed. (Rosset 1979, p. 64)

The statement above leads Rosset to a bold and uncommon philosophical position, namely, the rejection of intellectual as well as aesthetic values in music:

> Music is neither true nor beautiful. It is, let it be repeated, essentially other and only appears as foreign insofar as it is precisely not susceptible to let itself be represented, to lend itself to an intellectual or aesthetic adjudication; true this, not true that, beautiful this, not beautiful that. (Rosset 1979, p. 63)

However, Rosset recognizes the expressive power of music as constitution of real:

> […] music, even so it evokes nothing that be, is nonetheless expressive, to the utmost degree even, according to some: perhaps precisely insofar as it does not need to imitate the real because it suffices to constitute the real from scratch. The musician is like a prudent traveler, prepared for the vacant inns where they shall stay: he brings his real along with him. (Rosset 1979, p. 61)

Rosset makes clear that this "violent power" is not granted to other arts besides music:

> Irruption of real in its raw state, with no possibility of approach from the bias of representation: such is the musical effect, and the reason of its peculiar power. Hence music is creation of real in the wild, with no commentary nor replica; and the single art object which presents a real as such. (Rosset 1979, p. 63)

Adopting the position of Rosset on music has far-reaching implications for audio technology, some of which are drawn in a later essay: *Fantasmagories* (2006). First, and against the Platonic theory of *mimesis*, artworks are not remote from reality:

> Photography, sound reproduction, painting are products of art, that is, realities in their own right, therefore it would be vain to distinguish them from reality in general with which they share every privilege. (Rosset 2006, p. 11)

To talk of "real data", as we often do in computer science, is to say the same thing twice. If a sound is *datus* (from Latin, "given"), its *donatio* ("act of giving") is necessarily a thing independent from language, indeed an essential component of reality—regardless of sonic source or *dator/datrix* ("giver"). This point is worth belaboring in an age where, as i wrote before, the Turing test is being misconstrued as a test for reality of data.

Secondly, i take inspiration from Rosset to affirm that live acoustic performance and loudspeaker playback are commensurate yet not mutually identifiable. Commensurate: orchestras often employ a mixture of amplified and non-amplified instruments on the same stage, and it would be risible to contend that spectators would fail to perceive them *en masse*. Not mutually identifiable: the functioning of loudspeakers involves electromagnetic transduction, a physical phenomenon that is irreducible to wave mechanics. As a result, the loudspeaker is not a fake musical instrument but a real "singular object" for music duplication. Yet, failing to recognize this singularity causes anxiety and disappointment:

> The shadow of the double, bypassing the reality of particular objects, is cast on the fact of existence in general. Every reality that is potentially exposed to duplication thereby ceases to be credible. The thought of the double thus elicits a disappointment with respect to the most irrefutable real […]. (Rosset 1977, p. 14)

Such "disappointment with respect to the most irrefutable real" is the culmination of the misapplied Turing test, as described earlier. Under the anthropocentric paradigm, it explains the occurrence of *false negatives* in the judgments of the cohort: i.e., sounds which are labeled as "fake" even so they were produced without the use of the computer.

Rosset's conception of the double is non-anthropocentric in that it is not reduced to human-made artifacts, such as the equipment of a music studio. Instead, Rosset gives an example of double which reaches back in time by billions of years, thus preceding the appearance of life on Earth[5]:

> The echo, first kind of sound reproduction, may without a doubt be considered as a "resonance" of the reality […]. Could this archaic model of sound reproduction, one may even say prehistoric since there was echo in the atmosphere of our planet before there were men to perceive it, vouch for the authenticity of later processes

---

[5] A philosophical discussion on phenomena which occurred earlier than life on Earth is also present in Quentin Meillassoux's *Après la finitude* (2006), under the name of arch-fossil. Some implications for music are present in Hecker *et al.* (2010). It remains to be seen to what extent Rosset's philosophical conception of the echo is compatible with Meillassoux's.

> […] of sound reproduction? Obviously not: techniques of sound reproduction, born even later than photography, are prone to the same kind of suspicion of infidelity to the real.

For Rosset, this suspicion is caused by two factors: the possibility of doctoring (in French *trucage*) and the delay between audio acquisition and playback. Note that both of these factors are essential to current machine learning research. Therefore, it stands to reason that the suspicion will last in the future: not *despite* technological progress but *because* of it.

> I shall only mention two main reasons which forbid [sound reproduction] to aspire to more than conformity to the comparable real […]. The first reason is obviously the possibility of falsification, […] which is getting easier as sound reproduction techniques are making progress. […]
>
> I now come to the second of the two main reasons which condemn sound reproduction […] to differ from the sound […] it strives to acquire. Sound reproduction is, by definition, a reproducing gap. Thus it marks a temporal lag with respect to the sound it repeats. […] By considering repetition as rigorous repetition of what it repeats, it is manifest that every repetition is impossible […]. (Rosset 2006, p. 51–52)

This last statement is particularly crucial since it is indifferent to choice of technology. Near the end of his text, Rosset fleshes out his concept of phantasmagoria explicitly:

> Every phantasmagoria disappears on the brink of the real, just like phantoms disappears at dawn: "the sun dissipates it like a mist", writes Maupassant about fear. One may infer a first definition, somewhat vague of the real: one will say that the real is what dissipates phantasmagorias, doubles, fear. […] (Rosset 2006, p. 65)

Then, he warns the reader against the risk of treating real and fake as mutually exclusive and essential categories, and suggests that "the real without double is nothing":

> The real is thus first what remains when phantasmagorias dissipate. As Lucretius says: "the mask is torn off, reality remains". Inasmuch, of course, as anything remains. The real is perhaps the sum of appearances, images and phantoms which fallaciously suggest its existence. […] Hence, by wanting to clean the real from parasites which veil it, we run the risk of annihilating it. (Rosset 2006, p. 66–67)

This last sentence rings like a word of caution in our age of chronic audiophilia. Too much discourse on "artifacts", understood as the parasites of sound reproduction, runs the risk of reducing the act of listening to a search for the "authentic" source behind the veil. In her 2019 book *The Race of Sound*, musicologist Nina Sun Eidsheim has shown how pervasively the myth of authenticity has been ingrained into sound synthesis software, giving the example of Yamaha's *Vocaloid*. Alluding to Erik Satie, she writes: (emphasis in the text)

> When we listen, we do not simply automatically measure and compare to an a priori. We are not *phonometrographers* "weighing and measuring sound". Every measurement is preceded by countless decisions. (Eidsheim 2019, p. 182)

The Race of Sound concludes that it is only after a "critical performance practice" that such "countless decisions" may become apparent and challenged. This critical practice should not be misunderstood as a penchant for solipsism. Indeed, Eidsheim acknowledges the reality of what she calls the "thick event", i.e., "a continuous vibrational field with undulating energies" that is irreducible to "mere sound" (Eidsheim 2019, p. 8). However, the thick event is multiple and contingent. Even so it may be coerced into measurement, the result of the measurement is not predetermined by the thick event, nor does the result of the measurement necessarily determine the thick event. I wish to acquaint Eidsheim's concept of the thick with Rosset's concept of the real: it is "foreign to any characterization, and that is precisely its proper character to have no assignable characteristics" (Rosset 1979, p. 23).

Taking the microphone to be a kind of measurement apparatus, i deduce that Eidsheim and Rosset might agree upon three core tenets. First, all thick events are real through and through. Loudspeaker membranes belong to the same "continuous vibrational field" (Eidsheim) as vocal folds and thus "it would be vain to distinguish them from reality in general with which they share every privilege" (Rosset).

Second, digital audio is unfaithful at any level of technological advancement. This is because microphone and loudspeaker, via electromechanical transduction, introduce wormhole-like patterns in the vibrational field: hence a "disappointment with respect to the most irrefutable real" (Rosset), a "familiarity as strangeness" in style and technique (Eidsheim).

Third, the main ingredient to a musical phantasmagoria is not information (or lack thereof) but *timeliness*, defined as a historically cogent articulation of symbolic, material, and measured aspects (Eidsheim). Yet the misapplied Turing test proceeds to erase the symbolic and the material and only retains a few measured aspects: i.e., those which conspire to validate anthropocentrism *post festum*. To listen is to "never completely get accustomed" (Rosset), a renewed critical practice that is far more exigent than telling humans and machines apart.

5 Imagining matter, materializing imagination

The past three sections have challenged the opposition between real versus fake in sound synthesis. We have argued, against a certain kind of anthropocentrism, that reality and illusion may *coexist* in the same act of listening. Crucially, such coexistence is not primarily attributable to a feat of engineering: it was there long before us, in the echo of a distant cliff. In 2003, three years before the publication of Rosset's *Fantasmagories*, this point was already made clear by composer and researcher Jean-Claude Risset: (translation is mine)

> The echo, delayed copy of the initial sound, is produced naturally in certain conditions of propagation. Today, artificial devices allow to create sonic simulacra, replicas which reconstruct the audible appearance of the sound in the absence of its initial mechanical cause. Sound reproduction aims to recreate the auditory sensation, not the vibratory process from which it was born: its fidelity is not judged by objective criteria, but according to exigencies of hearing. (Risset 2003)

Furthermore, if a machine were to pass some "Turing test for music", even in its most rigorous variant—i.e., including a human opponent and a dialogue, see Wiggins (2021)—the news would have no bearing on the dividing line between reality and illusion. Indeed, the need for computation and electromechanical transduction implies that synthesizers are *untimely* compared to the "continuous vibrational field" (Eidsheim) in which they operate. If music is "irruption of real in the wild" (Rosset), then by definition, no data-driven process, however sophisticated, can keep up.

To recap, we have seen limitations to the powers of sound synthesis, both "from below" and "from above". From below: effects of acoustical duplication predate the invention of the computer, and even that of the loudspeaker. From above: a neural audio synthesizer may reconstruct past measurements of the past but "discounts enculturation, technique and style, and an infinity of unrealized manifestations in favor of preconceived essence and meaning" (Eidsheim 2019, p. 8). In short, we audio technologists do not hold a monopoly on the double, nor do we make any plus-value in the mechanical reproduction of the real.

Given these insurmountable limitations, it may seem that sound synthesis devices deserve no special status in critical performance practice. After all, if music played through loudspeakers casts a shadow on all "on the fact of existence in general" (Rosset), couldn't the same be said of music notation? Once music from the oral tradition is written down, the content is exposed to *exact* duplication, and thus subject to a structural organization which potentially alienates it from sensation and meaning (Rosset 1979). In this context, Babbitt (1965) has argued that music as stored on disc or as written on paper essentially belong to the same domain of representation, of the so-called *graphemic* kind.

For Babbitt, the graphemic domain forms a triangle with the auditory (cognitive) domain and the acoustic (physical) domain. Under such a tripartition, audio data could be defined as neither auditory nor acoustic; or, in philosophical terms, neither in us nor in the world. This privative definition has emboldened Wiggins *et al.* to present music as non-existent:

> We follow [Babbitt's] view here, taking the philosophical stance that the mysterious thing that is Music is actually something abstract and intangible, and which does not have real existence in itself, but which is described by all tree of these representational domains [graphemic, auditory, acoustic]. […]
>
> Since, we suggest above, the difference is in the listener, performer or other musician, we can only argue that the place where Babbitt's three domains come together is really in the human mind/brain. (Wiggins *et al.* 2010)

Even so i am sympathetic to the claim that models of human (or non-human animal) perception and cognition must play a central role in the scientific study of music, i am wary of the risk of false equivalence between audio signals and notation. Wiggins *et al.* (2010) describes them as mere "traces in the real world […] which are themselves musical stimuli—much as light cannot itself be seen, but leaves traces everywhere around us […]". Granted, the context of their article is a special issue of the *Musicae Scientiae* journal which celebrated the 25th anniversary of Lerdahl and Jackendoff's *A Generative Theory of Tonal Music*. By claiming that music "cannot exist unless a mind is implicated", Wiggins *et al.* are opposing a certain "overly positivistic approach to the linguistic phenomenon itself", as

represented by Chomsky's conception of an ideal speaker–listener. In this context, the material constituents of the graphemic domain are justifiably presented as out of scope.

But to say, with Wiggins *et al.* (2010), that "neither music nor language can be studied as pure surface forms, because the cognition of both produces information which is not contained in the surface form", is not to say that the material traces of the surface form should be expunged from *all* scientific theories of music. When positing the non-existence of music, we should not go as far as to suggest the non-existence of the music industry.

Whether a musical stimulus is stored on clay, parchment, paper, shellac, plastic, or otherwise may be irrelevant to an inquiry on Babbitt's "auditory domain"; still, it has profound implications for human societies. Admittedly, Wiggins *et al.* (2010) concede this point when they write that "there is a larger meaning of Music which arises from the combination of all the domains in their diachronic sociological context". Notwithstanding the heritage of Lerdahl and Jackendoff in music theory, i deem it worthwhile to elaborate specifically on the material underpinnings of music in the age of digital audio.

The history of popular music in the twentieth century speaks eloquently for the rich connections between Babbitt's "graphemic domain" and its "diachronic sociological context" (Wiggins *et al.* 2010). However, these connections are too often reduced to who is making and listening to music, or buying and selling it. Yet, the people who make the phantasmagoria possible (or precisely, *believable*) are not only the songwriters, session musicians, lyricists, audio engineers, lawyers, and so forth. In his book *Decomposed* (2019), Kyle Devine has shown that the political ecology of recorded music must be world-scale:

> [A] much wider range of people and a much broader range of experiences have played much more central roles in the history of recorded music than is normally recognized. […] Plastic formats could not exist without the drillers and toolpushers that pumped oil from the ground or the chemical engineers and material scientists that cracked hydrocarbons and develop polymer compounds to the specifications demanded by the recording industry and its customers. Nor would they exist without the women and men that pressed these records in factories. Data files could not be stored or transferred without the software engineers that develop algorithms or the IT workers who build and maintain internet infrastructures. Such files could not be accessed without miners in places such as the Democratic Republic of the Congo, who extract the rare minerals and metals that make up our listening devices. Nor could those files be heard without the solderers and line-workers who assemble these accessory electronics in places such as China. Moreover, all recording formats and listening devices need dump sites and communities willing or willed to absorb to absorb these technologies as they break down and obsolesce. Such workers are not typically thought of as musicians. Indeed, they are not usually thought of at all. At best, they might be considered support personnel. Yet their labors are essential parts of the musical world. (Devine 2019, p. 17)

Although Devine does not explicitly reference Wiggins *et al.* in his book, i believe that a putting the two texts side by side is helpful. Wiggins *et al.* (2010) stated that "because music […] only has existence in the mind, the very notion of a scientific theory of Music,

distinct from mind, is suspect". Crucially, the position of Devine is not to evade the suspicion of Wiggins *et al*., and for good reasons: it is wholly unassailable by empirical refutation[6]. Rather, Devine boldly decides to plead guilty as charged, or so to speak, and affirm his research practice as *musicology without music*: (emphasis in the text)

> If the turn toward performance in music research emphasizes that music is not a thing but an activity, the lesson here is that things are activities too. […] While the central aim of this book is to describe (and critique) the conditions of music's political ecology, a parallel aim is to critique (and describe) a particular conception of music that encourages us to take those conditions for granted in the first place. The two perspectives require each other. Together, they suggest the need for musicology without 'music'. […] A musicology without music suggests that researchers should not be so sure that they know in advance what counts as "properly" musical practice or a "properly" musicological object of study. […] The point is to develop a version of music research that does not begin as a musicology *of* music—that does not begin in a tautology where the force of preconstructed definitions of music delimit what musical culture can be or where music researchers should focus their attention. […] The goal is to describe and improve the messy associations of biology, geology, capital, and culture that define our collective musical life on this planet. (Devine 2019, p. 21)

In a previous essay, I have followed the footsteps of Kyle Devine by summarizing the state of current knowledge on the ecology of digital music (Lostanlen, 2023). Much remains to be understood on this topic, particularly in an age where sound synthesis is not only done by professionals in music studios but by a growing number of people on a growing number of devices. In this essay, my goal has been to warn against a certain discourse of weightlessness and transparency, as entertained by the music streaming industry, regarding their ecological impact. My position is that presenting neural audio models purely from the lens of "artificial creativity", whether beneficial or detrimental to human creativity, runs the risk of missing an equally important controversy: that of the material, energetic, social, and—ultimately—political implications of machine musicianship.

---

[6] To my understanding, the suspicion of Wiggins *et al.* (2010) may be radicalized into a critique of dogmatism in scientific knowledge. For Wiggins *et al.*, we cannot think about Music is in the absolute but necessarily in relation to the conditions of the donation of the Music in the mind at present. This position is known in speculative realism as *correlationism*, defined by Meillassoux (2006) as "the idea according to which we only ever have access to the correlation between thinking and being, and never to either term considered apart from the other". Meillassoux has offered a materialist and non-metaphysical alternative to correlationism, but this is beyond the scope of this article.

6 Conclusion

Modern renditions of the Turing test have reversed the position of humans from active interrogator to passive interrogatee. In the words of Eidsheim, the new Turing test asks an "acousmatic question" of the form: *Who is this?* i.e., human or machine? The invention of the phantasmagoria has demonstrated that humans are perfectly able to suspend this question, even when the presence of the machine is conspicuous. For this to happen, they should be given enough agency to navigate the material and symbolic aspects of the technical apparatus—be it "magic lantern" or variational autoencoder.

What is a stake is neither to improve "fidelity" or "realism" or even "alignment", but to cultivate the imaginary in the production of the double. This ethical duty was eloquently stated by philosopher Gaston Bachelard, and later quoted by Clément Rosset:

> In fact, the way in which we escape the real neatly designates our intimate reality. A being that is deprived from the *function of the unreal* is just as neurotic as one deprived from the *function of the real*. We may thus say that a disorder in the function of the unreal backfires onto the function of the real. If the function of openness, which is properly that of imagination, is done badly, perception itself remains obtuse. Thus one will have to find a regular kin from real to imaginary. (Bachelard 1943)

We have seen that artists should not underestimate the willingness and ability of listeners to imagine that a sound is machine-generated even so it isn't. To restate the words of Eidsheim: "every measurement is preceded by countless decisions". Or analogously: "we must reflect in order to measure and not measure in order to reflect" (Bachelard 1938).

Lastly, we have called for urgent action towards "making infrastructures audible" (Devine 2021), which can only be completed via some coherent articulation of computer sciences, Earth sciences, and social sciences. But before then, a prerequisite is to recognize the reality and non-repeatability of all music and the central role of collective listening. "The modern interest for the historicity of the real", Rosset writes, "is evidence among others of the difficulty that one experiences to take into consideration the real *tout court*" (Rosset 1979, p. 9).

Our ecological duty is to dispel the smoke in which the phantasmagorical music-making machine projects its ghostly apparitions. Let us pause the algorithmically recommended playlist like one pauses a diorama of dissolving views. Having recognized that music, like light, "cannot itself be seen, but leaves traces everywhere around us" (Wiggins *et al.*, 2010), we shall go on to open the magic lantern of sound synthesis in search for a combustible.

Acknowledgements. This article was written on the road. I am grateful to David John Baker, Théis Bazin, Jonathan Bell, Lucie Bouchet, Ashley Burgoyne, Huihui Cheng, Nina Sun Eidsheim, Alice Eldridge, Han Han, Florian Hecker, Louise Hochet, Henkjan Honing, Amy Ireland, Maya B. Kronic, Jennifer Laura Lee, Lia Jiaxin Li, Michel Lostanlen, Mathilde

Monjanel, Clémence Prévost, Orian Sharoni, Dan Stowell, and Cyrus Vahidi for thought-provoking conversations.

References

Agres, Kat, Forth, Jamie, & Wiggins, Geraint A. 2016. "Evaluation of musical creativity and musical metacreation systems". *Computers in Entertainment (CIE)*, *14*(3), 1–33.

Ariza, Christopher. 2009. "The interrogator as critic: The Turing test and the evaluation of generative music systems". *Computer Music Journal*, 33(2), 48–70.

Babbitt, Milton. 1965. "The Use of computers in musicological research". *Perspectives of New Music*, 3(2), 74–83.

Bachelard, Gaston. 1938. *La Formation de l'esprit scientifique*. Paris: Vrin.

Bachelard, Gaston. 1943. *L'Air et les Songes*. Paris: Joseph Corti.

Carlini, Nicolas, Matthew Jagielski, Christopher A. Choquette-Choo, Daniel Paleka, Will Pearce, Hyrum Anderson, Andreas Terzis, Kurt Thomas, Florian Tramèr. 2023. "Poisoning web-scale training datasets is practical." In *Proceedings of the IEEE Symposium on Security and Privacy (SP)*.

Castoriadis, Cornelius. 1990. "Nouveau millénaire, défi libertaire". Entretien lors des Décades de Cerisy-la-Salle.

Castoriadis, Cornelius. 2007. *Fenêtre sur le chaos*. Paris: Seuil.

Conner-Simons, Adam. 2016. "Artificial intelligence produces realistic sounds that fool humans". *MIT News*.

Dennett, Daniel. 1998. *Brainchildren*. Cambridge, MA: MIT Press.

Devine, Kyle. 2019. *Decomposed*. Cambridge, MA: MIT Press.

Devine, Kyle, Boudreault-Fournier, Alexandre. 2021. "Making infrastructures audible". In *Audible Infrastructures*, 3–55, Oxford: Oxford University Press.

Eidsheim, Nina. 2019. *The Race of Sound*. Durham: Duke University Press.

Gill, N. S. 2019. "What's the origin of the term Pyrrhic victory?". In *ThoughtCo*.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. 2014. "Generative adversarial nets". In *Advances in Neural Information Processing Systems (NeurIPS)*.

Hecker, Florian, Quentin Meillassoux, Robin Mackay. 2010. "Speculative solution: Quentin Meillassoux and Florian Hecker talk hyperchaos". In *Urbanomic Document*, UFD001. London: Urbanomic.

Lostanlen, Vincent. 2023. The ecology of digital music. Paris: Centre national de la musique. https://cnmlab.fr/en/short-wave/the-ecology-of-digital-music/

Meillassoux, Quentin. 2006. *Après la finitude*. Paris: Seuil.

Plant, Sadie. 1995. *Zeros and Ones*. London: Fourth Estate.

Risset, Jean-Claude. 2003. “Illusions musicales”. In *Pour la science*, 39, p. 66–73.

Rosset, Clément. 1979. *L’Objet singulier.* Paris: Minuit.

Rosset, Clément. 2006. *Fantasmagories.* Paris: Minuit.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot. “The curse of recursion: Training on generated data makes models forget”. 2023. *arXiv preprint arXiv:2305.17493*.

Steyerl, Hito. 2017. *Duty Free Art: Art in the Age of Planetary Civil War*. London: Verso.

Sterne, Jonathan. 2020. “The software passes the test when the user fails it: Constructing digital models of analog signal processors”. In Viktoria Tkaczyk, ed., *Testing Hearing*. Oxford: Oxford University Press.

Turing, Alan. 1950. “Computing machinery and intelligence”. *Mind*, 59(236), p. 433–460.

Wiggins, Geraint, Daniel Müllensiefen, Marcus Pearce. “On the non-existence of music: Why music theory is a figment of the imagination”. *Musicae Scientiae*, 14(1). 2010.

Wiggins, Geraint. 2021. “Computational creativity and consciousness: Framing, fiction and fraud”. *Proceedings of the International Conference on Computational Creativity*.