aletheia
Composing with and hacking neural audio synthesis

Axel Chemla–Romeu-Santos

chemla@ircam.fr 4, place Igor Stravinsky 75004 Paris, France

Abstract

Aletheia is a musical piece based on exploring, transforming, hacking, bending, and interplaying with generative audio models. Referring to the pre-socratic notion of aletheia, this piece illustrates the discovery of reality of neural audio synthesis by going from a pure phenomenological perception to an imitation game, yielding to a deranging and coercive version of reality. Besides attempting a reflexive work between philosophical aspects of reality and exploration of model's rules through deviation, aletheia is motivated by the experimental development of new approaches to compose and interact with neural generative models, and by which new aesthetics they can provide.

1 Description

Program notes. Aletheia is a part of a cycle of musical pieces, *Daath*, that explores several ways of composing and performing with neural audio and image synthesis. This musical piece echoes the concept of *aletheia* a pre-socratic allegory of reality later described as an *unveilement* by Heidegger, by revealing the inner processes subsumed by machine learning techniques by exploring, transforming, hacking, bending, and interplaying with the models.

While every sound present in the piece are generated with neural audio synthesis, Aletheia organizes a game with reality by decomposing and deconstructing sounds the real world by alteration and manipulation of neural audio synthesis models, illustrating the discovery of reality by going from a pure phenomenological perception to an imitation game (corresponding to origin of the word $\alpha - \lambda\eta\theta\epsilon\iota\alpha$, extraction from oblivion). However, the more these models are pushed to their limits and submitted to complex performative manipulations, the more they produce a deranging equivalent of reality, unravelling the act of modelling and the aporetic aspect of extracting the inner nature of phenomena through observation and quantification. Hence, this piece is a reflexive work between philosophical aspects of reality and exploration of the model's rules through deviation, motivated by the experimental development of new approaches to interact, compose and perform with neural generative models, and which new aesthetics they can provide.

Technical notes. We think that our proposition strongly correlates with the *Sound design and synthesis systems* section of this conference as it one of the first attempts to compose an entire electro-acoustic piece only by using and hacking neural audio synthesis models. This musical piece is accompanied with a corresponding article explaining both technical details and the compositional process of this piece, besides a description of the overall experience and the ethical statement that grounded the composition. The creation of this piece, motivated by the desire to experiment with the hijacking of neural audio synthesis models, involved through three different steps :

- *Step 1*: developing a workbench to unify neural synthesis audio models for encoding/decoding generation, weight alteration, and latent space exploration / modification ; training the models.
- *Step 2*: experimenting with the models and developing generation / hacking routines.
- *Step 3*: composing the form of the piece, and refining iteratively each part of the piece.

The models used in these piece (SampleRNN [4], variational auto-encoders[3, 2], RAVE [1]) have been trained on several custom datasets including additive/FM generated sounds, vocal/instrument samples, and also trained without inputs by feedback. The experimentation step consisted in testing and selecting several generation / hacking routines, by exploring the *latent space* of audio generative models and developing several weight alteration techniques. Five interaction methods were finally used : *forwarding* unknown data to the model, *stretching* latent trajectories to dilate the incoming sounds, *interpolating* between several examples to create unknown sounds, *free exploration* of the model's latent space, and *feedbacking* the model's generations to create a *vanishing* effect. Regarding weight corruption, several noising / permutation techniques were applied to the weights/biases of convolution layers. This exploration step yielded the required experience to compose the global form of the piece :

- Begin the piece with subsequent corruptions of a SampleRNN trained on the recorded voice, surrounded by VAE / SampleRNN trained without inputs (0:00 - 2:00)
- Pursuing with VAEs, exploring successively the models with the additive, FM corpus, the orchestral corpus, and the voice corpus (2:00 - 13:00)
- Concluding the piece with the RAVE model, exploring successively the orchestral corpus, voice corpus, and drums corpus. (13:00 - 18:18)

Then, regarding the completion of the global form, artistic choices were then guided by respecting the original intention, experimenting, and personal aesthetics. The composition have been made using the Ableton Live software, and mainly consisted in assembling and panning the samples (sometimes refining the generations towards precise perceptive goals). We also made the choice of using few audio effects to preserve the original aesthetics of the generations, except parsimonious EQing and reverberation.

Ethical statements. The creation of this piece has been motivated by several purposes, linked by the will to propose a philosophically-positive critic of machine learning through artistic practices. The correspondence between the notion of aletheia and machine learning methods, initially proposed by the French philosopher Éric Sadin, intended to demonstrate the contradiction between the observative posture that grounds machine learning and the coercive actions of their use by modern capitalism, having very concrete repercussions on society and economy [5]. Hence, we think that such research creation approaches would then allow to develop a counter-discourse on the positive, elitist, and techno-capitalist discourse widely broadcasted by the digital-industrial complex, and to develop inclusive works and tools that allow for people to really apprehend, seize, and control the techniques that surround them. Furthermore, all the models were trained on a single GPU and used on standard CPUs, and all of the models will be available for free on the corresponding open source repository on the author's GitHub.

2   Link

The corresponding audio file can be found at this address : `https://www.dropbox.com/s/ptwk0yrjogtghe5/Aletheia.wav`, or using the Soundcloud link `https://soundcloud.com/user-859797836/aletheia`.

3   Biography.

Axel Chemla–Romeu-Santos is a researcher, composer, and performer. After a double undergraduate degree in Engineering Sciences & Music Theory in Paris-Sorbonne, he graduated in Computer Science / acoustics at IRCAM and pursued as a PhD student between IRCAM (Paris, France) and LIM (Milano, Italy), that he defended in 2019. He is then post-doctorate at IRCAM, working on interfaces between AI and musical creation and computational creativity. Besides, he graduated in computer music at the CRR93 and was involved in several musical endeavors with theater, as composer and sound engineer in the *Théâtre de la Suspension* company, composer and sound engineer for the post-digital and maximalist trio Daim™, and co-founder of the w.lfg.ng collective. His main interests are co-creativity, post-digital aesthetics, epistemology, and hybrid forms between popular and experimental music.

References

[1] Antoine Caillon and Philippe Esling. Rave: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011*, 2021.

[2] Axel Chemla-Romeu-Santos, Philippe Esling, and Stavros Ntalampiras. Représentations variationnelles inversibles: une nouvelle approche pour la synthèse sonore. In *Journées d'Informatique Musicale 2020 (JIM 2020)*, 2020.

[3] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[4] Soroush Mehri, Kundan Kumar, Ishaan Gulrajani, Rithesh Kumar, Shubham Jain, Jose Sotelo, Aaron Courville, and Yoshua Bengio. Samplernn: An unconditional end-to-end neural audio generation model. *arXiv preprint arXiv:1612.07837*, 2016.

[5] Éric Sadin. *L'Intelligence artificielle ou l'enjeu du siècle: Anatomie d'un antihumanisme radical*. L'échappée, 2018.