



HAL
open science

ciel

Hugo Scurto

► **To cite this version:**

Hugo Scurto. ciel. Proceedings of the 20th International Conference on New Interfaces for Musical Expression (NIME 2020), Jul 2020, Birmingham, United Kingdom. hal-02985956

HAL Id: hal-02985956

<https://hal.science/hal-02985956>

Submitted on 7 Nov 2020

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.

ciel

HUGO SCURTO, Reflective Interaction group, EnsadLab / EnsAD—PSL Research University, France

Additional Key Words and Phrases: deep neural networks, noise, streaming culture, sustainable musical expression

ACM Reference Format:

Hugo Scurto. 2020. *ciel*. 1, 1 (July 2020), 4 pages.

ABSTRACT

On a web page, an oval map continually streams noise sounds and images generated by two deep neural networks. These deep nets were deliberately cracked by myself, using eight years of personal audiovisual archives collected with my smartphone to create low-quality datasets from which the networks would fail to learn. The meditative experience of this perpetual and distributed flow of neural noise invites critical reflection on AI technology in relation to musical expression. On the one hand, the standardized, anonymized, and automated music production processes enabled by deep learning networks could be eventually adopted by mainstream streaming platforms, continuing their eroding of economic bonds between musicians and the public by dematerialising norms of music listening and consumption. On the other hand, the symbolic and aesthetic interpretations of AI-generated media remain in the hands of people, with musicians regaining power over their own data, and communities producing cultural value from the new musical experiences. Rather than as an inescapable technology horizon, *ciel* intends to situate AI within such political struggle for sustainable musical expression.



Fig. 1. Image excerpt of *ciel*.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

Online Installations of the International Conference on New Interfaces for Musical Expression
NIME'20, July 21-25, 2020, Birmingham, United Kingdom

INSTALLATION DESCRIPTION

ciel is an online AI installation offering a critical meditation on AI technology and sustainable musical expression. Its noise aesthetics results from three artistic choices related to AI and music: (1) the recycling of low-quality, personal audiovisual data, (2) the cracking of deep neural networks, and (3) the allegorical displaying in an online “all-sky” map.

The recycling of low-quality, personal audiovisual data

Most of AI applications to music often rely on high-quality, large musical datasets. These datasets often dematerializes the links between musical data and musicians, which raises ethical and legal issues—*e.g.*, the one-million song and lyric dataset recently scraped by *OpenAI* [3]. To reappropriate AI, musicians and researchers explored data customisation practices, for example building accessible interfaces for machine learning—*e.g.*, Rebecca Fiebrink’s *Wekinator* [8]—, or situating data in cultural contexts—*e.g.*, Holly Herndon’s collecting of traditional singing recordings in *PROTO* [10].

I was interested in recycling low-quality, personal audiovisual data as an accessible and material data customisation practice for AI in music. I searched my *BlackBerry*, a mobile productivity device that let me capture poor audio recordings and pictures over eight years of my time—a somehow coherent data collection protocol for building a personal archival dataset. More than their individual qualities as experimental music media, it is their collective aspect that I wanted to explore, to counter AI applications and their perpetuation of digital dematerialisation and high-quality standards [16].

The cracking of deep neural networks

From the mid-twentieth century, many musicians and artists have searched for experimental sounds and images by hijacking music technology, from the manipulation of programming languages and web interfaces—*e.g.*, Sang Won Lee’s live coding of *YouTube* [13]—to the breaking of physical apparatuses generating media—*e.g.*, *Fluxus*’s Wolf Vostell and its *TV Burying* [14]. As an intermediary practice, Caleb Kelly introduced cracking as the exploitation of undesirable behaviors of playback media for artistic expression, often in a critique of the recording music industry [11].

I was interested in hijacking deep neural networks to challenge their promoted applications to music production. Rather than manipulating the parameters of fully-trained nets—*e.g.*, Mario Klingemann’s *neural glitch* approach [12]—, I wanted to crack the learning process of deep neural nets so that they would fail to learn musical representations—while not breaking enough to still generate sound and image. I thus deliberately trained the networks over my low-quality datasets. The generated neural noise reveals the powerlessness of state-of-the-art AI applied to idiosyncratic media.

The allegorical displaying in an online “all-sky” map

Automatic music generation is often framed as the main issue to be addressed with deep neural networks [5]. While more and more studies attempt to evaluate its application to music [17], the application case—music generation—remains barely debated across research communities. In our times where streaming platforms weaken economic bonds between music and the public [7], will a new creativity support tool genuinely help musicians sustain their musical expression? Eventually, could we think of debating such political and infrastructural issues within AI music technology research?

I was interested in summoning the sky symbol—*ciel* in French—to frame these political issues, as an allegory of the modern myth of unlimited progress detached from its material conditions. I thus displayed the generated images inside an oval map, in reference to all-sky maps used for scientific representation of sky radiations [2]. The online broadcasting of the installation eventually makes the dystopian horizon of streaming platforms continually producing AI-generated songs and sleeves accessible for musicians, research communities, and the broad public to debate collectively [4].

TECHNICAL DETAILS

Below are provided details on *ciel*'s dataset specifications, deep neural network training, and online exhibition.

Dataset specifications

I used data collected with a *BlackBerry Curve 9320*, which let me capture a heterogenous set of casual, descriptive, performative, private and public situations within human and nonhuman environments between 2011 and 2018. The smartphone's microphone supported mono recording with a 16,000 Hz sampling frequency. Its 3.15 megapixel camera rendered a 2048 x 1536 pixel resolution with low-quality grain as of mid-light setups. I extracted nearly 3 hours of audio from all voice notes and video records, and cropped all 2,615 camera pictures to a 256 x 256 format for preprocessing. I decided not to label the resulting datasets to cede interpretation of these experimental media to the networks alone.

Deep neural network training

I used a Variational Auto-Encoder (VAE) for audio generation [6] and a Generative Adversarial Network (GAN) for image generation [9]. I used *vschaos*, a library for VAE sound synthesis¹, and a GAN implementation for image generation². As my intention was to only rely on my dataset to crack the networks' learning, I decided to take the networks' architecture and hyperparameters provided in the libraries' templates as granted. Training was led on my laptop's 2.8 GHz Intel Core i7 CPU³, and was interrupted regardless of the networks' losses. The two resulting latent spaces, exhibiting light learning artifacts, enable generation of noise sounds and images—which I propose to call neural noise. Postprocessing solely consisted in normalising sound amplitude to prevent hearing damage due to high-frequency artifacts of the VAE.

Online exhibition

Online implementation was done with *A-Frame*, a web framework for virtual reality experiences⁴. Broadcasting currently consists in playing a new sound and projecting a new image into the oval map every thirty seconds—in reference to the 30-second shortening of songs caused by streaming platforms and their norms of music consumption [1]—from a folder of short sounds and images randomly sampled from the networks' latent spaces. Real-time generation from latent spaces may possibly be implemented by end of July—based on my experience with machine learning and SSH protocol [15]. Online visitors may use their laptop to experience the installation, using headphones or loudspeakers, setting their volume with care in relation to noise. They may click the grey circle at the bottom of page to read a program note.

FEASIBILITY

If accepted, *ciel* would be premiered at this year's *NIME* conference. As such, no documentation of previous deployment is available. Still, a live demo of its current implementation is provided in the next section to demonstrate feasibility.

MEDIA LINK

- Online installation: <http://wolfgang.wang/ciel>

¹https://github.com/domkirke/vschaos_package

²<https://github.com/eriklindernoren/PyTorch-GAN>

³The technical inadequacy of standard CPU-based computers to train deep neural networks—compared to massive GPUs required by AI applications—makes cracking a straightforward practice for artists wanting to inquire the economic and ecological assumptions over which AI is built.

⁴This choice was made for aesthetic coherence with the first part of the installation, *terre*, created for VR: <https://hugoscurto.com/portfolio/ciel-terre/>

ARTIST'S BIO

Hugo Scurto (1993, Marseille, Fr) is a researcher, musician, and designer. His research employs art, design, and science approaches to inquire, develop, and study machine learning technology within an ecology of music. His practice consists in creating interactive sound and music dispositifs in which machine learning seeks to reveal and reconfigure the entanglement of individuals and their material environments. Hugo is currently postdoctoral researcher at *EnsadLab*, and co-founding member of *wolfgang*, a Paris-based music design collective. Before this, he completed a doctoral thesis in Machine Learning and Music Interaction at *IRCAM* (supervised by Frédéric Bevilacqua), and was visiting researcher at *Goldsmiths University of London* (supervised by Rebecca Fiebrink). He graduated in Physics from *Ecole Normale Supérieure Paris-Saclay*, received the MSc in Engineering from *Sorbonne Université*, and trained in Popular Music at *Cité de la Musique de Marseille*. With the support of a diversity of people, Hugo has publicized works and co-organised workshops in international conferences and academic reviews such as *NIME*, *SIGGRAPH*, or *PLOS One*, as well as in cultural places such as *Friche la Belle de Mai*, *Le Cube*, or *Lutherie Urbaine*.

ACKNOWLEDGMENTS

The author would like to thank Carine Le Malet and *Le Cube*, Axel Chemla-Romeu-Santos and *wolfgang*, and Frédéric Bevilacqua and *IRCAM* for their caring support in prototyping, implementing, and submitting this installation.

ETHICAL STANDARDS

ciel was entirely funded by the author. The author certifies that no person appearing in the original sounds and images are recognisable in the sounds and images generated for the online installation. The author knows of no way to reproduce the original sounds and images from the trained networks.

REFERENCES

- [1] 2019. How streaming affects the lengths of songs. <https://www.theverge.com/2019/5/28/18642978/music-streaming-spotify-song-length-distribution-production-switched-on-pop-vergecast-interview>. Accessed: 2020-04-15.
- [2] 2020. Cosmic microwave background. https://en.wikipedia.org/wiki/Cosmic_microwave_background. Accessed: 2020-04-14.
- [3] 2020. OpenAI: Jukebox. <https://openai.com/blog/jukebox/>. Accessed: 2020-05-04.
- [4] Samuel Bianchini. 2015. Towards Organogenesis: For an Instrumental Approach in Research in Art. *Media-N, Research-Creation: Explorations* (2015).
- [5] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet. 2017. Deep learning techniques for music generation—a survey. *arXiv preprint arXiv:1709.01620* (2017).
- [6] Axel Chemla-Romeu-Santos. 2020. *Manifold representations of musical signals and generative spaces*. Ph.D. Dissertation. Sorbonne Université & Università degli studi di Milano.
- [7] Maria Eriksson, Rasmus Fleischer, Anna Johansson, Pelle Snickars, and Patrick Vonderau. 2019. *Spotify teardown*. Mit Press.
- [8] Rebecca Fiebrink, Dan Trueman, and Perry R Cook. 2009. A Meta-Instrument for Interactive, On-the-Fly Machine Learning.. In *NIME*. 280–285.
- [9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- [10] Holly Herndon, Lily Anna, Jenna Sutela, Martine Syms, and Jlin. 2019. *Proto*. 4AD.
- [11] Caleb Kelly et al. 2009. *Cracked media: The sound of malfunction*. MIT Press.
- [12] Mario Klingemann. 2018. Neural Glitch. In *Proceedings of the Machine Learning for Creativity and Design Workshop at NIPS*.
- [13] Sang Won Lee, Jungho Bang, and Georg Essl. 2017. Live coding youtube: organizing streaming media for an audiovisual performance.. In *NIME*. 261–266.
- [14] Nam June Paik. 1963. Exposition of Music–Electronic Television. *Leaflet printed for the show* (1963).
- [15] Hugo Scurto, Wanyu Liu, Benjamin Matuszewski, Frédéric Bevilacqua, Jean-Louis Frechin, Uros Petrevski, and Norbert Schnell. 2019. Entrain: encouraging social interaction in collective music making. In *ACM SIGGRAPH 2019 Studio*. 1–2.
- [16] Hito Steyerl and Berardi Franco. 2012. In Defence of the Poor Image. *E-Flux Journal The Wretched of the Screen*, 244 (2012), 31–45.
- [17] Bob L Sturm, Oded Ben-Tal, Úna Monaghan, Nick Collins, Dorien Herremans, Elaine Chew, Gaëtan Hadjeres, Emmanuel Deruty, and François Pachet. 2019. Machine learning research that matters for music creation: A case study. *Journal of New Music Research* 48, 1 (2019), 36–55.