Machine Learning for Computer Music Multidisciplinary Research: A Practical Case Study

Hugo Scurto, Axel Chemla–Romeu-Santos

Machine Learning for Computer Music Multidisciplinary Research: A Practical Case Study

Hugo Scurto*[1] & Axel Chemla–Romeu-Santos*[2,1]

[1] STMS IRCAM–CNRS–Sorbonne Université
[2] Laboratorio d'Informatica Musicale, Università degli Studi di Milano
{scurto,chemla}@ircam.fr

Abstract. This paper presents a multidisciplinary case study of practice with machine learning for computer music. It builds on the scientific study of two machine learning models respectively developed for data-driven sound synthesis and interactive exploration. It details how the learning capabilities of the two models were leveraged to design and implement a musical interface focused on embodied musical interaction. It then describes how this interface was employed and applied to the composition and performance of *ægo*, an improvisational piece with interactive sound and image for one performer. We discuss the outputs of our research and creation process, and expose our personal reflections and insights on transdisciplinary research opportunities framed by machine learning for computer music.

Keywords: Machine Learning, Interface Design, Composition, Performance, Transdisciplinarity

1 Introduction

Machine learning is a field of computer science that studies statistical models able to automatically extract information from data. The statistical learning abilities of the models induced a paradigm shift in computer science, which reconsidered mechanistic, rule-based models, to include probabilistic, data-driven models. Recent applications of machine learning led to critical advances in disciplinary fields as diverse as robotics, biology, or human-computer interaction. It also contributed to new societal representations of computers through the loosely-defined notion of Artificial Intelligence (AI).

Computer music also witnessed an increased interest in machine learning. Research has mostly been scientific in focus, using and studying models to automatically analyse musical data—*e.g.,* extracting symbolic information related to pitch or timbre from audio data. This led to technical advances in the field of music information retrieval [1], while also benefiting the field of musicology, notably through large-scale computational analysis [2]. In parallel, machine learning also enabled the building of many automatic music generation systems, which are currently being invested by the industry in the wave of AI [3].

* Equal contribution.

Scurto H., Chemla–Romeu-Santos A. (2021) Machine Learning for Computer Music Multidisciplinary Research: A Practical Case Study. In: Kronland-Martinet R., Ystad S., Aramaki M. (eds) Perception, Representations, Image, Sound, Music. CMMR 2019. Lecture Notes in Computer Science, vol 12631.
Author-created version. The final authenticated version is available online at:
https://doi.org/10.1007/978-3-030-70210-6_43.

Importantly, these scientific investigations of machine learning have also enabled the birth of new musical practices. For example, gesture modelling, as a scientific challenge, opened new design perspectives on body-based musical interfaces that adapts to one's way of playing it [4]. Similarly, symbolic sequence modelling created new human-machine improvisational situations where the machine learns to imitate a musician's style [5]. Reciprocally, artistic investigations of machine learning began taking a complementary approach, using the models themselves as material for composition of sound [6] and image [7].

We are interested in adopting a *joint scientific and musical approach* to machine learning research. We are inspired by the computer music pioneer Jean-Claude Risset [8], whose research and creation approach to computer science enabled new scientific understandings of sound as a physical and perceptual phenomenon, jointly with an artistic commitment toward computing aesthetics. His work and personal approach gave insight to both scientists—ranging from formal science to humanities—and artists—ranging from composers and performers to interface designers. Our wish is to perpetuate his multidisciplinary impetus toward contemporary computer music issues related to machine learning.

The work that we present here is a step toward this direction. We led a *scientific* investigation of two machine learning models that jointly frame new data-driven approaches to sound synthesis. We then adopted a *musical* approach toward these models, leveraging their interactive learning abilities to design a musical interface, for which we created an improvisational piece. Rather than seeking general abstractions or universal concepts, our wish was to test these models through a practical case study to develop a personal reflection that inquires, or even challenge, their current applications to computer music. Our hope is that our idiosyncratic research and creation process will help open new perspectives for computer music multidisciplinary research on machine learning.

The paper is structured as follows. We start by the scientific foundations of our work, describing the two models that we developed for two musical issues—sound analysis-synthesis, and sonic exploration. Next, we present the design of our musical interface, describing its embodied musical interaction workflow and implementation. We then describe *ægo*, an improvisational piece with interactive sound and image for one performer, which we created for our interface. Finally, we discuss our research and creation process, and share our personal reflections as computer music practitioners and researchers to draw insight on contemporary machine learning from crossed science, design, and art perspectives.

2 Scientific Modelling

In this section, we describe our two machine learning models, based on *unsupervised learning* and *reinforcement learning*, from a computer science perspective. We explain how they respectively address two specific musical issues: sound synthesis-analysis and sonic exploration.

2.1 Unsupervised Learning for Sound Analysis and Synthesis

Musical Issue. Most sound analysis-synthesis techniques, such as the phase vocoder [9] or the wavelet transform [10], are based on invertible transforms that are independent of the analyzed sounds. Such transforms provide frameworks that can be applied regardless to the nature of the signal, but in return impose a determined structure such that the extracted features are not corpus-dependant. Conversely, could we think about a method retrieving continuous parameters from a given set of sounds, but rather aiming to recover its underlying structure?

Model. The recent rise of *unsupervised generative models* can provide a new approach to sound analysis-synthesis, by considering each item of a given audio dataset $\{\mathbf{x}_n\}_{n\in 1\ldots D}$—here, a collection of spectral frames—as draws from an underlying probability distribution $p(\mathbf{x})$ that we aim to recover. The introduction of latent variables $\mathbf{z}$ allows us to control a *synthesis* process by modelling the joint distribution $p(\mathbf{x},\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$, such that these variables act as parameters for the generative process $p(\mathbf{x}|\mathbf{z})$. The full inference process, that would here correspond to the *analysis* part, leverages the Bayes' rule $p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}$ to recover the distribution $p(\mathbf{z}|\mathbf{x})$, called the posterior.

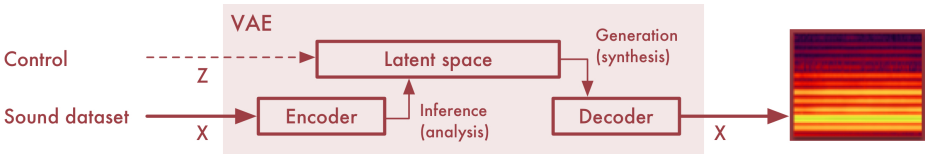

Fig. 1. Unsupervised learning for sound analysis and synthesis. The variational auto-encoder (VAE) encodes a sound dataset into a high-dimensional latent space, which can be parametrically controlled to synthesize new sounds through a decoder.

To improve expressiveness of inference and generation, we propose to investigate variational learning, a framework approximating the true posterior $p(\mathbf{z}|\mathbf{x})$ by a distribution $q(\mathbf{z}|\mathbf{x})$, such that both inference and generative processes can be freely and separately designed, with arbitrary complexity. The variational auto-encoder (VAE) is representative of such methods [11]. In this model (Fig. 1), inference and generation processes are held by two jointly trained separated networks, respectively the *encoder* and the *decoder*, each modelling respectively the distributions $q(\mathbf{z}|\mathbf{x})$ and $p(\mathbf{x}|\mathbf{z})$. The inherent Bayesian nature of variational learning enforces the smoothness of the *latent space*, a high-dimensional, non-linear sonic space, whose parametric dimensions can be freely explored in the manner of a synthesizer.

In related work, we show how this latent space can be regularized according to different criteria, such as enforcing perceptual constraints related to timbre [12]. We refer the reader to the latter paper for technical details on the model and quantitative evaluation on standard sound spectrum datasets.

2.2 Reinforcement Learning for Sonic Exploration

Musical Issue. Sonic exploration is a central task in music creation [13]. Specifically, exploration of digital sound synthesis consists in taking multiple steps and iterative actions through a large number of technical parameters to move from an initial idea to a final outcome. Yet, the mutually-dependent technical functions of parameters, as well as the exponential number of combinations, often hinder interaction with the underlying sound space. Could we imagine a tool that would help musicians explore high-dimensional parameter spaces?

Model. We propose to investigate *reinforcement learning* to support exploration of large sound synthesis spaces. Reinforcement learning defines a statistical framework for the interaction between a learning agent and its environment [14]. The agent can learn how to act in its environment by iteratively receiving some representation of the environment's state S, taking an action A on it, and receiving a numerical reward R. The agent's goal, roughly speaking, is to maximize the cumulative amount of reward that it will receive from its environment.

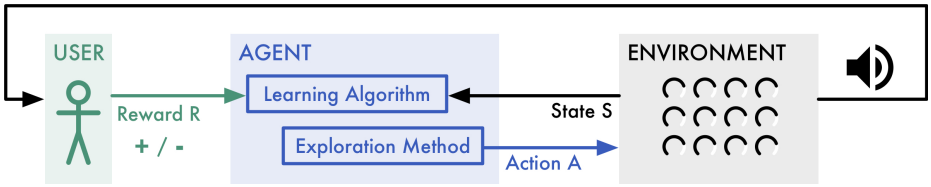

Fig. 2. Reinforcement learning for sonic exploration. The agent learns which actions to take on a sound synthesis environment based on reward given by the musician. The agent implements an exploration method to foster discovery along interaction.

For our case of sonic exploration, we propose that the musician would listen to the agent exploring the space, and teach it how to explore by giving reward data (Fig. 2). Formally, the environment's state is constituted by the numerical values of all synthesis parameters. The agent's actions are to move one of the parameters up or down at constant frequency. Finally, the musician communicates *positive or negative reward* to the agent as a subjective feedback to agent actions. We implemented a deep reinforcement learning model to support learning from human reward signal in high-dimensional parametric spaces [15].

A crucial requirement for reinforcement learning agents is to *autonomously explore their environment*, to keep on discovering which actions would yield the most reward. We developed a statistical method, based on intrinsic motivation, which pushes the agent to "explore what surprises it". The resulting interactive learning workflow was found to be useful to relax musicians' control over all synthesis parameters, while also provoking discoveries by exploring uncharted parts of the sound space. We report the reader to [16,17] for technical details on the model and qualitative evaluation from expert sound designers.

3 Interface Design

In this section, we present our musical interface that combines our two models and leverages their learning capabilities from a design perspective. We describe how interaction design was framed in joint coordination with hardware and software engineering to support embodied musical interaction.

3.1 Interaction Design

Motivation. Our main design motivation was to use our reinforcement learning agent to support musical exploration of high-dimensional latent sound spaces built by our unsupervised learning model.

Specifically, our aim was to exploit the exploration behaviour of our reinforcement learning agent to support *improvisation by feedback* inside the spaces. Instead of acting as a tool, we used machine learning as an expressive partner [5] that would be playable by musicians using positive or negative feedback.

A complementary aim was to use the generative abilities of our unsupervised learning model to support *customization* of synthesis spaces. Instead of accurately modelling sounds, we used machine learning as a creative interface [18] supporting experimentation with the intrinsic non-linearities of latent spaces.

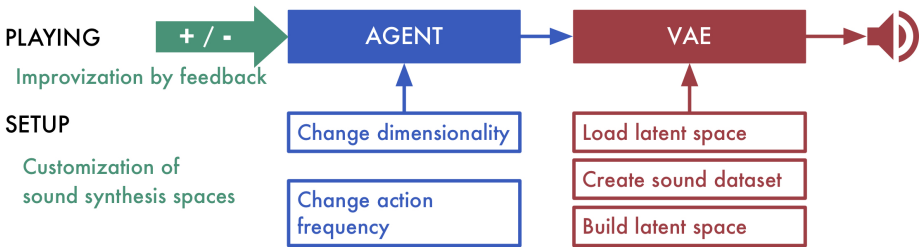

Fig. 3. The interactive workflow that we designed for our interface.

Workflow. We designed a two-phase interactive workflow, shown in Fig. 3.

The *setup* phase allows musicians to configure the interface. They can create a customized sound dataset for the unsupervised learning model, experiment with various training parameters, or also load a previously-built latent sound space. They can also change dimensionality of the reinforcement learning agent to explore specific dimensions of the latent sound space, as well as the frequency at which it would take actions inside the latent space.

The *playing* phase allows musicians to improvise with the agent by means of feedback. The agent produces a continuous layer of sound from the spectrum output of the VAE. Musicians can either cooperate with its learning to attain a sonic goal by giving consistent feedback data. Or, they can obstruct its learning to improvise in sonic exploration by giving inconsistent feedback data.

3.2 Engineering

Implementation. Technically (see Fig. 4), the reinforcement learning agent receives a representation of the environment's state S as a position in the latent space $\mathbf{z}$. Then, it takes an action A corresponding to a displacement along some dimension of the latent space. The resulting position has the unsupervised learning model generate a sound spectrum $\mathbf{x}$. Based on the sound, the musician would communicate reward R to the agent. The latter would progressively learn to explore the latent space in relation to the musician's feedback data.

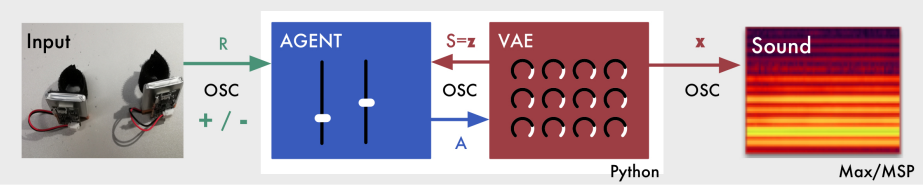

Fig. 4. Schematic representation for the engineering of our interface.

Hardware. We designed a hardware prototype to support embodied musical interaction (see Fig. 4, left). It consists in two velcro rings, each of them equipped with a wireless inertial measurement unit[1]. We took each unit angular rotation about each forearm axis and summed them to compute a single, normalized numerical reward signal. This, combined with the lightweight, nonintrusive velcro rings, lets musicians experiment with a wide range of gesture vocabulary [19] to communicate positive or negative feedback to the agent.

Software. We implemented our two machine learning models as Python libraries[2,3]. We developed a Max/MSP patch to implement a user interface for the setup phase, as well as a hardware data converter for the playing phase. We leveraged the OSC protocol to bridge hardware data, reinforcement learning agent, unsupervised latent space, and sound spectra together into the patch.

4 Musical Artwork

In this section, we present *ægo*, an improvisational piece that we created for our musical interface, premiered at the *14th International Symposium on Computer Music Multidisciplinary Research* on 16 October 2019, in Marseille, France. We describe how its aesthetics intend to challenge current views on AI and music, and detail how composition and performance were handled within our interface.

[1] `http://ismm.ircam.fr/riot/`
[2] `https://github.com/domkirke/vschaos_package`
[3] `https://github.com/Ircam-RnD/coexplorer`

4.1 Description

Intention. Our artistic intention for *ægo* was to emphasize the human learnings that machine learning could enable toward sound and music—rather than the opposite, as is often framed in contemporary AI applications.

We opted for a performance format showing a human and a machine improvising together—respectively using feedback, and an exploration method—to learn to interact with latent sound spaces—on an embodied level for the performer, and on a computational level for the machine. The slow-paced spectromorphologies, synthesized and projected in real-time over the stage and the performer, encourages meditation on this joint human-machine learning.

Crucially, we directed the performance so that the human would progressively relinquish communication of accurate feedback to the machine, thus leaving the machine's learning indeterminate on purpose. Released from the obligation of teaching and controlling its artificial alter ego, the human is allowed to let his or her embodied mind unify with sound, eventually learning to interact with music.

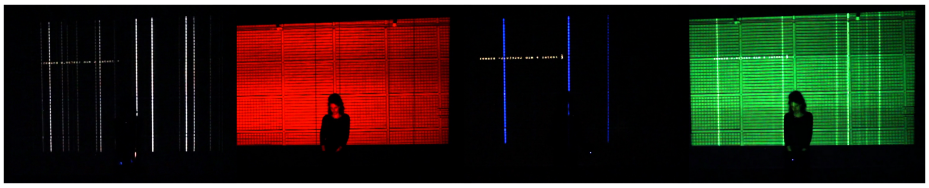

Fig. 5. Pictures taken from *ægo*.

Aesthetics. The piece's aesthetics result from two artistic choices, which conceptually and technically intertwine sound, body, and image (see Fig. 5).

Our first choice consisted in exploiting *artifacts of sound synthesis* produced by the unsupervised learning model to compose unheard-of timbral spaces for the piece. We built latent sound spaces using datasets of sounds that were commonly used in pioneering works of computer music to accentuate audience perception of sonic artifacts produced by learning. In addition, we projected the spectrogram image over the stage and the performer in real-time to provide the audience with a visual representation of artifacts. The blending of sound and performer representations symbolically accounts for the unification of performer and sound.

Our second choice consisted in creating *indeterminacy of composition* using the exploration behaviour of the reinforcement learning agent. We used the performer's body as a symbolic element to communicate kinesthetic information to the audience on how indeterminacy may be experienced while performing with sound. We also added raw textual information on the machine's learning at top left of the projected image to reinforce audience perception of machine's unpredictability. The indeterminacy pushes the performer to relinquish control over the machine's learning to fully focus on sound and its timbral attributes.

4.2 Writing

Composition. The piece was composed at three temporal scales (see Fig. 6).

The first scale is that of *exploration*. It consists in the improvisational paths taken by the reinforcement learning agent in response to performer's feedback data. We set the frequency of agent actions between 30 and 100 milliseconds. This choice resulted in slow and continuous evolution of spectromorphologies, which let the performer improvise at similar temporal scales than the agent.

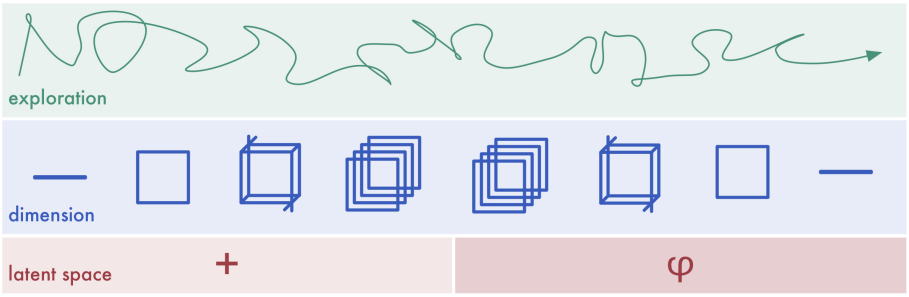

Fig. 6. Temporal structure composed for the piece.

The second scale is that of latent space *dimensionality*. It consists in defining the axis of the latent spaces that the reinforcement learning agent will explore. We set the dimensions to 1, 2, 4, and 8, respectively. This allowed us to write a specific kind of musical form inside the latent space: the more dimensions we open to the agent, the more sonic variance the performer and audience may experience—the harder it may be for the performer to teach the agent.

The third scale is that of latent space itself. It consists in connecting the reinforcement learning agent to another type of latent space. We built two latent spaces using synthesis sounds (additive and frequency modulation) and one using physical instruments recordings (flute, saxophone, piano, violin, bassoon [20]). This enabled us to build a narrative through the use of different soundscapes (here, going from elementary sinusoidal spectra to richer instrumental timbres).

Performance. While the piece is intended to be improvised, our sole direction toward the stage performer is that she or he may perform with the machine with deep attentiveness toward sound[4]. We proposed that the performer would start the piece facing the audience, relaxed, using small wrist rotations to communicate feedback through our interface. As the piece unfolds, the performer would freely adapt its gestures in response to sound, possibly forgetting the presence of the machine, as well as the mapping between gesture and feedback.

A second contributor is required to manage the two remaining temporal scales of the piece—*i.e.,* changing dimensionalities, and switching latent spaces.

[4]See these video excerpts from early rehearsals: `https://vimeo.com/418787133`

5 Discussion

In this section, we discuss our research and creation process, starting by providing contextual information about our case study. We then expose our personal reflections emerging from music practice with machine learning, and present insight for future multidisciplinary and transdisciplinary approaches to computer music practice and research.

5.1 Contextual Information about our Case Study

Process. The work presented here relates a practical case study with machine learning in the frame of computer music. We leveraged both conceptual and technical aspects of machine learning to jointly produce *scientific knowledge* with our two models for sound synthesis, as well as *musical creations* through the design of our interface and the writing of our improsivational piece. In this sense, our work emerged from a research and creation process, in which we closely articulated a creation project within a research methodology.

We followed a sequential multidisciplinary agenda (see Fig. 7, solid lines and arrows). We started by the scientific modelling of sonic exploration and sound synthesis, which took us two years to date. We then planned a one-month residency to design the interface and write the musical piece. This research and creation agenda was opted for because our work occupation at that time—doctoral researchers in machine learning applied to computer music—required a more important focus on computer science than on music creation.

While many researchers of our laboratory were involved in scientific modelling, we (the two coauthors) managed interface design and musical artwork as a pair. Importantly, both of us have professional experience in music composition and performance, and followed a dual training in science and music. These dual skills were central to individually work, as well as to effectively collaborate, on conceptual and technical aspects related to music and machine learning throughout the process.

Output. The relatively short period dedicated to music creation pushed us to take pragmatic decisions about the form of outputs, notably by relinquishing certain technical developments. For example, using the unsupervised learning model to learn temporal features of sound spectrums could have improved the dynamical richness of the generated sounds, as well as supported other musical forms than slow spectral evolution. Other agent parameters could have been used to create quicker or more discontinuous exploration behaviours, as well as other commands than feedback data to give the performer more control over reinforcement learning. Finally, many other musical forms could have been conceived, using other sound datasets—*e.g.,* voice corpora or environmental sounds—and investigating other temporal writings for dimensionality and exploration. Future continuation of our work may consider addressing these research questions to evolve the created outputs. Meanwhile, we do believe that interesting insights have already thrived out of the present case study.

5.2 Authors' Personal Reflections on Practicing with Machine Learning for Computer Music

Beyond the created outputs, the process of practicing with the two machine learning models gave us insight on the particular interests that they may have for computer music. In the next two sections, we successively share our personal reflections on composing with unsupervised learning (Axel Chemla–Romeu-Santos), and performing with reinforcement learning (Hugo Scurto). We use first-person narratives to make it clear that our personal approaches as musician-researchers will be exposed here, rather than general analyses or evaluations.

Axel Chemla–Romeu-Santos. *(On composing with unsupervised learning.)* The topic of my doctoral work, initiated in September 2016, targeted the investigation of machine learning-based generative models as a novel method of sound synthesis. This project was innovative, as most approaches developed so far were mainly focused on symbolic generation, due to the challenging density of audio signals. However, such symbolic approaches were rather aiming to model specific genres or authors and had, to my opinion, modest creative interests and ambiguous motivations. We decided to rather address the generation of audio signals, positing that the high-capacity modelling capacity of neural networks could disclose a novel approach with sound synthesis, nonconflicting with existing musical practises. This postulate hence enforced the use of representation-based methods, such as variational auto-encoders, allowing to directly control the generation through higher-order features, used as automatically extracted synthesis parameters (contrary to systems like adversarial methods, whose generation were initially only based on sampling). This choice was also partly inspired by my parallel practice of composition in electroacoustics, where I discovered among various composition processes (fortunately non-exclusive) the specificity of what I would call an *experimental* approach. This approach can be described by focusing on a physical (mechanical, analogical, digital...) or abstract (symbolic systems, generation rules...) object, and realizing them into whether compositions (hence allowing an iterative workflow, delineating composition and realization times) or performances (entangling composition and execution time, emphasizing the reflective interaction between involved agents). This approach, mandatory for the research and creation process I was coveting, drove my activity during the three years of the doctoral work.

This positioning, jointly with the musical interest aroused by the development of these methods, motivated simultaneously the writing of this paper and the composition of the piece. Hence, using these models to conceptualize a musical performance raised two ontological questions: first, how to *compose* with the developed models (distributing musical elements through time), and how to *interact* with it. Hugo and I quickly drew the conclusion after some initial experiments that the architecture of the variational auto-encoding system presented an inner *explorational* creativity (in the sense of Boden [25]), proposing a generative space that could be interestingly navigated by an agent (human, machine, or hybrid as we did in our performance). Hence, we chose to let the navigation of the latent space to the performer (Hugo), the compositional aspect then consisting in the dynamical determination of the performance *frame*. Therefore we had to split architectural decisions between, from the one hand the *free parameters* that

can be handled through time, and from the other hand the *fixed parameters* that are kept fixed among the performance. We left free the decisions that we found most decisive for both the diversity and the morphology of the produced output: the amount of explored dimensions, that had a direct impact on the complexity of the space (and hence on the performer's choices), and the explored models, trained on different datasets and then providing different spectromorphologies. This step, quite common in experimental music (that we can call *setup design*), then drove the subsequent experiments about the precise composition of the piece. We adopted a recursive compositional process, first by exploring generative spaces and several projections "at hand", and then including the navigation with the exploratory agent. This procedure naturally led us to a distribution of live actions between the performer and an operator, setting the refreshment rate of the agent and triggering the transitions between the successive episodes, then amounting to a three-agent improvisational setup. This choice has been made to extend the flexibility of the piece, allowing to dynamically adapt the frame of the improvisation to its realization, but also to face hypothetical technical issues arising from the prototype interface.

Hence, the compositional process adopted in this piece was rather close to the *experimental* method I described, first crafting models that were trained on different datasets, exploring their properties jointly with the performer, and giving the composition a macro-structure distributing in time the parameters considered as the most determining for the performance. If we analyze the shift that recent machine learning techniques proposed in the domain of scientific knowledge, that we can describe as modelling functions by with automatic determination techniques rather than an explicit formulation of targeted dependencies, what would mean the transposition of this shift in musical practises? Clearly, our work is more based on the *objectisation* of these techniques for its use in existing musical paradigms (that I call here experimental), rather than a compositional processed based on automatic generation of musical content through high-level attributes. I think that this question would be very interesting to investigate more deeply into artistic and scientific communities, in order to reconcile "AI-luthery" with "high-level composition" approaches.

Hugo Scurto. *(On performing with reinforcement learning.)*

> Then the answers, instead of coming from my likes and dislikes, come from chance operations, and that has the effect of opening me to possibilities that I hadn't considered. Chance-determined answers will open my mind to the world around. (John Cage, 1982 [22])

Rather than a fortunate introduction, this quote on composition and indeterminancy by John Cage actually embodies my very own reflections on performing with reinforcement learning—that is, switching from instrumental control of sound to *spiritual unification with music*. These reflections drove the artistic direction of our musical artwork—showing a human favouring unification with sound over the control of a machine's learning—, and were further fostered through improvisational practice with reinforcement learning. Below is an attempt to describe how these reflections progressively crystallised for me through experimentation within the setup designed with Axel.

Reinforcement learning enables humans to interact with sound using positive or negative feedback—a standardized form of likes and dislikes. The agent may explore and learn how to synthesize sound based on this feedback data, eventually providing humans with a certain degree of instrumental control over sound. In relatively small parameter spaces (for example the one- and two-dimensional spaces composed for our musical artwork), I was able to rapidly teach the agent my preferences toward sound, and gain control over the synthesis process. In spaces of higher dimensions (where the agent needed more feedback data to properly learn to behave), I was not necessarily able to tell whether I could teach the agent, or if it was acting by chance toward a desired sound—thus convincing myself of having some influence, instead of control, over sound synthesis.

This "mind game", as I would call it, pushed me to open my expectations as a performer away from gaining instrumental control over sound. I began mindfully listening to timbral attributes of generated sound, as timbre was the only clue for me to actually know if the agent was learning from my likes and dislikes. Entering this state of heightened listening, I observed myself oscillating between two mental postures toward sound: one that was performative—where I attempted to grasp control over timbre by producing very precise feedback—, and one that was meditative—where I carefully listened to sound as if it existed by itself, detached from my very own influence. In both cases, heightened listening almost had me forgetting about the technicality of the agent for the benefit of sound and its timbral attributes. This mental exercise eventually freed my physical movements from the task of being performative toward feedback, which unexpectedly let me contemplate new bodily sensations in relation to timbre over time—such as the apparent interdependence between my inner breathing motions and the perpetual unfolding of sound.

The enabling of these mental and physical practices by reinforcement learning paved the way, I believe, to a spiritual practice that I regularly undertake within musical performance, which I may refer to as *unification with music*. Unification with music seeks to relinquish instrumental control of sound in performance and cultivate awareness that its organisation over time is already part of one's self—echoing Cage's definition of music as an "affirmation of [the very] life" that we are living [23]. Of course, unification with music may be witnessed and practised through performance and improvisation with many other interactive music systems. However, I would argue that the intrinsic operations of reinforcement learning facilitate awakening in unification with music, compared to the logical, verbal, and embodied operations conventionally used in interactive music systems—*e.g.,* parametric, note-based, or gestural control of sound synthesis.

By releasing my mind from technical conventions, feedback allowed me to experiment with basic forms of nonverbal communication with sound. Reflecting on the symbolic and performative aspects of these communication forms, I ended up thinking of them as *invocation rituals for sound*, which may be characterized by the following sequence: first, focusing the mind on timbral attributes, then using the body to summon acoustic presence, and eventually letting one's self identify with sound. In a complementary manner, by systematically yet unpredictably responding to my acts of communication, the reinforcement learning agent—*i.e.,* its algorithmic operations and

exploration methods—helped me awaken to the *affirmation of an external agency* in the organisation of sound over time. Assuming that reinforcement learning remained a tool for performing a piece of music, I learned from this awakening that music *was* that actual affirmation of agency. Through invocation of a rapport with this agency—*i.e.,* through feedback on sound synthesized by the agent—I was able to witness and cultivate unification with music in ways I had not experienced in performance yet.

On a lighter note and to come full circle, I must agree that performing with reinforcement learning certainly opened my mind to the world around.

5.3 Insight for Computer Music Transdisciplinary Research

Our personal reflections gave us insight on the artistic, design, and scientific aspects of computer music research on machine learning (see Fig. 7, dashed arrows), whose multidisciplinarity may be rethought as transdisciplinary.

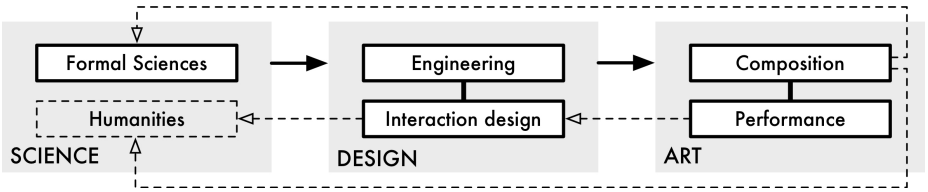

Fig. 7. Our case study. Solid arrows: The sequential research and creation process that we took to lead multidisciplinary research on machine learning. Dashed arrows: Insight gathered for a transdisciplinary approach in the frame of computer music.

Critical Music Practices with Machine Learning. Composition and performance of our musical artwork (see Section 4) allowed us to challenge current applications of machine learning to computer music, in a way that could have not been experimented within the standard scientific approach. Many applications of AI to music arguably seek to produce impressive results in terms of automatic generation of music, often leaving musical and aesthetic aspects behind. Conversely, our unconventional use of machine learning in our improvisational piece was intentionally deceptive toward these shared expectations. We deliberately composed with sound synthesis artifacts produced by unsupervised learning, as well as with the indeterminacy of reinforcement learning, to realise novel musical forms linked to our personal spiritualities before seeking to obtain innovative scientific results through the lens of machine learning. Also, we purposefully displayed a music performer progressively relinquishing control over a machine's learning to promote attentive musical listening over the fast-paced quest for technological progress typical of many contemporary AI applications to music. Our artistic choices could thus be described as *critical music practices with machine learning*, inquiring the musical representations and experiences that the formalism of machine learning models may implicitly encapsulate.

Intrinsic Design of Machine Learning for Music. Designing our musical interface let us reflect on our peculiar design approach to machine learning for music (see Section 3). Standard engineering of machine learning usually employs quantitative evaluation frameworks, mostly focused on measuring a model's performance regarding a set of explicit tasks, generally also involved in the training—and then raising legitimate suspicions about their intrinsic tautology. Such evaluations, that we call *extrinsic*, tend to prune out the emergent behavior of the trained system in favor to a measurable idea of efficiency, hence denoting a certain statistical materialism that is regularly castigated in this new trend of computer science. At the opposite, our use of latent spaces as customizable sound spaces, as well as our use of feedback as modality for improvisation, rather employed the *intrinsic* properties of such models, hence redefining their original purpose. Such qualitative, creativity-oriented evaluations targeted different interaction design properties, detached from the idea of measurable efficiency, but rather fostering high-level attributes—*e.g.,* expressiveness, compliance, richness, or empowerment. While marginally investigated so far within machine learning engineering, these interactive properties are actually substantially solicited within computer music design, such as in gesture modelling and symbolic sequence modelling applications to music practice. Our musical interface could thus be related to such an *intrinsic approach to the design of machine learning for music*.

The Formal and Humanistic Dimensions of the Sciences of Computer Music. In the present case study, we took a multidisciplinary approach to machine learning, successively assuming the roles of scientists, engineers, designers, and musicians along research. As a consequence, we do not pretend to provide a formal, quantitative, or universal evaluation of machine learning for computer music, as we did in our two scientific modelling studies (see Section 2). Rather, we believe that our research approach does constitute one example of machine learning research led by specific computer music practitioners—a complementary type of qualitative and humanistic evaluation, perhaps sharing similarities with the joint scientific and musical approach to computers of Jean-Claude Risset [24]. We hope that the present paper convinced the reader of our diligence toward switching these roles and approaches throughout research and creation.

More generally, we believe that this multidisciplinary approach to machine learning could be likened to a *transdisciplinary* approach to computer music research, considering the current social and industrial context surrounding AI. Historically, multidisciplinary collaboration between engineers and musicians has enabled discoveries and innovations that jointly benefited scientists and computer musicians [8]. Nowadays, rapid advances in digital technology—especially in machine learning engineering—put strong infrastructural pressures on computer musicians, arguably not leaving substantial time for equitable scientific and musical contributions as framed by standard multidisciplinary collaboration.

As researchers in computer science upon leading this case study, we took a modest step toward countering this trend, by letting our computer music practices and personal reflections reassign the scientific ontology of machine learning models, possibly at the expense of standard evaluation approaches of computer science and engineering.

Without depreciating nor seeking to relinquish multidisciplinary collaboration at all, we believe that such transdisciplinary approaches are increasingly becoming crucial nowadays, not only to build new practices for the development and evaluation of machine learning models, but also to construct a collective discourse about these technologies that critically considers their ecological integration in human practices—and philosophically speaking, a phenomenological understanding of their behavior. We hope that these insights will resonate with other computer music practitioners and researchers wishing to further contemporary cultivation of the *formal and humanistic dimensions of the sciences of computer music*.

6 Conclusion

We presented a practical case study of machine learning for computer music. We studied two machine learning models, from which we designed a musical interface, and wrote a musical piece for it. We discussed our research and creation process and our personal reflections and insight as computer music practitioners and researchers. Future work may explore transdisciplinary music research approaches that complement computer music multidisciplinary collaboration.

Acknowledgements

We thank Frédéric Bevilacqua, Philippe Esling, Gérard Assayag, Goffredo Haus, and Bavo Van Kerrebroeck for their broad contributions to scientific modelling.

References

1. Hamel, P., & Eck, D.: Learning features from music audio with deep belief networks. In 11th International Society for Music Information Retrieval Conference (2010)
2. Meredith, D. (Ed.): Computational music analysis (Vol. 62). Berlin: Springer (2016)
3. Briot, J-P., Hadjeres, G., and Pachet, F.: Deep learning techniques for music generation-a survey. arXiv preprint arXiv:1709.01620 (2017).
4. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Gudy, F., & Rasamimanana, N.: Continuous realtime gesture following and recognition. In International gesture workshop, Springer, Berlin, Heidelberg, pp. 73-84 (2009, February).
5. Assayag, G., Bloch, G., Chemilier, M., Cont, A., & Dubnov, S.: Omax brothers: a dynamic topology of agents for improvization learning. Proceedings of the 1st ACM workshop on Audio and music computing multimedia (2006)
6. Ghisi, D. Music across music: towards a corpus-based, interactive computer-aided composition. Doctoral dissertation, Paris 6 (2017)
7. Akten, M., Fiebrink, R., & Grierson, M.: Deep Meditations: Controlled navigation of latent space. Goldsmiths University of London (2018).
8. Risset, J.-C.: Fifty Years of Digital Sound for Music. In: Proceedings of the 4th Sound and Music Computing Conference (SMC) (2007)
9. Rodet, Xavier and Depalle, Philippe and Poirot, Gilles : Speech analysis and synthesis methods based on spectral envelopes and voiced/unvoiced functions. European Conference on Speech Technology (1987)

10. Kronland-Martinet, R.: The wavelet transform for analysis, synthesis, and processing of speech and music sounds. Computer Music Journal, 12(4), 11-20 (1988)
11. Kingma, D., Welling, M. : Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
12. Esling, P., Chemla–Romeu-Santos, A., Bitton, A. : Bridging audio analysis, perception and synthesis with perceptually-regularized variational timbre spaces. DAFx2018 (2018)
13. Ystad, S., Aramaki, M., & Kronland-Martinet, R.: Timbre from Sound Synthesis and High-level Control Perspectives. In: Siedenburg, K., Saitis, C., McAdams, S., Popper, A.N., Fay, R.R. (eds.) Timbre: Acoustics, Perception, and Cognition. SHAR, vol. 69, pp. 361389. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-AQ3 14832-4 13
14. Sutton, R. S., & Barto, A. G.: Reinforcement learning: An introduction. MIT press (2018)
15. Warnell, G., Waytowich, N., Lawhern, V., & Stone, P.: Deep TAMER: Interactive agent shaping in high-dimensional state spaces. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018, April).
16. Scurto, H., Bevilacqua, F., & Caramiaux, B.: Perceiving Agent Collaborative Sonic Exploration In Interactive Reinforcement Learning. In: Proceedings of the 15th Sound and Music Computing Conference (SMC) (2018).
17. Scurto, H., Van Kerrebroeck, B., Caramiaux, B., Bevilacqua, F.: Designing Deep Reinforcement Learning for Human Parameter Exploration. ACM Trans. Comput.-Hum. Interact. (TOCHI) 28(1), 135 (2021)
18. Fiebrink, R., Caramiaux, B., Dean, R., & McLean, A.: The machine learning algorithm as creative musical tool. Oxford University Press (2016)
19. Tanaka, A., & Donnarumma, M.: The body as musical instrument. The Oxford Handbook of Music and the Body (2018)
20. Ballet, G., Borghesi, R., Hoffmann, P., & Lvy, F.: Studio online 3.0: An internet "killer application" for remote access to Ircam sounds and processing tools. In: Journées d'Informatique Musicale (JIM) (1999)
21. Chowning, J. M.: The synthesis of complex audio spectra by means of frequency modulation. Journal of the audio engineering society, 21(7), 526-534 (1973).
22. Montague, S.: John Cage at Seventy: An Interview. American Music (1985): 205-216.
23. Cage, J.: Experimental music. In: Silence: Lectures and Writings, vol. 7, p. 12 (1961).
24. Risset, J. C., & Wessel, D. L.: Exploration of timbre by analysis and synthesis. In: The psychology of music, Academic Press, pp. 113-169 (1999)
25. Boden, Margaret A.: Computer models of creativity. In: AI Magazine, 30(3) (2009)