MUSICKING DEEP REINFORCEMENT LEARNING

Hugo Scurto
EUR ArTeC / Université Paris 8 / EnsadLab
Paris, France
hugo.scurto@ensad.fr

ABSTRACT

In this paper, I relate an auto-reflexive analysis of my practice of designing and musicking deep reinforcement learning. Based on technical description of the Co-Explorer, a deep reinforcement learning agent designed to support sonic exploration through positive or negative human feedback, I discuss how deep reinforcement learning can be seen as a form of sonic comprovisational agent, which enables musicians to compose a parameter sound space, then to engage in embodied improvisation by guiding the agent through sound space using feedback. I then relate on my own musicking experiments led with the Co-Explorer, which resulted to the creation of the *ægo* music performance, and build on these to sketch a music representation for deep reinforcement learning, highlighting its original aesthetics, as well as its ontological shifts between performer and agent, and epistemological tensions with engineering-oriented representations. Rather than discrediting the latters, my wish is to create space for practice-based approaches to machine learning in a way that is complementary to engineering-oriented approaches, while contributing to further music representations and discourses on artificial intelligence.

1. INTRODUCTION

Reinforcement learning defines a computational framework for the interaction between a learning agent and its environment [1]. The framework provides a basis for agents that learn an optimal behaviour within their environment by taking actions in it, then receiving positive or negative feedback from it, as a reward or punishment signal. Recent advances in deep learning enabled reinforcement learning to be applied to high-dimensional spaces, through the so-called deep reinforcement learning framework [2]. Such a framework actively contributed to the growing field of artificial intelligence, with application domains ranging from robotics and finance to healthcare and science [3].

Deep reinforcement learning was recently explored in the domain of music. Kotecha used deep reinforcement learning to generate symbolic polyphonic music [4]. Karbasi *et al.* explored deep reinforcement learning to create rhythms for a collective of interactive robots [5]. Ramoneda *et al.* applied deep reinforcement learning to learn optimal piano fingerings based on simulated piano performances [6]. Yet, all these works privileged an engineering-oriented approach to deep reinforcement learning, using it as a computational model for existing symbolic music representations. In addition, they did not include musicians in the research and design of these generative models, leaving both analytic and performative aspects of music practice aside.

As a musician, designer and researcher, I was interested in adopting a design-oriented approach to deep reinforcement learning. I was especially interested in exploring novel forms of musicking where a deep reinforcement learning agent would learn interactively from a musician, that is, by receiving positive or negative feedback from them. I was expecting that such a creative process could in turn lead to new designs and representations for deep reinforcement learning that originate from music practice as much as from engineering. I was notably inspired by previous works from Bevilacqua *et al.*, who pioneered such interactive approaches to machine learning for gestural control of sound [7], and by Fiebrink *et al.*, who highlighted musical attributes of machine learning by leading in-depth studies of the creative process of musicians creating gesture-sound mappings with machine learning [8].

In this paper, I relate an auto-reflexive analysis of my practice of designing and musicking deep reinforcement learning. In Section 2, I describe the Co-Explorer, a deep reinforcement learning agent that supports sonic exploration based on positive or negative human feedback, designed in collaboration with sound designers. In Section 3, I discuss how deep reinforcement learning may be seen as a form of sonic comprovisational agent, enabling musicians to compose a parameter sound space, then to engage in embodied improvisation by guiding the agent through sound space using feedback. In Section 4, I relate on my own musicking experiments with the Co-Explorer, which resulted in the creation of *ægo*, a music performance for one human improviser and one learning machine, presented at this year's TENOR music track. I end by discussing in Section 5 how musicking deep reinforcement learning helped me sketch a music representation for this computational framework, highlighting the epistemological, ontological and aesthetic shifts produced by musicking compared to its standard, engineering-oriented applications. Rather than discrediting the latters, my wish is to create space for practice-based approaches to machine learning in a way that complements engineering-oriented approaches, with the hope that it will contribute to further music representations and discourses on artificial intelligence.

2. CO-EXPLORER

In this section, I describe the Co-Explorer, a deep reinforcement leaning agent designed to support sonic exploration based on positive or negative feedback provided in real-time by a musician. The Co-Explorer was developed as part of my doctoral thesis, which sought to approach machine learning as design material in the context of new interfaces for musical expression [9]. Specifically, we adopted a human-centred design approach to deep reinforcement learning, involving sound designers in diverse steps of our design process, and studying their creative processes with our software agent [10]. The next sections describe technical foundations of deep reinforcement learning, and more specifically, the exploration method and interaction modalities that we developed within the Co-Explorer [1]. I refer the reader to my previous papers for technical details and qualitative evaluation of implementation.

2.1 Interactive Deep Reinforcement Learning

Deep reinforcement learning is a generic computational framework for the interaction between a learning agent and its environment. Our first design step thus consisted in defining a model of the environment and the agent that could be adapted to the use case of sonic exploration.

We opted for an elementary model of the environment, consisting of a parameter space of arbitrary dimension (*e.g.,* a synthesis space). Technically speaking, let $\mathcal{S} = \{S\}$ denote the state space constituted by all possible parameter configurations $S = (s_1, ..., s_n)$ reachable by the agent, with n being the number of parameters, and $s_i \in [s_{min}, s_{max}]$ being the value of the i^{th} parameter living in some bounded numerical range. Let $\mathcal{A}(S) = \{A\}$ denote the corresponding action space as moving up or down one of the n parameters by one step a_i, except when the selected parameter equals one boundary value. The resulting agent would thus iteratively explore the parameter space while producing continuous sound synthesis variations.

Crucially, we assumed that a musician observes the state-action trajectories taken by the agent in real-time, and interactively provides positive of negative feedback, or reward R, to the agent. As such, the agent would progressively learn a mapping between states and actions, leveraging deep learning to tackle learning and generalisation in high-dimensional parameter spaces. The resulting trained model can be used as a representation of a musician's subjective preferences toward a parameter space.

2.2 Exploration Method

In addition to learning musician's preferences, reinforcement learning agents have a second aim, which is to maximise feedback received from the musician. As such, they may help musicians find the best state-action in the parameter space as they explore it.

To do so, agents rely on exploration methods that enable them to find optimal state-action trajectories in their environment. Intuitively speaking, an agent has to balance exploitation of their computational knowledge (*e.g.,* taking the best actions as defined by previous feedback to maximise future feedback) with exploration of their environment (*e.g.,* taking sub-optimal actions in terms of previous feedback, possibly leading to better actions in the future).

For the Co-Explorer, we developed a novel exploration method that builds on an intrinsic motivation technique, which pushes the agent to "explore what surprises it". Specifically, it has the agent direct its exploratory actions toward uncharted parts of the space, rather than simply making random moves, as in most reinforcement learning [1].

Thus, deep reinforcement learning agents may have a dual role: on the one hand, their learned model can be used as a representation of a musician's subjective preferences toward a parameter space; on the other hand, their exploration behaviour can be used to foster the creative process of a musician toward some parameter space.

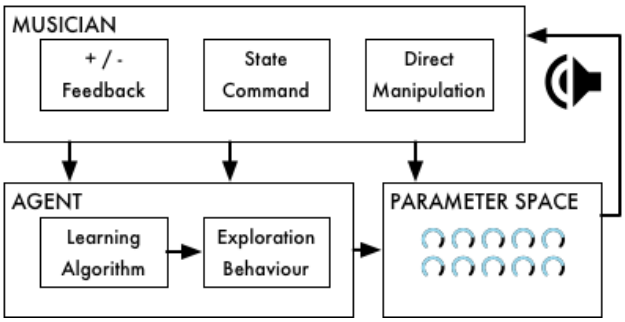

Figure 1. Co-Explorer workflow.

2.3 Interaction Modalities

Our next step consisted in designing interactions with deep reinforcement learning to let musicians experiment with both its learning and exploration abilities. We collaborated with sound designers to iteratively design and implement these interactions within the Co-Explorer (see Figure 1).

The first interaction modality is positive or negative feedback. We distinguished guiding feedback, which enables to provide feedback toward actions taken by the agent, and zone feedback, which enables to provide feedback toward states reached by the agent. While each type of feedback relies on a different implementation, they both consist of a scalar with continuous positive or negative value.

The second interaction modality is state commands. State commands enable to control the agent's trajectory more directly, that is, without relying on feedback. A first state command is changing zone, which enables to command the agent to make an abrupt jump to an unexplored parameter state. Another state command is start/stop autonomous exploration mode. In autonomous exploration mode, the agent takes actions in the parameter space at a regular time interval, whether the musician provides feedback (thus learning in real-time), or not (thus relying on its learned model and exploration behaviour).

The third and last interaction modality is direct parameter manipulation. It enables to explore the parameter space by hand, as in most sound synthesis workflows. Additionally, it enables to choose a given parameter state from which the agent would start its autonomous exploration.

[1] https://github.com/Ircam-RnD/coexplorer

3. DEEP REINFORCEMENT LEARNING AS SONIC COMPROVISATIONAL AGENT

In this section, I discuss how deep reinforcement learning may be seen as a form of sonic comprovisational agent, enabling musicians to compose a parameter sound space, then to engage in embodied improvisation by guiding the agent through sound space using positive or negative feedback. I detail possible strategies to compose such parametric sound spaces, as well as possible configurations for musical improvisation through positive or negative feedback, which let me argue for deep reinforcement learning as a form of technology for comprovisation [11].

3.1 Composing Sound Spaces

Composing a sound space consists in defining the timbral features and diversity of sounds to be produced by the learning agent. Sound spaces are context-independent, and may be fixed before interacting with the agent. As we will see in Section 4.1, they may be successively explored by the agent to create dramaturgy along a musical piece.

Technically speaking, composing sound spaces consist in linking the parameters of the model to parameters of a sound synthesis engine. Concretely, one may first choose one given synthesis engine, then curate n synthesis parameters from it, and set the numerical bounds within which the agent would lead exploration. As the environment's model is generic, one may connect the Co-Explorer to any parameter sound synthesis engines, including commercial VSTs, physically-inspired sound synthesis, descriptor-based sound synthesis, or custom Max/MSP patches. The resulting sound morphologies would continuously evolve across time, as the agent would iteratively take actions on parameters and thus reach new parameter states.

While the Co-Explorer was initially designed to explore sound spaces, the genericity of its environment model makes it theoretically applicable to other music representations. For example, the Co-Explorer was used to explore rhythmic structures, by approaching states as discrete rhythmic patterns of size n, and agent actions as activations or deactivations of beats within the pattern [12]. Other musical applications could lie in the creation a chord parameter space, and have the agent learning to modify note combinations.

3.2 Improvising Through Feedback

Beyond sonic exploration for sound design, I argue that deep reinforcement learning opens new approaches for musical improvisation due to its relying on positive or negative feedback. Below I detail how feedback may be used as a contingent element of a performance, supporting real-time instructions toward sound, symbolic communication with sound, and embodied responses toward sound, all contributing differently to the agent's learning.

3.2.1 Feedback as Instructions Toward Sound

A first musical use of feedback follows that which is technically defined by deep reinforcement learning: namely, enabling performers to provide instructions toward sound to guide the agent's learning and exploration of the space.

As a generic, scalar value, feedback may be directed toward various dimensions of sound. For example, positive or negative feedback may be used to evaluate timbral attributes of sound, so that the agent learns a model of timbre from its parameter environment. Or, feedback may be used to communicate subjective preferences toward sound, so that the agent learns a model of the composer's or performer's tastes toward sound.

In both cases, feedback-based instructions toward sound may be fixed before the performance by the composer. In this case, improvisation would be led by the agent, essentially through its exploration behaviour, while the performer would communicate accurate feedback to teach the agent to reach some goal sound fixed by the composer.

Alternatively, such instructions toward sound could be opted for in real-time by the performer. In this case, improvisation would be essentially led by the performer as they would guide agent exploration in real-time through feedback. Specifically, the performer may use feedback to convey spontaneous subjective preferences toward sound, or rely on some sonic scenario, decided before, or emerging from, improvisation, to guide agent exploration. In this case, the reaching of a goal sound may both depend on accurate feedback provided by the performer, as well as on agent learning and exploration of the parameter space.

3.2.2 Feedback as Symbolic Communication With Sound

Rather than sound-oriented instructions, feedback may be reappropriated by the performer to communicate with sound at a symbolic level. For example, a performer may use positive or negative feedback to express personal semantics or imagery toward sound, rather than to evaluate timbral features of sound. In this case, the performer might start to imagine that they are controlling sound production, even if the agent may not be able to properly learn such a high-level representation. Alternatively, a performer may consciously communicate contradictory feedback as a way to hijack the agent's learning, and thus, its trajectory in the sound space. In this case, the performer may have no pre-conceived scenario toward improvisation, except that of discovering unexpected sounds, due to the agent's struggling in interpreting the performer's feedback.

3.2.3 Feedback as Embodied Response To Sound

In addition to instructions or symbolic communication, feedback may be produced by the performer as an embodied response to sound generated by the agent. For example, a performer may produce feedback involuntarily, as errors toward instructions provided by a composer, or as an emotional response toward timbral or symbolic features of sound. Or, a performer may produce feedback to expressively accompany sounds generated by the agent, in a way similar to ancillary gestures produced by musicians with their instruments [13]. In the latter case, the performer may approach feedback as an abstract thread that connects them with the agent, thus creating space for expressive improvisation with sound, in a way similar to dance, where movements that accompany music can lead performers to feel that they have control over sound production [14].

3.3 Comprovising with Deep Reinforcement Learning

I believe that the combination of context-independent with contingent elements makes deep reinforcement learning a new technology for computer-based comprovisation. In its current formalisation, deep reinforcement learning highlights sound listening as a main feature, where it be in the composition of sound spaces, or during improvisation with the agent. Its second feature is the enabling of musical improvisation through a high-level communication channel, that is, positive or negative feedback, which can be used as either an indirect control modality toward sound generation (in the case of instructions and symbolic communication), or as a direct engagement modality with sound (in the case of symbolic communication and embodied responses). While recent, the framework was already explored by other musicians, specifically, to compose and improvise with musical gestures [15].

4. MUSICKING EXPERIMENTS

In this section, I relate my own musicking experiments led with deep reinforcement learning, made in collaboration with composer-researcher Axel Chemla–Romeu-Santos between 2019 and today. They resulted in the creation of *ægo*, a music performance for one human improviser and one learning machine. The piece was performed one time in 2019 [16]; we produced a reworking in 2022, which we will premiere at this year's TENOR music track.

ægo started by the wish to experience comprovisation with the Co-Explorer, possibly leading to the discovery of alternative music representations for deep reinforcement learning. We adopted a practice-based approach to the Co-Explorer, that I propose to describe as musicking [17], since it essentially relied on listening to, and performing with, the sounds and music produced by deep reinforcement learning, without assuming any pre-established musical form. The next sections details the compositional, improvisational, and comprovisational experiments led through *ægo*. I refer the reader to our previous paper for aesthetic and technical details on the performance itself [16].

4.1 Composing Latent Sound Spaces

A first aspect of musicking deep reinforcement learning lied in composing parameter sound spaces that the agent will navigate through. For *ægo*, we opted for latent sound spaces, that is, sound spaces created by generative deep learning, another machine learning framework that enables to produce new data that resembles existing data [18]. Latent sound spaces have interesting musical features for comprovisation. Specifically, their parameters are not necessarily interpretable as technical synthesis parameters, such as frequency, amplitude, or modulation. Rather, they should reflect perceptual variations of timbre of sound datasets used for learning. Thus, improvising in a latent sound space should generate continuous timbre variations, interpolating between recognisable timbres contained in the training dataset, while also generating ambiguous timbral artifacts typical of generative deep learning [19].

For *ægo*, we opted for two latent sound spaces built over two training datasets: synthesis sounds and acoustic instrument recordings. We stress that we consciously chose these latent spaces in terms of the training datasets they relied on to be created. Yet, we underline that we could not exactly define, nor control, the types of sounds contained in these latent spaces, due to the intrinsic generativity of deep learning [18]. As such, we opted for an experimental approach to composing sound spaces, first crafting generative models through their sound dataset, then curating the latent dimensions to be explored by Co-Explorer. This process was highly recursive, as composing latent sound spaces required improvising through gestural feedback to fully grasp their musical attributes.

4.2 Improvising Through Gestural Feedback

Indeed, a second aspect of musicking deep reinforcement learning consisted in improvising through feedback with the agent. As a performer, I opted to develop a gestural controller to communicate positive and negative feedback. Specifically, I used inertial measurements units, placed on top of velcro rings, to measure my hands' orientations. I added both angular values and scaled the resulting numerical scalar so that it goes from -1 to 1. My wish was that such a bodily interface would allow for more intuitive and creative musicking with deep reinforcement learning.

In early experiments, I was able to discover the bodily vocabulary enabled by this gestural controller to communicate feedback. The most elementary and illustrative gestures consisted in turning my hands front to communicate positive feedback, and turning them back to communicate negative feedback, by only pivoting wrists. Through improvisation, I discovered other gestures to be explored to provide instructions toward sound, as well as to symbolically communicate with sound. Asymmetric hand postures, for example, enabled to obtain neutral feedback, since the sum of the two angular values would be zero. Yet, the resulting gesture would not be neutral, and would produce expectation and tension for both the performer and the audience. I also explored somatics-based gestures, focusing on internal bodily sensations as I was listening to sound, and producing free-form aerial gestures as embodied response to sounds, resulting in varying feedback values.

All along our experiments, I witnessed myself entering in a state of heightened listening toward sound. Specifically, I observed myself oscillating between two approaches: one that was performative, where I attempted to grasp control over sound by producing precise instructions or symbolic communications, and one that was meditative, where I carefully listened to sound as if it existed by itself, detached from my very own influence, even if my body responding to it in spite of me. Both cases almost had me forgetting about the agent's learning abilities for the benefit of discovering novel sound morphologies, at times witnessing my light influence on it. In short, feedback-based improvisation pushed me to consider both optimal and non-optimal behaviours of deep reinforcement learning as relevant for music performance, while simultaneously contributing to a feeling of spiritual identification with music [16].

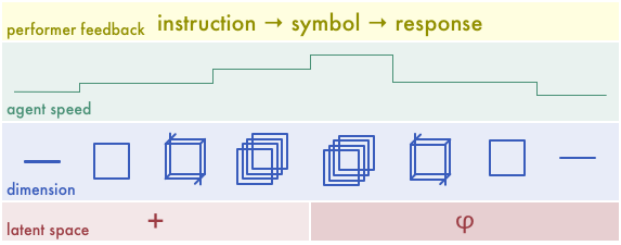

Figure 2. *ægo* musical structure (2022 reworking).

4.3 Comprovising a Musical Structure

Based on our musicking experiments, we sought to compose a musical structure for *ægo* (see Figure 2).

We chose to start the performance with the latent sound space built over synthesised sounds, since its timbral variations were less chaotic than those produced by the instrumental latent space. We also wrote series of latent dimensions for each space, going from one to eight for the first latent space, then eight to one for the second. As such, widening or narrowing timbral richness in sound spaces let us create dramaturgy across performance for both the performer and the audience.

For our premiere of *ægo* in 2019, the composer was responsible for choosing the moments where the agent would switch latent sound spaces. In our reworking, we renounced to this idea, and had the agent autonomously change zone within a sound space based on its exploration behaviour. This decision was shown to create space for improvisation for the performer, since they would discover the new sound space at time of performance only, thus dealing with even more indeterminacy from the agent during performance.

Similarly, for our premiere of *ægo* in 2019, we had set the time interval between agent's action at a fixed value, which resulted in the agent navigating sound spaces at a slow speed, thus producing a continuous drone sound. In our reworking, we wrote this time interval to produce diverse spectromorphologies along performance, alternating between continuous drone sounds through slow speeds, and glitchy sounds through higher speeds. This let us compose musical tension for both the performer and the audience.

Last but not least, we directed the performance so that the performer progressively relinquishes communication of accurate feedback to the agent, that is, going from instructions to symbolic communication and eventually to embodied response to sound. On the one hand, this choice aimed at improving audience comprehension of the performance, as they would first witness the agent being optimally guided by the performer, then progressively observe the blurring threads of influence between the agent and the performer. On the other hand, this choice enabled us to critically engage with deep generative learning, as embodied responses to sound will lead the agent to produce non-optimal behaviours within the sound spaces. Displaying such indeterminate behaviours as one defining musical attribute of deep reinforcement learning do not conform to established engineering-oriented applications; yet, from our perspective, it speaks of their material engagement with musicians and the world.

5. FROM COMPUTATIONAL FRAMEWORK TO MUSIC REPRESENTATION

In this section, I sketch contours of a music representation for deep reinforcement learning, which emerged from my musicking experiments with the Co-Explorer. I detail the epistemological tensions over learning models, the ontological shifts between performer and agent, and the aesthetics of feedback-based improvisation, produced by such a music representation for deep reinforcement learning.

5.1 Epistemological Tensions over Learning Models

As described in Sections 1 and 2, deep reinforcement learning defines a computational framework for agents that learn by interacting with their environment. Typical applications of deep reinforcement learning seek to learn an optimal behaviour in relation to the goal of a task. Engineering-oriented approaches thus seek to optimise an agent's learning by constructing some synthetic reward function that will yield the best results in terms of learning [2]. Thus, every other feedback functions can be seen as sub-optimal, or even incorrect, from this engineering perspective [1].

The music-oriented approach to deep reinforcement learning suggests that imperfect human feedback functions for engineering may in turn yield rich forms of improvisation for music research and practice. In fact, I argue that "optimal behaviour" may be a dynamic and emergent attribute of musicking and improvisation, as opposed to the static and pre-existing definition of engineering sciences. For example, one could argue that indeterminacy, as a musical feature of deep reinforcement learning, is what contributes the most to musicking, beyond agent learning or exploration behaviour. Yet, indeterminacy remains a variable that needs minimising in engineering-oriented approaches to deep reinforcement learning. Thus, goals of music and engineer practices may sometimes be opposed.

I believe that such epistemological tensions should be taken seriously by music researchers and practitioners, especially in the current growing applications of artificial intelligence to music, which often reinforce static representations of music through symbolic modelling of existing languages [4]. I suggest that musicking can be one such practice-based approach to discover material attributes of machine learning and illuminate their emerging properties. Rather than discrediting engineering-oriented approaches, I see this highlighting as an opportunity for interdisciplinary collaboration, enabling to iterate the design and implementation of learning models that are entangled with music.

5.2 Ontological Shifts Between Performer and Agent

As described in Section 4, musicking deep reinforcement learning enabled me to enter in a state of heightened listening toward sound. This heightened listening had me witness my oscillation between two different postures toward sound and the agent. On the one hand, I would aim at instrumental control over both sound and the agent, using feedback as both sonic instructions and symbolic communications. On the other hand, I would witness the existence

of sound beyond myself and the agent itself, as feedback would only help me believe that I control sound.

I argue that this oscillatory phenomenon reveals an ontological displacement of the notions of performer and agent toward sound. In fact, this displacement may drastically differ from other computational frameworks for music improvisation based on machine learning [20, 21, 22]. The latters often rely on anthropomorphic representations of sound and music, such as MIDI signals, and inject these in the design of the agents. Simultaneously, the performer may also rely on their joint technical and embodied knowledge of music to produce sound with their instrument and interact with the agent. As such, the role of the performer remains clearly defined as that of a musician. The agent, on the other hand, can be described as intelligent, or even as creative, as it builds on the same anthropomorphic music representation than that of the performer, while simultaneously being equipped with a greater musical agency compared to other software for music composition.

In deep reinforcement learning, however, no anthropomorphic representation of sound or music are injected in the design of the agent. Simultaneously, the role of the performer slightly moves away from that of a musician, since they do not rely on their instrumental knowledge to interact with the learning agent, nor do they actually produce sound directly. As such, the role of the performer oscillates between that of a musician and that of a listener, while also creating space for observation of the fluid boundaries that operate between these two roles, thus fostering spiritual identification with the produced sound. In parallel, I suggest that the role of the agent may progressively move away from that of an anthropomorphically-creative agent, to that of a non-human form of intelligence that produces music by conveying temporal form to sound. I believe that such an analysis should be deepened from a musicological perspective to produce alternative discourses toward artificial intelligence: rather than seeking to imitate or replace musicians, machine learning may enable to produce rich forms of musicking that foster human creativity while heightening their listening to their environments.

5.3 Aesthetics of Feedback-based Improvisation

As described in Section 3, deep reinforcement learning enables to engage in sonic improvisation by only relying on positive or negative feedback. The resulting interactions include instructions toward sound, symbolic communication with sound, as well as embodied response to sound. In a sense, they move away from standard instrumental techniques to music, as they privilege material attributes of sound over languages usually employed to describe it; indirect influence on sound over precise control and mastery of it; but also and crucially, identification with music over actual sound production. As a result, the values encapsulated in feedback-based music improvisation may differ from values of improvisation found in more traditional written music. In fact, they may lead to ethical encounters with certain communities of music practice and research, for example debating the quality or "truthfulness" of the produced music, or criticising the entertaining aspect or "seriousness" of the designed interactions with sound.

I argue on the contrary that feedback-based music improvisation create novel embodied interactions with sound that produce as "true" music as any other approaches to music composition or performance, and as "serious" interactions with sound than other physical or computer technology for music. If required, I would situate the aesthetics produced by deep reinforcement learning in line with the experimental music movement, in the sense that they push boundaries of existing genres, definitions, or disciplines of music, through their ontological shifts of performer and agent, epistemological tensions over learning models, but also and essentially, through their concrete approach to sound, and their reliance on free improvisation and indeterminacy processes to produce music [23]. In this sense, I suggest that debates toward aesthetics produced by musicking deep reinforcement learning may not fundamentally differ from those opposing conventional and nonconformist music practices with computer technology, long before the current trend for artificial intelligence.

Furthermore, I believe that feedback-based music improvisation summon "serious" social, cultural, bodily and spiritual phenomena related to embodied interaction with sound. Feedback shares similarities with gesture, in the sense that both may be used to translate sound using embodied knowledge and somaesthetic appreciation. In fact, feedback may be closely linked to biosignals, such as muscle tension or heart beats, in the sense that both may sometimes reflect involuntary responses of a musician toward sound. In this sense, feedback as an interaction modality for sound may be shared among diverse communities of people, be they musicians or non-musicians. Going further, I would suggest that the belief of controlling sound, as fostered by feedback, could be of interest for music practices that engage with people with disabilities [24]. Rather than just a basis for entertainment, identification with music have been at the heart of musicking for centuries: I suggest that it may be actively summoned within machine learning design and engineering to empower people toward both music practice and artificial intelligence technology.

6. CONCLUSION

In this paper, I have reported an auto-reflexive analysis of my practice of designing and musicking deep reinforcement learning. I have described the computational framework of deep reinforcement learning, along with the Co-Explorer, an agent designed to support sonic exploration through positive or negative feedback. I have discussed how deep reinforcement learning may be seen as a form of sonic comprovisational agent, enabling musicians to compose a sound space, then to engage in embodied improvisation by guiding the agent through sound space using feedback. I have reported musicking experiments made with the Co-Explorer, which led to the creation of *ægo*, a music performance for one human improviser and one learning machine. This enabled me to sketch a music representation for deep reinforcement learning, attempting to make its epistemological, ontological, and aesthetic aspects explicit for practitioners and researchers in music and ma-

chine learning. I hope that the present work will contribute to further music representations and discourses on artificial intelligence within the TENOR community.

Acknowledgments

I would like to thank Axel Chemla–Romeu-Santos for collaborating in the composition and performance of *ægo*, as well as Bavo Van Kerrebroeck, Baptiste Caramiaux and Frédéric Bevilacqua for collaborating in the design and implementation of the Co-Explorer, and Raphaël Imbert for pushing me to write on my such musicking experiments with deep reinforcement learning.

7. REFERENCES

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[3] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *arXiv preprint arXiv:1811.12560*, 2018.

[4] N. Kotecha, "Bach2bach: Generating music using a deep reinforcement learning approach," *arXiv preprint arXiv:1812.01060*, 2018.

[5] S. M. Karbasi, H. S. Haug, M.-K. Kvalsund, M. J. Krzyzaniak, and J. Torresen, "A generative model for creating musical rhythms with deep reinforcement learning," in *2nd Conference on AI Music Creativity*, Online, 2021.

[6] P. Ramoneda, M. Miron, and X. Serra, "Piano fingering with reinforcement learning," *arXiv preprint arXiv:2111.08009*, 2021.

[7] F. Bevilacqua, R. Müller, and N. Schnell, "Mnm: a max/msp mapping toolbox," in *Proceedings of the 5th International Conference on New Interfaces for Musical Expression (NIME)*, Vancouver, CA, 2005, pp. 85–88.

[8] R. Fiebrink, D. Trueman, N. C. Britt, M. Nagai, K. Kaczmarek, M. Early, M. Daniel, A. Hege, and P. R. Cook, "Toward understanding human-computer interaction in composing the instrument," in *ICMC*. New York, USA: Citeseer, 2010.

[9] H. Scurto, "Designing with machine learning for interactive music dispositifs," Ph.D. dissertation, Sorbonne université, 2019.

[10] H. Scurto, B. V. Kerrebroeck, B. Caramiaux, and F. Bevilacqua, "Designing deep reinforcement learning for human parameter exploration," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 28, no. 1, pp. 1–35, 2021.

[11] S. Bhagwati, "Notational perspective and comprovisation," *Sound & Score. Essays on Sound, Score and Notation*, pp. 165–177, 2013.

[12] H. Scurto, B. Caramiaux, and F. Bevilacqua, "Prototyping machine learning through diffractive art practice," in *Designing Interactive Systems Conference 2021*, 2021, pp. 2013–2025.

[13] R. I. Godøy and M. Leman, *Musical gestures: Sound, movement, and meaning*. Routledge, 2010.

[14] M. Leman, *Embodied music cognition and mediation technology*. MIT press, 2007.

[15] F. G. Visi and A. Tanaka, "Interactive machine learning of musical gesture," in *Handbook of Artificial Intelligence for Music*. Springer, 2021, pp. 771–798.

[16] H. Scurto, A. Chemla *et al.*, "Machine learning for computer music multidisciplinary research: A practical case study," in *Perception, Representations, Image, Sound, Music. 14th International Symposium, CMMR 2019, Marseille, France, October 14–18, 2019, Revised Selected Papers*, vol. 12631. Springer, 2021, pp. 665–680.

[17] C. Small, *Musicking: The meanings of performing and listening*. Wesleyan University Press, 1998.

[18] P. Esling, A. Chemla-Romeu-Santos, and A. Bitton, "Bridging audio analysis, perception and synthesis with perceptually-regularized variational timbre spaces." in *ISMIR*, Paris, Fr, 2018, pp. 175–181.

[19] D. Ghisi, "Music across music: towards a corpus-based, interactive computer-aided composition," Ph.D. dissertation, Paris 6, 2017.

[20] G. Assayag, G. Bloch, M. Chemillier, A. Cont, and S. Dubnov, "Omax brothers: a dynamic topology of agents for improvization learning," in *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, Santa Barbara, USA, 2006, pp. 125–132.

[21] J. Nika, M. Chemillier, and G. Assayag, "Improtek: introducing scenarios into human-computer music improvisation," *Computers in Entertainment (CIE)*, vol. 14, no. 2, pp. 1–27, 2017.

[22] J. Borg, "Somax 2: A real-time framework for human-machine improvisation," Internal Report-Aalborg University Copenhagen, Tech. Rep., 2019.

[23] M. Nyman, *Experimental music: Cage and beyond*. Cambridge University Press, 1999, vol. 9.

[24] S. T. Parke-Wolfe, H. Scurto, and R. Fiebrink, "Sound control: Supporting custom musical interface design for children with disabilities," in *Proceedings of the 19th International Conference on New Interfaces for Musical Expression (NIME 2019)*, Porto Alegre, BR, 2019.