Qualitative Inquiry on Machine learning in Musical Performance

THÉO JOURDAN, Sorbonne Université, CNRS, ISIR, France
BAPTISTE CARAMIAUX, Sorbonne Université, CNRS, ISIR, France

As in many artistic sectors, Machine Learning (ML) has also become part of music performance practice, although still little studied. Such inquiry can provide important insights into modes of expression and interaction, and their collective practice as a community. We conducted an interview study with 14 musical artists about their relationship with ML. We first find that artists developed new interaction strategies with ML to enable musical agency, by familiarizing themselves with the technology, control its behavior and explore its limits into live performances. This strategies are developed through data curation, real-time interaction and long-term practice. Secondly, artists have a practice of remixing and assembling musical material that extends to the collective level through the sharing of knowledge and content. Drawing from our findings, we discuss how artistic community provide knowledge to support new forms of interaction with ML.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Machine Learning, Music Performance, AI Art, Qualitative Study, Interview

ACM Reference Format:
Théo Jourdan and Baptiste Caramiaux. 2024. Qualitative Inquiry on Machine learning in Musical Performance. 1, 1 (July 2024), 22 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The term "AI Art" has been widely used to characterise the artistic movement that involves the latest advances in Machine Learning (ML) and artificial intelligence into artistic creation. Concretely, this denomination mainly includes visual artists who began using generative algorithms in the wake of advances in deep learning. At the same time, sound and music artists, evolving in the field of experimental music, have long been incorporating ML techniques into their artistic processes to produce interactive sound installation, musical performances, digital musical instruments, or dj sets among others. The performative nature of this musical practice and its avant-garde trait offers a unique perspective to the analysis of innovative interactions with ML and a better understanding of this community of practice.

In music performance, ML has been historically seen as tool to respond to specific musical and interaction needs. For instance, early works have demonstrated the use of ML to build gesture-based interactive systems [32]. Gestural inputs were mapped to sound synthesis parameters by providing examples of gesture executions for a given set of sound synthesis parameters. Although it is tedious to construct such expressive interaction scenarios manually, using learning methods that capture part of the expressive intention in the gesture and associate it with sounds is a more suitable solution for artists, performers or composers [19]. More recently, motivations behind the use of ML

Authors' addresses: Théo Jourdan, theo.jourdan@sorbonne-universite.fr, Sorbonne Université, CNRS, ISIR, Paris, France; Baptiste Caramiaux, Sorbonne Université, CNRS, ISIR, Paris, France, baptiste.caramiaux@sorbonne-universite.fr.

XXXX-XXXX/2024/7-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

have diversified and the techniques used have evolved, especially after the advances in generative models based on deep learning, using techniques such as generative adversarial networks [28] or variational auto-encoders [16]. The emerging field of neural audio synthesis, which relies on deep neural networks for audio generation (whether of raw audio samples [28] or spectro-temporal audio representation samples [16]), has given rise to new ways of exploring sound practices and developing rich interaction styles. For example, artists explored the control of deep learning-based methods [87], their potential to aggregate sounds and produce alternative listening experiences [77], or their use as front-person of a metal band [24]. However, except for a few idiosyncratic artistic uses of ML [17, 35], the literature on ML and music focuses mainly on the underlying techniques and instrument design and less on practice. In other words, we know little about how to get to grips with the technology and how to familiarise ourselves with it, nor do we know much about the benefits and challenges associated with this technology in the context of musical performance or its place in the broader community of practice.

In a larger context of the arts, research in HCI, and AI and Society, have started to study the practice of ML. This was conducted mostly in visual arts, which is undoubtedly the most represented practice in AI-assisted art making [18, 44, 49, 76]. For example, HCI research [18] analysed how visual artists managed to appropriate deep learning-based generative algorithms in their practice, meaning how artists tinker with the code, how they managed to train these models on their data, making them expressive and making them go beyond a potentially normative aesthetic. More recently, the latest advances in multi-modal techniques for the generation of visual content through text description allowed a wider range of people to try AI-assisted visual art making [69, 76]. These techniques gave rise to tools such as Midjourney [47], Dall-E [71] or Stable Diffusion [75], which offer an easy access to powerful image generation capacity without lines of codes and through text input. However, advances in the underlying ML techniques have accentuated the need for massive amounts of data and computing resources, making the design and development of these AI methods accessible only to a small number of private players. Consequently, a growing number of works also highlighted the harm due to the takeover of AI-based technology for the arts by tech companies, with dramatic consequences in terms of hype, technology governance, copyright infringements and human labor [45, 49, 63].

In line with this previous work on AI and visual art in HCI, in this article, our objective is to better understand the practice of ML for musical performance. Through an interview study with fourteen artists, we aim to investigate the following research questions (Q1) What strategies are implemented by music artists to handle ML techniques in music performance? (Q2) To what extent is ML technology compatible or a source of tension in the community of musical performance practice? Our contributions are threefold. We first found that music performers implement a variety of strategies to handle ML and include it into live performance. This involves hijacking it for creative purposes and highlights the importance of developing a musical agency over the technology, through data curation, real-time interaction and long-term practice of the technology. Secondly, we found that artists in the community are drawing on the ML ecosystem to remix and resample musical material. This practice is aware of and attentive to the current problems of misuse of this material when it is considered as data for commercial purposes. These results thus allow us to draw insightful lessons for designing expressive interactions with ML that extend to broader applications beyond music.

2 BACKGROUND AND RELATED WORK

In this section, we start a brief presentation of the use of ML for musical performance. Then we report previous work investigating interaction for musical expression in order to identify its

prerequisites. Finally, we report previous work investigating more generally the use AI in artistic practice, highlighting methodological underpinnings in qualitative research.

2.1 Background on ML technology for music performance

This first wave of work was initiated in late 80s. For instance, in 1988, Lewis [59] proposed the use of neural networks for automatic music composition. In 1993, Lee et al. [57] created complex mappings from high-level control parameters to sound synthesis parameters. In this background, we propose to present previous work involving ML technology for musical performance through waves of ML models, as these have a direct impact on the types of musical scenarios that artists can imagine and on the set of constraints that artists have to deal with. We chose to present the waves in ML architecture as: shallow learning, deep learning, foundation models.

2.1.1 Shallow Learning. We call shallow learning models a set of models (e.g. Random Forest, K-Nearest Neighbors) that require domain expertise to engineer features from data, i.e. write domain-specific logic to convert raw data into higher-level features (e.g. raw audio signals to pitch, loudness and brightness parameters). These features are used to train the model, typically for pattern recognition or input-output regression. In the context of musical performance, the input data used is often values from IoT (*Internet of Things*) sensors to record gestures of a body's performer [83] or movements of an object [61]. For example, the model is then used to control the variation of a set of parameters to generate sounds [88]. Many ML algorithms have been applied to real-time gesture recognition, for example Hidden Markov Models (HMMs) [54], Dynamic Time Warping (DTW) [10], and shallow Artificial Neural Networks (ANNs) [14]. Understanding the behaviour of these models remains simple, as the user chooses how to extract the features and the amount of data used is generally small. This makes these models fast and easy to use.

2.1.2 Deep learning. Deep learning refers to models that are trained on low-level data representation (e.g. image pixels, audio samples or audio spectro-temporal representation), and higher-level features are automatically calculated by the model through training. Deep learning models are generally relying on neural network architecture [56]. A deep neural network contains multiple layers of processing that extract progressively higher level features from raw data. So there is no need to pre-calculate features as shallow learning. Deep learning models have higher capacity than shallow learning models, which means more parameters (millions for deep learning models vs. tens for shallow learning models) to tune during training. Therefore these models are more demanding in terms of quantity of data to use for training and computation.

With the increasing development of deep learning techniques over the last decade, we have seen a development of music generation application. On one side, it applies on melodic pattern generation, generally in MIDI format, such as MIDINet [90]. The model is able to generate melodic patterns from scratch, by following a chord sequence or by relying on the melody of previous bars. Google developed MusicVAE [72], a model able to generate sequences with long-term structure thanks to a recurrent Variational Auto-Encoder (VAE). Beyond symbolic generation, new models are able to generate audio representation used for synthesis, which is called Neural Audio Synthesis. The sound is generally represented either as a waveform like WaveNet [85] or as a spectrogram like RAVE [16]. Some techniques also rely on learning a synthesis model to approach a sound as an additive model including oscillators and noise in DDSP [30]. Several models proposed for sound synthesis, and especially those used by artists, incorporate an Auto-Encoder [7] comprised of an *encoder* and a *decoder*. The *encoder* compresses the information (non-linearly), while the *decoder* has to reconstruct, as accurately as possible the initial information. The Auto-Encoder is then forced to find significant features to encode useful information through the *encoder*. The output of the encoder is generally called *latent variable* or *latent space*. These models have been used by several

artists in the field of musical performance. Some use it as an autonomous generation tool, like Dadabots [24]; as an abstract synthesizer to navigate during performance [80, 87]; or as a medium for listening experiences [77].

2.1.3 Foundation model. Foundation models can be defined as models trained on broad data that can be adapted to a wide range of downstream tasks. From a technical point of view, the foundation models are made possible by transfer learning and scaling. The idea of transfer learning is to take the 'knowledge' gained from one task and apply it to another task [11]. What has made these models powerful is the change of scale (the number of parameters of these models are over a hundred of billions) made possible by improvements in computer hardware, the development of the Transformer model architecture [86] which exploits hardware parallelism to train more expressive models than before and the availability of much more training data. Today, foundation models are mainly represented by Large-language models such as GPT-4 [15] or visual synthesis models such as Dall-E [71]. Following the example of visual synthesis, the models are generally trained on billions of images-text pairs obtained from the Internet, making the training process opaque for the public, sometimes impossible to know the extent to which training data contains copyright protected images [49]. Despite this, many artists have reported finding their works in the training data of foundation models without their consent [6]. Foundation models are not yet mainstream in musical application, but within the last year several models have been released such as MusicGEN [22] by Meta, MusicLM [1] by Google, and at the time of write StabilityAI released Stable Audio [78], all text-to-sound models relying on this type or architecture. As far as we know, we have not yet observed any artists developing a practice of these tools, although we have noted that some musicians are starting to incorporate these techniques into their process, as observed at the last AI song contest, where several artists used MusicGEN in the creation of their sound, such as the team Melody Makers[1].

2.2 Interaction for Musical Expression

Started in 2001 as a CHI workshop [70], The New Interfaces for Musical Expression' conference (NIME) is the primary venue for scientific research on designing new musical instruments mediated by technology. Because of its link to HCI, a number of works have been inspired by HCI-related tools and applied to digital musical instruments (DMI) [67], and reciprocally [51]. Contributions cover a variety of aspects related to musical controllers, such as design and technology, frameworks and interfacing protocols, reports on performance and composition, but also artistic, cultural and social impact [48]. At the heart of research into new interactions with Digital Musical Instruments (DMIs) is the study and design of interactive systems enabling musicians to develop musical agency over a DMI. When related to instrumental music making, the notion of musical agency is defined as the physical change and control needed for an individual to produce the desired musical output [82]. Musical agency has also been studied in musical pedagogy and instrument learning [52]. Roberts and Krueger [73] define two key characteristics that lie at the heart of musical agency, which also characterises interaction for musical expression: embodiment and emotional expression.

An embodied interaction with the instrument refers to an instrument as an extension of the body (notion that has also been explored in HCI through the works by Dourish [29], Kirsh [53] or Loke [60] to cite a few). In simple words, it implies that the practitioners do not focus on the instrument to make music but rather on the music itself, through the instrument [81]. Ihde [46] describes an embodied technology as transparent, in a sense that such a technology does not draw attention to itself but to the world it gives access through it. In music, the notion of embodiment has also been shown to create co-adaptive processes at a cognitive level [58]. Guidi and McPherson [41]

[1]https://www.aisongcontest.com/participants-2023/melody-makers

transposed this concept to embodied Digital Musical Instrument (DMI) arguing that manipulating an instrument is automatic when the performer can adjust their actions *rapidly* and *precisely* and then expressively shape their performance. Eventually, it leads to an appropriation of the instrument capacities by the performer [92].

A musical agency is also characterised by emotional expression. Expression in a performance is characterised by intentional corporeal nuances executed while performing [27, 39, 40]. The techniques used to create such variations involve subtle control of aspects such as accents or timbre, acquired over practice with the instrument [27, 68]. Control is not sufficient for expressivity, Fels et al. [31] argue that expressivity also depends on the transparency of the correspondences between the input variation induced by the performer and the output sound, both for the performer and for the audience, and their interrelation [42].

In summary, designing interaction for musical expression involves a transparent relationship with the technology, allowing expressive nuances on the part of the performer. One of the aims of this article is therefore to understand how performers and music composers can achieve this goal of musical agency with ML.

2.3 AI Art studies

The practice of "AI Art", which is based on the use of artificial intelligence technology as a tool in artistic practice or as an artistic object in its own right, is not new and early works were shown in the 90s, such as the "*aglaopheme*" by Nicolas Baginsky (from [5]). However, while in its early days this practice was considered experimental, in recent years it has developed into the socio-economic fabric of contemporary art and has been more widely disseminated in the mainstream media. In this section, we present research in HCI, or at the intersection between HCI and other fields, on this practice. We will use the term "AI" because it is the one commonly used to designate the movement and practice of "AI art", bearing in mind that, technically, AI refers mainly to ML methods.

From the point of view of HCI research, the study of this practice is in its infancy and is interesting in several respects. First, the use of AI in the arts have triggered misleading narratives about the fact that a machine can become autonomously creative, which has always been thought to be a human characteristics. Recent work in critical algorithm studies has analysed and deconstructed such a narrative [25], promoting the importance of an interdisciplinary dialogue (in cultural and technical terms) on technology [43]. The idea is that by looking at how creative practice engages with emerging technologies, such as artificial intelligence and machine learning, we can better understand what technology is capable of and how to communicate those capabilities in a way that is more accurate than a misleading narrative about a super-intelligent machine surpassing human capabilities [18]. Eventually, the approach could be fruitful to improve AI literacy [44].

Second, the study of AI Art can push forward the current research on AI ethics, although this research is also at its infancy and exploratory [26, 79]. Stark and Crawford argued that "The lack of attention on artists and the cultural sector in this burgeoning literature [of AI and data ethics] is surprising given that artists have historically deployed new technologies in unexpected and often prescient ways and have been interpreted as vanguards: of new ideas, techniques, and cultural practices" [79]. As explained in [26], works that examine the interplay between AI ethics and art practice are usually done in two ways. First it can be done through art-works that use AI as a way to excavate biases, discrimination or glitches in AI. Examples are "*Image Roulette by*" Crawford and Paglen [23], "*Deconstructing Whiteness*" by Meshi [62], or "*Learning to See: Hello World!*" by Akten [2]. Second it can be done through the understanding of the socio-cultural basis of the arts through field work [26]. This is this second approach that we are also following in this article.

Thirdly, the study of 'AI art' can also provide innovative design ideas. By documenting how artists integrate AI and ML systems into their practice, this knowledge can be translated into design

specifications. For example, in a recent work [18], the authors have documented how the learning process becomes design material for visual artists: the evolution of the model's learning was visualised in real time, resulting in an artwork. But this idea of interfacing with the learning process can be transferred to other applications to understand the behaviour of models. The methodology used is a field study based on interviews. It should be noted that other studies have followed similar objectives where the process of artistic creation with ML is documented by the artists themselves [17, 35, 91].

Finally, the study of 'AI art' can allow for documenting and describing the community of practice gathering artists involving AI in their work. These works have looked at visual artists hacking deep generative models [18], or artists using prompt-based approach to image generation [20, 76]. For example, the typology of artists using prompt-based approach was investigated highlighting that it is a recreational activity conducted by specific socio-demographic groups [76].

In this article, we aim to investigate the practice of AI and ML by music performers and composers. Previous work presented in this section has mainly focused on the visual arts, with a few exceptions in the field of music [17, 35]. Therefore, we wish to follow an interview-based approach, similar to [18, 20, 26], in order to highlight the complementary knowledge that can be gained (along the lines defined in this section) by focusing on this particular community.

3 METHOD

In this section we present our method. We describe our recruitment for semi-structured interviews, followed by our analysis method, where we used thematic analysis to form themes and subthemes and discuss their implications.

3.1 Participants

In this research, our intention was to target artists who explicitly used ML techniques to produce music, performances or concert. We conducted a purposive sampling strategy [74] for selecting participants. This strategy is based on the researcher's a-priori theoretical understanding of the groups of cases targeted, on the fact that these groups have important perspectives on the phenomenon in question and that their presence in the sample must be ensured. In this purpose, the recruitment targeted events at the crossroad of technology and art, namely MUTEK Montréal, MUTEK Barcelona, MUTEK Buenos Aires, MUTEK Mexico, MUTEK Tokyo[2], Ars Eletronica[3], S+T+ARTS [4] and New Interfaces for Musical Expression (NIME) conference[5]. 18 artists were contacted by email to be informed about the purpose of the study, what participation entailed (e.g time of the interview, voice recording), how anonymity was protected and any other information that might help them make an informed and consensual decision before agreeing to take part of the study. 14 artists accepted the invitation, the remaining 4 artists either negatively responded or did not responded. Participants are mainly close to research circles in the field of interfaces for musical expression probably due to the fact that it is a place where avant-garde practices are developed

3.2 Ethical statement

Before conducting the interviews, the authors sought ethical clearance from the Research Ethics Committee of their university. The committee considered the methods, recruitment strategies and treatment of personal data.

[2]https://mutek.org
[3]https://ars.electronica.art
[4]https://starts.eu
[5]https://www.nime.org

The consent was given by each interviewee after explicitly giving all the information about the project and the data processing. It was also specified to the interviewees that they could withdraw from the project at any point in time.

All the artists interviewed are anonymized in this article. They are designated by a number between one and fourteen. Demographic and background information are given in Table 1 in the Appendice. As the interviewees are anonymized, all pronouns in this paper will be non-gendered, so they/them.

3.3 Interviews

We conducted 14 semi-structured interviews using Zoom or Jitsi Meet video conferencing application. The interviews were conducted between 02/12/2022-16/12/2022 and between 12/07/2023-20/07/2023. They lasted between 36 minutes and 73 minutes (mean duration of 57 minutes). The interviews were designed to investigate artists' practice in using ML and in particular how advances in ML for music, from industry and academia, link or collide with their personal practice of the technology. More precisely, the interviews were structured according to the following items:

- **Introduction.** After introducing the project by the interviewer, the interviewees were invited to give a general introduction of themselves, explaining the role of ML in their work and, if possible, describe an example of projects that they consider important.
- **Practices and methodology.** We questioned the interviewees about the way they use ML, specifically the models and techniques they choose to use and to not use, the dataset they produce or existing dataset they integrate in the training process, the way they implement their system or adapt for their own purpose existing system.
- **ML developments.** We asked interviewees if they follow discussions, actions, and published works concerning the recent developments of ML whether in the industry, research community or other community (e.g artistic) and how they keep up-to-date. Then we asked them if these recent developments and the digital practices promoted by technology owners go in the same direction as what they conceive of as their musical practice.

We recorded the audio of the interviews, then interview's audio recordings were processed following two steps. First, the interviews were automatically transcribed using Whisper[6] run on a local computer. We proofread each automatic transcription while listening audio recordings. Audio recorded were destroyed once we finalized the final transcribed interviews

3.4 Inductive coding and data analysis

We conducted a thematic analysis, relying on the methodology proposed by Braun and Clark [13]. Analysis of qualitative data began with familiarisation with the data by reading the transcripts twice. Then the two co-authors separately highlighted the relevant parts of the interview through an inductive coding analysis. Inductive analysis is a process of coding data without trying to fit them into a preexisting coding framework or into the researcher's preconceived analytical ideas [66]. A code represents "the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon" [12]. Then the codes of both authors were put together and discussed. Codes common to both authors were retained. Codes identified by only one of the two authors were discussed and retained if they were collegially considered relevant. Then, the authors organised separately the codes into themes. The development of the themes represented the interpretation of the authors made from the codes. Finally these themes were put together and discussed until finding a consensus. We converged toward **3** themes for this analysis, comprising **6** sub-themes reported in Section 4.

[6]https://github.com/openai/whisper

4 FINDINGS

Our results contribute to the understanding of the practice of ML in music performance. We established two themes from our data. Firstly, music performers imagine new interaction strategies with ML to enable musical agency, which means that musicians find ways to familiarize themselves with the technology, control its behaviour, explore its limits, and favor long-term practice. Secondly, the use of ML technologies leads to the reuse of data and software tools, which highlights a new form of sampling practice in music and, at the same time, raises concerns related to the general context of the use of ML for art.

4.1 Artists are finding musical agency with machine learning

Musical agency is characterised by the musician's ability to control an instrument with sufficient precision and subtlety to express the artist's intention in the output [52, 82]. Drawing a parallel with the interaction between a musician and their instrument, we can understand how the artists interviewed develop this type of agency with their system and how this translates into practice with ML, in particular through different strategies to understand the model's behavior and develop idiosyncratic practices, and long-term practice.

4.1.1 Strategies for familiarizing with machine learning models. Understanding the behaviour of ML models can be challenging, particularly with deep neural synthesis models, which involve complex relationships between control parameters and the sound output. Understanding this complex representation often leads artists to undergo an iterative trial-and-error exploration. For example, P4 makes an analogy between the latent space of a neural synthesis model and a synthesizer with unlabeled knobs: *"The only way you're doing it is trial and error. It's like having a huge synthesizer in which you have a bunch of knobs, none of the knobs are labeled, so you just have to try each of them.* The absence of labels on *knobs*, as described by P4, shows the lack of semantics in the abstract representation created by deep generative neural networks. These semantics, on the other hand, are present in most of the classic audio synthesizers that artists are used to working with. This trial and error form of work pushes the practitioners to develop idiosyncratic strategies to create meanings in this techniques.

A strategy to facilitate the exploration of sounds synthesized by a deep learning-based model (in this case a RAVE model [16]) is proposed by P6. They define the sound space learned by the model by including mostly *elementary* sounds in the training data:

> *"The approach is working with [...] simple things, you know? Like only sine waves, noise, blips and sweeps and things like that are very simple. So you know that you can create something new among those."*

Using elementary sounds allows one to start with a simple timbre (or set of simple timbres) and explore how the model can create more complex sounds in terms of timbre by combining or morphing them.

Similarly, P4 proposed to link the sounds produced by a deep learning-based sound synthesis and the data it uses to be trained by building a database of personal sounds of which they have full knowledge:

> *"Sometimes what has been pretty useful and meaningful to me is working with a data set of sounds that I know very well. For example, your own music. So I have a data set of eight hours of my music, so I know exactly all the data that is there. And so by learning a model of that, I can know if what it's been decoded in a specific point in the space, is something new or something that was unexpected."*

Through the use of familiar sonic data, P4's strategy allows them to build a more accurate mental model of the system's behaviour. A similar idea was raised by P3 through the concept of "*predictability index*". This index refers to the width of predictability measured as the variability of the sound outputs in the context of the input-output mapping design. If the output sounds are very different from each other, small changes in the input control will result in large changes in the output, thus increasing the unpredictability. In P3's words:

> *"Predictability index is a new term I've invented. (...) The goal is to create areas of exploration within the synthesis. For example, in terms of the size of the terrain, if you make a model where you say this input makes this sound, this other input will make a sound that is barely modified, your terrain of predictability is very narrow. It's just going to make a very narrow modulation. If I give very different examples of synthesis, it will go into these unpredictable spaces. And at the end of the day, the work I do is to see if these points in the field, when they are very different, produce sounds that I find really interesting, and then to limit the sound around them."*

Rather than a mathematical or quantitative metric, this concept can be seen more as a perceptual representation of the sound synthesis "terrain" that the artist can explore with the model used. Then the artists change how the training is modeled in order to scale or define the limits of this terrain. This way of perceiving the artist's exploration of the model can be linked to the concept of "timbre space" developed by Wessel [89]. A timbre space is a representation of a set of sounds according to the timbral dimensions that characterise them. In this space, sounds that are close from the point of view of timbre perception will be close from a geometrical point of view. Designing this interpretable layer upon the latent space of a deep generative model would help potential users to better understand the behavior of a neural synthesis model, and make it easier to get engaged with the model.

4.1.2 Strategies for controlling machine learning models. Control is essential in musical performance, and can be seen as a specificity of this practice in its need for precision and expressivity. When using deep learning, control parameters are usually abstract and describe a latent space where each point of this space corresponds to a different sound. Following P4, a way to create meaning in their performance is to have bookmarks in this latent space, capture different points in the sound representation of the model in order to be able to navigate between them is a way for the practitioner to create a musical structure:

> *"How many times have we explored modular synthesis? It could be pretty boring, except for the person who sort of handles It's a hit or miss, you know? So the same happens with the latent model exploration. It could be hit or miss. But if you want to put that in the context of an actual performance, usually you want the most interesting moments. So by applying another layer on top of this, I can go to the points in space, which are the most relevant for me for a specific performance. It's like adding bookmarks. So I can interpolate between those points, for example."*

P4 therefore proposes an additional information layer that acts as an instrument over the latent space and improve the control over it. Then, controlling the model's behaviour to be better interpreted also means creating meaning for artists. Particularly with the inherent non-linear behavior of deep learning models, a small change between two input given to a same model can potentially generate really different output. P11 sees the sense of control as a reduction of this feeling of randomness that can be perceived, so that a single action of the musician is linked to a single response of the system:

> *"We are aiming to make these things to be more controllable, but more interpretable by human musicians, so that they could understand or have an idea how these systems response. What it was doing, that it was randomly jumping in between different vector points in the latent space and generating audio samples. So in order to be have more control in that space, we applied a dimensionality reduction technique, PCA which allowed us to use Gansynth in more controlled manner. So we can actually really put that into a musical instrument, like in a way that when you pull a guitar string, you don't want it to change each time when you pull.*

The strategy employed here has been to structure the latent space around a few defined components (thanks to a PCA) that the musician can more easily understand in order to be able to play with them.

4.1.3 Looking for the anti-program to the technology. ML generally involves standard pipeline which consists in constructing model architecture, training the model on some data and evaluate its performance. However, while in conventional ML each of these steps follow a set of established norms and metrics (for instance using accuracy to test the model performance, and then using accuracy on a validation test to monitor over-fitting and stop the learning process accordingly), these tend to be deconstructed in music making. P9 described this approach in their words:

> *"We can have an anti-program to the technology, we take it and we misuse it and we reject the script of the technology. And that's where artists come in, in a very interesting way and misuse the technology."*

The term "anti-program" refers to the terminology used by Latour [55] to describe the means by which users can circumvent the use of a technology that does not conform to the original prescription for its use. Interestingly, the process starts from using the technology in order to latter be able to misuse it, similarly to the practice of music making with an instrument: first one has to learn the way to play it before developing their own style.

For example, in conventional ML, one objective has often been to test the generalisation of a model on data that have not been seen during training. This idea of generalisation is also deconstructed by artists using ML in music performance because their goal is, on the contrary, to look for idiosyncratic behaviours. One way for artists to do this is to curate personal training data. P12 describes this as follows:

> *"You know, this is small data. This is my data. This is myself as one person, not a whole population. Because on top of that, in performance, there's a tabula rasa in terms of training set at the beginning. That's my personal point of view, but I think there are general challenges for the NIME field. If you had access to big data and deep learning, what could you do? What would you do? What are the constraints of NIME? And how would they interface with the way those algorithms are made?"*

P12 highlights the fact that certain needs in terms of artistic practice with ML do not require large models, which are sometimes claimed to be universal, but an individual local model that reflects each person's aesthetic and expression. In their case, starting with a blank model and fulfill this model with live data, is a part the performance that would not be possible with deeper model. This raises questions that have not yet been resolved for them: for what purposes should they use deep learning methods and which interface would be best suited to this objective.

Another example is provided by P3, who described their way of producing music with gesture-sound mappings by playing with gesture non-recognized by the model: *"What interested me was what was not recognised. So recognising a gesture was fine, but what interested me was everything in between, that recognition."*. P3 uses regression models that create a continuous space between

gestures used for learning. Navigating in this space generates interpolations between the sounds initially associated with these gestures. As a result, P3 trains the system with specific gesture-sound pairs so that they can then navigate between the gestures, looking for what is not recognised but filling the gap.

In a similar way, P7 reflects in the idea that in a classification system there is always data taken into account by the model, and data outside the system that they call "waste". By giving importance into this waste, the outliers, the misclassified, the artist is subverting the normal use of a classification model, in the same time questioning more broadly what is considered as classifiable or as waste:

> *"I'm thinking about what are the politics of classification and taxonomies and, in a way, to challenge that through the way I am building model and training them. So I worked a lot around images of waste and waste being matter outside of classification. Thinking of what do we mean by a classification system and what is outside and what's inside"*

4.1.4 There is a need for long-term practice of ML, which conflicts with the speed of innovation. Mastering ML systems for musical practices requires skills and the development of strategies. Some artists are therefore interested in developing a practice of the technology on the long run, implying the fact that the technology remains almost unchanged. For example, P5 explained: *"I am very much a supporter of a slow motion or sort of a slow music approach. So I think that the time you spend with the tools is what actually makes an instrument"*. In the same way P3 described their long-term relationship with an instrument as way to find the limits of the instrument and developing a sense of agency:

> *"I have to keep that and develop the limits, understand the limits. I find that a lot of instrument design is about understanding the limits. And this is the first time in 10 years that I've started to do something that seems to really respond to the system."*

The need expressed by artists to take the time to develop deeper understanding with ML-embedded musical systems may relate to their individual experience in musical instrument practice: musical skills develop through deliberate practice, to take the words of Ericsson [4], and deliberate practice is a long-term process. However, in the context of ML, long term practice with this technology may be in tension with the pace of innovation which gives a pressure to use the last innovation (which will be further described within the following theme). Following P12 there is a necessity to have a critical attitude about that aspect:

> *"to resist the temptation, you know, the sort of technodeterminist temptation to take anything new and try to do something with it, the AI temptation to say it's in fashion, you know, it literally took me 15 years to get to a sort of a configuration of the technology that was musically and scientifically interesting to me."*

P12 refers to the contextual pressure when working with ML. Massive investments in both the private and public sectors have dramatically increased the pace at which new techniques are released and deployed. This pace at which the field of ML and AI innovates can be perceived by some artists as a pressure, pushing them to take a step back from the use of new techniques. P9 described their position regarding the contextual pressure in ML as follows:

> *"When there's huge hype happening, I need to digest stuff properly before I use anything. I try to just carve out little spaces where when I have time for it, I'll try something new. But I've stopped feeling the pressure of knowing what is the latest new thing happening, because I'm always going to be behind now"*

4.2 Reusing musical material and tools in a context of tension

Almost all the practitioners interviewed used sound synthesis models in some of their work. These generative models are based on the use of existing data, i.e. existing musical material. The interviews elicited how some artists are inclined to use other people's data in their creative process. Interestingly, some interviewees related the practice of data reuse to remixing and sampling, common practices in music making especially in the DJ culture. This practice of reuseextends to software tools, facilitated by open source material. But the current context of ML as applied to artistic creation, and in particular the abusive use of copyrighted material for the development of commercial products [49], also prompted respondents to comment on this practice of re-use in this context. As P7 put it: *"I'm also very much aware of the economy of media within it. (...) Big scale projects tend to be very kleptocratic."*. Interviewees stressed the idiosyncratic nature of musical material and its ownership when considered as data.

4.2.1 Reusing data and software tools is in the culture of music performers and composers. Typically, artists can train ML models from scratch on their own personal musical data or using other artists' sounds. They can use pre-trained ML models in order to make new use of other people's data, or they can fine-tune ML models in order to reuse the knowledge gained from one musical database, potentially from someone's else, to another one by adapting a pre-trained model to a new database. In this context, P1 sees as inevitable and even desirable to use the sounds of other artists to distort them and produce new creations:

> *"As an independent artist, I'm trying to think a bit more and understand that it's inevitable. So I asked myself what I could do from my point of view. I'm helping to make the data set as extensive as possible. How can I, as an independent artist, contribute to other datasets. Because I work with song mixes and conventional song mixes. Or immersive narratives with a different language. And how all my data can feed into other datasets for other artists. That's the most important thing for me. Because other people can re-use my data in different ways."*

The practice of reusing other's data to one's own practice is also very specific to music performance. Coming from the DJ culture, P6 confirms this: *"I'm totally fine if somebody else trained their model on my music because I grew up in like DJ culture and hip-hop culture so I do sample somebody else"*. In this way, generative arts enabled by deep learning technologies belong to a new range of remix practices that come from music sampling used live by DJs, first spinning records on turntables, then using digital samples integrated in producer's music tracks.

Artists' practice also relies on reusing existing software tools. Like many technological tools, ML requires technical skills to develop new tools from scratch, which not all artists want to do and can do. This practice is illustrated by the tendency of bringing new techniques from outside the musical community for reuse, as explained by P11 : *"There's a lot of ML that is incorporated into the community that don't necessarily come from the community"*. Talking about specific technical aspects, P11 says that the goal is to *"reuse and hack stuff that was already kind of built and customize those to the specific use"*. Quite naturally, this approach is facilitating by an open source culture: giving users the opportunity to make their own version of a system is a way of guaranteeing a certain degree of sustainability. This system can be maintained over time, even if it is no longer supported, as long as it remains available. As P5 put it: *"make something sustainable is to make it open so that people eventually can repair, can make their own versions"*.

4.2.2 The current context of copyright in the ML applied to artistic creation. The question of copyrighted data and copyright infringements in the context of ML applied to artistic creation is an important problem that gives rise to a collective awareness about the practice of tech companies and

the need to take actions against abusive use of copyrighted data to build commercial products [49]. This has been particularly visible in the field of visual arts. The way the artists interviewed elicited this issue shows that this awareness spread beyond visual arts and it is very present in music performance. Certain artists, conscious of data rights prefer therefore to use exclusively their own data, as P8 stated: *"I was very intentional early on to use my own data set for training, because I had so many questions about copyright"*. Others, like P2, avoided the use of copyright data and ensure that the licences for use are clear: *"You have MIDI files from 90s and it's not clear who built them and who owns the dataset. I'm not using them, so I try to use a dataset where it's very clear, the licenses are very clear."* P4 mentioned the case where a fair use of the data can be acceptable, as long as there is no commercialisation:

> *"I'm a very copy left person. So if we all agree that we can reuse the piece of media that we generate to train models and the access to those models is free, I'm totally open for that. But if some people or corporations are making money out of other people's work, I'm not up for that. Of course. (...) So that's a full round or it's a full circle to say that since the access is not equal for everyone, I would say that I'm not okay with having my music to be used for training in models that will be commercially distributed."*

That being said, the music industry still highly relies on record labels that are often the right holders of large part of the music content. Therefore, this delegation of ownership makes them able to involve this content as training set for ML-based products, or to sell the content to third-party willing to train their model on these data. As P8 put it:

> *"I think all the major labels are currently trying to figure this out, how to regulate. And I think there's a lot of fear too. And it's actually, I feel even better that I'm an independent artist rather than like signed to a major label, because if I was a major label artist, I would be worried right now of like what labels will decide to do with my stems. And maybe they ask for my permission and then monetize all."*

As a way to question the problem of ownership and its abusive use in commercial products, artists emphasise the importance of the idiosyncratic nature of the musical material. P6 explained: *"I don't want to imitate existing music so I try to avoid mimicking existing materials even though"*. As a result, P6 is taking a stance on this issue by refusing to re-use data in order to nurture their own aesthetic. To go further on this point, certain artists also consider their work as too personal to be replicable, because it takes into account subtleties inherent to the artist that cannot be transposed, such as the voice for P7:

> *"I do not think that someone can replicate my work computationally. I feel that the way I treat the voice is very mercurial, it's always changing, so harder to fix it"*

Another example is that what makes the sound produced by the artist is linked to the unique practice of an instrument, as is the case for P3: *"I think it's impossible, if you like, in the sense that the data coming from the instrument is really attached to the instrument. And the synthesis is completely linked to the instrument. You can't separate the two"*. Consequently, in the case of P3, it makes no sense to use the data alone, as it is indiscernible from the instrument.

5 DISCUSSION

In this article, we have interviewed fourteen artists about how they integrate ML techniques for musical performance. Our research questions were: (1) What strategies are music artists implementing to handle ML techniques in music performance? (2) To what extent is ML technology compatible or a source of tension in the community of musical performance practice? We found that artists in ML have found innovative ways of exploring and navigating sound spaces. In addition, the interactions

elicited are often in real-time and built up over a long period of deliberate practice. We found that the community of practice develops musical agency over the technology, which implies the forms of interaction outlined above, and an agency over the data and its sharing. Assembling musical material is common in music making (e.g., sampling), which remains when using ML. However, concerns accompany this practice linked to the more global context of using ML in the arts. In this section, we discuss the notion of musical agency over ML related to the concept of instrument and Interactive Machine Learning. We then examine the challenges posed by the analogy between sampling musical material, common in music creation, and data reuse, common in ML. We then give implications for the design of expressive interactions with ML. Finally, we present a few limitations that help frame future research directions.

5.1 Musical agency as a way to have an interactive instrumental control over ML

We have found that artists seek to develop subtle and expressive long-term control with ML-based models, and this in the case of sound generation, or bodily interaction with sound. This is achieved in three ways: finding strategies in data curation that shape the behaviour of the model; interacting in real time with the model and its learning procedure; and acquiring specific skills related to the control of the system through deliberate long-term practice. We interpreted this approach as the expression of musical agency on ML-based systems, which means the ability to reach a situation where the system becomes a transparent instrument for playing music [82], creating an embodied emotional expression [58]. Here, we would like to discuss two aspects related to musical agency as illustrated in our results: the idea of technology becoming an instrument and the need to increase agency over the ML system for musical performers in ML.

Regarding the notion of instrument, for the majority of the artists interviewed, the use of ML technology serves as a mediation means between their musical intentions and the musical material. In other words, ML can be seen as an instrument acting on musical objects. The concept of instrument finds an echo in HCI through the notion of instrumental interaction, as introduced by Beaudouin-Lafon [8, 9]. The main idea behind instrument interaction is to consider domain objects, which can be text, data, or images, and design instruments acting on those objects. Instruments can be activated or selected by users and can become themselves domain objects on which other instruments could act, pretty much as a tool in the physical world (e.g. a knife) can become the object on which a tool is acting (e.g. sharpening machine). It is therefore interesting to re-examine this notion from the perspective of musical agency (or instrumental agency). Firstly, although there is no specific mention of embodiment, the idea of instrumental interaction implies that users focus on the task at hand (manipulating domain objects) rather than the instrument itself. A new property we see is therefore one of transparency. If there is no transparency, there can be no instrument. Secondly, since the original notion of instrument in instrumental interaction was introduced in the context of post-WIMP interfaces, the final tasks were not exploratory. However, we have shown in our results that exploration (of sound spaces) is essential in musical action, which means that the instrument gives access to a wide range of possibilities and that these have to be discovered or learned. As a third point, the notion of (musical) instrument implies the notion of learning, which is not developed in instrumental interaction as proposed in the articles cited. It would therefore be interesting to revisit the notion of instrument in HCI from the angle of its learnability and the capacity to acquire skills over the long term.

In terms of increasing agency over the machine learning system, artists have developed strategies to act on the system's behaviour through interactive learning, immediate feedback and an intertwined training and testing procedure, as in the example given where the artist trains an online gesture-sound regression model during their performance. These properties of interaction with the ML are those emphasised in Interactive Machine Learning (IML) [3, 38]. It is not surprising

that IML has found an echo in the NIME community as it allows the use of the ability to create a communication channel between artists and music through data (motion data, sensor data, other sounds, etc.) [33, 37]. Although the IML approach has had some success with shallow learning models [36], deep learning has reduced interaction capabilities, particularly with training procedures. However, our results show that interactive strategies, even with deep models, need to be devised in order to facilitate an understanding of these models as well as a fine-grained control. Bridging communities and working with artists seem to be fundamental to make progress on these research questions at the intersection between HCI and ML.

5.2 Challenging the analogy of remixing musical material through data reuse

When music performers and composers train ML models, we have seen that the materials on which the models are trained may come from the artist's own production or from the production of other artists. Our results have shown that some artists find the practice of reusing other people's musical material inevitable and that it creates interesting new assemblages. We pointed out that this reuse approach could be linked to the practice of remix, that has its history in music in avant-garde music, DJing or the hip-hop culture. In this section, we would like to challenge this analogy with remixing when it comes to reuse musical material to train generative audio models.

In his book, "Art in the age of machine learning" [5], Audry presents a set of remix practices linked to generative art, that they call *deep remixes*. These remixes are either occuring in the creation of the dataset (through an aggregation of materials from various sources), or by allowing the access of a trained generative model to a range of applications (remixing the model), or by re-training part of a generative model through fine-tuning and transfer learning (remixing part of the model or several models). However, the book emphasizes some aspects of visual art and music, such as *deepfaking* based on face swap in videos, or creating synthetic voices indistinguishable from original one, based on the use of pre-trained models or fine-tuning techniques. It does not take music as a specific domain to discuss the practice of remix and the distinctive features that emerge. Our results in the music context allow us to bring the argument forward as certain artists have elicited the analogy with the remix culture through sampling. Sampling is the process of reusing sound fragments in a new musical arrangement, while potentially applying effects to them. We could therefore think of integrating sound fragments in a training set as a form of sampling. However, the practical use of what the model has extracted from this fragment remains largely opaque. If artists use samples to feed a generative model, but they have no way of knowing how these samples are used (i.e. they cannot use the characteristics of the sounds they have chosen) it will certainly be not useful for them. Our results have shown that some artists develop strategies to have control on that, by, for example, meticulously choosing sounds they know well or by playing on the dissimilarity between sounds and linking this dissimilarity to a measure of predictability of the model's outputs. These results are the beginning of what we could imagine as a strategy for a practice of *neural sampling* (in analogy with *neural audio synthesis*), which constitutes a research theme open and promising, both for HCI and for ML. This practice has also arisen because the specificities and objectives of music often diverge from the objectives pursued by the ML or HCI scientific communities. Datasets are generally created with the aim of making the models developed as efficient as possible, following performance criteria, for the task assigned. This is not necessarily the case in an artistic practice where, for example, using one's own data is a more common practice. For example, Morreale et al. [64] adopt a critical position regarding the constitution of music dataset in academia arguing that it generally relies on an extractivist approach, an exploitation of human labour.

From there, we feel it is important to separate the musical material from the data. Sampling is based on carefully curated sound fragments while generative model training is based on an

exhaustive use of many sounds aggregated in a dataset. In a recent article, Jo and Gebru [50] highlighted the differences that exist between the creation of ML dataset and the archival practice. They identify a spectrum of intervention where the former lies at one extreme where data collection is done without rigorous guidelines, while the latter lies at the other side of the spectrum where information (and artefacts) are filtered according to their significance and value. In the latter case, archivists have experience of selection and representation bias in the collections created, namely they are aware that the act of avoiding biases in datasets by finding means to de-bias them becomes another way to introduce biases [21]. Similarly, we believe that musical artists, composers, or DJs have a curation practice that exhibits differences compared to those noted in the act of creating ML datasets. A model trained on a dataset blurs the selection work that may have been carried out when the dataset was created. We think it would be interesting to study this analogy in more depth as well as how the data curation skills of musicians practising sampling can be transferred to the creation of datasets for training generative models. A few articles in the field of HCI have also raised the challenge of data curation within communities of practice [65, 84]. Still, its study in music and the arts, although essential, remains to be explored.

5.3 Interaction design implications

In recent years, interacting with deep learning models, especially neural synthesis models, has faced recurrent and common challenges when exploring interaction and practical applications. As observed in the previous sections, the artists interviewed have developed different interaction strategies, some may be interesting in an interaction design context beyond musical performance, particularly with recent neural synthesis models.

Firstly, controlling the model's behaviour via the latent space was regularly discussed. Bookmarking regions of interest in this latent space is a strategy to avoid too much trial and error in finding zones of interest but also to better structure performance by moving from one zone of interest to another or returning to previously explored zones. This suggests **investigating ways to develop an informational layer** identifying specific zones associated with sound descriptors. Natural language is a popular option today, but not necessarily the one preferred in embodied interactions. Supervised (e.g., regression) or unsupervised (e.g., dimension reduction) methods need to be better understood from an interaction design point of view.

Secondly, the perceived randomness of a model's behaviour is also seen as a barrier to its practical use in a musical context, either for the performer or the audience attending the performance. Small changes, sometimes unperceived by the performer or the audience, in the model's input data (such as a gesture made with an instrument) will induce significant changes in the output data, resulting in a sense of unpredictability. From an HCI point of view, mitigating this phenomenon poses several problems. As mentioned in the interviews, one interesting avenue is the idea of perceptual predictability metrics. The **formalisation and assessment of perceptual predictability metrics** can help the design of expressive and more controllable embodied interactions mediated by deep learning models.

Finally, for the sonic exploration of deep generative models, we found that the artists interviewed use ad-hoc interfaces for their practice. For example, regarding shallow learning models, Wekinator [34] is regularly cited, but similar interfaces are not used for deep generative models applied to music. Generally speaking, we noted that the artists interviewed talk about models, not tools, to describe the ML technologies used. Some interfaces have however been designed as VST plug-ins (as is the case for RAVE[7] or Neutone[8]), but **tools for musical performance using**

[7]https://forum.ircam.fr/projects/detail/rave-vst/
[8]https://neutone.space

neural synthesis have yet to be studied and developed. The research presented in this paper provides some ideas for designing this type of tool.

5.4 Limitations

In this article, we interviewed 14 artists working in the field of musical performance, composition and sound art. Each of these artists has a specific practice, which can vary greatly from one artist to another. For example, not all the artists interviewed use their bodies (through movement) in their performance. In addition, not all the artists use the same type of ML techniques (e.g. deep neural networks) in their practice. This wide range of practice allowed us to have wider view over the community of practice. However, one limitation of our study is that we are unable to make a more specific analysis by type of practice, which would allow us to obtain more granular results on modes of expressive interaction. For instance, more details could have been extracted on the exploration of generative models' latent spaces if interviewees share this common practice.

Another limitation we have identified concerns the representation of our sample. We interviewed a number of artists who were mainly based in Western countries. It would therefore be interesting to continue this research with artists working in musical performance in non-Western musical communities, i.e. communities that may not have the same approach to technological and artistic advances in the field.

6 CONCLUSION

In this article, we interviewed artists who use ML in their musical performances. The goal was to better understand the expression of the interaction with ML for musical performance, and the characteristics of this community of practice. We first found that music performers develop multiple new strategies for familiarising with, understanding and controlling the models and their behavior . The community stresses the importance of having a musical agency for technology, implying agency for data and data sharing build up over a long term practice. Finally, we found that the artists using ML are part of an ecosystem based on a culture of remixing and assembling musical material, raising tensions in a global context of the use of ML in the arts. We believe that this line of research can bring a new way of conceptualizing interactions with ML by revisiting the notion of instrument in HCI and develop new interactive strategies in the light of the artists' exploratory practice.

REFERENCES

[1] Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text.

[2] Memo Akten. 2017. Learning to See: Hello World! https://www.memo.tv/works/learning-to-see-hello-world/

[3] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. Ai Magazine 35, 4 (2014), 105–120.

[4] K Anders Ericsson. 2008. Deliberate practice and acquisition of expert performance: a general overview. Academic emergency medicine 15, 11 (2008), 988–994.

[5] Sofian Audry. 2021. Art in the Age of Machine Learning. The MIT Press.

[6] Andy Baio. 2022. Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator. https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/

[7] Dor Bank, Noam Koenigstein, and Raja Giryes. 2020. Autoencoders. CoRR abs/2003.05991 (2020).

[8] Michel Beaudouin-Lafon. 2000. Instrumental interaction: an interaction model for designing post-WIMP user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems. 446–453.

[9] Michel Beaudouin-Lafon and Wendy E Mackay. 2018. Rethinking interaction: From instrumental interaction to human-computer partnerships. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 1–5.

[10] Frédéric Bettens. 2009. Real-time DTW-based gesture recognition external object for MAX/MSP and puredata. (01 2009).
[11] Rishi Bommasani and et al. 2022. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258
[12] Richard Boyatzis. 1998. Transforming Qualitative Information: Thematic Analysis and Code Development. (01 1998).
[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3 (01 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[14] Bernd Bruegge, Christoph Teschner, Peter Lachenmaier, Eva Fenzl, Dominik Schmidt, and Simon Bierbaum. 2007. Pinocchio: Conducting a Virtual Symphony Orchestra. In Proceedings of the International Conference on Advances in Computer Entertainment Technology. Association for Computing Machinery, New York, NY, USA, 294–295.
[15] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv:2303.12712 [cs.CL]
[16] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. CoRR (2021).
[17] Baptiste Caramiaux and Marco Donnarumma. 2020. Artificial Intelligence in Music and Performance: A Subjective Art-Research Inquiry. In Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Crea
[18] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. "Explorers of Unknown Planets": Practices and Politics of Artificial Intelligence in Visual Arts. Proc. ACM Hum.-Comput. Interact. 6, CSCW2, Article 477 (nov 2022), 24 pages.
[19] Baptiste Caramiaux and Atau Tanaka. 2013. Machine Learning of Musical Gestures. In Proceedings of the International Conference on New Interfaces for Musical Expression. Graduate School of Culture Technology, KAIST.
[20] Minsuk Chang, Stefania Druga, Alexander J Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The Prompt Artists. In Proceedings of the 15th Conference on Creativity and Cognition. 75–87.
[21] Catherine Nicole Coleman. 2020. Managing bias when library collections become data. International Journal of Librarianship 5, 1 (2020), 8–19.
[22] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2023. Simple and Controllable Music Generation.
[23] Kate Crawford and Trevor Paglen. 2021. Excavating AI: The politics of images in machine learning training sets. Ai & Society 36, 4 (2021), 1105–1116.
[24] Dadabots. 2017. AI in metal music. https://dadabots.bandcamp.com/. [Online; accessed 8-12-2023].
[25] Antonio Daniele and Yi-Zhe Song. 2019. Ai+ art= human. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 155–161.
[26] Ajay Divakaran, Aparna Sridhar, and Ramya Srinivasan. 2023. Broadening AI Ethics Narratives: An Indic Art View. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2–11.
[27] Christopher Dobrian and Daniel Koppelman. 2006. The 'E' in NIME: Musical Expression with New Computer Interfaces. 277–282.
[28] Chris Donahue, Julian J. McAuley, and Miller S. Puckette. 2018. Synthesizing Audio with Generative Adversarial Networks. CoRR abs/1802.04208 (2018).
[29] Paul Dourish. 2001. Where the action is: the foundations of embodied interaction. MIT press.
[30] Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. 2020. DDSP: Differentiable digital signal processing. arXiv preprint arXiv:2001.04643 (2020).
[31] Sidney Fels, Ashley Gadd, and Axel Mulder. 2002. Mapping transparency through metaphor: towards more expressive musical instruments. Organised Sound 7, 2 (2002), 109–126. https://doi.org/10.1017/S1355771802002042
[32] Sidney Fels and Geoffrey Hinton. 1995. Glove-TalkII: An Adaptive Gesture-to-Formant Interface. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95).
[33] Rebecca Fiebrink and Baptiste Caramiaux. 2016. The machine learning algorithm as creative musical tool. arXiv preprint arXiv:1611.00379 (2016).
[34] Rebecca Fiebrink and Perry Cook. 2010. The Wekinator: A System for Real-time, Interactive Machine Learning in Music. Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010) (2010).
[35] Rebecca Fiebrink and Laetitia Sonami. 2020. Reflections on Eight Years of Instrument Creation with Machine Learning. In International Conference on New Interfaces for Musical Expression (NIME).
[36] Rebecca Fiebrink, Daniel Trueman, Perry R Cook, et al. 2009. A meta-instrument for interactive, on-the-fly machine learning. (2009).
[37] Jules Françoise, Norbert Schnell, Riccardo Borghesi, and Frédéric Bevilacqua. 2014. Probabilistic models for designing motion and sound relationships. In Proceedings of the 2014 international conference on new interfaces for musical expression. 287–292.

[38] Marco Gillies, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, Saleema Amershi, Bongshin Lee, et al. 2016. Human-centred machine learning. In Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems. 3558–3565.
[39] Donald Glowinski, Sélim Yahia Coll, Naëm Baron, Maëva Sanchez, Simon Schaerlaeken, and Didier Grandjean. 2017. Body, space, and emotion: A perceptual study. Human technology 13 (2017).
[40] Rolf Inge Godøy and Marc Leman. 2010. Musical gestures: Sound, movement, and meaning. Routledge.
[41] Andrea Guidi and Andrew McPherson. 2022. Quantitative evaluation of aspects of embodiment in new digital musical instruments. In NIME 2022.
[42] Michael Gurevich and Jeffrey Treviño. 2007. Expression and its discontents: toward an ecology of musical creation. In Proceedings of the 7th international conference on New interfaces for musical expression. 106–111.
[43] Drew Hemment, Ruth Aylett, Vaishak Belle, Dave Murray-Rust, Ewa Luger, Jane Hillston, Michael Rovatsos, and Frank Broz. 2019. Experiential AI. AI Matters 5, 1 (apr 2019), 25–31. https://doi.org/10.1145/3320254.3320264
[44] Drew Hemment, Morgan Currie, SJ Bennett, Jake Elwes, Anna Ridler, Caroline Sinders, Matjaz Vidmar, Robin Hill, and Holly Warner. 2023. AI in the Public Eye: Investigating Public AI Literacy Through AI Art. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 931–942.
[45] Francis Hunger. 2023. Unhype Artificial 'Intelligence'! A proposal to replace the deceiving terminology of AI. https://doi.org/10.5281/zenodo.7524493
[46] Don Ihde. 1990. Technology and the Lifeworld: From Garden to Earth. Indiana University Press.
[47] Midjourney Inc. 2022. MidJourney. https://midjourney.com
[48] Alexander Refsum Jensenius and Michael J Lyons. 2017. A nime reader: Fifteen years of new interfaces for musical expression. Vol. 3. Springer.
[49] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and Its Impact on Artists. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery, 363–374.
[50] Eun Seo Jo and Timnit Gebru. 2020. Lessons from archives: Strategies for collecting sociocultural data in machine learning. In Proceedings of the 2020 conference on fairness, accountability, and transparency. 306–316.
[51] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In Proceedings of the 1st international conference on Tangible and embedded interaction. 139–146.
[52] Sidsel Karlsen. 2011. Using musical agency as a lens: Researching music education from the angle of experience. Research Studies in Music Education 33, 2 (2011), 107–121.
[53] David Kirsh. 2013. Embodied cognition and the magical future of interaction design. ACM Transactions on Computer-Human Interaction (TOCHI) 20, 1 (2013), 1–30.
[54] Paul Kolesnik and Marcelo Wanderley. 2005. Implementation of the Discrete Hidden Markov Model in Max / MSP Environment. 68–73.
[55] Bruno Latour. 1990. Technology is Society Made Durable. The Sociological Review 38 (1990), 103–131.
[56] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. nature 521, 7553 (2015), 436–444.
[57] Michael A. Lee, Adrian Freed, and David Wessel. 1991. Real-Time Neural Network Processing of Gestural and Acoustic Signals. In Proceedings of the 1991 International Computer Music Conference, ICMC. Michigan Publishing.
[58] Marc Leman. 2007. Embodied music cognition and mediation technology. MIT press.
[59] Lewis. 1988. Creation by refinement: a creativity paradigm for gradient descent learning networks. In IEEE 1988 International Conference on Neural Networks. 229–233 vol.2. https://doi.org/10.1109/ICNN.1988.23933
[60] Lian Loke and Toni Robertson. 2013. Moving and making strange: An embodied approach to movement-based interaction design. ACM Transactions on Computer-Human Interaction (TOCHI) 20, 1 (2013), 1–25.
[61] Mclean J Macionis and Ajay Kapur. 2018. Sansa: A modified sansula for extended compositional techniques using machine learning. NIME.
[62] Avital Meshi. 2020. Deconstructing whiteness. In SIGGRAPH Asia 2020 Art Gallery. 1–1.
[63] Fabio Morreale, Elham Bahmanteymouri, Brent Burmester, Andrew Chen, and Michelle Thorp. 2023. The unwitting labourer: extracting humanness in AI training. AI & SOCIETY (05 2023), 1–11.
[64] Fabio Morreale, Megha Sharma, I Wei, et al. 2023. Data Collection in Music Generation Training Sets: A Critical Analysis. ISMIR 2023 (2023).
[65] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How data science workers work with data: Discovery, capture, curation, design, creation. In Proceedings of the 2019 CHI conference on human factors in computing systems. 1–15.
[66] Lorelli S. Nowell, Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. International Journal of Qualitative Methods 16, 1 (2017).

[67] Nicola Orio, Norbert Schnell, and Marcelo M. Wanderley. 2001. Input Devices For Musical Expression : Borrowing Tools From Hci. (2001).
[68] Victor Paredes, Jules Françoise, and Frederic Bevilacqua. 2022. Entangling Practice with Artistic and Educational Aims: Interviews on Technology-based Movement-Sound Interactions. In NIME 2022. PubPub.
[69] Roelof Pieters and Samim Winiger. 2016. Creative AI: On the Democratisation & Escalation of Creativity. Technical Report. https://medium.com/@creativeai/creativeai-9d4b2346faf3
[70] Ivan Poupyrev, Michael J. Lyons, Sidney Fels, and Tina Blaine (Bean). 2001. New Interfaces for Musical Expression. In CHI '01 Extended Abstracts on Human Factors in Computing Systems. Association for Computing Machinery, 491–492.
[71] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. arXiv:2102.12092 [cs.CV]
[72] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. 2018. A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. In Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80), Jennifer Dy and Andreas Krause (Eds.). PMLR, 4364–4373.
[73] Tom Roberts and Joel Krueger. 2022. Musical Agency and Collaboration in the Digital Age. 125–140. https://doi.org/10.5040/9781350197725.ch-007
[74] Oliver C. Robinson. 2014. Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. Qualitative Research in Psychology 11, 1 (2014), 25–41. https://doi.org/10.1080/14780887.2013.801543
[75] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV]
[76] Téo Sanchez. 2023. Examining the Text-to-Image Community of Practice: Why and How do People Prompt Generative AIs?. In Proceedings of the 15th Conference on Creativity and Cognition. 43–61.
[77] Hugo Scurto, Axel Chemla, et al. 2023. Deeply listening through/out the deepscape. In 28th International Symposium on Electronic Art (ISEA 2023).
[78] StabilityAI. 2023. Stable Audio. https://www.stableaudio.com/
[79] Luke Stark and Kate Crawford. 2019. The work of art in the age of artificial intelligence: What artists can teach us about the ethics of data practice. Surveillance & Society 17, 3/4 (2019), 442–455.
[80] Koray Tahiroğlu, Miranda Kastemaa, and Oskar Koli. 2021. AI-terity 2.0: An Autonomous NIME Featuring GANSpaceSynth Deep Learning Model. In NIME 2021.
[81] Koray Tahıroğlu, Thor Magnusson, Adam Parkinson, Iris Garrelfs, and Atau Tanaka. 2020. Digital Musical Instruments as Probes: How computation changes the mode-of-being of musical instruments. Organised Sound (2020).
[82] Atau Tanaka. 2006. Interaction, Experience and the Future of Music. Vol. 35. 267–288 pages. https://doi.org/10.1007/1-4020-4097-0_13
[83] Atau Tanaka, Balandino Di Donato, Michael Zbyszynski, and Geert Roks. 2019. Designing gestures for continuous sonic interaction. NIME.
[84] Alex S Taylor, Siân Lindley, Tim Regan, David Sweeney, Vasillis Vlachokyriakos, Lillie Grainger, and Jessica Lingel. 2015. Data-in-place: Thinking through the relations between data and community. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2863–2872.
[85] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. CoRR abs/1609.03499 (2016).
[86] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention Is All You Need. arXiv:1706.03762 [cs.CL]
[87] Gabriel Vigliensoni, Rebecca Fiebrink, et al. 2023. Steering latent audio models through interactive machine learning. (2023).
[88] Federico Visi and Luke Dahl. 2018. Real-time motion capture analysis and music interaction with the modosc descriptor library. NIME.
[89] David L. Wessel. 1979. Timbre Space as a Musical Control Structure. Computer Music Journal 3, 2 (1979), 45–52.
[90] Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang. 2017. MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation.
[91] Shuntaro Yoshida and Natsumi Fukasawa. 2022. How Artificial Intelligence Can Shape Choreography: The Significance of Techno-Performance. Performance Paradigm 17 (2022), 67–86.
[92] Victor Zappi and Andrew P McPherson. 2014. Dimensionality and Appropriation in Digital Musical Instrument Design.. In NIME, Vol. 14. Citeseer, 455–460.

A QUESTIONNAIRE OF THE INTERVIEWS

A.1 Introduction

Thank you for having accepted this exchange with us. This interview is conducted as part of a research project that we are running at Sorbonne Université. This research project investigates the use of ML and AI for musical creative practices.

(1) Can you give a general introduction of yourself ?
(2) Can you present your work using ML with one significant example for you ?

A.2 Main questions

(1) What types of ML techniques are you using?
(2) What specifically are you looking for in these technologies?
(3) Which ones are you not using? For what reason ?
(4) What about the use of datasets, do you re-use existing dataset or construct your own dataset ? If so, how ?
(5) Are you concerned by the fact that others can re-use your musical data for their own purpose (such as companies that train their model with artist's music) ?
(6) Do you implement everything by yourself or reuse parts of existing projects / adopt complete systems made by research teams or companies ?
(7) Concerning the existing technologies that you use, have you encountered technical problems emerging with one of them ? At which step of your creative process ?
(8) How did you handle these problems ? Could you handle them all by yourself ?
(9) Still concerning the technologies you use and those you have consciously chosen not to use, have you felt that they might at some point conflict with your personal way of making music, but also your personal values or ethic?
(10) Again, how did you solve each of these conflicts?
(11) There is a lot of advancement in ML for music each year. To continue talking about your interaction with this technology but beyond the technical aspects. How do you keep up to date with new technologies? Are you active or passive in finding out about these ?
(12) Talk about it in social media, in a specific community, and in which way generally ?
(13) Do you think that recent developments and the digital practices promoted by technology owners go in the same direction as what you conceive of as your musical practice, or do you propose a counter-voice to these digital practices?
(14) Can you explicit which aspects of the proposed practices are in line with or opposed to your musical practice?

A.3 Conclusion

We are reaching the end of the discussion, do you have any concluding remarks or subject you would like to talk about?

B DESCRIPTION OF THE ARTISTS

P1	Argentina	Composer, sound designer	Combines experimental electronic soundscapes, multilingual voice sample processed with ML to create immersive experiences for virtual ecosystems and live performances.
P2	Austria	Composer, performer, researcher	Use AI-based collaborative music composition system based on symbolic generation
P3	United States	Sound artist, performer, researcher	Design an play instrument based gesture-sound mapping and real-time audio synthesis for performances
P4	Canada	Music artist, producer, performer, and researcher	Implement a custom machine learning-driven musical instrument to make and explore rhythms and sound textures
P5	Italy	Sound designer, researcher	Develop interactive musical systems and instruments for music improvisation, audio-visual composition, sonic interaction design
P6	Japan	DJ, performer, composer, researcher	Use AI models on stage to generate real-time symbolic music(i.e., MIDI) and control generative AI models and drum machines.
P7	Lebanon	artist, writer, technologist, producer, DJ	From composition to machine learning, performance, visual art. Adopt a critical exploration of machine learning tools for speech synthesis, vocal encryption and DJing
P8	Germany	Singer, songwriter, producer, and an interactive installation artist	Create musical and visual works, co-create lyrics and melodies with an AI version of their voice
P9	Iceland	Researcher, performer	musical performance, improvisation, new technologies for musical expression, live coding, musical notation, artificial intelligence and computational creativity
P10	Norway	Interdisciplinary artist, researcher, performer	Investigates real-time sound and music control distributed among human performers and artificial agents
P11	Finland	Musician, performer, researcher	Create and perform with AI musical instrument, explore embodied approaches to sonic interaction, participative music experience, multimodal physicality in sound and interaction
P12	England	Researcher, composer, performer	Do musical performance where the human body becomes musical instrument thanks to gesture-sound mappings
P13	England	Computer-musician and musical instrument designer	Performs with custom-made instruments including malleable foam interfaces, touch screen software, machine learning and interactive sculptures
P14	United States	Composer and cellist	Focuses on gesture in music, the sustainability of technology in art, and sonification of data

Table 1. Description of the artist with demographic information, status and a short description of their work