AIMC 2024 (09/09 - 11/09)

Learning to Learn: A Reflexive Case Study of PRiSM SampleRNN

URL: https://aimc2024.pubpub.org/pub/fnpykfdv

ABSTRACT

The emergence of neural audio synthesis technology has opened up many new creative and collaborative avenues for musical practitioners in recent years. With a growing number of software tools becoming openly accessible, many composers and sound artists start to map their music-making processes into a nebulous, data-informed collaborative framework. This often puts the practice of data curation, generative machine-learning models, as well as the artistic usage of machine-generated outputs into a state of play, whereby much of the idiosyncrasy of the resultant work is shaped by fine-tuning deep-learning algorithms. However, issues surrounding agency, distributed creativity, and access to computational resources / specialists tend to surface. This paper looks at these issues within the existing infrastructure of a Music Conservatoire, where to engage creatively and strategically with data and artificial intelligence tools becomes an increasingly important skill for artists to adopt outside their conventional musical training. Through the lens of the work of PRiSM (The RNCM Centre for Practice & Research in Science & Music) and the rollout of PRiSM SampleRNN between 2020-2022, we identify an emergent model of musical training and research that institutionally facilitates knowledge exchange and collaborative dialogues between practitioners, pedagogues, as well as research software engineers who are often not considered part of the existing conservatoire establishment.

List of Abbreviations

AA - Anna Appleby
AI - Artificial Intelligence
BM - [Author 1]
CM - Christopher Melen
DT - Dawn Tse
EH - [Author 4]
ES - [Author 2]
HF - Hongshuo Fan
JdA - José del Avellanal
LG - Lara Geary
NOVARS - The NOVARS Research Centre at the University of Manchester
ML - Machine Learning
ML4M - The Machine Learning for Music working group and collaboration between PRiSM and NOVARS
PRiSM - The RNCM Centre for Practice & Research in Science & Music
RL - Robert Laidlow
RNCM - Royal Northern College of Music
RNN - Recurrent Neural Network
RSE - Research Software Engineer(ing)

SS - Sam Salem
TA - Tasos Asonitis
TR - Tywi J.H. Roberts
VC - Vicky Clarke
ZL - Zakiya Leeming

Introduction

Since the release of many open-source machine-learning (ML) frameworks such as TensorFlow[1] (Abadi et al., 2015) and PyTorch[2] (Paszke et al., 2019), a growing community of artists and technologists have become interested in researching how artificial neural networks could facilitate nebulous compositional and creative tasks (L. Sturm & Ben-Tal, 2017; Melen, 2020b; Salem, 2020; Sturm et al., 2019). This has included interest in neural audio synthesis – and specifically research (from both academic and private sectors) in using deep-learning techniques to recognise, analyse, and manipulate patterns and features of any raw audio data, as well as to generate new audio samples based upon them (Civit et al., 2022; Clancy, 2022; Dyer, 2022; Kehagia & Moriaty, 2023; Laidlow, 2024; Mehri et al., 2017). This research produced deep-learning models and algorithms such as WaveNet, SampleRNN, Jukebox and RAVE, and encouraged many practitioners to start utilising machine-generated audio samples in their work.

This way of working has brought software engineers and creative practitioners together in new ways, calling attention to the collaborations emerging between artists and technologists. Despite this, a review of recent studies in musical AI suggests that the dialogues between research software engineers (RSE) and the artistic users of the tools developed tend to be overlooked (Ben-Tal et al., 2021; Civit et al., 2022; Clancy, 2022; Dyer, 2022; Engel et al., 2017; Kehagia & Moriaty, 2023). Yet these emerging dialogues are suggestive of a novel approach towards software design, refinement and dissemination, one that is iterative and collaborative where the artistic incorporation of tools that are in-development helps to identify technical advantages and shortcomings; consequently tools are built largely upon a negotiation between conceptual artistic aims and a statistics-based solution-finding. This negotiation often leads to new practices and highly personalised strategies to curate, optimise and interpret audio datasets, effectively enabling ‘innovative forms of expression whilst problematising traditional practitioner roles, methods of praxis, and epistemologies’ (Kanga et al., 2024, p.5).

PRiSM SampleRNN: A Case Study

Through this paper, we offer a highly positional point of view of collaborative dialogues between artists and RSEs within PRiSM, the Centre for Practice & Research in Science & Music at the Royal Northern College of Music (RNCM)[3], during the development and rollout of the neural audio synthesis software *PRiSM SampleRNN* (Melen, 2020a) between June 2020 and October 2022.

Refined and reimplemented for TensorFlow upon the pre-existing SampleRNN architecture (Mehri et al., 2017)[4], PRiSM SampleRNN is a deep-learning software tool that operates in and through audio-based recurrent neural networks (RNN)[5]. It performs "unconditional audio generation in the raw acoustic domain" (ibid, p. 9), specialising in generating new audio outputs through modelling the probability of a sequence of waveform samples encapsulated in any existing corpus of sound or music. Launched in June 2020, the software was PRiSM's first major open-source contribution to the field of neural synthesis. The project was initiated by composer-researcher Sam Salem and coded by musical technologist Christopher Melen, PRiSM's Research Software Engineer between 2019 and 2023.

This paper provides a case study on the collaboration between artists and Melen through a selection of creative projects undertaken by artists who used PRiSM SampleRNN between June 2020 and October 2022. Significantly, this study includes artists who helped initiate the project and those who informed the iterative design and refinement of the tool at varying stages. This study sought to understand these artists' chosen methods to engage PRiSM SampleRNN (contextualised into their various pre-existing knowledge of programming and neural synthesis); their interaction with the RSE and the impact this had on their process. Taking place in a conservatoire, this project additionally sought to understand how new models of shared learning could engender new creative and teaching strategies in the age of machine learning (ML) and AI.

Methodology

The objectives of this study were:

1. To observe how PRiSM SampleRNN has been deployed to fulfil diverse creative objectives;
2. To understand various artistic needs when engaging neural audio synthesis technologies;
3. To identify gaps of knowledge concerning the curation and management of custom audio datasets, emerging from dialogues with both the RSE and creative practitioners;
4. To collectively share new strategies of working with trained deep-learning models and machine-generated audio outputs, and thus creating data-rich musical / compositional works;
5. To extrapolate key advantages of engaging with the RSE and various iterations of software development in and through creative processes, and to discuss any logistical / access barriers around this way of working and possible workarounds.

The authors of this paper were involved as artist-researchers throughout this study, which was conducted over four interconnected phrases:

1. Participatory narrative inquiry / group workshops;
2. Preliminary data collection;
3. Semi-structured interviews;
4. Themed data analysis, interpretation, and report.

Using elements of participatory research, as well as collective and collaborative auto-ethnography, this study foregrounded a 'co-construction of research and interview questions with co-narrators' (Karalis Noel et al., 2023, p.2) as the primary device to facilitate intersubjective data collection and to foster the "narrative interpretation (analytical) rather than interpretative narration (evocative) of autobiographical data" (Chang et al., 2016).

The study commenced in March 2022 through a series of narrative inquiry workshops involving the core research team [Author 1] (BM), [Author 2] (ES) and [Author 4] (EH) as well as participants associated with the iterative development of PRiSM SampleRNN - RSE Christopher Melen (CM); Sam Salem (SS), Robert Laidlow (RL), José del Avellanal (JdA), Zakiya Leeming (ZL), Hongshuo Fan (HF), Tywi J.H. Roberts (TR), Tasos Asonitis (TA), Vicky Clarke (VC), Anna Appleby (AA), and the artist-mathematician duo Lara Geary (LG) and Dawn Tse (DT). These workshops drew on characteristics of group interview, where all participants gathered as co-researchers to discern commonalities between their highly varied practice research narratives during a common research time period. These workshops were chaired by BM and ES and considered as preliminary data collection for this study.

Informed by this data collected, we formulated a set of questions to ask each artist-researcher through a series of semi-structured interviews.[6] Contrarily, the interview with the RSE (CM) took place at a later stage of the series and was informed by patterns that had surfaced from the semi-structured interviews with the artist-participants. All interviews were informal and all participants had the opportunity to engage in dialogues with BM and ES concerning their own experiences, as well as to ask questions relating to other aspects of the study. Each session was video recorded, and transcribed by the two interviewers in a hybrid form of selective to intelligent verbatim[7] (McMullin, 2023). Each transcript was then interrogated as part of a themed analysis carried out by the core research team. This paper thus serves as a summary report, interweaving self-reflections of each participant-researcher with key findings emerging from the analysis.

Discussion

Artistic Motivation

In order to best understand their nebulous working relationships with PRiSM SampleRNN, we started by looking at the context of each artistic project, and what drove artists to incorporate PRiSM SampleRNN in their creative processes.

The Programmatic: Institutional Encouragement

Our dialogues suggested that, in most cases, AI-related artistic inquiries tend to correlate strongly to the participants' pre-existing knowledge of, and/or their prior interaction with digital musical technology. For participants focused on an acoustic (primarily instrumental) compositional practice, the programmatic context of the project was one of the most pronounced driving factors:

Table 1: Programmatic Motivations
“There was a call for proposal (for RNCM composition students) through PRiSM [...] for new solo pieces with electronics connecting to the research PRiSM was doing”;
“I attended a meeting with CM and heard RL’s works [exploring AI music] and learnt much about the software, and I was then encouraged by EH to use it for my solo singer-songwriter project”.

The ML4M working group provided another crucial programmatic context for this work. The group was founded in late 2020, following the official launch of PRiSM SampleRNN. At first, the group was solely joined by RNCM doctoral composer-researchers and NOVARS-based doctoral researchers focused predominantly on electroacoustic musical practices. With the premises of jointly producing an event showcasing new ML-driven audiovisual works (as part of PRiSM’s Future Music #3 Virtual Festival in June 2021), many members naturally turned to PRiSM SampleRNN, given that they were ideally positioned to explore it in-depth and to connect it with their own research interests. And as CM commented, “With ML4M initiated, people were finally made aware of SampleRNN’s potential and I was suddenly being introduced to new people all the time who wanted to know how they could use it in their work”.

The Technical: Prior Encounter with Generative AI

Much of this programmatic context also benefitted participants who had already engaged generative AI to various extent, and those that had already been researching pre-existing examples of projects exploring neural synthesis. These practitioners often specialise in sound-art, multi-media production, and/or interactive performance systems. They tend to share an interest in specific technical challenges which could not have been solved by other tools.

Both SS and RL had utilised the earlier WaveNet algorithm in their work prior to SampleRNN. For them, to reimplement SampleRNN and to “modernise” it with the most up-to-date Python dependencies proved a good step forward considering that “no one had really set up WaveNet back in 2018 and it was really complicated to use”. Working closely with CM throughout the code modernisation process would also enable more powerful and customisable features to be added to the resultant tool, making it more accessible to a wider artist-practitioner community.

Sound and electronic media artist VC approached PRiSM and ML4M especially for an Artistic Residency centred upon using PRiSM SampleRNN - having had a “taste of neural synthesis” while working on another project previously. An audiovisual artist with an interest in the “black box phenomenon” around AI and data extraction / mapping, TA aspired to “have access to the evolution of SampleRNN’s audio generation process”, and to “let SampleRNN to collaborate with other synthetic neural network systems[8]”.

LG presented the first known use case of PRiSM SampleRNN outside of the RNCM/NOVARS communities. As an interdisciplinary artist exploring generative media, her work with ML has been frequently facilitated through collaborating with mathematician DT, who only began to “dive deep into the mathematics behind ML technologies” and learning to program after “coming across a research paper by Catherine and Desmond Higham discussing the current understanding of Deep Learning from the perspectives of applied mathematicians [Higham & Higham, 2018]”. Similar to SS’s and RL’s motivations, they turned to PRiSM SampleRNN upon realising that “the GitHub entries for WaveNet were not maintained properly”, and that “CM’s constant maintenance of the SampleRNN code [resulting from his working dialogues with other participants at the time] helped the community tremendously”.

The Art-anthropological: A Shared Curiosity

In discussion of works incorporating neural synthesis technologies by SS and the composer-vocalist Jennifer Walshe, composer-researcher Mark Dyer (2022) identifies an emergent network of musical practices that posits art-anthropological inquiries through recognising and collaborating with “the cultural status and agency of algorithms (and thus neural networks themselves)” (p. 224).

Dyer maps the technological anthropologist Nick Seaver’s proposition of “algorithms as culture” (Seaver, 2017) onto Tim Ingold’s conception of “learning to learn” (Ingold, 2013), whereby a dynamic, proactive process of “working with machine learning, whose unknown logic has the potential to show new and unpredictable versions of the world it is trained upon” (Dyer, 2022, p. 225) drives “a back-and-forth, continuous flow of cultural exchanges” that enables the subjects to “know the world and perhaps (the multiple versions of) themselves a little differently” (ibid, pp. 224-5). And this art-anthropology is, as Dyer delineates, often also autoethnographic by nature. It extrapolates from data closely associated with the autobiographical and the lived encounters with “the world in the twenty-first century” (ibid, p. 225), and through artistic discourses that “‘respond’ rather than describe” (Ingold, 2013, p.7, as cited in Dyer, 2022, p. 225) to invite further critical problematisation of the socio-cultural environments around contemporary lives.

This art-anthropological approach, as we found across the interview data, is also widely shared amongst our participants. Despite pronounced disparities in their technological literacy and prior programming experiences, nearly all participants - through their creative work - made inquiries about the multifarious agencies afforded through a human-algorithm collaboration as well as enquiring after a technologically mediated articulation of their own artistic priorities:

Table 2: Art-anthropological Motivations
“[...] about the industrial revolution, and the mechanisation of work, and of course this is all parallel to AI [...]. I suppose [...] these algorithms - their very loose generation of ‘what music is’ somehow meets how an experimental composer might want to reassess the definition of music”;

“[...] to see what would happen if human and the machine answer to the same thing, [...] to hear variations of what I thought I could hear, [...] it’s not conceptual but it’s just a way of seeing things differently”;
“[...] about metamorphosis - [...] this peculiar Welsh myth talking about the transformation from human to other creatures, about amalgamation, which to me is a great analogy to how AI works”;
“[...] interested in the alphabet and rudimentary manipulation of emergent meaning out of my own spoken voice - I’m not interested in natural language processing, but in making sound that isn’t quite sound”;
“[...] keen to see what would happen if I train an audio generating algorithm on a very large dataset, similar to how they would do it in Silicon Valley with these large commercialised models, [...] to see how musicians can imitate AI sounds [...] like ‘a weird whistling project’”;
“I was only aware of the ML audio examples around that time such as Dadabots and CM’s earlier Beethoven dataset. [...] curious to see how a completely different dataset would bring up interesting results - how the subtlety of the performance and conversation-led processes of research can be informed by the materiality of machine-generated audio samples: how, and why, and what are we investigating about the physically rich sound when the communications had to be made and mediated by technology?”;
“[...] interested in how sound objects can be translated statistically, and the materiality of it. [...] I heard about SampleRNN, understood from the text how this ML stuff works but wanted to be really hands-on with it - I wanted to see whether it’s a tool, or a collaborator”;
“Following closely the conversations between SS and CM, I was curious to see how this ML software can provide new collaborative ways of working with electronics. I wanted to see what musical features the algorithm picks up and how it picks it up - I think it has a ‘musique concrète’ bent to it.”
“[...] keen to play with the socio-cultural overtones of an object or material, and an act of translation that doesn’t quite work. I’m curious to see how [the algorithm] facilitates this - how it helps communicate my intentions in ways that [these intentions] are bound to get misunderstood, and how [working this way] changes my original idea”.

We recognised that these three types of motivation are, in most cases, interwoven across the whole of the creative process and are actively impacting on the participants’ creative decision-making. They imply a networked, collaborative negotiation between the artists, the data they decided to work with in the first place, the ways in which the dataset is assembled, the ways in which the assembled dataset is learnt and processed, and a technologist/RSE who is often not considered part of the artistic practice but whose intervention leaves multifarious footprints on the output.

Collaborating with Data, RSE, and the Algorithm

For most participant-researchers, the creative process involved building collaborative relationships with two elements previously *unknown* to them - the technology and the RSE. The latter connotes not only the impact of

working with RSE for the first time, but also of working with a particular technologist. That said, no participants - at least through their conversation with us - articulated that collaborating with a / the RSE was an artistic motivation for them, despite many (including ES and BM) engaging in critical collaborative inquiry within their artistic practices[9].

Our use of the term *collaboration* is intended to encapsulate many types of working relationship, which scholarship over the last two decades has categorised into labels such as ‘integrative’, ‘family’, ‘complementary’ and ‘distributed’ collaboration (John-Steiner, 2000); ‘directive’, ‘interactive’ and ‘collaborative’ working (Hayden & Windsor, 2007); ‘hierarchical’, ‘consultative’, ‘cooperative’ and ‘collaborative’ working (Taylor, 2016) and, turning these composer-centred models on their heads, a model categorising the performer’s role as either an ‘interpreter’, ‘advisor’ or ‘deviser’ (Torrence, 2018).

Our analysis considers these relationships from the ground up by looking at the RSE’s role in each project and at each stage of the artistic processes. Drawing upon its impact on a broader, distinctive interaction between human and nonhuman actors (Born, 2005), we strive to uncover patterns emerging from the triangulation between the artists, RSE, and their collective yet differently positioned exploitation of custom audio data and the SampleRNN algorithm.

Dataset Curation

We noted that the majority of our participants only began to collaborate with their RSE after curating the dataset for their envisioned artistic projects. To assemble a corpus of training audio data was often viewed as a task completed by the artist (or artists) as part of their creative process. Despite this project taking place in an educational institution, little ‘learning from’ or ‘studying with’ (Ingold, 2013, p.2) the RSE happened at the point of dataset curation. This is despite CM’s own active learning across the immediate time period after releasing PRiSM SampleRNN:

> “People often picture ML as a dry mechanical process, but it’s not. It’s very fluid - a lot like composing, like improvisation. I was re-writing bits of the code as a result of working with JdA. I realised I was going to be constantly responding - learning about its limitations and what improvements needed to be made.”

As a result, for many participants working around this period, guidance towards their provision of training data was often simply to produce “at least one hour of wav files in a 16kHz sample rate[10]”. From there, many developed their own criteria for curating the dataset. We found these criteria fell into four areas of artistic concern:

1. The origin of sound sources;
2. The quality of audio files;
3. The organisation of audio files in the dataset;

4. Critical appraisal of the relationship between audio files.

Whilst all participants prioritised A when collecting their data, fewer considered D, and only a selected few - including those who had worked with audio neural synthesis (e.g. WaveNet) previously - critically considered C. These participants showed a nuanced understanding of how the RNN would process their audio data, which can be surmised as “I was kind of aware of the need for a more timbrally diverse, eventful collection of sounds [...]”, evidenced in other processes by comments such as “[I used] contrasting datasets featuring music of different genres, presenters’ voices etc”.

This led participants to reflect that their work might have benefitted from collaborating with the RSE at an earlier stage. In some cases, there was seemingly a missed opportunity for an in-depth conversation with the RSE regarding a critical understanding of how “the training might have been made more ‘successful’” and what ‘success’ means to both sides. Comments were suggestive of the need for an interactive knowledge exchange amongst practitioners - who understand the technology at varying degrees - to identify potential misconceptions of the attributes of the algorithm, particularly as the software tool undergoes constant refinements and therefore improves.

Algorithmic training

Collaborative dialogues, as our study reveals, predominantly took place during the RSE-facilitated training processes of PRiSM SampleRNN. These tended to yield possibilities whereby the artists were better informed about the training mechanics, and/or both the artist and RSE could engage in discussions around the hyperparameters of the algorithm[11]. Artists could participate in this process through listening to, or reading visualisations (e.g. on the TensorBoard) of the training progress and outcome - not needing an advanced understanding of the technical specifics of the RNN. The following comment from RL shows how decisions over data processing were distributed between him and CM:

> “It took me a good while to convince CM that [increasing the sample rate] was a good idea [...] if you are interested in the metrics of the training session then you would perhaps accept 16kHz, but not if you are interested in the audio. We then increased it to 44.1k [and it] took about 36 hours to complete one epoch, and we had 10 epochs in total [...] The more details contained in the dataset resulted in [SampleRNN] generating files with more musical interest - no longer being very drony [what SampleRNN tended to generate in the past]. And each Epoch has a different sound in it - it just sounds increasingly better”.

We categorised the relationships artists described with their RSE as:

1. Consulting RSE about the code but working to train models independently;
2. Handing over the training process to RSE but with distributed direction of the training process based on listening to sonic outcome and/or technical criteria such as determining the output was underfitting/overfitting;

3. Handing over the training process to RSE who independently directed the training process (based on listening to sonic outcome and technical knowledge);
4. Indistinguishable relationship.

Modes A, B, C occurred fairly evenly across the interview data we studied.[12] Where some of the working relationships fell exclusively into one of the three categories, others were more fluid, moving between patterns at different stages of the work. Notably, some of those artists who had not worked with neural synthesis in the past were able to participate in a distributed model through listening to or reading visualised training results and discussing this with the RSE, which in many cases helped facilitate a reflexive knowledge exchange (and collaborative relationship) towards the definition of an idiomatic dataset.

It was also significant that across all of the four categories, comments pertaining to ownership over the training process and generated outputs emerged. In some cases, the artist exhibited a desire for ownership within what had become a necessarily distributed process:

> "[I] tried to run the code directly first but it didn't work. [...] I never got anything usable from Colab Notebook either. [I then] needed to create more patches/audio to train the model, and [RSE] had to be creative in playing around with the hyperparameters. Got to trust [RSE] in the end."

And in other cases the artists reflected on how moving into Mode C impacted on a process that they had originally envisaged:

> "[RSE] would filter out stuff that 'didn't work' but I always wanted to hear everything even if it's rubbish."

The following comments, when discussing expectations from a specific training session, best articulate this reciprocal mediation from the standpoints of both an artist and a RSE "wearing the hat of a pure data scientist":

Table 3: Differences in training objectives	
The Artist	**The RSE**
"I wanted to see if I can hear stuff differently and how it elicits new ways of listening".	"I wanted to see if [SampleRNN] can generate something similar to the original - hence [I was] adjusting the hyperparameters like a religion".

While in isolation, these comments suggest tensions in some of the collaborative relationships, CM exhibited a reflexive awareness of his position as a collaborator and how this changed over time as he balanced his own desires for an optimal SampleRNN training with those of his collaborators:

> "[I] tried to be agnostic [...] didn't make judgements but could have made a lot of judgements about all sorts of things [...] I liked complex stuff but [the artist] liked glitchy, ghostly whispers. [...] didn't want it to go in that direction and thought of encouraging [the artist] to go for the opposite but [...] you can't do that in the end. It would be unethical."

He went further to describe the type of working relationship he favoured, which would allow him to be further "detached" from the process of determining artistic usability of the generated material:

> "The composer does have to understand they are not just getting generated audio [from SampleRNN] but are getting a model. If I could give the model to the composer, then they could make the decisions about what material to generate - SS does that - [...] to work with the model."

That said, by studying these RSE-artists dialogues, we recognised that to creatively and collaboratively utilise PRiSM SampleRNN often foregrounded - echoing composer-technologists Artemi-Maria Gioti and Aaron Einbond (Gioti et al., 2023) - a proactive "questioning of common understandings of ML processes as closed tasks oriented to optimization, instead reframing them as open-ended experiments, serving purposes of aesthetic experimentation and imaginative critique" (p. 16). This also calls for further investigation concerning the ontological and epistemological framework of the distributed decision-making throughout a RSE-facilitated ML training process, in order to further theorise the directional structure of these highly dialogical interactions.

Approaching Generated Outputs

Similarly, to work with machine-generated audio samples is often, as asserted by SS, "not a set recipe". Despite our study not focusing on detailed analyses of any of the artistic products resulting from exploiting PRiSM SampleRNN, learning about how samples were approached during diverse musical processes helped reveal novel patterns and considerations enacted by one of the latest - as the anthropologist Georgina Born (2022) posits - "epochal shifts in the appearance and prevalence of media technologies" (p. 4).

Our line of inquiry here concerned the participants' chosen methods of auditioning sounds generated by PRiSM SampleRNN, their ever-shifting relationships with these sounds informed by working with RSE throughout the earlier training processes, and their reflexive interpretation of this "human-music-technology assemblage" (ibid, p. 223) in relation to artistic autonomy, iterative learning, and "the inherently aesthetic nature of music data and the distinctive qualities of material engagement with ML" (Gioti et al., 2023).

As we discussed previously, the output auditioning process often started as early as algorithmic training and was woven with an ongoing critical reassessment of the training audio data. Akin to artists having to define what an optimally organised training set means to them, they also had to come up with their own rule of thumb to determine the artistic value/usability of every audio clip generated. This presented a nuanced disruption to decision-making, which was shared by many of our participants:

> “It was really, really hard to work with these materials. It’s all different but it’s all the same.”

SS compared this to “working with field recordings and found sounds”, where the focus was ideally placed upon “sound rather than concept - separate from concepts that lead into the sound”. To some extent, this implied an attentive listening to “the dynamics of more distinct sound objects” and “the uncanniness - the horror and disgust that you feel towards something that seems familiar but it’s not”.

Analogously, others mapped this listening exercise onto a more performative assumption of machine creativity. For TA, it was to represent “the journey [SampleRNN] went through during the [training] process” by rigidly spotlighting samples generated out of every five epochs, regardless of how much the material endorsed his original speculation that “it [would] improve over time”. For VC, it was a back-and-forth process of “trying to become a machine myself” and “to listen back in a human way”, in order to “make sense of the material [...] and to find the commonality - tonal information, events etc.”.

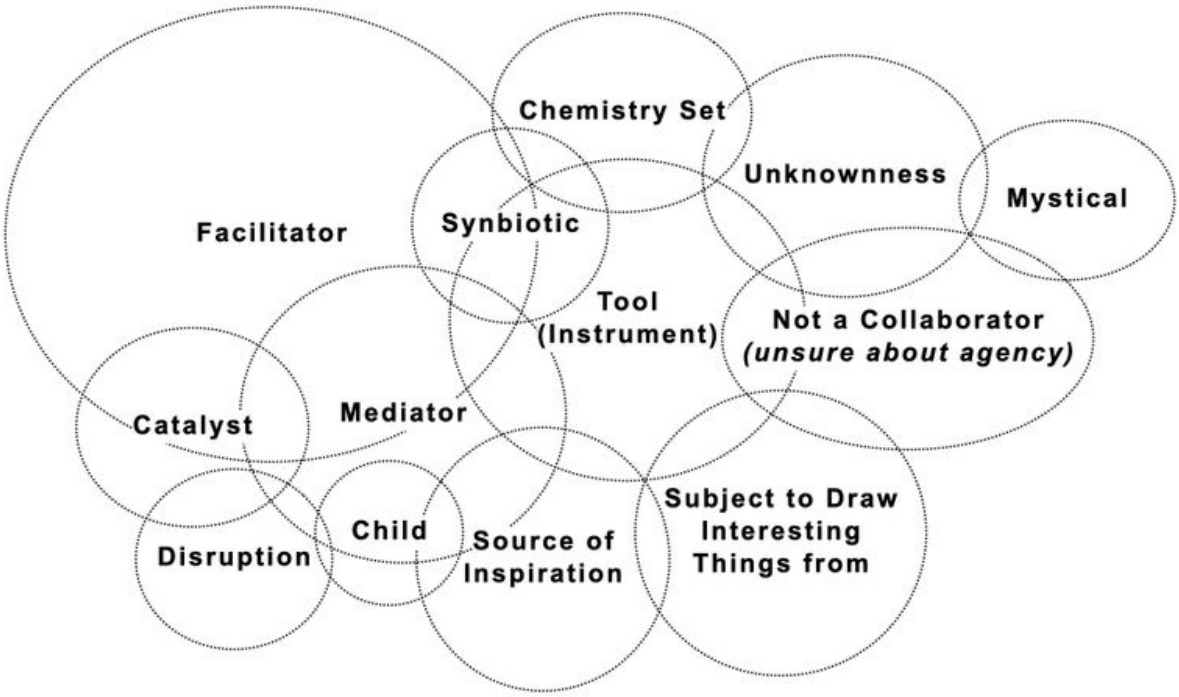
Although CM was often not directly involved with the artistic projects after helping artists to fulfil ML tasks, his methods for validating training progress were impacted by learning how artists listened to the generated samples, which in turn further consolidated the novelty of a RSE-facilitated artistic practice:

> “I used to decide on what material to generate in the first place by not listening - [I] didn't use to test audio, but to use the charts to see where things were going. [...] But I now listen to everything - from the position of the artist I work with [...] as far as I can”.

Emerging from these reflections was also a multifaceted interrogation of artistic autonomy, encompassing diverse approaches to manipulate AI-generated material; the aesthetic, ethical concerns around being “faithful to the training outcome”. This was especially relevant to those who worked with PRiSM SampleRNN immediately after its release, given that they had to work around low sample rates and mono outputs “with a distinctive Lo-fi quality”:

> “Perhaps it’s useful to think of [SampleRNN outputs] as video footage if I were working on a film. [...] didn’t want the piece to be about my own knowledge of producing electronics but rather about the AI itself”.

These statements prompted us to question a multifarious role that PRiSM SampleRNN played in each of the processes we examined. We asked each artist-participant to summarise, using one word or a short phrase, what the algorithm meant to them as they developed their projects:



The role of PRiSM SampleRNN according to participants' self-reflection

As the above figure shows, the answers we collected hinted at a tangled network concerning a flux of agency and materiality of the software tool, implicitly foregrounding a collective search for - echoing the artist-technologist Oliver Bown (Bown, 2021) - "the central role of the human as artist" (p. 298); "the basic *use* of art as a human social pursuit based around individual authorship" (ibid); as well as "new ways to think about what it means to be an artist, including extending the term to machines and transforming it as we do" (ibid, p. 319).

It is worth noting that, although such labels as "child", "chemistry set" occurred on the spectrum, they were discussed with particular regard to the amount of care and attentiveness required - from both the artists and RSE - to make PRiSM SampleRNN- / AI-inflected processes work. By going through these highly labour-intensive, iterative learning processes, as VC surmised, "the idea of automation is debunked [...] not as simple as saying that robots are taking over", which was also mirrored by TA's final remark during our conversation:

> "Initially I was trying to remove the human from the process, but probably the piece ended up being more human than any of my other pieces."

Conclusions

Tools, Institution, and *Learning to Learn*

Throughout the period of 2020-2022, RNCM PRiSM was in the unique position as the only UK specialist musical conservatoire to have employed a RSE to work with students, staff and artists-in-residence on developing new tools and processes concerning audio neural synthesis and data-rich musical practices.

This institutional structure enabled both learners and professional practitioners who had no prior experience of working with ML to embed the technology into their work: of the thirteen interviewees, eight had never directly engaged generative AI tools in their musical work. These were facilitated for the first time through CM (RSE) and via deep-learning facilities that PRiSM acquired in late 2019, both of which were only made possible through a Research England E3 fund for the centre to establish “a unique approach within UK HEPs [Higher Education Providers] and distinct worldwide” (UKRI, 2023), bringing together “researchers and practitioners in composition, performance, mathematics, artificial intelligence, music perception and big data to engage in creative research collaborations between the sciences and music” (ibid).

Significantly, this context afforded ripe conditions for novel collaborative interactions between artists and RSE to emerge, and with a distinctly specialised focus on music-making. These interactions were, echoing earlier propositions from both Kiri Wagstaff (2012) and Sturm et al. (2019), precisely what the majority of ML researches in recent years tended to neglect - “how the technology [that the researcher is] developing actually impacts practitioners, and how that in turn can inform the research pursuit” (ibid, p. 37).

Through our highly positional discussions around these interactions resulting in and from this PRiSM SampleRNN case study, we discovered a pivotal interdisciplinary model of *learning to learn.* This, enacted upon concurrent scrutinies of the technoscientific, art-anthropological, programmatic and ontological merit, as well as upon material agency surrounding the development of ML-driven musical software tools and the data they operate upon, extends a collaborative, open transfer of knowledge to every stage of the creative process. This brings an audio-based ML research closer to its targeted end users, but also urges musical educators to critically engage, assess and adapt to how new tools complement the way they - alongside their students - might approach music-making in the age of AI.

Building on this, our summary report calls for further empirical studies on the knowledge flow between RSE and musical practitioners, revolving around an ‘agonistic-antagonistic’ (Barry & Born, 2013) negotiation - stemming from “a commitment or desire to contest or transcend the given epistemological and/or ontological assumptions” (ibid, p. 12) - of the ownership over ML training processes, as well as the distributed creativity manifesting across any RSE-facilitated musical collaborations. This might further unpack the ways in which artists and technologists can come together in shared inquiry.

In the meantime, we acknowledge a multitude of logistical barriers around this way of working. While embedding a RSE into a musically specialised institutional structure engenders critical, reflexive, humanities-led interdisciplinary inquiries on ML technologies, to sustain this model also necessitates a more systematic review of the existing knowledge exchange frameworks around musical education, research funding, and distribution of resources. For example, the absence of a more comprehensive faculty of Computer Science in the vicinity of a conservatoire-based RSE often connotes inadequate access to, or investment in specialised computational resources, manpower, hardware / cloud-based data processing infrastructure, and a more informed understanding of the research in ML-driven musical systems across the globe. This in turn tends to

compromise on the possibilities whereby further resources can be drawn to help crystallise new research in the field. This was reflected by, given the context of the project initiation back in 2019, that to reimplement SampleRNN was perhaps one of the only gateways for PRiSM to make contributions towards the growing scholarship around neural audio synthesis. It also presented challenges around the day-to-day upkeep of tools developed, particularly around staff expertise and additional time required for designing / maintaining accessible means / interfaces for dissemination (e.g. Colab Notebook, multimedia / interactive tutorials, adaptive UI for artists with disabilities).

It is thus our hope that this study invites practitioners, academics, technologists and institutions to partake in unpacking these emergent dialogues and processes; to collectively nurture possibilities whereby these dialogical interactions are accessed, scrutinised and ultimately prioritised. We believe that these human-centred, interdisciplinary approaches - drawing down on all the ethical, structural, and logistical challenges around an embodied, collaborative *learning to learn* - are as significant as the singular act of software development, as we progress further into a technologically mediated musical future.

Acknowledgements

The authors would like to thank Research England for funding the PRiSM SampleRNN project through an Expanding Excellence in England (E3) grant. They would also like to thank all participants who took part in this study, as well as the collaborating institutions such as the NOVARS Research Centre at the University of Manchester. Special thanks goes to Dr Christopher Melen (PRiSM RSE 2019-2023) who developed the PRiSM SampleRNN software tool, and whose dedication on maintaining the code made all the projects discussed in this study possible.

Ethics Statement

This research project was conducted with full compliance of research ethics norms, and more specifically the codes and practices established by the British Psychological Society and the Royal Northern College of Music concerning Human Research Ethics and Internet-Mediated Research. The research involved human participants, incorporating hybrid qualitative research methods drawing on elements of collaborative auto-ethnography, group interview and semi-structured interviews, all of which were scrutinised and approved of by the Royal Northern College of Music’s Research Ethics Committee prior to project commencement. The authors took core responsibility to explain, in appropriate detail, what the research was about and by what means, at every stage of the study, to all participants. Every research participant was given a two-page ‘participant information sheet’ that outlined the purpose of the study, who were undertaking the study, and how their data would be collected, managed, and used for any academic dissemination informed by their participation of this study. The sheet included contact information should participants require additional information or wish to retract information or withdraw participation at any point during the study. It also explained how anonymity and confidentiality would be ensured throughout the process, and that the

information provided by the participants would only be attributed to them by name upon their explicit consent. Although all participants consented to the data they provided being attributed to them by name, the authors were critically reflexive about the ethical complexities around an implied dichotomy between artists and RSE throughout the process of writing up this summary report. As a result of this, the authors paid extra attention to all the data - and the transcription of the data - discussed in the write-up and made sure that any direct quotes or comments connotative of such a dichotomy be anonymised (and in some cases neutralised) in order to best prevent any such comments from being possibly mis-interpreted as to imply any interpersonal conflict.

Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2015). *TensorFlow: A system for large-scale machine learning*.

Barry, A., & Born, G. (2013). *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences*. Routledge.

Ben-Tal, O., Harris, M. T., & Sturm, B. L. T. (2021). How Music AI Is Useful: Engagements with Composers, Performers and Audiences. *Leonardo*, *54*(5), 510–516. https://doi.org/10.1162/leon_a_01959

Born, G. (2005). On Musical Mediation: Ontology, Technology and Creativity. *Twentieth-Century Music*, *2*(1), 7–36. https://doi.org/10.1017/S147857220500023X

Born, G. (2022). Music and Digital Media: A planetary anthropology. In G. Born (Ed.), *UCL Press: London, UK. (2022)*. UCL Press. https://doi.org/10.14324/111.9781800082434

Bown, O. (2021). *Beyond the Creative Species: Making Machines That Make Art and Music*. The MIT Press. https://doi.org/10.7551/mitpress/10913.001.0001

Chang, H., Ngunjiri, F., & Hernandez, K.-A. C. (2016). *Collaborative Autoethnography* (1st ed.). Routledge. https://www.perlego.com/book/1567603/collaborative-autoethnography-pdf

Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*, *209*, 118190. https://doi.org/10.1016/j.eswa.2022.118190

Clancy, M. (Ed.). (2022). *Artificial Intelligence and Music Ecosystem*. Focal Press. https://doi.org/10.4324/9780429356797

Dozier, R. (2019, April 19). This YouTube Channel Streams AI-Generated Death Metal 24/7. *Vice*. https://www.vice.com/en/article/xwnzm7/this-youtube-channel-streams-ai-generated-black-metal-247

Dyer, M. (2022). Neural Synthesis as a Methodology for Art-Anthropology in Contemporary Music. *Organised Sound*, *27*(2), 219–226. https://doi.org/10.1017/S1355771822000371

Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., & Simonyan, K. (2017). Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. *Proceedings of the 34th International Conference on Machine Learning*, 1068–1077. https://proceedings.mlr.press/v70/engel17a.html

Gioti, A.-M., Einbond, A., & Born, G. (2023). Composing the Assemblage: Probing Aesthetic and Technical Dimensions of Artistic Creation with Machine Learning. *Computer Music Journal*, 1–43. https://doi.org/10.1162/comj_a_00658

Hayden, S., & Windsor, L. (2007). COLLABORATION AND THE COMPOSER: CASE STUDIES FROM THE END OF THE 2OTH CENTURY. *Tempo*, *61*(240), 28–39. https://doi.org/10.1017/S0040298207000113

Higham, C. F., & Higham, D. J. (2018). *Deep Learning: An Introduction for Applied Mathematicians* (arXiv:1801.05894). arXiv. https://doi.org/10.48550/arXiv.1801.05894

Ingold, T. (2013). *Making: Anthropology, Archaeology, Art and Architecture*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9780203559055/making-tim-ingold

John-Steiner, V. (2000). *Creative Collaboration*. Oxford University Press.

Kanga, Z. (2014). *Inside the Collaborative Process: Realising New Works for Piano* [Doctoral, Royal Academy of Music]. https://www.zubinkanga.com/s/Zubin-Kanga-PhD-thesis-compressed.pdf

Kanga, Z., Dyer, M., Rowley, C., Packham, J., Benjamin, M., Climent, R., Gioti, A.-M., Gorton, D., Hayden, S., Howard, E., Hunt, E., Laidlow, R., McLaughlin, S., Nickel, L., & Redhead, L. (2024). *Technology and Contemporary Classical Music: Methodologies in Practice-Based Research* [Working Paper]. National Centre for Research Methods. https://eprints.ncrm.ac.uk/id/eprint/4945/

Karalis Noel, T., Minematsu, A., & Bosca, N. (2023). Collective Autoethnography as a Transformative Narrative Methodology. *International Journal of Qualitative Methods*, *22*, 16094069231203944. https://doi.org/10.1177/16094069231203944

Kehagia, N., & Moriaty, M. (2023). Recurring patterns: An ethnographic study on the adoption of AI music tools by practitioners of electroacoustic, contemporary and popular musics. *Journal of Pervasive Media*, *8*(AI, Augmentation and Art), 51–64. https://doi.org/10.1386/jpm_00004_1

L. Sturm, B., & Ben-Tal, O. (2017). Taking the Models back to Music Practice: Evaluating Generative Transcription Models built using Deep Learning. *Journal of Creative Music Systems*, *2*(1). https://doi.org/10.5920/JCMS.2017.09

Laidlow, R. (2024). *Artificial Intelligence and the Symphony Orchestra*. 355–378. https://doi.org/10.11647/obp.0353.18

McMullin, C. (2023). Transcription and Qualitative Methods: Implications for Third Sector Research. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *34*(1), 140–153. https://doi.org/10.1007/s11266-021-00400-3

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., Courville, A., & Bengio, Y. (2017). *SampleRNN: An Unconditional End-to-End Neural Audio Generation Model* (arXiv:1612.07837). arXiv. https://doi.org/10.48550/arXiv.1612.07837

Melen, C. (2020a). *PRiSM SampleRNN* [Python]. https://www.rncm.ac.uk/research/research-centres-rncm/prism/prism-collaborations/prism-samplernn/

Melen, C. (2020b, May 22). A Short History of Neural Synthesis—Royal Northern College of Music. *PRiSM Blog*. https://www.rncm.ac.uk/research/research-centres-rncm/prism/prism-blog/a-short-history-of-neural-synthesis/

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library* (arXiv:1912.01703). arXiv. https://doi.org/10.48550/arXiv.1912.01703

Roe, P. (2007). *A phenomenology of collaboration in contemporary composition and performance* [Phd, University of York]. https://doi.org/10/02-Track_02.mp3

Salem, S. (2020, June 15). A Psychogeography of Latent Space—Royal Northern College of Music. *PRiSM Blog*. https://www.rncm.ac.uk/research/research-centres-rncm/prism/prism-blog/a-psychogeography-of-latent-space/

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 2053951717738104. https://doi.org/10.1177/2053951717738104

Sturm, B. L., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., Hadjeres, G., Deruty, E., & Pachet, F. (2019). Machine learning research that matters for music creation: A case study. *Journal of New Music Research*, *48*(1), 36–55. https://doi.org/10.1080/09298215.2018.1515233

Taylor, A. (2016). ‘Collaboration’ in Contemporary Music: A Theoretical View. *Contemporary Music Review*, *35*(6), 562–578. https://doi.org/10.1080/07494467.2016.1288316

Thủy, N. T., Östersjö, S., Laws, C., Brooks, W., Gorton, D., Thuy, N. T., & Wells, J. J. (2019). Arrival Cities: In *Voices, Bodies, Practices* (pp. 235–294). Leuven University Press; JSTOR. https://doi.org/10.2307/j.ctvmd83kv.9

Torrence, J. (2018). Rethinking the Performer: Towards a Devising Performance Practice. *VIS - Nordic Journal for Artistic Research*, *0*. https://www.researchcatalogue.net/view/391025/391476/25/26

UKRI. (2023, September 12). *Expanding excellence in England (E3) funding decisions*. UK Research and Innovation (UKRI). https://www.ukri.org/what-we-do/what-we-have-funded/research-england/expanding-excellence-in-england-e3/

Wagstaff, K. (2012). *Machine Learning that Matters* (arXiv:1206.4656). arXiv. https://doi.org/10.48550/arXiv.1206.4656

Footnotes

1. *https://www.tensorflow.org/about/bib* ↩
2. *https://pytorch.org/* ↩
3. *https://www.rncm.ac.uk/prism* ↩
4. The SampleRNN architecture was publicly introduced at the 2017 International Conference on Learning Representations (ICLR). It attracted great public attention in 2018 through the release of *Relentless Doppelganger*, a continuous livestream on YouTube of SampleRNN-generated death metal audio tracks created by Dadabots (developed by CJ Carr and Zack Zukowski) (Dozier, 2019). This led to many similar SampleRNN-based generative experiments to emerge. Although many of these codes were made publicly accessible, they were mostly programmed upon the Python 2 codebase, which has been officially unsupported from 2020 onwards. This, together with the lack of active upkeep of the original SampleRNN code as well as its dependencies on several obsolete software packages, effectively render the re-implementation of SampleRNN a great challenge since late 2019 (Melen, 2020b). ↩
5. RNN refers to a type of neural networks developed to process sequential, time-based data. In other words, they are to process arbitrary sequences of inputs (data) through retaining “a kind of internal ‘memory’ of their previous states” (Melen, 2020b). ↩
6. The interviews loosely followed these questions: ↩
7. This transcription method was chosen based on interviewers’ familiarity with the projects discussed and their first-hand engagement with the study. ↩

8. These include a collection of algorithm-driven software synths that TA had programmed in Max previously. ↩

9. Collaboration in these contexts often centre around working relationships with other creative artists, practitioners, or individuals whose contribution often directly shapes the resultant artistic outputs, as identified in existing models of collective creative collaboration theorised by performers Nguyễn Thanh Thủy, Stefan Östersjö, Paul Roe, and Zubin Kanga (Kanga, 2014; Roe, 2007; Thủy et al., 2019) in the last two decades. ↩

10. 16kHz was the default sample rate for the algorithm to perform optimal training efficiency at the time. ↩

11. These include, by using various Python dependencies (e.g. Keras Tuner, Ray Tune), to automate / determine variables that control the training process such as the batch size, frame sizes, GPU/CPU allocation, number of epochs, sequence length, learning rate/momentum, number of RNN layers, among others. ↩

12. This was anticipated to a certain extent, considering the diverse prior technical knowledge and access to computational resources at participants' home studios. ↩