PATIENCE X: Extended artistic expression in AI-assisted music composition

Lois Macdonald
School of Digital Arts (SODA)
Manchester Metropolitan University
Lois.macdonald@mmu.ac.uk

Abstract

PATIENCE X is a collection of seven audio compositions exploring the integration of machine learning tools into compositional and performative workflows of alt-pop. This practice-led research examines real world AI-assisted music composition. Through qualitative experimentation with selected AI tools and original datasets of female vocals collected by the researcher, PRiSM SampleRNN emerged as most artistically effective for generating unique samples for selection and manipulation. The findings emphasise the need for a balance between artistic control and algorithmic unpredictability, ensuring AI functions as a creative catalyst rather than a tool merely for extensive replication. Additionally, accessibility remains critical. For AI tools to be practically useful to musicians, they must accommodate non-coding artists without compromising usability and creative flow. Composed between 2022 and 2023, PATIENCE X serves as a document of accessible systems during this period as well as demonstrating their sonic aesthetics. Through the process of discovery the project led to the formation of a new artist persona, 'PATIENCE', a genuine representation of extended musical expression through AI use.

1 Introduction

This 10-minute performance presents an arrangement of compositions from PATIENCE X, a body of work created using unique vocal samples generated largely via PRiSM SampleRNN. These samples are derived from original datasets featuring female vocals, and the resulting tracks embody an exploration of the authentic integration of AI systems into a compositional workflow for electronic alt-pop music. Through this practice-led research a new artist persona, 'PATIENCE', came into existence, the personification of the human/machine vocalisations generated throughout the project. As a musician with little prior knowledge of coding I aim to contribute to the discourse around the implementation of AI tools into creative music composition workflows in terms of accessibility, authorship, and artistic control. Here, I outline the creative process, contextual influences, and the methodological framework behind this work.

2 Context

The work relates to the theme of AIMC 2025 through exploration of original datasets that uncover new sonic materials and drive extended creative expression in music composition (and related artistic fields). The project also addresses the issue of accessibility of AI tools for musicians and how these tools can be integrated authentically into existing practices, both for in studio and live workflows.

Proceedings of the 6th Conference on AI Music Creativity (AIMC 2025),
Brussels, Belgium, September 10th-12th

There is a growing body of research exploring the gap between technological innovation and its practical adoption by artists. Sturm and Ben-Tal (2017) address this issue by inviting musicians to engage critically with CharRNN and FolkRNN. Similarly, Ma, Sargen, De Roure, and Howard (2024) examine the application of the PRiSM SampleRNN model by conservatoire-based artists who possess a degree of prior familiarity with the technology. However, access to and autonomous engagement with such tools remains limited for musicians operating outside of specialist contexts. This performance offers an alternative perspective by reflecting on the application of PRiSM SampleRNN from an artist situated within the alt-pop genre. As a creative project built from raw vocal audio datasets processed through neural networks, PATIENCE X is in dialogue with works such as Second Self (2019) by Dadabots, Reeps One and Bell Labs (2019), PROTO (2019) by Holly Herndon, and Future Chorus (2023), curated by Eleni Ikoniadou.

3 Methodology

Taking an iterative approach, I initially experimented with 17 AI tools, each with various methods of audio generation. These were: AIVA, Beatbot.ai, ChatGPT, Elf Tech (Grimes, 2023), Genny, Holly+, Kits.ai, Lovo, Magenta Studio, MaxMSP (ml.star), PRiSM SampleRNN, RAVE, Soundful, These Lyrics Do Not Exist, TTS Maker, Wekinator, and Word2Wave. Tools were assessed based on three criteria: accessibility (technical skill and cost), creative potential, and their practical ability to support and sustain flow (Csikszentmihalyi, 1990). I found MIDI and data-driven tools introduced a mechanical or predictable quality to the outputs whilst producing extensive variations with little usable content. In contrast, working with a more selective amount of raw audio clips produced by neural networks was more productive. I was able to access new sonic materials that introduced me to a timbral space that is both emotionally resonant and uncanny. This was characterised by microtonal warbles, glitches, and vocal anomalies that evoke a primal quality. This is where I was able to connect creatively with these tools.

PRiSM SampleRNN, ElfTech, TTS Maker and ChatGPT were used for the final compositions. Most significantly, PRiSM SampleRNN was used to create a library of 249 unique samples. Five original datasets of 30 minutes were used to generate these samples.

Table 1: Datasets used

Dataset	Description
1	Electric guitar (written and performed by artist)
2	Synth improvisation (MicroKorg, Korg MS20 and Roland TB3)
3	Spoken storytelling (Lead artist telling personal anecdotes in conversational tone)
4	"Vocal dérive" (improvised vocalisations by lead artist)
5	Mixed vocal set (10 minutes each from lead artist and two additional female vocalists performing original melodies)

Datasets 1 and 2 yielded limited musical interest as the outputs were similar to the original audio. The datasets using female vocals (3, 4, 5) provided varied and anomalous results that initiated a creative impulse, with outputs from dataset 5 most musically engaging. Created from three different female voices in a similar range, the audio files used were coherent stylistically as to

generate fluid sounds, whilst small details such as accent, phrasing and articulation allowed the model to form a more diverse output than with a single vocal. In total, over 100 minutes of audio files were generated from these data sets. Musically interesting fragments of these sonic outputs were then edited into short samples and categorised by theme to create a unique sample library to compose with. Eg. 'Kick', 'Hits', 'Rhythm', 'Singing'. The limitation of working with these selected samples was essential to the creative development of the project.

Audio 1
Example of audio generated from dataset 3 (Conversational)

Audio 2
Example of audio generated from dataset 4 (Vocal Dérive single vocal)

Audio 3
Example of audio generated from dataset 5 (Mixed vocal)

Audio 4
Example of percussive audio generated from dataset 5 (Mixed vocal)

In addition to generating samples using PRiSM SampleRNN, ElfTech was used to create vocables from nonsensical or percussive sounds. ChatGPT was used to write lyrics 'collaboratively', and TTS maker was used to voice these lyrics where needed.

Audio 5
Example of original sample from PRiSM SampleRNN

Audio 5.1
Example of original sample from PRiSM SampleRNN with distortion

Audio 5.2
Example of original sample from PRiSM SampleRNN with distortion processed with ElfTech

Tracks were composed using Ableton Live or Logic Pro X. The resulting works are performed using Ableton Live, hardware drum machine Roland TR8 (uploading unique samples) and the inclusion of female vocalists to further characterise the sounds and persona of 'PATIENCE'.

4 Conclusion

In adopting an open-ended, exploratory approach, I have engaged in practitioner-led processes that allow a nuanced brokering of our relationship with emerging technologies. Sitting adjacent to formalised academic research, this practice-based method plays a critical role in shaping the design of future research agendas.

Through exploring the outputs of the stated systems, I was drawn toward the generation and manipulation of human-like sounds. This manifested either through the expansion of my own vocal characteristics by blending them with those of other performers, or through the personification of percussive and melodic elements taken from generated clips. As a result, the project's title evolved into a new artistic persona, 'PATIENCE', embodying the uncanny, hybrid quality of the voice-based samples. This alterego has led to several new projects that extend past the initial intentions of the research.

References

Grimes (2023) 'Elf Tech' 'https://elf.tech/connect' [accessed on 5 May 2023]

ChatGPT (2022) Available at: <https://chatgpt.com/> [accessed on 14 December 2022]

Csikszentmihalyi, M., 1988. The flow experience and its significance for human psychology. *Optimal experience: Psychological studies of flow in consciousness*, *2*, pp.15-35.

Herndon, H., (2019) *Proto,* Available at: Spotify [Accessed: 10 May 2022]

Ikoniadou, Eleni et al., (2023) *Future Chorus*, Digital album compilation, Available at: Spotify [Accessed: 10 October 2023]

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). doi: https://doi.org/10.5920/JCMS.2017.09

Ma, B., Sargen, E., De Roure, D. and Howard, E., 2024. Learning to Learn: A Reflexive Case Study of PRiSM SampleRNN. *AIMC 2024 (09/09-11/09)*

Melen, C. (2020) *PRiSM SampleRNN*. Available at: https://www.rncm.ac.uk/research/research-centresrncm/prism/prism-collaborations/prism-samplernn/ [Accessed: 4 November 2021]

Dadabots, Reeps One and Bell Labs (2019) *Second Self*. Available at: https://www.youtube.com/watch?v=q981cTdL0Y&t=67s [Accessed: 11 March 2022]

TTS Maker (2023) Available at: <https://ttsmaker.com/> [Accessed: 21 April 2023]