Embodied Exploration of Latent Spaces and Explainable AI

ELIZABETH WILSON, Creative Computing Institute, University of the Arts London, UK

DEVA SCUBERT, Independent Artist, DE

MIKA SATOMI, KOBAKANT, DE

ALEX MCLEAN, Then Try This, Cornwall, UK

JUAN FELIPE AMAYA GONZALEZ, Independent Artist, DE

In this paper, we explore how performers' embodied interactions with a Neural Audio Synthesis model allow the exploration of the latent space of such a model, mediated through movements sensed by e-textiles. We provide background and context for the performance, highlighting the potential of embodied practices to contribute to developing explainable AI systems. By integrating various artistic domains with explainable AI principles, our interdisciplinary exploration contributes to the discourse on art, embodiment, and AI, offering insights into intuitive approaches found through bodily expression.

1 INTRODUCTION

In this position paper, explainable AI is examined through embodied latent space exploration, referencing a performance at Transmediale Studios, Berlin in March 2023. Details of creating an AI-informed artistic performance are provided, challenging conventional notions of intelligence and exploring AI's esoteric and social dimensions through rituals. The performance connected the AI model 'RAVE' (Real-Time Audio Variational autoEncoder [5]) to three separate artistic practices: e-textiles, live coding, and performance art. Through their interplay, the exploration of the latent space of the model trained on the performers' vocal input became apparent, where the model became material, and bodies in space were the main vectors with which to explore potential possibilities offered by the latent space. In terms of AI concepts, this piece aimed to challenge mainstream ideas of AI, such as the emulation of human neural pathways as the method of creating new intelligences. A collective view of intelligence was given as the impetus for the performance: by treating intelligence as collaborative, the aim was to undermine the dualist thinking of body/mind as well as science/fiction, and open ourselves to a more human-centric approach to AI, in terms of pattern generation, pattern recognition and connected-ness, but also resonance and textility.

Authors' Contact Information: Elizabeth Wilson, e.j.wilson@arts.ac.uk, Creative Computing Institute, University of the Arts London , London, UK; Deva Scubert, deva.schubert@gmx.de, Independent Artist, Berlin, DE; Mika Satomi, mika@kobakant.at, KOBAKANT, Berlin, DE; Alex McLean, alex@slab.org, Then Try This, Cornwall, UK; Juan Felipe Amaya Gonzalez, info@jfag.co, Independent Artist, Berlin, DE.

2 XAI FOR THE ARTS

The field of eXplainable AI (XAI) investigates methods to enhance the comprehensibility of machine learning models for individuals, thereby expanding their practicality and enabling non-specialists to employ them across diverse scenarios. Specifically, XAI experts delve into strategies for understanding complex and opaque AI models, like neural networks and deep learning methods, which are often challenging for laypeople to grasp.

Systems that are easy to use and understandable are desirable to artists for many obvious reasons, such as the need to avoid overtly steep learning curves or opaque technical processes. Bryan-Kinns et al. [2] were among the first to directly tackle the issue of eXplainable Artificial Intelligence (XAI) in the arts, with some particular works focused on music (e.g. [8], [3]) are applicable to this work. Overall, within the community the ideas that AI models may be exploited as both an artwork and a form of explanation is not uncommon. The latent space often provides artistic impetus, sometimes providing the not-yet-conceptualised to artists. Explainability becomes embedded where the navigating latent space can be used to explain its limits [3].

3 THE CONTEXT

E-Textiles are fabrics that integrate electronic components like batteries, lights, sensors, and micro-controllers. For this performance, e-textiles were used to create soft and flexible sensors that were worn on the body to sense movements. Pressure sensors were explored [9] made of Eeontex stretch resistive fabric to sense bend and stretch of garments caused by body movements. The Bela mini board was used which allowed us to read up to 8 analogue sensors per performer and 1) directly trigger sound from a single sensor data, 2) process multiple sensor data with machine learning to be used as a reduced number of synthesis parameters. It also used ml.lib externals on Pure Data open source software [4], classifying specific gestures and enabling the system to recognise repetitions in the performers' movements.

Live coding, in this context, refers to artistic programming onstage as performance with dedicated systems, often in part or wholly improvised [1]. TidalCycles was the main software used for live coding sounds [10], however a few custom elements were employed to alter the territories that live coders were used to, pushing them into new modes of interacting with sound through software. The two live coders were tasked with working independently when manipulating sensorial information from the materials but also collaboratively in response to each other.

One artistic use of AI in the performance was to create an artificial voice that could be explored with both live coding and the sensor information gathered from the textiles. To create the sounds from an artificial voice, the Neural Audio Synthesis model RAVE [5] was employed. In short, this uses a Variational Auto Encoder (VAE) structure but is applied as a real-time model (able to generate 48kHz audio signals, while simultaneously running 20 times faster than real-time on a standard laptop CPU). The RAVE-model is trained on a corpus of audio generated by the performers voices, and the VAE structure decodes and then encodes the signals. The encoding and decoding can be controlled internally from the MaxMSP software, or via external messaging protocols. Data from the sensors on the face of the four performers was sent using the zero-mq (zmq) protocol [7]. An overview detailing the specific interactions between the performers and the AI model can be seen in figure 1.

4 THE PERFORMANCE

The performance took place in March 2023 at Transmediale Studios, in Berlin, DE. It was designed as a durational performance, lasting four hours, with visitors invited to come and go throughout. It was an inter-disciplinary collaboration
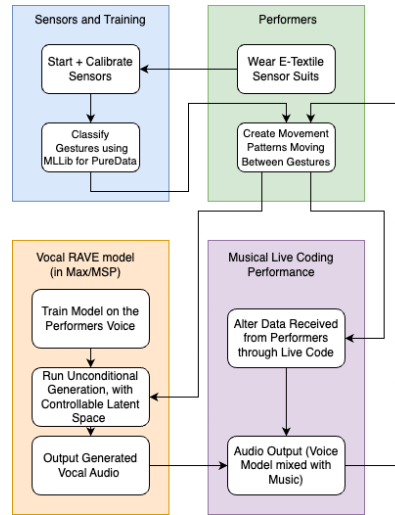
Fig. 1

105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156

between two live coders (who use computer code as the means for artistic expression [11]), five performers, and one e-textile artist.

Fig. 2. Performer Manoela Rangel exploring latent decoding with sensors on the body. Image Credit: Frank Sperling, https://franksperling.net/Info.

A video summary of the performance can be found at https://vimeo.com/882219756.

5 REFLECTIONS

An interesting case for explainable AI emerged through developing this performance. The aim was to integrate AI into artistic practice, but experimentation revealed that performers' bodies were crucial for understanding. The performers engaged in a tactile dialogue with the algorithms, with their movements influencing the traversal of the latent space. Repetitive movements helped explore the voice model's outcomes, and performers often wanted to relate the sounds of the voice to their movements, especially when phrases or words became more intelligible.

VAE models are not inherently deterministic. In a traditional autoencoder, the encoder network maps the input data to a fixed-length latent vector, resulting in a deterministic encoding process. However, in a VAE model, the encoder network maps the input data to a probability distribution in the latent space, typically Gaussian. This means that instead of producing a single fixed-length vector, the encoder outputs parameters (mean and variance) of a distribution from which latent vectors are sampled. During training, the VAE models learns to encode input data into the latent space such that similar data points are mapped close together in the latent space. Sampling from the learned latent space during inference introduces stochasticity, as each sample drawn from the latent space will generate different output data when decoded by the decoder network [6]. This property of the VAE model allowed connections between bodily expression and new ways of learning about the intricacies of the system itself. Where the focus of the piece was on integrating AI into artistic practice, the performers' interactions with their bodies unexpectedly emerged as a pivotal aspect of the exploration. Through this embodied interaction, the performers not only contributed to the artistic expression of the piece but also helped provide a use case for the body as a means with which higher-dimensional spaces can be explored. This unexpected synergy between performers' bodies and AI technology highlights the potential for embodied practices to inform and enrich the development of explainable AI systems.

REFERENCES

[1] Alan F Blackwell, Emma Cocker, Geoff Cox, Alex McLean, and Thor Magnusson. 2022. *Live coding: a user's manual.* MIT Press.

[2] Nick Bryan-Kinns, Corey Ford, Alan Chamberlain, Steven David Benford, Helen Kennedy, Zijin Li, Wu Qiong, Gus G Xia, and Jeba Rezwana. 2023. Proceedings of The first international workshop on eXplainable AI for the Arts (XAIxArts). *arXiv e-prints* (2023), arXiv–2310.

[3] Nick Bryan-Kinns, Bingyuan Zhang, Songyan Zhao, and Berker Banar. 2024. Exploring Variational Auto-encoder Architectures, Configurations, and Datasets for Generative Music Explainable AI. *Machine Intelligence Research* 21, 1 (2024), 29–45.

[4] Jamie Bullock and Ali Momeni. 2015. Ml. lib: robust, cross-platform, open-source machine learning for max and pure data.. In *NIME*. 265–270.

[5] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011* (2021).

[6] Carl Doersch. 2016. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908* (2016).

[7] Pieter Hintjens. 2013. *ZeroMQ: messaging for many applications.* " O'Reilly Media, Inc.".

[8] Maria Teresa et al. Llano. 2020. Explainable Computational Creativity. *International Conference on Computational Creativity* (2020).

[9] Yifei Luo, Mohammad Reza Abidian, Jong-Hyun Ahn, Deji Akinwande, Anne M Andrews, Markus Antonietti, Zhenan Bao, Magnus Berggren, Christopher A Berkey, Christopher John Bettinger, et al. 2023. Technology roadmap for flexible sensors. *ACS nano* 17, 6 (2023), 5211–5295.

[10] Alex McLean. 2014. Making programming languages to dance to: live coding with tidal. In *Proceedings of the 2nd ACM SIGPLAN international workshop on Functional art, music, modeling & design.* 63–70.

[11] Click Nilson. 2007. Live coding practice. In *Proceedings of the 7th international conference on New interfaces for musical expression.* 112–117.