Echo Chamber: Generative Music AI within a Participatory Museum Sound Installation

MONICA LIM, University of Melbourne, Australia

JARROD KNIBBE, University of Queensland, Australia

BINGQING CHEN, University of Melbourne, Australia

YING SIMA, University of Melbourne, Australia

Large transformer deep learning models have, for the first time, enabled fast and high-fidelity text-to-music generation directly in the audio domain, sparking frenzied debate about the impact of generative Artificial Intelligence (AI) on musicians' rights and job security, as well as its potential for computational co-creativity. This paper describes a participatory sound installation at the Grainger Museum in November 2023 that allowed visitors to experience multiple AI-generated versions of their own piano-playing. The installation prompted critical reflection on the ethics of data scraping, AI's role in amplifying bias against underrepresented musical cultures and the creation of socio-cultural echo chambers through an embodied experience of music-making, listening and moving in physical space. Its engagement with the museum collection and history provided an additional subtext to the visitor experience, allowing a recontextualizing of historical archives through the lens of technology.

CSS Concepts: • Human-centered computing • Human computer interaction (HCI) • Interaction devices • Sound-based input / output

Additional Keywords and Phrases: Generative AI, Participatory Sound Art, Interactive Art, AI Ethics, Creativity in AI, Living Archives and Novel Interfaces

1 INTRODUCTION

In the past year, we have seen a proliferation of generative AI systems being deployed for artistic purposes. In the field of music, recent developments in large transformer models have enabled the generation of sound directly in the audio domain by text prompts and musical input, creating new possibilities for generative AI to be used in real-time participatory art. Beyond creative possibilities, generative AI has led to many ethical questions surrounding copyright, bias and human creativity [1][2]. Creative human-computer interactions informed by contemporary arts practice can foster space for the general public to self-discover and explore some of these issues through the use of embodied experience and ambiguity – creating multi-layered experiences that are evocative rather than didactic [3].

Echo Chamber[1] was a recent work that invited participants to question the ethics and aesthetics of music generative AI through experiencing generative AI. It was part of a broader exhibition and residency by the first author entitled *Simulacra* at the Grainger Museum in November 2023 that speculated about the role of AI in the Creative Arts and its intersection with the museum collection. Jean Baudrillard's philosophical conception of *simulacra* on the relationship between reality, symbols and society was adopted as the theme of the exhibition. Although generally applied to media culture, Baudrillard's claim that current society has replaced all reality with symbols and signs [4] seemed particularly relevant to a world in which AI is used to generate fake news, fake identities and fake art. In conceptualizing the work, we were inspired by Percy Grainger's pioneering work with Free Music experiments in the 1930s-1940s, creating musical instruments unbound by conventional rhythm and note pitches [5]. We looked to the contemporary frontier of technological development in

[1] Video documentation available at https://vimeo.com/889822278/.

sound (arguably generative music AI) and asked whether it could assist in the creation of 'free music', or music unbound by conventions of our time, or whether it would conversely lead to a feedback cycle of monoculture and dominant musical paradigms.

Echo Chamber was designed as a participatory installation where visitors played a 10-second melody on an acoustic piano. A generative music AI developed by Meta called MusicGen [6] was used to generate multiple solo piano or orchestral versions of the melody, which were layered, looped and played back through a 16-channel speaker configuration (Figure 1). Percy Grainger's idiosyncratic musical instructions from his musical scores became part of the text prompt for the AI, providing a connection to the museum archives and raising questions about prompt engineering and semantic meaning for AI music generation. Participants could choose to play their own melody or choose from unfinished musical ideas found in the Grainger archives.

Figure 1: Physical installation including participatory piano and screen user interface, and 16-channel speaker configuration that occupied two different spaces of the museum. Photos by Peter Casamento.

2 RELATED WORK

Text-to-music generation is a recent phenomenon with the release of Google's MusicLM [7] in March 2023 and Meta's MusicGen in June 2023. Generative AI that generates audio samples directly is itself very recent, beginning with OpenAI's Jukebox model released in April 2020 [8]. Therefore, there is very little literature on their use for creative outcomes, let alone in a real-time participatory setting.

Generative AI has, however, been used in the symbolic domain (e.g. music information representation systems such as MIDI or MusicXML) to assist in compositional co-creativity [9] [10][11], as non-human performance agents [12] and as speculative tools for completing unfinished musical works [13]. More broadly, deep learning AI tools have been developed for music production purposes such as audio separation [14], neural voices [15], mastering [16] and sound synthesis [17].

The use of generative AI for artistic practice is more developed in the visual arts field, with AI-generated images winning prestigious awards [18][19] and spawning new media art superstars such as Refik Anadol [20]. In a participatory art context, Memo Atken's *Learning To See* [21] allows visitors to manipulate visual images captured by a camera, which affects the output of a generative AI. Increasingly, artistic works incorporate AI technologies not only for their creative capabilities, but also as a means to interrogate the sociocultural impact of these technologies and the ethical and existential questions they give rise to [22][23].

3 SYSTEM DESIGN

3.1 Participatory User Interface

We chose the acoustic piano as the main interface and source of sound as it is an instrument that many people can play. Regardless of proficiency, its percussive nature meant that most people could produce sound with reasonable control over dynamics and rhythm through depressing the keys. The piano also had a historical and thematic link to Percy Grainger, a world-famous pianist in his time with a history of experimentation with the piano as means to generate microtonal sounds. Lastly, the piano's role as a symbol and tool of colonization [24] gave another nuance to the work and its questioning of generative AI as a new form of colonization of culture. We used the right-most note of the piano as a trigger for the participant to start the recording process to capture their piano-playing. A contact microphone was attached to the soundboard of the piano to capture the participant input.

A visual interface was designed to sit on the piano music stand, providing instructions to guide the participant through the recording and generation process. The visual interface was implemented in Processing and included text as well as animation derived from texture maps of 3-D scans from the museum collection (Figure 2).

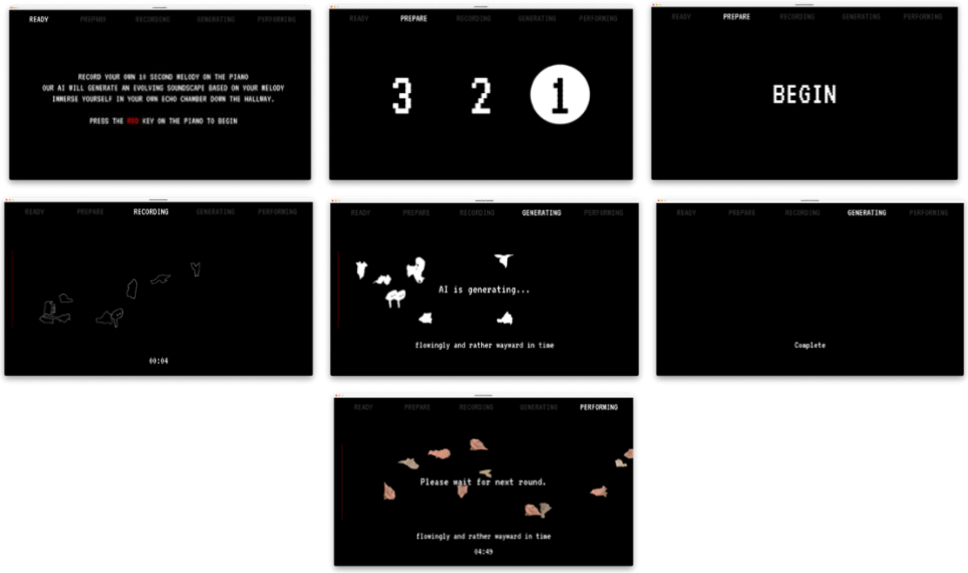

Figure 2: User interface design

3.2 Generative AI

Our backend was built in Python. This was used to capture the 10-second participant input and send the audio file to the pre-trained MusicGen model together with a text prompt which specified the instrumentation and musical instructions derived from Percy Grainger's scores (e.g. 'clingingly' and 'sharp, chippy and dead rhythmic'). We requested an output duration of 20 seconds (the default is 15 seconds) with a temperature of 1. We experimented with various temperature settings and found that while temperatures below 1 produced a much more predictable output that had more similarity to the original melody, there were fewer creative results that could generate an element of surprise. Temperatures above 1, on the other hand, tended to produce results that often had very little cognitive resemblance to the original melody.

Once the system produced a generated audio sample, the sample was then input into the model with another random text prompt for the next generation, creating a new output from the synthetic input. We limited each participant experience to 4 minutes, with 5 generations that evolved from the previous generation plus the original participant sample.

One of the challenges we faced in using MusicGen for real-time applications was the time required to generate each audio sample. Using cloud-based GPUs, the time required for each generation was too long to be practicable (approximately 40 seconds to generate a 20 second sample) and carried risks of internet-dependence and reliability. Therefore, we used a local computer with an RTX 3090 GPU card to run the model, reducing processing time for each generation to approximately 20-25 seconds. Given that it took longer to generate an audio sample than the length of the sample itself, we used strategies such as looping and layering to provide a continuous 4-minute audio experience for each participant.

The audio samples generated by MusicGen were stored in a local file, which was accessed by MaxMSP, a programming language commonly used for digital audio processing. The main Python file communicated with MaxMSP via Open Sound Control (OSC) messages and with Processing via a User Datagram Protocol (UDP) socket.

3.3 Sound Design

After experimenting with a range of audio input and text prompts with the MusicGen API, we found that the generated output tended to fall into conventional note pitches, melodies, harmonies and genres – perhaps due to the prevalence of such conventions in the training data [3]. Even where atonal or microtonal input was fed into the AI, the resulting output tended to be more melodically and harmonically conventional and within the Western equal-tempered scale. We used this 'limitation' of MusicGen to create an evolving soundscape where the participant prompt was used to generate a melody, which was then used to generate the next melody and so on (up to 6 generations due to a 4-minute time limit for each participant). Because of the tendency of the AI to stay within the same harmonic and melodic structure from generation to generation, layering multiple generations over each other was possible while still sounding musically coherent. The layering of different yet similar audio samples also alluded to the exhibition theme, where generative AI's simulation and regurgitation of culturally-dominant human artistic output was problematized.

Despite the relatively predictable harmonic & melodic output, there was uncertainty due to variable generation times (usually between 20-25 seconds) as well as unexpected silences, particularly towards the end of the generated sample. Therefore, we used extensive looping and layering of audio samples to mask unexpected silences. We also implemented a fade out at the end of every loop.

As Python generated each AI sample, it would overwrite the existing file in a local drive with the new sample. This ensured that we never ran out of local memory storage and did not store any participants' output for ethical reasons. Two buffers were created in MaxMSP - one for the original recording of participant input, and one for the latest AI generated sample. Throughout the 4-minute audio journey, these files were looped and played back at variable amplitudes, octave ranges and spatialization to create a whirling vortex of sound.

3.4 Transcorporeal Design

In designing the physical layout of the installation, we deliberately placed the speakers in a different room to the piano, requiring the participant to walk through a hallway of approximately 10 meters length to be immersed in the sound. The physical distance created a sense of rupture, or displacement, between real and unreal, simulated and simulation.

The number of speakers used were strictly speaking not required by the sound design; we could have easily created a similar sonic effect with 4 speakers. However, the physical form of the speakers was used as a metaphor to give the effect

of a chorus of people / agents – a personalized echo chamber created by algorithmic curation, feedback networks and social media 'like' culture.

4 EVALUATION

Given the very recent release of text-to-music AIs, there was very little literature or prior works that we could refer to in designing an installation with this technology. As a community, we are still grappling with its impact, both positive and negative, on the music industry and society in general. In this regard, *Echo Chamber* is a pioneering work in the field of real-time participatory sound art using text-to-music generative AI.

We interviewed 9 participants during the development phase of the work, 8 of whom had not previously experienced AI-related interactions in a museum. All participants recorded enjoyment from interacting with our participatory system. Most participants were able to recognize their input from the generated output. A number of participants commented that the generative AI made their piano playing sound 'better'. Most did not feel that the AI generated music gave the impression of 'free music' but rather 'sounded somewhat similar to traditional music'. Many were surprised by the quality of audio output generated and their similarity to real instruments. Unexpectedly, the aesthetic quality of output was often unrelated to the piano-playing skill reflected in the input.

Almost all participants said that they would prefer to hear music 'fully made by humans' rather than AI-generated music; however, some participants were excited by the potential of AI to help them create music. A number of participants commented that the experience made them think about ethical issues related to copyright, as they wondered if the AI-generated music 'belonged' to them. Curiously, there was no questioning of what type of music the AI model was trained on and how this might have influenced the generated music. Most participants also felt that the wait time (usually between 20-25 seconds) for the AI-generated music to start playing was too long, perhaps attesting to raised expectations around the capabilities of generative AI.

Overall, *Echo Chamber* allowed participants to self-discover their own opinions on AI-generated music through this embodied experience of playing, listening and being in the space - to evaluate its potential for creativity and for democratization of music against the risks of homogenization, artistic cliché and intellectual property plundering. By placing the participant body as a central focus of the experience and utilizing the physical space, acoustic instruments and physical objects, we elevated the work from an intellectual exercise to an immersive experience that participants could share with others as they listened and conversed about their own output and that of other participants.

5 DISCUSSION

There is an increasing body of research in HCI on interaction design, evaluation methodologies and the ethical implications of generative AI. However, there is a lack of practical examples in the literature where the human-computer interactions with generative AI are driven by artistic outcomes, particularly in the participatory sound field. The focus on artistic outcomes encourages design to embrace constraints, uncertainty and ambiguity, which can lead to more interesting and varied participant experiences capable of multiple interpretations. For example, ambiguity and inconsistency in the aesthetic and creative value of the sound output in *Echo Chamber* mirrors shifting opinions around AI as co-creative systems. The layering of participant input and AI-generated output created ambiguity around memory and provenance, blurring boundaries between human and synthetic output and re-enacting posthuman anxieties around human-AI dystopias. Rather than looking at technological constraints as something to be 'solved', we harnessed them in the design strategy, for example by relying on harmonic and melodic cliché. In addition, the placement of the artistic work within a real-life

museum setting allowed generative AI to interface with archival practice, recontextualizing historical narratives with speculative futures.

ACKNOWLEDGMENTS

We would like to thank Angus Donald for completing the MIDI augmentation of the piano, and Melanie Huang, who designed the Processing user interface. Thank you also to Dr Antony Chacon from the University of Melbourne for their support, and to the Grainger Museum.

REFERENCES

[1] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 82, 1–8. https://doi.org/10.1145/3544549.3585657

[2] Oppenlaender, Jonas. "The creativity of text-based generative art." *arXiv preprint arXiv: 2206.02904* (2022).

[3] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03). Association for Computing Machinery, New York, NY, USA, 233–240. https://doi.org/10.1145/642611.642653

[4] Baudrillard, Jean. 1994. *Simulacra and simulation*. Ann Arbor: University of Michigan Press.

[5] Suzanne Robinson (2012). '"Mr. Percy Grainger: A Prophet of Modernism" – How Grainger Brought Ultra-modernism to Australia, 1934-1935', *Musicology Australia*, 34:2, 207-231, DOI: 10.1080/08145857.2012.681777.

[6] Copet, Jade, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre D'efossez. 2023. "Simple and Controllable Music Generation." *ArXiv* abs/2306.05284.

[7] Agostinelli, Andrea et al. "MusicLM: Generating Music From Text." *ArXiv* abs/2301.11325 (2023): n. pag.

[8] Dhariwal, Prafulla, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford and Ilya Sutskever. "Jukebox: A Generative Model for Music." *ArXiv* abs/2005.00341 (2020): n. pag.

[9] GoogleAI. "Magenta: Make Music and Art Using Machine Learning." Accessed 18 January 2024. https://magenta.tensorflow.org/

[10] Zhou, Yijun, Yuki Koyama, Masataka Goto and Takeo Igarashi. "Generative melody composition with human-in-the-loop Bayesian optimization." *arXiv preprint arXiv:2010.03190* (2020).

[11] Louie, Ryan, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. "Novice-AI music co-creation via AI-steering tools for deep generative models." In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1-13, 2020.

[12] NPR. 2021. Team uses AI to complete Beethoven's unfinished masterpiece. NPR (Oct 2021). https://www.npr.org/2021/10/02/1042742330/team-uses-ai-to-complete-beethovens-unfinished-masterpiece.

[13] Roberts, Adam, Jesse Engel, Curtis Hawthorne, Ian Simon, Elliot Waite, Sageev Oore, Natasha Jaques, Cinjon Resnick and Douglas Eck. "Interactive musical improvisation with Magenta." In *Proc. NIPS*. 2016.

[14] Lalal.ai. "Extract vocal, accompaniment and various instruments from any audio and video." Accessed 18 January 2024. https://www.lalal.ai

[15] Voicemod Inc. "Voicemod AI Text To Song Generator v1.0." Accessed 18 January 2024. https://www.voicemod.net/text-to-song/

[16] Landr. "Unlimited Instant AI Mastering." Accessed 18 January 2024. https://landr.com

[17] Engel, Jesse, Cinjon Resnick, Adam Roberts, Sander Dieleman, Mohammad Norouzi, Douglas Eck and Karen Simonyan. "Neural audio synthesis of musical notes with wavenet autoencoders." In *International Conference on Machine Learning*, pp. 1068-1077. PMLR, 2017.

[18] Katz, L. (2023). *See The Bizarre Image That Just Won An Inaugural AI-Art Award.* Forbes. https://www.forbes.com/sites/lesliekatz/2023/10/08/see-the-surreal-image-that-just-won-an-ai-art-contest/?sh=71c4b3f64df8

[19] Seymour, T. (Ed.). (2023, April 18). *The camera never lies? Creator of AI image rejects prestigious photo award*. The Art Newspaper - International Art News and Events. https://www.theartnewspaper.com/2023/04/18/the-camera-never-lies-creator-of-ai-image-rejects-prestigious-photo-award

[20] Anadol Refik and Kivrak, Pelin, Machines that Dream: How AI-Human Collaborations in Art Deepen Audience Engagement (2023). Management and Business Review, Vol. 3, No. 1 & 2, Winter and Spring 2023, Available at SSRN: https://ssrn.com/abstract=4597317

[21] Akten, M. (2023). *Learning to See*. Onassis.org. https://www.onassis.org/whats-on/festival-you-and-ai-through-the-algorithmic-lens/artworks/learning-to-see#:~:text=%E2%80%9CLearning%20to%20See%E2%80%9D%20is%20an

[22] Juxtapoz. "The Art of Interrogation: An Interview with Bill Posters." 19 June 2020. https://www.juxtapoz.com/news/street-art/the-art-of-interrogation-an-interview-with-bill-posters/

[23] Iregular. 2021. *Our Common Home*. https://iregular.io/work/our-common-home/

[24] Moffat, Kirstine. 2009. "The Piano as Cultural Symbol in Colonial New Zealand." *History Compass* 7 (3): 719-741. https://doi.org/https://doi.org/10.1111/j.1478-0542.2009.00612.x.