Perception of control in generative AI music user interfaces

RUPERT PARRY, Audio Intelligence Agency, Australia
CAROLINE PEGRAM, Audio Intelligence Agency, Australia
ERIC J DUBOWSKY, Audio Intelligence Agency, Australia
BETH SHULMAN, Audio Intelligence Agency, Australia

Machine learning tools are growing more and more capable of producing creative works, especially since the advent of deep learning. In the field of music, musicians are beginning to encounter tools that allow them to leverage computer creativity in songwriting & audio synthesis. This presents a new problem for human-computer interaction: how to provide a sense of control over the creative process, when a separate artificial creative entity is contributing to it or creating it on the user's behalf.

ACM Reference Format:
Rupert Parry, Caroline Pegram, Eric J Dubowsky, and Beth Shulman. 2022. Perception of control in generative AI music user interfaces. 1, 1 (March 2022), 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

"I see the studio must be like a living thing, a life itself. The machine must be live and intelligent... I put my mind into the machine and the machine perform reality... The jack panel is the brain itself, so you got to patch up the brain and make the brain a living man, that the brain can take what you sending into it and live." — Lee Scratch Perry.

Music creation has always been a partnership between humans and technology: whether that technology is a resonant wood cavity, a metal string in a magnetic field, or 16 billion tiny transistors running Ableton. Musicians directly and deliberately manipulate tools, and in turn these tools provide predictable and explainable sonic output.

But what if we could get beyond technological tools, to technological collaborative partners? True creative collaboration with a machine is more possible today than ever before, with the advent of deep learning neural networks. This collaboration takes different forms, whether it's singing through a new voice that's being generated on the fly [4], or experimenting with generative rhythms and melodies [9]. This raises a challenge for the field of Human-Computer Interaction. Up until today users of music software have been used to a user-directed paradigm. How do we adapt to a collaborative software agent, and how do we do so while still maintaining user agency over software?

Authors' addresses: Rupert Parry, rupert@audiointelligence.co, Audio Intelligence Agency, 67 Great Buckingham St, Redfern, NSW, Australia, 2016; Caroline Pegram, caroline@audiointelligence.co, Audio Intelligence Agency, 67 Great Buckingham St, Redfern, NSW, Australia, 2016; Eric J Dubowsky, eric@audiointelligence.co, Audio Intelligence Agency, 67 Great Buckingham St, Redfern, NSW, Australia, 2016; Beth Shulman, beth@audiointelligence.co, Audio Intelligence Agency, 67 Great Buckingham St, Redfern, NSW, Australia, 2016.

XXXX-XXXX/2022/3-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 AI AS "MAGIC", AND USER AGENCY

One of pioneering computer scientist Ben Shneiderman's basic Rules of Interface Design is that designers ought to create interfaces that give users a locus of control [10]. Users, according to Schneiderman, benefit when they feel like they are in charge of a software system, that it can be reliably manipulated, and that the system responds to their input appropriately. In creative work with machine learning models, this is especially vital: a recent cognitive study suggested that when humans were paired with a computer for a collaborative joint-action task, they feel as if they had no agency at all [7]. In one sense, neural networks *do* respond to user input: users need to initiate the inference process based on supplied input to get output. However, the connection between input and output is much more abstract and less direct than in traditional computing. A neural network may give unexpected results when it is asked to do something and, because the model is a "black box", it can be difficult to understand why it behaved that way, or how it can be manipulated into producing better outputs.

For example, take Google AI's popular *Magenta Studio* product, a set of plugins for Ableton that allow users to generate MIDI based rhythms and melodies based on one-shot learning from user-created MIDI clips. While the tools are technically impressive, they are limited in terms of interactions that foster real-time human-computer collaboration. Users can control the models by changing their input or altering sliders that control model hyper-parameters, but it is unclear exactly how these parameters will affect the output. In addition, the interaction is turn-based, with no immediate feedback for the user to guide or alter the behaviour of the model and conform it to their desires. While this can be useful for finding surprising new melodies, and other serendipitous discoveries, it doesn't facilitate user agency if they want to get to a particular sound or progression, even if this is loosely defined [5].

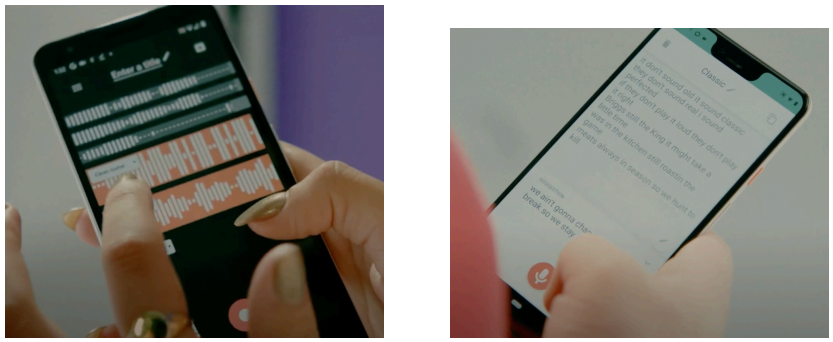
This pattern is replicated across machine learning tools in music, and is compounded by the popular conception of AI as a magical, mysterious form of computing that is beyond comprehension [1]. For machine learning to become a capable partner in the music creation process, it needs to be trusted, predictable, manipulable, provide timely feedback, and be understood by the artists operating it.

3 DEMO MEMO & AD LIBBER

We explored approaches to human-AI interfaces in a 2019 collaboration with Google's Creative Lab, production studio Uncanny Valley, and Australian musicians Briggs, Milan Ring, and Cosmo's Midnight. The project, called *Machine Learning Tools for Musicians* [3], culminated in two Android apps produced through experimentation with the musicians to see what forms of interaction were fruitful for the creative process (see Fig. 1).

The first tool, *Demo Memo*, is an AI-assisted recording app that allows users to create a demo of a complete song using only their voice. Locally recorded audio is transmitted to a server, where it is processed using a DDSP style-transfer model developed by Google's Magenta team [2], and sent back to the client. As a result, it's possible to sing a bassline or trumpet melody, and convert this into a realistic instrument sound, preserving idiosyncratic features of the raw input audio such as legato slides, vibrato and microtonal pitch deviation.

The second tool, *Ad-Libber*, is a collaborative and spontaneous lyric writing app. Users can rap or sing into their device, and see realtime suggestions for the next lyric, which adheres to rough verse cadence and rhyming words. Lyric suggestions are generated with a fork of Open AI's GPT-2 model [8] that was trained on a text corpus of song lyrics across a wide range of genres. Users can either continue singing and disregard the lyric suggestion, accept the suggestion, or use it as inspiration to sing an altered lyric.



(a) Demo Memo (b) Ad-Libber

Fig. 1. Machine Learning Tools for Musicians, mobile applications

In both of these tools, a subjective feeling user agency was a key performance indicator of the success of the designs, established by self-reported user feedback from the participating artists. This led to us designing the *Demo Memo* interface to conform to expectations and affordances of a standard multi-track recording tool, until the machine learning capabilities are deliberately requested, with users selecting a new instrument from a drop-down menu. For *Ad-Libber*, we chose to treat the tool as something that could run in the background without requiring direct input from the user. Users can continue to sing or rap, regardless of what the algorithm is processing or suggesting, without needing to explicitly accept or reject the generated line.

4 THE FUTURE

While our development process for *Machine Learning Tools for Musicians* optimised for user agency through direct feedback from our musical collaborators, we didn't set out to establish a set of principles to guide the development of agency in future human-AI creative collaboration tools.

Future work in this area should establish a set of principles that enable user control to persist even when working with complex deep learning networks, where outputs don't correlate strongly with inputs. One promising avenue is establishing how agency can be exerted on different hierarchical levels. Moore et al. have shown that when users can effectively establish higher goal-level control over a system, then lower control levels (like motor control) cease to be important [6]. How might this be adapted to allow high-level control over outputs of a machine learning composer, without feeling a lack of agency about the specific notes or rhythms played? It is only by experimenting with new forms of human-computer interaction that we will establish new standards for collaborative software design in machine learning and music.

REFERENCES

[1] Kate Crawford. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press, New Haven.
[2] Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. 2020. DDSP: Differentiable Digital Signal Processing. (2020), 19.
[3] Google Australia. 2020. Machine Learning Tools for Musicians. (Nov. 2020).
[4] Holly Herndon. 2021. Holly+. https://holly.mirror.xyz/54ds2IiOnvthjGFkokFCoaI4EabytH9xjAYy1irHy94. (July 2021).
[5] Hannah Limerick, David Coyle, and James W. Moore. 2014. The Experience of Agency in Human-Computer Interactions: A Review. *Frontiers in Human Neuroscience* 8 (2014).

[6] James W. Moore, Daniel M. Wegner, and Patrick Haggard. 2009. Modulating the Sense of Agency with External Cues. *Consciousness and Cognition* 18, 4 (Dec. 2009), 1056–1064. https://doi.org/10.1016/j.concog.2009.05.004

[7] Sukhvinder S. Obhi and Preston Hall. 2011. Sense of Agency in Joint Action: Influence of Human and Computer Co-Actors. *Experimental Brain Research* 211, 3-4 (June 2011), 663–670. https://doi.org/10.1007/s00221-011-2662-7

[8] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2018. Language Models Are Unsupervised Multitask Learners. (2018), 24.

[9] Adam Roberts, Jesse Engel, Yotam Mann, Jon Gillick, Claire Kayacik, Signe Nørly, Monica Dinculescu, Carey Radebaugh, Curtis Hawthorne, and Douglas Eck. [n. d.]. Magenta Studio: Augmenting Creativity with Deep Learning in Ableton Live. ([n. d.]), 7.

[10] Ben Shneiderman and Catherine Plaisant. 2004. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (4th ed ed.). Pearson/Addison Wesley, Boston.