Musical Agents, Agency, & AI: Towards a Phenomenological Understanding

Ted Moore
Johns Hopkins University, Peabody Institute
tmoore97@jh.edu

ABSTRACT

Creating AI musical agents is an interest to many electronic music artists and researchers. This paper suggests a phenomenological understanding of "musical agency" determined by four criteria: in order to perceive a technological system as a collaborative agent it must (1) be perceived as separate from the user, (2) be surprising to the user, (3) be intended as a collaborative agent, and (4) mirror the user's own musical intentions. Each criterion is explored through practitioners' writings, with modular synthesizers as an example of using complexity to fulfill these criteria. The aim is to pursue a broad identification of "AI" in music technology towards any systems that fill the criteria, thereby avoiding definitions that use specific categories of algorithms or implementations.

1. INTRODUCTION

Using AI to create musical agents is a current topic of interest to many electronic music artists and researchers. "Musical agency," a "musical agent," or an "AI improviser" is often defined subjectively using criteria along the lines of "you know it when you see it." This paper examines the mechanisms by which technical or musical objects are perceived as musical agents, why practitioners pursue musical agency, and what it reflects about the human user.

This paper suggests a phenomenological understanding of agency in music-technology systems that provides a framework of subjective criteria that can flesh out the "you know it when you see it" approach. I suggest that in order to perceive a technological system as a collaborative agent it must (1) be perceived as separate from the user, (2) be surprising to the user, (3) be intended as a collaborative agent, and (4) mirror the user's own musical intentions. This phenomenological approach suggests that "artificial intelligence" (at least in music technology systems) should be a broad, inclusive grouping based on perceptions rather than technological implementations.

2. CONTINGENCY & SURPRISE: WHY WE WANT TECHNOLOGICAL COLLABORATORS

Collaboration with music technologies has been part of creative practices from the discipline's beginning. As seen in the accounts below, the goals of these collaborations are often to solicit *surprising* sonic forms from technological processes in the hope of exploring new sonic ideas. Once surprised, a user can then choose to pursue, extend, or discard the results from the technology.

2.1 Analog Contingency

For electronic musicians, the desirable moments of surprise while using technology have often been rooted in the imprecision of analog systems. Harry Collins and Trevor Pinch describe the concepts of control and contingency playing out between the desires of synthesizer engineers and musicians saying, "The history of the synthesizer can be seen as a battle ground between the engineers' desire for control and repeatability and the artists' desire for contingency...The engineer wants the machine to be reliable...The artists...want an *instrument* not a *machine*–something that will play something unique, something which, although subject to control...can surprise them." (emphasis mine) [1] Collins and Pinch further describe this relationship, writing of early analog synthesizers, that, "Musicians would talk about getting an incredible sound at night in the studio only to return to the instrument the next morning to find they couldn't reproduce it. This imprecision was a source of constant delight for some musicians." [1]

2.2 Improvisational Contingency

In their ethnography of group-improvisation from a human-computer interaction (HCI) perspective, Kang, Jackson, and Sengers found that many performers described the "tension" from improvisational uncertainty "as a source of both fragility and potential failure, but also energy and creativity." [2] They go on to describe the pursuit of this contingency using "coder" terminology, saying, "musicians and artists may seek to exploit or create uncertainty as a mechanism of discovery and expression, making breakdown in effect a *'feature'* rather than a *'bug'*." (emphasis mine) [2]

Similarly, when describing her improvisational practice, New Renaissance Artist The Honourable Elizabeth A. Baker explains, "I always have one piece of gear that will probably blow up or I make a new configuration at the beginning of a show that will probably go wrong and I specifically do this so that when it goes wrong...I have to think on my feet...I find the joy of de-escalating the bomb." [3]

2.3 Time-Varying Contingency Towards Musical Expression

In discussion of his computer program *Voyager*, George Lewis equates unpredictability-as-agency on multiple time scales, saying, "If the computer is not treated as a musical instrument, but as an independent improvisor, the difference is partly grounded in the form of program responses that are not necessarily predictable on the basis of outside input... *Voyager*'s response to input has several modes, from complete communion to utter indifference...while tendencies over a long period of time exhibit consistency, moment-to-moment choices can shift unpredictably. It is a fact, however, that the system is designed to avoid the kind of uniformity where the same kind of input routinely leads to the same result." [4] Lewis has explicitly programmed into *Voyager* an added layer of agency: the ability to autonomously change the degree of its contingency throughout a performance.

2.4 Contingency in Machine Learning Systems

The desire for unpredictability, or systems "difficult to control" [1], that these artists exhibit can also be seen in contemporary artists working with artificial intelligence. Rebecca Fiebrink argues that machine learning transforms 'bugs' into creative opportunities, stating "One of the great things about using machine learning as opposed to coding, as I mentioned it's faster, but also the kinds of mistakes it makes are different. If you make a mistake while you're writing code, often you're going to get a compiler error or...silence, or...a filter blowing up. With machine learning the way that a lot of these systems are configured, if it gives you something unexpected...it's going to do something and that's often just more creatively useful than having nothing happen and that can lead you to experiment further." [5] Fiebrinks's description makes a strong argument for why a creator interested in employing contingent technologies as a collaborator would want to use machine learning: the exercise of trial and error is primed for creative surprise and potential instead of discouragement.

Laetitia Sonami–a sound artist, performer, and researcher based in Oakland, CA–says of her work with machine learning, "...in a way, you don't want the instrument to perform like a well-trained circus, you kind of want it to be a little wild, and you want to adapt to it somehow, like riding a bull...I think the machine learning allowed more of this". [5] Furthermore, in an interesting turn on the control vs. contingency paradigm of Collins and Pinch [1], Laetitia Sonami describes using machine learning to *control* the *degree* of *contingency* in her instrument, stating, "The unpredictability...depends on how 'wide' the machine learning is. If I feed the system training examples whose sounds encompass wide changes...the trained models will move through all these points in unpredictable ways as the springs settle to a resting place...I can thus easily scale the instrument between predictable and unpredictable results by changing how I train....This 'predictability index' is very easily modified and unique to ML." [6] Even with this control over contingency, Sonami still chooses to inject the imprecision of analog mechanisms into her instrument, adding, "I was looking for more complex inputs and opted for a partially chaotic system which would 'fight' the intention of ML and not learn (!)." [6]

2.5 Balancing Control and Contingency

These accounts reveal how others have identified technology as perceived, described, and designed to be agential. These writers also acknowledge the fluidity of these distinctions and the precarity of the relation between them. Collins and Pinch hedge against their contingency-vs.-control dichotomy saying, "Life is compromise and artists sometimes long for control just as scientists dream of the serendipitous discovery." [1] Lewis' *Voyager* performs this interplay and balance by varying the degree of it's "communion" and "indifference" towards its interlocutor [4]. Sonami balances the degree on contingency in her performance instrument by adjusting what she calls machine learning's "predictability index." [6]

3. PERCEIVING AGENCY

3.1 Collaborative Agency Requires Separation

In order to perceive a technology as agential, it must first be perceived as having musical actions separate from a user, designer, or performer. This perception essentially amounts to creating the moment of "surprise" described by many practitioners above.

3.1.1 Solo vs. Duet

In 1991, Robert Rowe offered a few classification systems for interactive music technologies, including the "instrument" and "player" paradigms. [7] He describes "instruments" as being "concerned with constructing an extended musical instrument: performance gestures from a human player are analyzed by the computer and guide an elaborated output exceeding normal instrument response. Imagining such a system being played by a single performer, the musical result would be thought of as a solo," which is contrasted with, "Systems following a player paradigm try to construct an artificial player, a musical presence with a personality and behavior of its own, though it may vary in the degree to which it follows the lead of a human partner. A player paradigm system played by a single human would produce an output more like a duet." [7]

3.1.2 Intention Bonding

Another way of considering how systems are be perceived as agents is by analyzing the *sonic* relationship between the human and technology in real-time performance. Sam Pluta presents a two-dimensional continuum in which to position relationships between human sound inputs and computer music outputs. A computer's response with no delay and no distortion relates closely to the human's intention, while increased temporal offset or timbral distortion indicates greater perceived agency.

Pluta's continuum recalls the agency of Lewis' *Voyager* into which was programmed agency to behave at different positions on Pluta's continuum at different moments during a performance.

3.1.3 First- and Third- Person Descriptions

In his chapter on musical intentionality, Marc Leman offers a useful distinction for phenomenologically determining if one is identifying music as agential saying, "third-person descriptions are about repeatable measurements of phenomena", while "first-person descriptions in musicology draw upon interpretations of intentions attributed to music...moving sonic forms receive the status of actions to which intentionality can be attributed". [8]

Leman's third person descriptions ("repeatable measurements of phenomena") are useful for identifying non-agential technologies. [8] A system's measurable repeatability based on similar human inputs (e.g., every time the human does x, the technology does y) will be perceived as a non-agential, "solo" relationship. The repetition of this relationship further establishes and maintains it as non-agential, preserving the perception of shared intention even as musical content transforms over time. This perception of non-agential technology is precisely what Lewis was aiming to avoid with *Voyager*, stating, "It is a fact...that the system is designed to avoid the kind of uniformity where the same kind of input routinely leads to the same result." [4]

Using a first person description requires a listener to first *perceive* a moving sonic form as an "action": only then can intentionality be attributed. The attribution of intentionality depends on the performance context. According to Rowe's paradigms, if the intentionality is perceived as separate from the human's, creating a collaborative "duet," the technological collaborator would be agential. If the intentions behind the moving sonic forms are perceived to originate from and extend the human performer's intentions, the technology would not be seen as agential, only assisting the performer. [7]

First-person descriptions suggest perceptions of collaborative, agential technologies, while third person descriptions suggest predictable extensions of a user's intentions.

3.2 Collaborative Agency Requires Surprise

Modular synthesizer performer Richard Devine describes using modular synthesizer systems as, "Things would happen unexpectedly...It was like a living organism that sort of does its own thing. These circuits would come to life...it's almost like it has its own personality." [9] By using these first-person descriptions, [8] Devine is expressing his perception of agency in his modular system.

This moment of "surprise" is also described by Deniz Peters in his study using motion tracking with dancers. [10] When the dancers would experience "motion tracking glitches", "a foreign agency would seem to gain presence." [10] Peters says, "the instrument [(the motion tracking software)] may...turn into [an] agency, particularly if its response is less predictable than that of a static object." [10]

3.3 Collaborative Agency Depends on Intention

Devine's description makes the moment of "surprise" a *goal* (i.e., desirable and/or planned) rather than a *glitch* (i.e., undesirable and/or unplanned). In order to experience a moment of surprise from a system (as a goal *or* glitch), one must have an established facility with it, such that one's interactions can create predictable outcomes. When one's interactions then produce an unpredictable outcome, one can be surprised. Peters describes this necessary facility saying, "the instrumental action becomes transparent, disappearing as a resistance to our sonic intentions. The instrument seems to become part of one's body. This transparency is a facet of the technical mastery attributed to virtuosity" [10]. Peters' description of the technology as being "part of one's body" echoes Rowe's perception of a "solo" performance [7] as it extends and elaborates the dancers' intentions, and therefore only when it "glitches" does it take on a sense of agency. The dancers' perceptions of agency reflect Leman's first-person descriptions, [8] a consequential distance from the origin on Pluta's continuum, [11] as well as Rowe's perception of a duet. [7] Perceiving a surprising result as a glitch implies that the technology's *intention* is non-agential, while perceiving a surprising result as a desired outcome implies the *intention* to be an agential collaborator.

3.4 Collaborative Agency Requires Mirroring

In addition to an agent being *surprising*, *separate*, and *intended*, the last step in the perception of a collaborative agent is described by Leman, saying, "Attribution of intentionality is likely to occur on the basis of mirroring, that is on the basis of a simulation of the perceived action in the subject's own action. Actions of others are understood as intended actions because the subject can simulate them and understand them as its own intended actions." [8] He clarifies that, "This intentionality can be attributed to subjects as well as to objects (or, rather, events)." [8] Regarding AI music systems, including modular synthesizers, in order to perceive the system as an agential collaborator it must be perceived as a separate, but *similar*, actor. The actions (i.e., sounds) it makes when it surprises the user must be recognizable as sounds that the user could make and might desire to make using such a system.

4. THE USE OF COMPLEXITY

While Peters' dancers experience surprise via unplanned "glitch", Devine is able to achieve his *goal* of surprise in spite of his virtuosity. I say "in spite of" because one may suppose that a "virtuosic" user performing virtuosic execution would be in control at all times (and therefore not-surprisable). How does Devine, a virtuosic user of these systems, achieve his "goal" of surprising himself? How can one create a transition from control to contingency?

Agency can emerge when a system becomes too complex for the user to keep track of all the interconnections and relations necessary to precisely predict the outputs of the system, *yet mirroring still occurs*. Furthermore, for many users, creating this agency is a goal of using the system, as one can then employ that agent as a collaborative partner in the creation of music. The value of this collaboration is heightened by the mirroring property of the agent: that the user and agent share a sense of musical intentionality.

When beginning to create a modular synthesizer patch, Devine says that he "just kind of start[s] from nothing and then patch[es] up and see[s] what happens" [12]. Early in this process Devine would be able to provide Leman's third person descriptions of the sound being produced, using objective, measurable descriptions such as, "the square wave oscillator is being filtered by the low pass filter, the cutoff frequency of which is being modulated by a triangle wave LFO." These descriptions point towards the system currently being non-agential, not offering surprise, only extending the user's intentions.

As one works with a patch, the complexity tends to grow to the point where it becomes difficult to keep track of all the interconnections contributing to the resulting sound.

At this stage, it becomes challenging to provide a third-person, objective description of the system. Instead, one starts using first-person descriptions that, as Leman says, "draw upon interpretations of intentions attributed to music." [8] Devine achieves his *goal* by pursuing complexity to the point where he is surprised by the sonic results, indicated by the use of first-person descriptions such as, "a living organism that sort of does it's own thing." [9] He also emphasizes the value he finds in collaborations with mirroring agents, stating, "It may give you something that you were looking for...it may give you something...that's even cooler than what you were trying to come up with." [12]

5. WHAT DO WE MEAN WHEN WE SAY AGENCY?: TOWARDS A PHENOMENOLOGICAL DEFINITION

At the 2020 inSonic conference, panelists discussed what is meant by "artificial intelligence", particularly in the context of creative music making. Rather than outlining technological categorizations, the panelists' responses about identifying AI all focus on the perception of intention or agency in collaborative technology. Lewis described that, while performing with *Voyager* "people have to feel that they can get the machine's attention, that they can dialog with it, that it, quote, 'understands' them somehow, and if that's AI, I'm prepared to go with that as one potential method of thinking about AI" [13] Palle Dahlstadt adds, "if there's a certain kind of complexity–and that threshold is really quite low–it can be perceived as an agent that actually plays with you." [13] He goes on to explain how low this threshold can be, stating that, "even such relatively simple systems that contain complex internal states and latency and feedback, they start to behave...like playing with another musician because it's so complex." [13]

For these practitioners who have been working with AI in musical creativity for decades, the identification of what is an "intelligent" system is not determined by a category of algorithm or even a high degree of complexity, but rather by a perception or attribution of agency in a technological system. The panel also reveals that a common strategy for inducing this perception is the use of complexity, that of which may involve a machine learning algorithm. This view is nicely summarized by Lewis at the end of his article about *Voyager* titled *Too Many Notes*, saying, "Rather than asking if computers can be creative and intelligent...*Voyager* asks us where our own creativity and intelligence might lie–not 'How do we create intelligence?' but 'How do we find it?"' [4] Lewis' suggestion places the onus for identifying artificial intelligence, not in the contents of the machine, but in the perception of the user.

6. CONCLUSION

In order for a technological system to be perceived a collaborative agent it must (1) be perceived as separate from the user, (2) be surprising to the user, (3) be intended as a collaborative agent, and (4) mirror the user's own musical intentions. For these criteria to arise a system must be sufficiently complex for a user to become unable to predict the system's outcomes (although, as Dahlstadt's observed, the required threshold of complexity is "really quite low"). [13] I propose that the identification of "Artificial Intelligence" in collaborative music technology systems should be a broad grouping based on phenomenological perceptions of agency. The inclusivity this definition offers connects the lineage of collaboration with contingent technologies from early synthesizers (and before) up through the current interest in AI systems. A phenomenological understanding avoids the apotheosis of ever newer technologies, instead rooting one's attention to the experience of artistic practice and the spirit of experimental creativity.

7. REFERENCES

[1] H. Collins and T. Pinch, "On Chance and Contingency," *Public*, no. 33, 2006.

[2] L. Kang, S. J. Jackson, and P. Sengers, "Intermodulation: Improvisation and Collaborative Art Practice for HCI," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–13.

[3] T. H. E. A. Baker, "Talking Free Music with Elizabeth A. Baker," YouTube, 10 2020. [Online]. Available: https://www.youtube.com/watch?v=5WjmVLd7XEM

[4] G. E. Lewis, "Too Many Notes: Computers, Complexity and Culture in Voyager," *Leonardo Music Journal*, vol. 10, no. 2000, pp. 33–39, 2000.

[5] CeReNeM, "Rebecca Fiebrink: Machine Learning as Creative Design Tool," YouTube, 08 2019, centre for Research in New Music. [Online]. Available: https://www.youtube.com/watch?v=Qo6n8MuEgdQ

[6] R. Fiebrink and L. Sonami, "Reflections on Eight Years of Instrument Creation with Machine Learning," *NIME 2020 Proceedings*, 2020.

[7] R. Rowe, *Interactive Music Systems: Machine Listening and Composing*. MIT Press, 1993.

[8] M. Leman, *Embodied music cognition and mediation technology*. MIT press, 2008.

[9] I. D. of Wires, "Richard Devine: IDOW Extended Interview #8 (Analog Voodoo Effect)," YouTube, 02 2013. [Online]. Available: https://www.youtube.com/watch?v=Z7naEUAYDfg

[10] D. Peters, "Haptic illusions and imagined agency: Felt resistances in sonic experience," *Contemporary Music Review*, vol. 32, no. 2-3, pp. 151–164, 2013.

[11] S. Pluta, "Laptop Improvisation in Multi-Dimensional Space," Ph.D. dissertation, Columbia University, 5 2012.

[12] SweetwaterSound, "Richard Devine Interview - The Sweetwater Minute, Vol. 256," YouTube, 09 2014. [Online]. Available: https://www.youtube.com/watch?v=K_5mb_utKdM

[13] Z. Karlsruhe, "inSonic 2020: Syntheses - Day 2," YouTube, 12 2020. [Online]. Available: https://www.youtube.com/watch?v=sooNxK6oQ4c