Embodied AI and Musicking Robotics

Craig Vear

Abstract This chapter discusses a hypothesis for embodied AI and its development with music robots. It proposes a solution to the challenge: *if we want AI/robots to join us inside the creative acts of music, then how do we design and develop systems that prioritise the relationships that bind musicians inside the flow of musicking?* This requires significant thinking around some core questions such as 'what does AI need to do in order to stimulate an embodied interaction in music?', 'what sort of intelligent agent does the AI need to be?' and 'what does the machine need to learn; what is to be modelled?'. I proffer a definition of embodied AI as *an intelligent agent whose operational behaviour is determined by percepts* (Percepts are objects of perception, or put another way, something that is perceived) *interacting to the dynamic situation within which it operates*, and outline foundational theories such as embodiment, embodied cognition, flow, musicking, meaning-making and embodied interaction to argue for such a concept in music. Following this, I outline the theoretical and conceptual structures that I have developed in order to get embodied AI and musicking robots to create music and meaning with human-musicians through which I offer examples of some of the works that have been developed using this approach.

Keywords Creative AI · Embodiment · Robotics · Music · Embodied cognition · Flow · Musicking

1 Pre-introduction: Music Heard So Deeply …

If you have ever listened to music so deeply that you have been taken away by it, then you will understand what I am to discuss in this chapter. It is the type of feeling where you leave behind notions of your body in a physical place, and venture into the space created by the music. Your sense of being shifts as you become more deeply attuned to the sounds and the melodies, the rhythms, the instruments and the voices that you

C. Vear (✉)
University of Nottingham, Nottingham, UK
e-mail: Craig.Vear@nottingham.ac.uk

find in this space. For me, I can feel the live presence of these sounds, and that it is happening in the here-and-now. This even happens when it is a recording that I have listened to lots over the years. In a sense, I become absorbed into the materiality of the music and attuned to the presence of those sounds in its flow through time: I have become embodied.

T. S. Eliot described this sensation poetically in his *Four Quartets.* In *The Dry Savages,* he wrote:

> Music heard so deeply
>
> That it is not heard at all, but you are the music
>
> While the music lasts.
>
> T.S.Eliot Four Quartets *The Dry Savages*

Here, Eliot describes this phenomenon as 'you are the music' and I would agree. Although this may seem an odd thing to say, in my experiences this is what happens: I do become the music. When this happens, I lose the awareness of my physical body as a shell that holds my sense of being, and I vaporise into the world created by the flow[1] of the music. Inside this world, I am free to explore and to mingle with the other presences; I feel the music and the other things within it, I can interact with them and feel their presence deeply, or I can zoom out and enjoy being in the flow as the music passes through time.

Although when I am listening, composing or dancing to music, the sense of inter-action is one way (my interaction with the sounds), when I play with live musicians this interaction becomes communal and social. It can be a dance of sorts: to touch, to feel, to work with, to play with, to hide and seek with, to flirt and subvert with others through the flow as they reach out and play with me. And it is through these embodied interactions inside the music that I find meaning.

Meaning, used here, does not refer to how I interpret the melodic line to be some symbolic container for an inner message, or how the harmonic structure makes me feel (sad, happy etc.); nor does meaning equate to how the sounds are generated: electronic or natural. On the contrary, from the embodied inside perspective meaning is constructed through the interactions I described above in the 'dance', and by becoming the music as Eliot portrays.

In my book *The Digital Score* (Vear 2019), I outline my understanding of meaning-making in music. At the heart of this is Christopher Small's concept of 'musicking', in that 'to music is to take part' (Small 1998, p. 9). Small wrote that taking part can happen 'in any capacity, in a musical performance, whether by performing, by listening, by rehearsing or practising, by providing material for performance (what we call composing)' (ibid). Small stresses that 'the act of musicking establishes in

---

[1] Is the experience of musicking from inside the activity. Within the context of this project, the flow of musicking defines how 'musicians become absorbed in the music through a sense of incorporation within their environment (the soundworld), a shared effort (with the digital, virtual, AI and robotic agents) and a loss of awareness of their day-to-day wakefulness and bodily self-consciousness (embodiment with their instrument and into their music)'.

the place where it is happening a set of relationships, and it is in those relationships that the meaning of the act lies' (Ibid., p. 13). Simon Emmerson clarified Small's principle of meaning to infer the 'what you mean to me' (Emmerson 2007, P. 29), and this subtle shift circumvents the significant issues of value and who is doing the evaluation of meaning. Therefore, meaning (or the *what-you-mean-to-me*) is to be found in the relationships formed between the embodied act of musicking and the materials therein, e.g. musicians, sounds, music space and time.

2 Embodied Interactions

For nearly four decades, I have enjoyed this experience as a professional musician who, before the call to academia, performed at a high level of musicianship in a variety of professional situations. This experience is central to my creative AI practice, and it is a foundation stone to understanding embodied intelligence in music. However, this presents a significant challenge:

> if we want AI/robots to join us inside the creative acts of music then how do we design and develop systems that prioritise the relationships that bind musicians inside the flow of musicking?

To illustrate this, imagine a scenario where two human-musicians are musicking together; they are embodied, in the flow and dancing as sound through the music. Their interactions are a mixture of creative, mischievous, surprising, familiar and playful. Of course, it is not like this 100% of the time, they will also feel ignored, isolated, stupid, rejected and inadequate; and at other times, they are simply getting on with their own journey aware of the other, enjoying any coincidences yet focussed on exploring the larger space of sound. But this is all part of the rich tapestry of the embodied experience within live music performance, an experience we can call *embodied interaction.*

Now let us replace one of those musicians with an AI-driven robot, and you will understand the goal that I have set myself: what does AI need to do in order to stimulate this embodied interaction in music? The key phrase here is 'stimulate this embodied interaction' and outlines the nature of the AI being a co-creative other inside music. This is opposed to using AI as a signal generator, or constructing the harmonic lattice of a song, or using a robot to play a musical instrument (although it may well need to do this as part of its role). Furthermore, for me, this question proposes that the perceptual focus is on the human: i.e. it is the human-musician who feels this sense of embodied interaction. But this is a personal rationale as my research aims to make humans more creative through AI, others may choose to make the AI conscious or for it to have embodied perception.

In order to deal with this challenge of AI stimulating (as opposed to simulating) an embodied interaction, the AI is tasked with operating in a specific way that feels intelligent within the situation that it is to be perceived, i.e. within embodied musicking. And, that it is interacting with the human-musician in ways that can be recognised

(by the human-musician) as the dance as described above. This is, in the context of musicking, intelligent behaviour, and an AI operating in such a way I call embodied AI and will discuss in the following section.

3   Embodied AI

With the above discussion in mind, I define embodied AI as:

> an intelligent agent whose operational behaviour is determined by percepts interacting to the dynamic situation within which it operates.

This definition is built on two principles:

(a)   that artificial intelligence is not limited to the thinking-mind model
(b)   that we understand meaning-making from the perspective of embodied cognition.

Percepts in this context are objects of perception, or put another way, something that is perceived i.e. the stuff that is found in the relationships formed between the embodied act of musicking and the materials therein, e.g. musicians, sounds, music space and time.

To unpack this first principle—*that artificial intelligence is not limited to the thinking-mind model*—we need to address a bias and prejudice still found in people's opinion that intelligence in AI is understood from a specific perspective of a thinking machine. This is in part a legacy from the early days of AI research that sought to position the study in line with formal reasoning philosophies and Turing's notion of the 'electronic brain' (Wooldridge 2020).

Thankfully, the field of AI has moved away from notions of intelligence being solely limited to cognitive operations of reasoning and logic. The mind-based models that predominated Good-Old-Fashioned AI (GOFAI) have, since the 1980s, accepted other areas of intelligence such as behavioural and embodied, and that these are understood within the context with which they operate. Landmark papers such as *Intelligence without Representation* Brooks (1987) and Anderson's *Field Guide* (Anderson 2003) to embodied cognition for AI research reinforce the limitations of building GOFAI systems that model the thinking processes like a human mind.

This shift in the definition of AI is not limited to a small niche in AI research. The widely respected textbook *Artificial Intelligence—A Modern Approach* also supports this approach. They state that the 'main unifying theme is the idea of an intelligent agent' (Russell and Norvig 2020. Preface pp. vii–viii), from which it defines AI as 'the study of agents that receive percepts from the environment and perform actions' (ibid.). They go on to say that

> Each such agent implements a function that maps percept sequences to actions, and […] we stress the importance of the task environment in determining the appropriate agent design.

For embodied AI in music determining the 'appropriate agent' that 'implements a function that maps percept's sequences to actions' involves understanding the type of cognitive and perceptual processes that occur when embodied musicking (the task environment); and these, as explained in Pre-Introduction (above), are very different to the everyday cognitive processes of the wakeful mind going about its business such as walking to work, or the rationale thinking mind as it strategizes a chess game. Russell and Norvig stress that developing AI in this modern way should be 'concerned mainly with rational action', and that ideally, 'an intelligent agent takes the best possible action in a situation'. In relation to this chapter, then the best possible action for embodied AI in music is to stimulate the dance between musicians that is discussed earlier; and the best possible intelligent agent is one that 'acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome' (Ibid., p. 4), i.e. the embodied interaction of musicking.

When unpacking the second principle—*that we understand meaning-making from the perspective of embodied cognition*—we can see that this also deals with the notion of 'rational action'[2] and guides this research into understanding what is to be modelled for 'an intelligent agent takes the best possible action in a [musicking] situation'. From this perspective, approaching embodied intelligence in music with a focus of behavioural and embodied cognition emphasises the close-coupled relationship between the *situation* that an intelligent agent is operating in (the embodied musicking space), and the *behaviour* that it exhibits to cope inside such a system (embodied musicking and interaction; i.e. the dance).[3]

As mentioned above, there is a shift in the nature of AI away from a 'thinking thing' as advocated by Descartes, towards an understanding of an intelligent agent that copes within a dynamic and interactive environment. The former proposes a split, or separation of the mind from the body, it prioritises the role of the thinking mind as intelligence, and foregrounds notions of thought and reason, logic and planning as primary facets of intelligence. The latter, however, understands the biological system as a situated agent that has evolved behaviour to cope with its dynamic environment. For innovators in embodied intelligence such as Rodney Brooks, these Cartesian aspects of intelligence 'cannot account for large aspects of what goes into intelligence' (Brooks 1999, p. 134).

While it is true to say that aspects of the thinking mind such as short-term planning, logic and reason are required to make music, this does not come close to accounting for the decisions that are made while in the flow of musicking. As such, any hypothesis for embodied intelligence in music needs to look beyond cognitivist priorities of 'representation, formalism and rule-based transformation' (Anderson 2003, p. 93) and consider the interactive, coping mechanisms at the heart of the

---

[2] There are other approaches such as thinking humanly, thinking rationally and acting humanly and are discussed in more detail in Russell and Norvig (2020).

[3] You can see this in action by watching any Roomba robot hoover navigate any room and any encounter with a rogue sock, cable, slipper, pet or moved chair.

embodied behaviour of musicians, and to understand the 'more evolutionary primitive mechanisms which control perception and action' (Ibid., p. 100) in this music world.

It should be obvious that embodiment is a central concern of embodied cognition and therefore embodied AI. While there are various definitions of embodiment, you will be familiar with the concept through learning to ride a bike, or having taught a child to ride a bike. For example, to successfully ride a bike—by that I mean get from location A to the desired location B—we must carefully and naturally navigate the dynamic mechanical system of the bike to such an extent that we are able to exercise our desires and wants through it to achieve the goal. Through this process, our whole biological system must adapt to this new dynamic environment and draw it into our senses to such an extent that we might (or start to in the case of our novice child) feel part of the bike, or the bike feels part of us. This attunement goes beyond the thinking mind alone (although that is part of it) and relies on other intelligences, senses and attributes of our being to successfully negotiate dynamic elements such as gravity, speed, acceleration, bumps, braking and balance. Once embodied, we start to forget about the separation of the two elements (bike and biology) and enjoy the experience of our new sense of self, and the sensation (perhaps thrill) of moving in a different way, and develop intelligent behaviours to deal with this new version of being.

In music terms, this is best described by Nijs et al. as:

> In music performance the embodied interaction with the music implies the corporeal attunement of the musician to the sonic event that results from the performance. The embodied experience of participating in the musical environment in a direct and engaged way is based on the direct perception of the musical environment and on a skill-based coping with the challenges (affordances and constraints) that arise from the complex interaction within this musical environment.

They continue to explain that this becomes:

> an optimal embodied experience (flow) when the musician is completely immersed in the created musical reality (presence) and enjoys himself through the playfulness of the performance. Therefore, direct perception of the musical environment, skill-based playing and flow experience can be conceived of as the basic components of embodied interaction and communication pattern. (Nils et al. 2009)

One of the key elements here is 'direct perception of the musical environment', and this goes back to the Eliot quote and the opening of this chapter.

While this chapter is not the place to explain the whole of phenomenology, it is helpful to understand the role that phenomenalist philosophers such as Heidegger and Merleau-Ponty can play in defining at least where and what this direct perception might be. Merleau-Ponty, for example, argues that 'perception and representation always occur in the context of and are therefore structured by, the embodied agent in the course of its ongoing purposeful engagement with the world' (Anderson 2003, p. 104). In basic terms, this means that the perception of, and the representation thereof, a world is experienced and expressed through the full body system; what Bachelard describes as the 'polyphony of the senses' (Bachelard quoted in

Pallasmaa 1996). These perceptions are not given form or content by the separate and autonomous mind, but rather are in themselves the form and content of the whole experience. To the phenomenologists, meaning is created through getting-to-grips with the experience of the world in flow. On this Anderson writes: 'at the highest level, what is at issue here is the fact that practical, bodily activity can have cognitive and epistemic *meaning*' (Anderson 2003, p. 109).

At the centre of embodiment is the recognition that not only does it involve acting within some physical world, but that the 'particular shape and nature of one's physical, temporal and social immersion is what makes meaningful experiences possible' (Ibid., p. 124). Embodied interaction, such as those described at the top of this chapter, is the 'creation, manipulation, and changing of meaning through engaged interaction with artifects' (sonic or physical) (Dourish 2001, p. 126). It is how the world reveals itself through our encounter with it, and in musicking terms, this happens inside the flow of musicking and the relationships that are encountered there. It is these percepts that are to be mapped onto actions, which should determine the operational behaviour of embodied AI.

4   Embodied AI and Musicking Robots

It should be obvious by now that the AI I am discussing here goes beyond a 'thinking engine' or an 'electronic brain' as posited in the origins of GOFAI research. As such, the AI that I design, build and deploy are not merely symbolic representation of thought processes (i.e. a schema for operation), or are used to construct the physical phenomena of music (i.e. the sound wave), or a trained neural network designs to output the meta-workings of music composition (i.e. the organisation and sequencing of music theory), although it may well do these as part of its role. Instead, embodied AI in music should:

> stimulate percepts leading to meaning-making in the human-musician so as to be believed to be rationally operating in the close-coupled relationships in the situated space of live music.

This is an immense challenge—life-long I would argue—and there are many ways to crack this nut; however, I have been developing these following strategies that get closer to an optimal experience in musicking. These are:

  i.   Creativity and the Flow
 ii.   Experiential learning and Recollection
iii.   Embodied dataset
 iv.   Coping and Beliefs
  v.   Embodied Percept Input: Affect module and Bio-synthetic nervous system

I will now introduce each of these in turn and discuss them in the context of the works that were created to develop and validate them.

4.1   Creativity and the Flow

As discussed above, meaning (or the *what-you-mean-to-me*) is to be 'found in the relationships formed between the embodied act of musicking and the materials therein'. I also mention that meaning is 'created through getting-to-grips with the world in flow' and that 'practical, bodily activity can have cognitive and epistemic *meaning*'. These of course are from the perspective of the human-musician; and although it is probably feasible to implement these meaning-making principles into an AI, my concern is in making humans more creative. As such, all following discussions about meaning are from the perspective of the human-musician in the flow. Flow in this sense is defined as 'the experience of musicking from the inside perspective of being inside the activity' (Vear 2019, p. 68).

The role of my embodied AI is to reach out, suggest, offer and shift connections and relationships as percepts (that which is 'taken-in') by the human-musician. It also needs to establish a world of creative possibilities for exploration through the flow of musicking that the human-musician is taken-into (another domain of percepts). This 'taking-in' and 'taken-into' structure is discussed in full in Vear (2019, Chap. 6) as a basic structure with which to understand the creative relationships in musicking that can form the *what-you-mean-to-me* to the human-musician. This basic structure is split into these two domains:

1. *Taking-in*—within the flow musicians make connections with the AI as they reach out, suggest, offer and shift through the tendrils of affordance experienced through notions of

   - *Liveness*: the sensation that the AI is co-operating in the real-time making of the music, and this meaningful engagement feels 'alive'
   - *Presence*: an experience that some*thing* is there, or I am *there*
   - *Interaction*: The interplay of self with environments, agents and actants[4]

2. Taken-into—the AI can establish a world of creative possibilities for exploration through the flow through the domains of

   - *Play*: the pure play of musicking happens inside a play sphere in which idea and musicking are immutably fused
   - *Time*: the perception of time (of now, past, future and the meanwhile of multiple convolutions of time) inside musicking plays a central role to the experience of the musician
   - *Sensation*: is an aesthetic awareness in the experience of an environment (music world) as felt through their senses

This basic structure was deployed in a collection of standalone music compositions entitled *Black Cats and Blues* (2014–18) which were recorded by the US cellist

---

[4] 'Actants denote human and non-human actors, and in a network take the shape that they do by virtue of their relations with one another'. (Online. Available https://en.wikipedia.org/wiki/Actor-network_theory (accessed 27 May 2022).

Craig Hultgren and released on Metier Records.[5] These compositions focussed on deploying and testing the 'taking-in' and 'taken-into' principles through narrow-AI within the flow of musicking. Narrow AI typically operates within a limited pre-defined range of functions and in this scenario worked to provide the tendrils and creative possibilities in the embodied dance of the flow (discussed at the top of this chapter), so that Hultgren could find and create meaningful relationships. The analysis of Hultgren's experience inside the flow with these compositions is discussed in detail in Chap. 6 of *The Digital Score* (Vear 2019).

4.2 *Experiential Learning and Recollection*

In an attempt to arrive at a suitable solution for the embodied AI to learn, I designed, developed and deployed a rapid prototype that explored notions of experiential learning and recollection. This dealt with three key questions that emerged from discussions about embodied AI and machine learning,[6] which are:

1. what is to be learnt/modelled in embodied musicking?
2. 'how can we be sure that our learning algorithm has produced a hypothesis that will predict the correct value for previously unseen inputs' (Russell and Norvig 2009, p. 713)?
3. how will the results have meaning for the embodied human-musician in the flow?

On Junitaki Falls a Trio for Solo Instrument and two AI Performers (2016–17)[7] was created to experiment with a propositional solution to address these questions. The central concern was to find new insights into the challenge posed by question 3, while also adhering to the stimulation of percepts as per my definition of embodied AI (above).

The solution I embedded was built around theories of behavioural AI with creativity philosophy inspired by David Gelernter's book *The Muse in the Machine* (Gelernter 1994). Behavioural AI [also referred to by Rodney Brooks as Nouvelle AI (Brooks 1990)] posits that:

> intelligence is demonstrated through our actions and interactions with the world. Critical to this is that the environment within which the robot operates must be independent of the robot design. (Jordanous 2020)

Gelernter theorises creativity based on attention, focus, affect and emotion. Although harshly criticised around its release, it is gaining interest in modern AI research as a gateway into new approaches to the nature of creativity and intelligence in embodied and behavioural AI systems. Of particular interest to *On Junitaki*

---

[5] https://divineartrecords.com/black-cats-and-blues-new-work-for-cello-and-electronics-from-metier/

[6] Thanks to Dr Fabrizio Poltronieri and the Creative AI Research group members for these.

[7] Available at https://vimeo.com/254090680.

Falls was Gelernter's notions of 'thought-trains' and the role of 'recollection' and 'affect linking', to creative thought, and his attempt to computerise the poetics of thought in its full richness. Coupled with core notions of p-novelty as defined by Margaret Bodem, these three facets can provide stimulation as percepts within the embodied interactions generated by my embodied AI in the flow (see Figs. 1, 2, 3 and 4 for a pictorial overview of these processes joined into a robot system). In short, these are defined as:

- *thought-trains*—are 'sequences of distinct thoughts or memories. Sometimes, our thought-trains are assembled—so it seems—under our conscious, deliberate
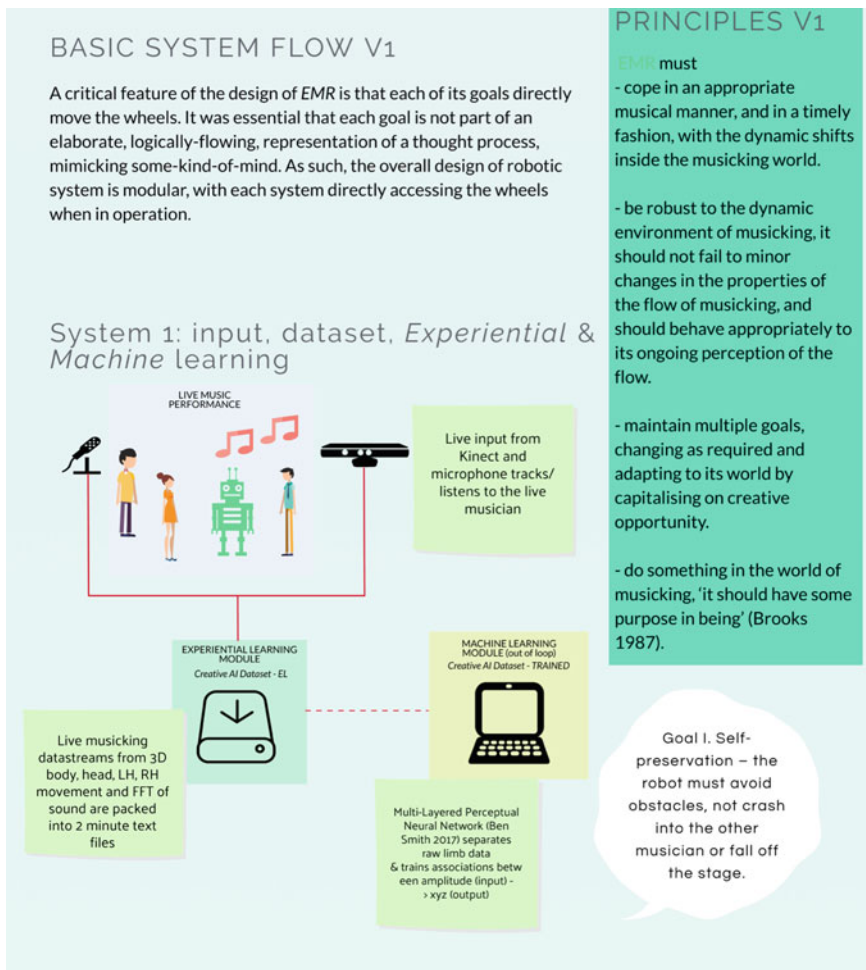

Fig. 1 Pictorial overview of the experiential learning system in a robot design
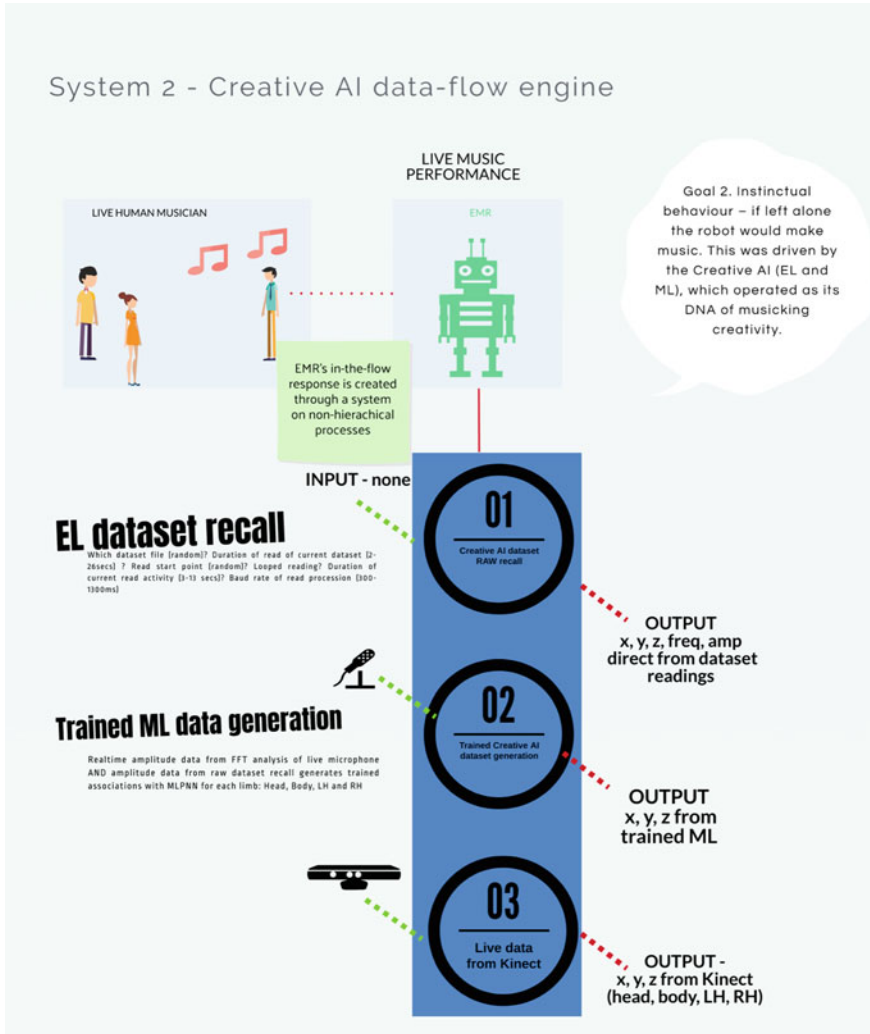
Fig. 2 Pictorial overview of the experiential learning system in a robot design

control. Other times, our thoughts wander, and the trains seem to assemble themselves' (Gelernter 2007).

- *recollection*—according to Gelernter 'when we pull out of memory a recollection associated with the same sort of feelings we're experiencing now… it's natural to apply the outcome or conclusion or analysis we arrived at then. And that's (in briefest outline) how emotions work as a 'parallel mind', how they lead us to fast conclusions we can't necessarily explain—but they feel right'.
- *affect-linking*—Gelernter argues that the mind can leap between mental images when 'two recollections engender the same emotion' (Gelernter 1994, p. 36).
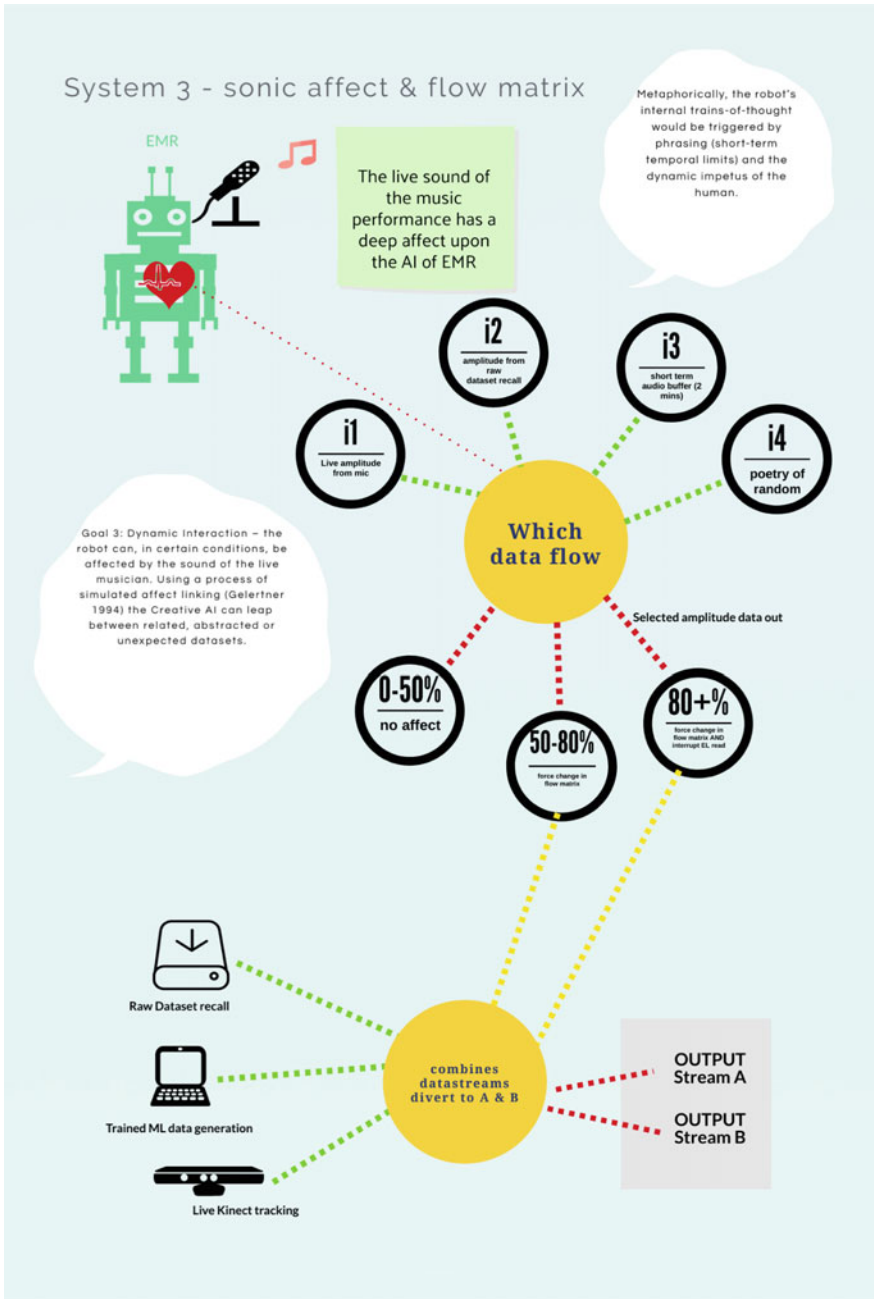
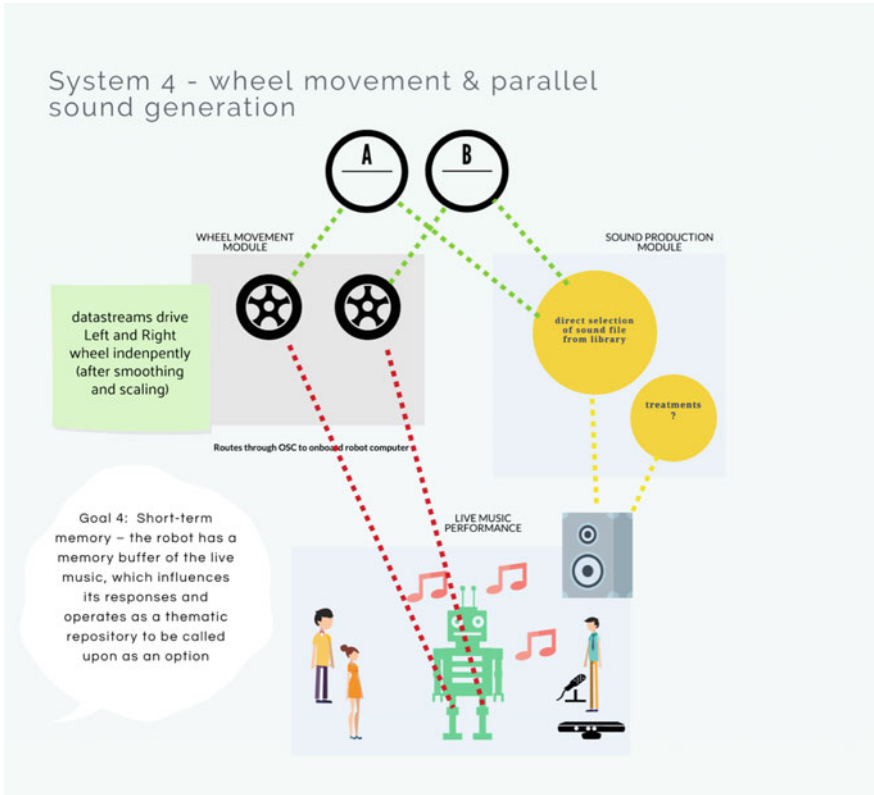**Fig. 3** Pictorial overview of the experiential learning system in a robot design

Fig. 4 Pictorial overview of the experiential learning system in a robot design

Each of these facets of the mind is part of a 'spectrum of thought' ranging between 'Upper-spectrum thought is abstract, full of language and even numbers; and lower-spectrum thought is concrete, full of sensation and emotion' (Ibid.). Furthermore, that these do not happen in steps but through a sudden awareness. 'It all depends not on a step-by-step logical sequence but on a step-by-step emotional one' (Friedersdorf 2017).

On Junitaki Falls was created for a solo musician, Christopher Redgate playing Oboe. It used a central director to control (a) a unique harmonic sequence with which it navigated a deconstructed version of a transcription of Eric Dolphy's *God Bless This Child*[8] (see Fig. 5), (b) a dynamic visual score shown on a laptop screen that mashed in real-time bars from the transcription that corresponded to the present harmonic sequence, (c) recorded and organised the recording of Redgate's performance as it corresponds to the current harmonic sequence; in a poetic sense, these became a repository of shared memories between the human-musician and the AI, and (d)

---

[8] The Roger Jannotta transcription was separated into bars and catalogued according to its underlying harmony. https://www.youtube.com/watch?v=m0-auuH1A-8.
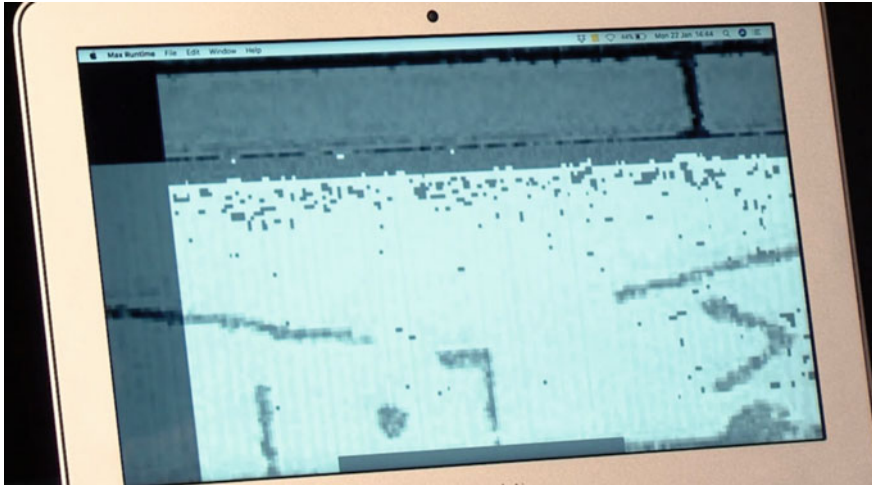
Fig. 5 An example of the AI-generated visual notation for *On Junitaki Falls*

coordinate the computer performers which recalled and manipulated the stored shared memories corresponding to the current harmonic sequence.

The technical system underlying this embodied AI was less concerned with symbolic ideas of logical thought, and more about supporting the process of it coping as an intelligent agent and for Redgate to feel its rational agency inside the music. The AI was not listening and analysing Redgate's performance, but instead used the central spine of the harmonic sequence as its perceptual core, as was Redgate. In practice, the human and AI agents operated concurrently within the shared environment of the live music. They are bound together by a shared goal: to navigate the pre-defined, randomly generated harmonic sequence. The percept manager (the general rule intended to regulate behaviour) of the embodied AI algorithm was to maintain a sense of familiarity with the human as both sounding material and logical adherences to the harmonic sequence. In this sense, the embodied AI for this project was listening into the harmonic sequence and problem-solving a response to it by controlling the two AI performers. The goal was to stimulate creativity in the human, who was simultaneously navigating the same harmonic sequence. In music terms, this concurrent relationship has less to do with improvisation and has more in common with pre-composed Western Art music, albeit, with a greater degree of freedom from the individual agents.

The embodied AI continually improves its relationship with the human by memorising and storing the human musicians' interaction with the AI over months, if not years, of shared experiences through the music. When presenting distorted sonic images of these memories back to the human within the flow of musicking, they operate like a distorted mirror of sorts, where the human feels an image of themselves in the flow and recognises a familiarity about these phrases leading to affect-linking. These distorted audio phrases working together with the visual score begets

another response from the human-musician, which is stored and catalogued as a shared acoustic memory through the memory folder system. This in turn creates a cycle of response and invention because of the thought-trains and affect-linking from within the embodied *situation* that they are operating in, and for this to be perceived by the musician as intelligent and creative from the situated space.

Crucially, the embodied AI behaviour needed to feel intuitive inside the live musicking (flow), be meaningful to Redgate (familiar/inspiring 'thought-trains') and do something creative in this realm. The result was a technical solution that emphasised the symbiosis in the behaviours of the code and the recorded media in real-time musicking, over symbolic traits of analysis, reasoning and logic. After 18 months of development, through a process of iterative design and deployment, increasingly complex beta versions were developed with Redgate to a point where he felt the embodied AI was 'being there with him'.

I would like to call this 'machine learning', but that term is already restricted to a form of statistical data analysis that automates and improves automatically, the use of data, so I call this experiential learning instead. Although *On Junitaki Falls* is restricted to experiential learning through gathering shared acoustic memories and recycling them through distorted recollection, it did provide enough evidence to support the premise that the human-musician was in a relationship with them and, crucially, engendered affect-linking and made meaning because of them. Furthermore, that by storing shared memories and recalling them did appear to Redgate to be learning an aesthetic/interpretative approach to each of the harmonies in the composition.

The dataset of audio memories captured as part of the *On Junitaki Falls* project now spans to over four years of performance and development with this individual composition. There are also, at the time of writing, two other musicians feeding and growing an individual dataset through their own nurturing of this project. The recollection processes within each of these versions are quite fascinating for the individual musicians; they are aware that recollections may be presented to them by the composition, and at the same time, their responses are being recorded, which in turn may form the raw material for a recollection later down the line. Crucially, these recollections stimulate meaning-making in the musician as they recognise an essence of themselves in the distorted recollection, that may be historic, or interpreted differently because of the distance in time. More so that they are not some superficial artefact of music-AI generation, but are a container of a type of felt embodied intelligence from which they offer percepts towards 'more evolutionary primitive mechanisms which control perception and action' (Ibid., p. 100).

As an extension of this, the next phase of experiential learning developed a dataset that captured bodily movement and audio memories of a human-musician in the flow of musicking. This created a proof-of-concept (PoC) embodied dataset that was implemented into the follow-on projects and is discussed next.

4.3 Embodied AI Dataset

The embodied AI dataset was designed to dig a little deeper into the challenge of *what is to be modelled* when deploying embodied AI in creative practices and to train neural networks from this dataset. The initial design focussed on a single principle of embodied musicking: that physical gesture and sound actuation are linked not by the sequence of *mental idea → physical gesture → sound-activation*, but that musician's think as sound and the sequence is closer to *sound-as-idea → sound-activation as physical gesture.* Therefore, if we are to model neural networks on embodied musicking, the physical gesture of the musician and the resultant sound production are implicitly linked not as part of a linear sequence of a thinking mind. This presents a more convincing set of features for modelling, than, say, harmonic construction.

The primary features of the dataset captured the complex of bodily movement of the musician with the physical properties of the sound. Defining the complex of bodily movements was an important aspect to this design, as I understood through personal experience and in conversation with other musicians, that, say, when an arm moves a cello bow, it is the whole body that is involved in the embodied music gesture, not just the isolated limb or joint. Therefore, the inter-relationship between the whole complex of bodily movements would need to feature in the dataset (see Fig. 6).

For the purposes of this PoC, the design for this first stage dataset was *Sequential ID; Body Part* (*Head, Body, Left hand, right hand*); *3-D position* (*x, y, z axes*); and *Audio Analysis* (*FFT fundamental, amplitude*). This data was then used to train four Multi-Layered Perceptron Neural Networks (MLPNN), one for each of the body
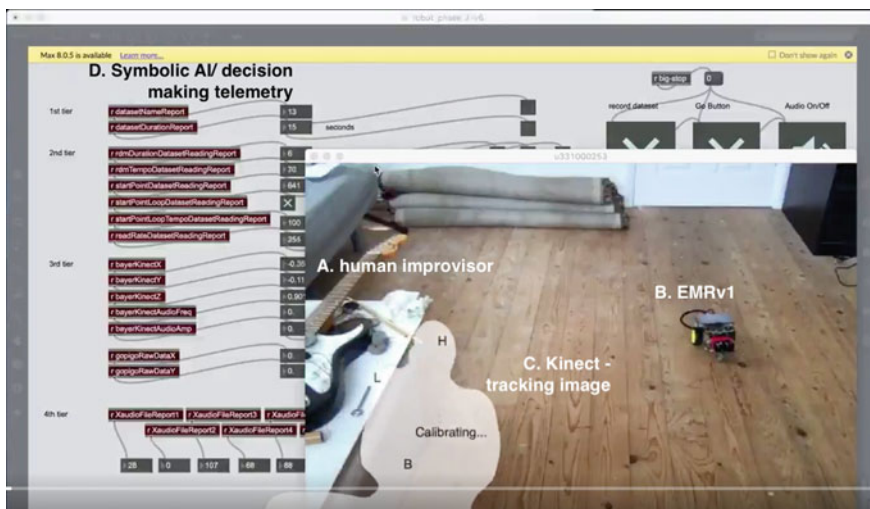

Fig. 6 Capturing the embodied dataset through in-the-loop methods and improvisation

parts captured, and implemented into two projects: *Seven Pleasures of Pris* (2019) and *Voight-Kampft Test* (2019). These projects were built around 2 then 3 separate AI's performing with each other. Each of the AIs contained each of the four MLPNNs. The behaviour of the AI was coordinated through a director that organised the streams of incoming signal (audio) with MLPNN response.

In the two examples, this MLPNN data and the raw dataset were used to control every aspect of the AI, movement, sound production choices and interaction goal. The raw dataset and the MLPNN outputs were also used to make choices about how the dataset is to be recalled and read by the algorithms (e.g. the read rate and ramp speed for each instance of wheel movement). This means that the direct application of data into wheel movement, and also the translations of that into sound-object choice and therefore as music in the flow is imbued with the essence of embodied musicking that has been embedded in the core of the dataset. The version of the dataset in this application was crude and small, but it has since been superseded by a larger project and more comprehensive embodiment approach to the dataset.

The inference I made through this process was that one isolated feature of this dataset, say the x coordinate of the right hand, might be imbued with a sense of embodied musicking, and this alone might be a viable data stream in the stimulation of percepts for the human-musicians. By that I mean, that any of the streams (raw data or neural network prediction) is embedded with a sense of embodied musicking and could therefore be used to control any parameter in the musicking AI behaviour. By extension and hypothetical proposition, that the essence of this dataset—imbued with creative and embodied musicking—could be used in a variety of creative applications not just music.

4.4 Coping and Belief

Earlier in this chapter, I discuss that the focus for embodied intelligence is on the coping behaviours that are required to maintain a balance of relationships within such a situated environment, and it is within this balance that the human-musician can find meaning. As music is a dynamic problem space, concentrating on the coping behaviours that are required to maintain a balance of relationships within such a situated environment is in my experience an optimal way of offering the human-musician an opportunity to find meaning. This is because the situated space of embodied musicking requires musicians to cope with a dynamic and changing world and that they need to do something in this world. If the AI is also to cope with dynamic changes and more importantly do something in this world that stimulates embodied interaction with a human-musician, then modelling it to some representational world-view is not the solution as this brings with it some significant questions, such as whose world-view are we to use for this representational model? and why would we want to have a strangle hold on the limits of its creative potential to known parameters?

Embodied AI needs to cope in real time within the realm of musicking and stimulate an embodied interaction within the flow. This requires a non-representational

approach to how it relates to the flow as the coping mechanisms needed to be open and dynamic enough to co-operate in any given musicking realm. Limiting the robot to a single representation of what musicking is, or might be, imposed onto the system by the human designer(s), would only work in a number of instances.

The design of the coping systems that I implement is informed by two early papers by the robot innovator Rodney Brooks, specifically *Intelligence without Reason* (Brooks 1991) and *Intelligence without Representation* (Brooks 1987). In these, he lays out the foundation of his approach to designing and building robots that are first and foremost able to cope and therefore adapt to a dynamically changing environment within the parameters of specific and multiple goals. Brooks' research considered these robots to be a sort of 'creature' that copes in a specific world-of-concern in real time. They do not have a model of representation of their world (such as building a 3-D model of the space through computer vision and object analysis), nor do they make one as it goes about its business, but uses goals and strategies to cope with whatever that world can throw at it.

Brooks' foundational theories guided the developed for my early embodied AI musicking robot (EMR) projects[9] and generated this set of principles[10]:

- EMR must cope in an appropriate musical manner, and in a timely fashion, with the dynamic shifts inside the musicking world;
- EMR should be robust to the dynamic environment of musicking, it should not fail to minor changes in the properties of the flow of musicking and should behave appropriately to its ongoing perception of the flow;
- EMR should maintain multiple goals, changing as required and adapting to its world by capitalising on creative opportunity;
- EMR should do something in the world of musicking, 'it should have some purpose in being' (Brooks 1987).

A subsumption architecture was designed to support these multiple goals. This is a control architecture innovated by Rodney Brooks as an alternative to traditional AI, or GOFAI. Instead of guiding robotic behaviour by symbolic mental representations of the world, subsumption architecture correlates sensory information to action selection in a 'bottom-up' way. These were (in order of priority):

I.   *Self-preservation*—the robot must avoid obstacles, not crash into the other musicians or fall off the stage.
II. *Instinctual behaviour*—if left alone the robot would make music. This was driven by the embodied AI dataset (discussed above), which operated as its DNA of musicking creativity.
III. *Dynamic Interaction*—the robot can, in certain conditions, be affected by the sound of the live musician. Using a process of simulated affect-linking the embodied AI could leap between related, abstracted or unexpected datasets. Metaphorically, the robot's internal trains-of-thought would be triggered by phrasing (short-term temporal limits) and the dynamic impetus of the human.

---

[9] https://twitter.com/i/status/1074617355345559552.

[10] Adapted from Brooks (1987).

A critical feature of the design of EMR is that each of these goals directly moves the wheels. It was essential that each goal is not part of an elaborate, logically flowing, representation of a thought process, mimicking some-kind-of-mind. As such, the overall design of robotic system was modular, with each system directly accessing the wheels when in operation. The overall design of the data flow is:

1. *Live data sensors (audio and visual) & SELF PRESERVATION*

|

2. *Data wrangler, neural net prediction, raw data recall/ recollection*

|

3. *Affect mixing (trains-of-thought): INSTINCTUAL BEHAVIOUR & DYNAMIC INTERACTION*

|

4. *Smoothing and deviation*

|

5. *Wheels move = Make sound*

In order for the embodied AI to have some purpose of being, it also needs to have a world view of music; in fact, I would argue that it needs a personalised world view rather than the whole entirety of music at its disposal. In short, the embodied AI need to have a belief system that frames its personalised world view through limitations, embedded aesthetics and behavioural traits (even glitches and bugs in the system).

Belief in this sense is used to describe an acceptance by the robot that something is true or that it has trust or confidence in something from its perspective especially in how it interprets percepts in the environment of musicking. Beliefs are not facts: they are subjective and individual, and they are not consensual and deal with conceptual parameters that may be biased, prejudiced or activated by affect and emotion. Beliefs may be based on subjective experiences and past episodes and may distort reasoning, knowledge and the processes of synthesising knowing (Cohan and Feigenbaum 1982, pp. 65–74). All of which I recognise as very musical traits.

The belief systems that I have implemented in my embodied AI are, at the moment, limited to three different attributes:

(A) *Personalised embodied AI dataset*—at the core of embodied AI belief system is a dataset that drives the individual AI. This PoC dataset has been represented in different formats across different project. The design of this dataset is discussed above, but it is important to stress that a new neural network was created for each of the AI, and that an individual character was embedded into each by randomising the raw dataset before training each model.

(B) *Movement behaviour*—The robot's movement operates within a behavioural system, designed to react openly to the dynamic soundworld and move the wheels accordingly. It is important to note that the interaction with the musician begets movement as its primary goal for musicking and that this movement is embedded with essence of embodied musicking because of the embodied dataset process. Following this, the movement begets sound, which begets music such that all relationships between human and AI are informed by phenomenon data captured within the embodied flow of musicking: either from the neural network, raw dataset or through live interaction.

C) *Soundworld*—the robot has a fixed sound library of sounds that amounts to its foundational aesthetic repository. These are different for each robot and are always extracted recordings created by humans engaged in embodied musicking, so as to embed them with an essence of musicianship. These are triggered only when the wheels move. These are then either presented to the world in their raw state or are treated in some way (time stretch, pitch shift or both) using the data streams as controlling parameters.

4.5   Embodied Percept Input: Affect Module and Bio-synthetic Nervous System

It has been necessary to rethink the sensory mechanisms and percept inputs for embodied AI. While standard instruments such as 3-D webcams (e.g. Kinect) and microphones can capture some of the physical properties of embodied musicking, it has been necessary to re-conceptualise aspects of sensory capture that help to emphasises belief and the inside nature of a type of perception in musicking.

Although the two solutions that I will introduce here—the affect module and the bio-synthetic nervous system—may read like attempt to bring the AI closer to a level of consciousness and human perception, they are in fact named metaphorically and are merely solutions that get the AI closer to a set of characteristics that I defined in order to stimulate such a relationship. These characteristics are:

1. listening
2. having my own train-of-thought
3. being surprised and maybe doing something with that
4. being surprising.

The first solution, the affect module, interrupts the continuity of internal data streams in response to the live sonic stimuli from the music. This module received all the data streams from the dataset query and parsing process, and the neural networks and mixed them to two outputs: left wheel data and right wheel data. The mixing was controlled by a special process designed to symbolically represent affect and affect-linking of a musician. I define affect as 'the mind's connecting response between sensorial input of external events with the internal perception of causation such as emotion or feeling, through time' (Vear 2019). This module translated this definition symbolically, the streams of amplitude data from the live input, the dataset parsing and a randomly generated 'drunk walk' would be used to trigger (a) local changes in the module such as mix and (b) global conditions such as dataset file selection. The basic process was:

1. randomly switch between input streams (1–4 s, or with a loud affect trigger)
2. if amplitude is < 40% do nothing
3. else if amplitude is between 41 and 80% trigger a new mix (see below)
4. else if amplitude is > 80% trigger condition changes across the architecture (new mix, new file read, restart reading rate and change smoothing rate, change audio read in following modules).

The mix function randomly selected which of the incoming data streams (x, y, z from dataset read, x, y, z, from live Kinect and x, y, z from neural network prediction) to be output to the following module for wheel movement. It was desirable that this involved multiple elements from these incoming streams being merged, metaphorically fusing different trains-of-thought into a single output.

The second solution, the bio-synthetic nervous system (BSNS), works as a sensory input for the embodied AI. In short, a container of moss is placed on top of the speaker used to transmit the AI's sound generation and relay the incoming sound from the other musician. Inside the moss container are positive and negative terminal that sense capacitance changes across the moss in response to vibration. When the speaker makes a sound, the levels of capacitance change, although there is no direct correlation between input value of a specific vibration, say 440 Hz, to the capacitance shift. In Fig. 7, you can see in an initial experiment how the capacitance generates a rhythmic response when I play a 50 Cent track through the speaker. The value from this system is then fed into the affect module as a 'self-awareness' data stream.

I must stress that I do not believe the AI to be self-aware (that I'm aware of), but this system is a way of achieving several connections between human-musician and the AI through musicking. There are three main connections that this system engenders:

1. the BSNS operates as a crude embodiment sensor. It is a direct connection between the sound vibration and some felt stream, that is different
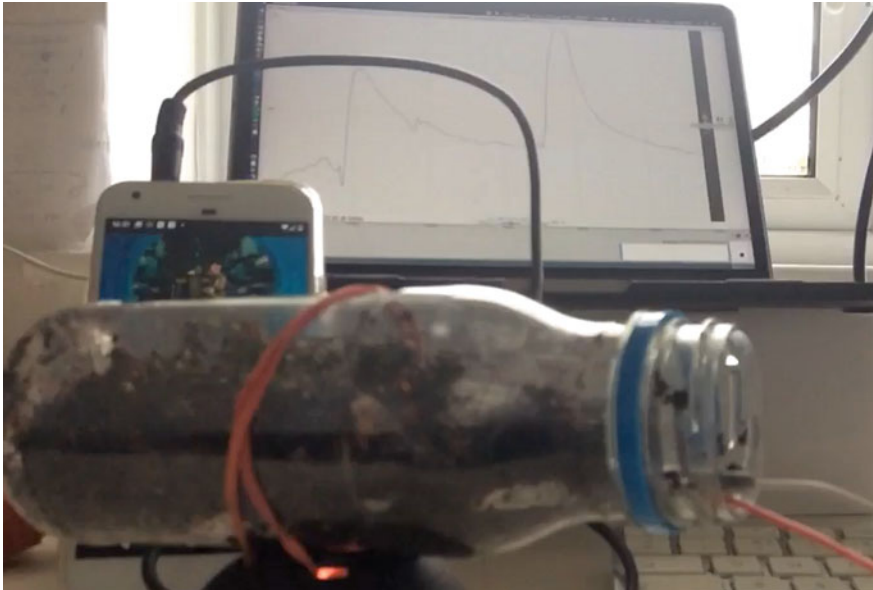

Fig. 7 A track from 50 cent playing into a speaker, on top of which the moss-based sensor system feels the vibration, which is turned into digital signal (on the screen behind)

from a microphone and introduces a sense of biological connection with the human-musician.

2. the BSNS stream is subjective as there is no true direct correlation between the moss's response and the sound frequency or amplitude. This feeds into the AI's belief system as these are based on subjective experiences and past episodes and can distort reasoning, knowledge and the processes of synthesising knowing, beyond empirical and into interpretation.

3. to paraphrase Breazeal's description of her metaphorical 'synthetic nervous system' (Breazeal 2004, p. 38) in her social robot *Kismet*, BSNS draws together inspiration from 'infant social development, psychology, ethology and evolutionary perspectives' to enable embodied AI to 'enter into a natural and intuitive social interaction' with a human-musician.

5   Conclusion

In embodied and creative pursuits such as creative AI with music, or any performing/interactive situation, it is important to address the meaning of artificial intelligence and to understanding it in its context, rather than imposing upon it an outsider objective viewpoint limited to mind-based models. While I am not trying to claim that this research presented here is the solution, or even that this approach is part of the solution, I am proposing that understanding embodied intelligence from these perspectives can unlock areas of creativity with AI, by enhancing the close-coupled relationships that stimulate a sense of meaning in the human-musician.

The solutions that I have introduced above prioritise a way of thinking about embodied musicking between human-musicians and AI/robots. This way of thinking can be outlined by the following principles:

- *the robot was not an extension of the musician; but should extend their creativity*
- *the robot should not be an obedient dog or responsive insect jumping at my commands or impetus, but a playful other*
- *it should not operate as a* simulation *of play, but as a* stimulation *of the human's creativity*
- *it is not a tool to enhance the human's creativity, but a being with presence in the world that they believe to be co-creating with them*
- *it should prioritise emergence, surprise, mischief, not expectation.*

These solutions discussed in this chapter do not operate like, or even resemble GOFAI, if you were to only judge them by 'looking under the hood' at the code. This is because the 'hood' in this case is not symbolic code that solves problems or learns data sequences but is the operational behaviour of the code in its situated environment.

There is much more for this research to do. This will take time as each step of the research journey builds a new AI and robot and evaluates it inside embodied

musicking. This is truly important as, to quote Rodney Brooks, 'one real robot is worth a thousand simulated robots' (Brooks 1989).

References

Anderson ML (2003) Embodied cognition: a field guide. In: Artificial intelligence, 149, pp 91–130

Bachelard G (1958, 1994 ed.) The poetics of space. Beacon Press, Boston

Breazeal C (2004) Designing sociable robots. MIT Press, Massachusetts

Brooks R (1999) Cambrian intelligence. The early years of the new AI. MIT Press, Cambridge

Brooks R (1991) Intelligence without reason https://people.csail.mit.edu/brooks/papers/AIM-1293.pdf. Last accessed 2020/10/23

Brooks R (1990) Elephants don't play chess robot. Auton Syst 6:3–14

Brooks R (1989) "Battling reality". Available online https://www.researchgate.net/publication/37597354_Battling_Reality. Accessed 12 May 2022

Brooks R (1987) Intelligence without representation. Artif Intell 47(1):139–159

Cohan PR, Feigenbaum EA (eds) (1982) The handbook of artificial intelligence, vol III. Addison-Wesley Publishing Co. Inc., New York

Dourish P (2001) Where the action is: the foundation of embodied intelligence. MIT Press, Cambridge

Emmerson S (2007) Living electronic music. Routledge, London

Friedersdorf C (2017) 'There's enough time to change everything': the polymath computer scientist David Gelernter's wide-ranging ideas about American Life. In: The Atlantic. Online available at https://www.technologyreview.com/2007/07/01/36908/artificial-intelligence-is-lost-in-the-woods/

Gelernter D (2007) Artificial intelligence is lost in the woods: a conscious mind will never be built out of software, argues a Yale University professor. In: Technological review. Online, available at https://www.technologyreview.com/2007/07/01/36908/artificial-intelligence-is-lost-in-the-woods/

Gelernter D (1994) The muse in the machine

Jordanous A (2020) Intelligence without representation. Systems 8:31. Online available at https://res.mdpi.com/d_attachment/systems/systems-08-00031/article_deploy/systems-08-00031-v2.pdf

Nijs L, Lesaffre M, Leman M (2009) The musical instrument as a natural extension of the musician. Online. Available http://www.academia.edu/205605/The_musical_instrument_as_a_natural_extension_of_the_musician. Accessed 30 Jan 2021

Pallasmaa J (2005) The eyes of the skin. architecture and the senses. Wiley, New York

Russell SJ, Norvig P (eds) (2020) Artificial intelligence—a modern approach, 4th edn. Prentice Hall, New York

Russell SJ, Norvig P (eds) (2009) Artificial intelligence—a modern approach, 3rd edn. Prentice Hall, New York

Vear C (2019) The digital score. Routledge, London

Wooldridge M (2020) The road to conscious machine: the story of AI. Pelican Books, London