

Article

# Emergent Interfaces: Vague, Complex, Bespoke and Embodied Interaction between Humans and Computers

Tim Murray-Browne <sup>1,\*</sup>  and Panagiotis Tigas <sup>2</sup> <sup>1</sup> Dyson School of Design Engineering, Imperial College London, London SW7 9EG, UK<sup>2</sup> Department of Computer Science, University of Oxford, Oxford OX1 3QD, UK; ptigas@robots.ox.ac.uk

\* Correspondence: t.murray-browne@imperial.ac.uk

**Abstract:** Most Human–Computer Interfaces are built on the paradigm of manipulating abstract representations. This can be limiting when computers are used in artistic performance or as mediators of social connection, where we rely on qualities of embodied thinking: intuition, context, resonance, ambiguity and fluidity. We explore an alternative approach to designing interaction that we call the *emergent interface*: interaction leveraging unsupervised machine learning to replace designed abstractions with contextually derived *emergent representations*. The approach offers opportunities to create interfaces bespoke to a single individual, to continually evolve and adapt the interface in line with that individual’s needs and affordances, and to bridge more deeply with the complex and imprecise interaction that defines much of our non-digital communication. We explore this approach through artistic research rooted in music, dance and AI with the partially emergent system *Sonified Body*. The system maps the moving body into sound using an emergent representation of the body derived from a corpus of improvised movement from the first author. We explore this system in a residency with three dancers. We reflect on the broader implications and challenges of this alternative way of thinking about interaction, and how far it may help users avoid being limited by the assumptions of a system’s designer.

**Keywords:** Human–Computer Interaction (HCI); Artificial Intelligence (AI); unsupervised machine learning; emergent representation; non-representational interaction; dance; music; creative coding; New Interfaces for Musical Expression (NIME)



**Citation:** Murray-Browne, T.; Tigas, P. Emergent Interfaces: Vague, Complex, Bespoke and Embodied Interaction between Humans and Computers. *Appl. Sci.* **2021**, *11*, 8531. <https://doi.org/10.3390/app11188531>

Academic Editor: Philippe Esling

Received: 31 May 2021

Accepted: 23 August 2021

Published: 14 September 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Building an interactive interface involves creating digital representations of the actions of a human, and defining the influence these have over the effects of the system. These representations are typically in the form of designed *abstractions*: bundles of qualities chosen by the system’s designer. These choices are inevitably infused with assumptions made by the designer about who will use the system and how, which in turn can limit the person who eventually interacts with it. This paper outlines artistic research into an alternative approach which leverages unsupervised machine learning to replace throughout the interaction loop designed *abstract representations* with contextually derived *emergent representations*. This approach allows us to create expressive gestural interactions without explicitly declaring input or output. This may help free the user from the ontological assumptions of the system designer, although, as we explore, it may introduce other biases.

Gillies [1] uses the term *movement interaction* as any technique for Human–Computer Interaction (HCI) that uses the body in a different and fuller way than a mainstream interface. Movement interaction has the potential to leverage our existing embodied skills, such as sensorimotor skills and environmental awareness. This is in contrast to graphical user interfaces (GUI), which leverage our skill at manipulating visual representations. Designing for embodied interaction is challenging, because the embodied knowledge it relies upon is implicit. We may know a gesture through doing and experiencing but this

knowledge is non-representational [2]. Designers cannot represent what they cannot access consciously so end up falling back on adapting graphical metaphors, such as buttons in 3D space, or falling back on easily calculated features of movement such as the position and velocity of a body part. This leads to gestural interaction where a symbolic gesture is arbitrarily mapped to a symbolic action [1]. It does not gain from the detailed awareness of our sensorimotor skills, but it loses the feedback and ease of discovery that a graphical interface brings [3].

While here we consider movement interaction in a creative context, our approach may hold value in types of interaction beyond movement interaction that leverages complex multidimensional input signals, such as Brain–Computer Interfaces [4].

### 1.1. The Duality of Embodied and Formalist Cognition

The thesis of embodiment posits that many features of cognition are deeply dependent upon characteristics of the physical body [5]. Traditional, non-embodied models of cognition assume mental representations to be symbolic, quasi-linguistic abstractions. Wilson and Foglia [5] write, ‘on the traditional view, not only are the internal representations employed in language, concept formation, and memory essentially distinct from those processed by the sensorimotor system, but their meaning is divorced from bodily experience.’ In contrast, the embodied perspective sees cognition as relying on qualities of representations more directly connected to perception and action [5]. Embodied knowledge is then defined as knowledge in which the body knows how to act, such as riding a bicycle, but which we may not be able to articulate [6]. The term *embodied* is often used more broadly to describe any knowledge that is grounded in and emerges from everyday experience, which may go beyond the physical body to include social interaction [7]. It rests on a phenomenological approach that Dourish describes as *preontological*:

*Perception begins with what is experienced, rather than beginning with what is expected; the model is to ‘see and understand’ rather than ‘understand and see’ [7] (p. 98).*

Storkerson describes ‘naturalistic cognition’ as an umbrella term to include embodied, unconscious, implicit, experiential and non-conscious thinking. In contrast, *formalist* cognition is rule-based, and deals with abstract entities, where abstract does not necessarily mean ‘divorced from reality’, but considered independently from its context [8].

While useful, the term ‘naturalistic’ may be conflated with the philosophical position of naturalism. Following Dourish, I will adopt the term *embodied cognition* in its broader sense to also encompass the unconscious, implicit, experiential and non-conscious.

In the context of HCI, Dourish describes *embodied interaction* as ‘the creation, manipulation, and sharing of meaning through engaged interaction with artefacts’ [7] (p. 125). For Dourish, meaning emerges through communities of practice and their shared histories, identity and practical activities, but an interactive system embeds the ontological assumptions of its designer, i.e., what is captured, how it is individuated and what qualities of it are relevant. By assuming these assumptions to be shared between designer and user, systems become ‘brittle’ to being adapted and appropriated to new practices [7]. According to Dourish, and others such as Rajko [9], embodied interaction is more a stance on interaction design rather than a form of it. It emphasises that meaning arises through action, and actions are situated in communities of practice.

Embodied thinking is critical to perception. However, while formalist thinking relies upon embodied in order to ‘ground’ its abstract symbols to reality, we should avoid underestimating the role of embodied thinking in human thought as merely a support mechanism to abstract reasoning [10]. For example, Stanovich and West’s dual process theory [11] models cognition through the interplay of two systems. System 1 is heuristic: fast, opaque and involuntary, while system 2 is slow, deliberative, reflective and more commonly associated with the subjective experience of agency. This framing underlies theories such as ‘mindless computing’ where an individual’s behaviour is shaped by digital interaction in ways they are not conscious of [12]. The term aptly describes the endless scroll through social media feeds where our behaviour seems to be operating beyond

our volition. However, few would describe the heightened state of focused, embodied consciousness of a dancer or athlete in line with the mindlessness of being hypnotised by scrollbait. There is presence, awareness, intelligence, focus and a sense of agency without the need for formalism or deliberation.

Both formalist and embodied cognition are fundamental to human cognition, but the former has tended to dominate within HCI. This perhaps reflects the layers of abstraction and formalism that underlie the development of software. For our discussion here, the key distinction of this duality is between *abstract* and *situated* mental representations which are native to formalist and embodied modes of thinking respectively. My use of the term *abstract* follows that of Storkerson [8] and McGilchrist [10] to emphasise a representation that may be reasoned about independently of its context. I use *situated* to describe knowledge whose meaning is unavoidably entangled with a context, which could include perceptual or sensorimotor capabilities. (Gillies and Kleinsmith [2] use the term ‘non-representational’ to describe knowledge that cannot be expressed symbolically, but I avoid this term as I feel it relies on a specific interpretation of ‘representation’ which would be confusing in the later discussion on emergent representations.)

### 1.2. The Challenge of Resituating Abstract Representations

A situated label becomes an abstraction when it is detached from the context it emerged from and is instead treated as a symbol that can be dropped into a new context. This enables rule-based reasoning and computation. However, when the abstraction is imposed back onto the world, we risk what Hayles describes as the ‘Platonic backhand’: deviations from the abstraction are regarded as noise or distortion masking this idealised form rather than as the ground truth the abstraction approximates [13]. For example, Rajko offers a critique from an embodied stance of Google Glass, a pair of internet-connected ‘smart’ glasses with a mounted camera. She quotes one of its designers reasoning that a camera on a pair of glasses was no different than the many security cameras surrounding us [9]. This reasoning considers the camera in the abstract. The camera becomes any camera, rather than the camera that I, of all people, have chosen to attach to my face to record specifically you while we hang out.

Friction between formalist and embodied thinking can become particularly apparent when we introduce computers into embodied activities. Much of my own artistic practice has involved collaborating as a coder with dancers in the studio. For the dancer, an idea may be immediately materialised into movement. Indeed, in many cases choreographic material is not so much invented as it is encountered through exploratory improvisation. This is often captured in the term *instant composition*, of which dancer Neuhaus writes:

*The performer’s body and mind need to be specifically tuned for perception, imagination, intuition, inhibition and action. At the same time one must be able to constantly read and respond to one’s body, the other dancers and the composition itself, including all its different layers. These seemingly elusive qualities are actually skills that can be learned through specific training and practice. [14]*

Her description reminds us that embodied thinking can be fully immersive and require the complete focus of the mind, much like Csikszentmihalyi’s concept of ‘flow’ [15].

In a dance/tech collaboration, this process can be frustrated by the slow pace of the coder, who must imagine, plan and execute, doing their best to work in short iterations. Of course, while slower to create, once it is written code becomes endlessly reusable (barring technical failure). However, this fixed nature means that any abstract representations of movement that have been conceived in the coding process, such as gestures and trigger zones, are now fixed constraints to work with. Yet, movement of the body remains ephemeral and in flux. For the dancer to ‘fix’ the material they encounter, should they so wish, additional work and skill is required. Even when working with non-tactile sensors such as cameras, coded abstractions can constrain the body’s physical movement.

For example, suppose we encounter a moment where one dancer reaches a hand out towards a person across the room. We want to return to this so we give it a name, ‘the reach’.

Our understanding of what we mean by ‘the reach’ is situated: the representation it triggers is embedded in the situation we all experienced when that material was encountered. In this way, the name is a label to an implicit representation we share that has emerged from the activity. If a new person joins the performance, we cannot expect them to understand what we mean by ‘the reach’. To share this, we would need to recreate the situation for them. What the label represents has emerged from a shared context. Gillies describes this as an *emergent representation* [1].

Now, suppose we want to represent this moment digitally, say to trigger a sound. The coder might abstract ‘the reach’ into ‘a hand pointing within  $\theta^\circ$  of direction  $v$ ’. If that is too general, then we might try a more complex abstraction, such as ‘a hand pointing within  $\theta^\circ$  of the line connecting its body to any other body’. The representation of the situation that we receive from the sensor defines which abstractions are simple and which are complex to implement, what we might call its *inductive bias* (a term borrowed from Machine Learning). To say it is ‘divorced from reality’ would be to suggest it was ever married to it, which it was not. Any relationship between the digital abstraction and the situated thing it represents needs to be invented from scratch. Therefore, it does not easily factor into the actions of a performer responding instantaneously to their perception of the situation. Working with this fixed digital abstraction becomes a skill that needs to be practised and internalised to factor into embodied thinking.

Rajko calls for HCI designers to practise embodiment first-hand ‘way of being’ to increase their contextual sensitivity and ability to process complex dynamics intuitively [9]. However, this is not enough. The friction I describe above is not because the coder is insensitive to the context, but because code is explicit while what we intend to capture is implicit. A simple abstraction in code can become a complex constraint for the situated human due to its inflexibility. However, a few simple words describing a situated movement can be intractably complex to define explicitly in code.

### 1.3. Emergent Representations

Emergence, in the context of interactive art, is defined by Seevinck as ‘occurring when a new form or concept appears that was not directly implied by the context from which it arose’ [16] (p. 1). To design for emergence is to design for openness and ambiguity, to create opportunity for forms to emerge that go beyond the designer’s vision. This emphasis on allowing what emerges to go beyond the designer’s model is similar to Dourish’s preontological model: understanding follows perception, rather than perception following understanding.

Gillies articulates the challenge presented by the traditional approach to programming interaction design through symbols and libraries of abstractions. Those designing for embodied cognition need a first person perspective, to be doing and moving. However, what they design does not easily translate to the symbolic realm of code [1].

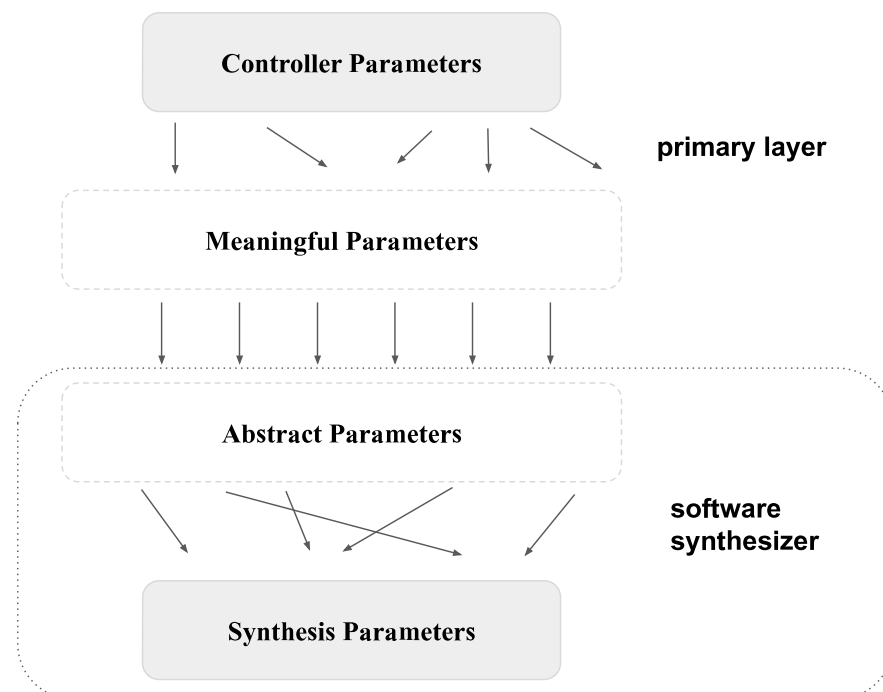
He, and others [17,18], propose Interactive Machine Learning (IML) as a solution. IML allows a creator to iteratively build up a set of examples of a gesture to train a model to identify it, testing as they go to gain feedback. Tools such as Wekinator [19] and MIMIC [18] have been created to support this process. The iterative process allows a shared representation of the gesture to emerge between creator and system, without the need for it to be made explicit. In this way, IML helps us escape the inductive bias of the sensor’s representation. We can define a vocabulary of input gestures by doing rather than explaining mathematically. This saves our creator from attempting to find an explicit representation of the gesture they understand implicitly. However, the output from a supervised model remains an abstract representation: either a symbolic event from recognising a specific gesture or else a predefined parameter. Instead of mapping symbolic gestures to symbolic actions, we are mapping emergent gestures to symbolic actions or parameters. While we no longer need to be explicit about *why* this is gesture  $X$  and that is gesture  $Y$ , we are still left with a discrete vocabulary of symbolic actions we can perform:

X and Y. IML helps us to bind our emergent gesture to an abstract representation in the system, but it does not let us eliminate that representation.

For example, Google Maps on the iPhone lets you indicate an error in the map by shaking your phone. One might argue that this is embodied design as there is a link between the frustration of finding an error and the mildly violent act of shaking the phone. However, the effect of the action is to open a prompt for feedback, a discrete event. How you feel or think when you shake the phone does not factor into the effect, so the action itself becomes an abstraction, a ‘shake’. We still act through the symbolic realm. Little value is added from our embodied intellectual capabilities.

This example highlights the difficulty in introducing embodied interaction into a traditional interactional activity. A file is a file, and it would not be useful for it to change based on the expression on my face or the room I am currently sat in. An operating system of abstractions demands abstract thinking which relies on precision and consistency. Traditional interfaces have evolved to suit these needs well. The prevalence of abstractions named after files, folders, desktops and menus belie the bureaucratic roots of everyday operating systems and the formalist thinking it is designed for.

Now, consider a Digital Musical Instrument where we are processing a high-dimensional input into a high dimensional output. This presents the *mapping problem*, familiar to researchers in the field of *New Instruments for Musical Expression*: how do we define a function that translates human action into the many possible parameters governing a synthesiser in a way that best supports someone making music. A common approach outlined by Hunt et al. [20] is to transform both gesture and synthesiser parameters into abstract parameters that we can make sense of, such as ‘energy’ and ‘wobble’ for gestural parameters and ‘brightness’ and ‘pitch’ for synthesis parameters. We are then in a position to decide rationally how these two sets of parameters ought to be related. This creates the three layered mapping model shown in Figure 1.



**Figure 1.** Hunt et al.’s three-layer mapping model [20]. Raw sensor parameters are mapped to ‘meaningful’ gestural parameters, which are mapped to ‘abstract’ synthesis parameters, which are mapped to raw synthesis parameters.

The intermediate state is defined by explicit abstractions much like the GUI of an operating system. However, unlike an operating system, Hunt et al. found musicians prefer more complex mappings:

*So, why should a complex and cross-coupled interface completely outperform a more standard sliders-only one? Most users of the multiparametric interface describe a moment when they ‘stopped thinking’ about the interface and began to ‘just play’ with it—almost as if their body was doing it for them, and their conscious mind was somehow ‘not in control’ [20].*

The musician’s description aptly matches our description of embodied thinking.

A preference for complexity was notably not replicated by Brown et al. [21]. In interviews with musicians integrating the MiMu Glove controller into their stage performances, they found a preference for consistency and simplicity. As observed above, what is simple from an engineering perspective can be the opposite of what is simple from an embodied perspective, which may go some way to explain this difference. Perhaps there was enough going on in the full context of their stage performance that there is less risk of losing their flow by overthinking things. There is also the issue of how the meaning of the interaction comes to be shared between musician and spectator [22] as, presumably, Hunt et al.’s musicians were playing in a lab context where a shared understanding of the interactional mechanics had been established.

However, the findings nonetheless suggest that even an opportunity to adopt a formalist approach to interaction is enough to undermine one’s capacity to act through an embodied approach. To me, this sounds similar to the difficulty we might find it to walk naturally once our attention has been drawn to our style of walking.

#### 1.4. Interaction without Abstraction

To summarise, the abstract representations that characterise most interactive systems may inhibit a person’s capability to interact through an embodied way of being. We may accommodate this by internalising the representations but we are still constrained by decisions made by the designer who is unlikely to be a part of the context of interaction. Even when the system designer is fully situated in that context, the explicit nature of coding makes it impossible to build systems with the same fluidity and sensitivity that embodied action is capable of.

Is it possible to create meaningful interaction between human and computer without designing abstract representations? In the next section, I propose that unsupervised machine learning offers the opportunity to build the entire system of interaction through emergent representations. To do so, we need to think of the system as an embodied agent in its own right rather than as an agent of an embodied human. We adopt Dourish’s preontological position to consider how the system can root its understanding in what it can perceive rather than basing its perception in what it is primed to understand.

## 2. Materials and Methods

### 2.1. Machine Learning and Emergent Processes

To consider how to build an interactive system that avoids abstract representations, we can consider how the distinction between formalist versus embodied thinking is somewhat analogous to that between traditional programming and machine learning.

Traditional programming involves abstracting behaviour into procedures, classes, databases and protocols. We look for generalisations that allow us to ‘factorise’ code into reusable components, broadening the range of contexts in which our code might be used. This is necessary to leverage the power of computation because the process of writing code is so much slower than the process of executing it.

However, machine learning (ML) offers an alternative paradigm to defining computation. An ML model defines a broad space of possible functions, and we train it by searching within that space for a function that best fits a dataset of example input/output pairs. It is often applied to *supervised* tasks, such as classifying whether or not a photo contains a dog. In such a task, we are training a model to map from the image of a dog situated in the context of a photo to the abstract symbol ‘is a dog’. This symbolic representation might then support a rule-based process such as presenting only photos containing dogs,

in much the same way our formalist capacity to reason with abstractions depends on our embodied capacity to identify those abstractions in the world. The features the model identifies to decide if the photo contains a dog *emerge* from the dataset and are bespoke to the classification problem at hand. We can reason about how it does this through the structural abstractions of neural networks like modules, layers, weights and activation functions, but how these combine to see a dog remains entangled in the full context of the photos in which they appear. Classifiers help us ground abstract symbols in reality. The input is a vague situation and the output is a specific abstraction.

To invert the abstracting process of the classifier would be to define a generative process that takes an abstract symbol such as ‘dog’ and generates an image of a dog. *Inceptionism* (also known as *DeepDream*) is such a model, where an image classifier is used to ‘hallucinate’ the objects the objects it has been trained to recognise.

However, unsupervised generative models, such as a Variational Autoencoder (VAE) [23] or a Generative Adversarial Network (GAN) [24], process data without any symbolic labels. These models are trained to find a *latent representation*, a compressed but smooth representation of samples in a dataset from which they can generate new samples that subjectively appear to belong to that dataset. The space of possible latent representations is called the *latent space*. Currently, GANs are associated among other things with generating photorealistic faces of non-existent people. Sometimes, the model finds a representation that successfully disentangles qualities corresponding to familiar abstractions, such as whether a face is smiling. On discovering such a component, we might label it accordingly, just like we labelled a certain movement of the body ‘the reach’ in the dance studio. We have labelled an emergent quality that remains situated within its context in the latent representation. However, if we now classify or generate images based solely on this quality, then the quality is detached from its context (the other components), and it becomes an abstraction. Emergent qualities may *afford* abstraction but within the latent representation they remain entangled in a manner reminiscent of the situated nature of embodied thinking.

If traditional programming is analogous to formalist thinking and a classifier analogous to our embodied capacity for symbol grounding, then we might consider the generative model to be analogous to embodied thinking, where input influences output without intermediate steps where it is represented through predetermined abstraction. I describe this as an *emergent process*.

This distinction between emergent and abstract representations defines four types of computational process.

- **Formal process:** *abstract*  $\rightarrow$  *abstract*. Traditional programming paradigms. Representations are designed.
- **Abstractive process:** *emergent*  $\rightarrow$  *abstract*. Supervised machine learning (e.g., classification and regression)
- **Reifying process:** *abstract*  $\rightarrow$  *emergent*. Parameterised generative machine learning (e.g., *DeepDream*)
- **Emergent process:** *emergent*  $\rightarrow$  *emergent*. Unsupervised generative machine learning (VAE and GAN), machine translation (seq2seq [25])

Note that this categorisation is concerned solely with how the representations used within a computational process are derived. A formal process may produce outputs in which we perceive emergence, such as Reynolds’s simulation of flocking ‘boids’ [26] or Seevinck’s interactive fractal drawing artwork [27]. In such a case, emergence describes a quality perceived by a human observer. In our case, emergence describes a quality perceived by a computational process.

The key distinction of the emergent process is that input is mapped to output without being constrained by the parameters of an abstract representation preconceived by the system designer. This is somewhat paradoxical as digital data is itself an abstraction of the state of the hardware that is storing it. However, our definition here is concerned with the relationship between input and output rather than the underlying medium upon which they are defined. For example, the pixels of a digital photo are an abstract representation

of light falling on a camera lens. However, within the domain of processes that map collections of photos to new collections of photos, the GAN satisfies the definition of an emergent process because it does not introduce any predetermined abstract representations in how it transforms input to output.

## 2.2. Emergent Interaction

With interaction, instead of simply considering the influence of input on output, we consider the cyclic flow of influence back and forth between human and computer, which may include multiple computational processes. If input is transformed to output purely through emergent processes that avoid a reduction to abstract representation, then we define that as an *emergent interface*. When the human is likewise responding to the output through an embodied mode of thinking, then we can consider the whole activity as *emergent interaction*. Emergent interaction is perhaps a theoretical ideal that cannot be fully attained in a real world system. For example, any sensor is a kind of abstracting process that removes context by design. However, it is an ideal that we can move towards, and it provides a criterion against which we can consider interaction.

An emergent process may be encapsulated within a system where it is controlled by abstract parameters through an interface. For example, even though the GAN itself is an emergent process, our opportunity to interact with it is often reduced to a single trigger that generates an image from a random point in the latent space of a pretrained model. In such a case, the influence from human to GAN does not fit our criteria of an emergent process, so we do not have an emergent interface.

Likewise, supervised machine learning offers a way to define gestural features as emergent qualities of situated gesture. However, this transforms these qualities into abstract parameters and symbols to form its output. We can train a gesture classifier to detect a clap which then triggers output from a new point in a GAN's latent space, but while both input and output are transformed through emergent processes, there remains the discrete trigger as a bottleneck of abstraction in the middle.

To truly leverage our embodied capabilities, such as our sensorimotor skills, we need the nuance and complexity of our input to be reflected in the system's output. However, the extent to which they can be is limited by the extent to which they can be captured within each intermediary representation. If a single intermediary representation is a designed abstraction then the overall interaction is limited to that abstraction's expressive potential, what we might call the *abstraction bottleneck*.

A suitable candidate for embodied emergent interaction are emergent processes trained on human action that might then be used as input to an interaction. Unsupervised generative models are often trained on aggregated datasets of incidental media such as text, photos, paintings and recorded music, but there are a number of examples involving embodied action such as dance. Crnkovic-Friis and Crnkovic-Friis recorded motion capture of a dancer to train a recurrent neural network to generate novel dance sequences [28]. Likewise, Pettee et al.'s *Mirror Exercise* uses a VAE to generate dance sequences, which are then rendered alongside the input sequence creating a fully rendered duet [29]. McDonald worked with Rhizomatiks Research and ELEVENPLAY to create the network 'dance2dance' to generate sequences recorded by a single dancer based on the 'seq2seq' architecture of text translation [25]. These were rendered alongside the dancer in a live performance to create a duet [30]. Here, the influences the model which then influences the dancer, but the cycle does not yet continue into full interaction.

Studio Wayne McGregor collaborated with Google to create *Living Archive* trained on the choreographer's own archive [31]. Google's online version allows someone to dance to their webcam to search the archive for similar movement. However, in a dance performance output from the project, McGregor talks of a 'live dialogue' where the system generates dance in response to the dancers, who are themselves responding to the system. Such a work may fit our definition of emergent interaction. A more detailed analysis would be of interest but is beyond our scope here.



### 2.3. Artistic Explorations

I argued in Section 1.3 that the embodied interaction of an emergent interface is unlikely to realise its value when performing the role of a conventional interface. That is more likely to happen in new forms of digital interaction that it makes possible. Accordingly, the exploratory work that follows is rooted less in comparative analysis with existing interfaces and more in artistic practice. As well as building an interactive system, we are searching for activities where that system might be uniquely valuable. More specifically, we seek to demonstrate the existence of a context and a person such that our system delivers a result unobtainable through other approaches.

The explorations that follow are an artistic collaboration between both authors. Beyond ourselves, the key stakeholders were Creative Scotland, a public funding body with a mission to support audience-facing artwork, and Present Futures, an arts festival where we publicly shared our work. Beyond our commitment to the Present Futures sharing, we were free to be guided by our sense of artistic value.

With an aim to maximise the unique capabilities of an emergent interface, we set out to build a system matching the following criteria:

- **Gestural.** Gesture and vocalisation are both forms of embodied communication familiar to humans. Non-tactile gesture is particularly well suited to multidimensional, ambiguous expression. We work with full body movement.
- **Emergent.** We want to avoid designing what a gesture is as far as possible, and allow this to emerge from how the body already moves.
- **Open ended.** We want to remain open to what emerges rather than trying to make it fit within a existing creative vision.
- **Bespoke.** The entire dataset is recorded of a single individual, moving creatively to expose the full range of their comfortable gestural vocabulary.
- **Small data.** To train our model on a single individual's movement, it needs to be simple enough to be trainable with a dataset small enough to be generated by a single individual. Smaller datasets and simpler models also let us train locally on our own hardware, highlighting the opportunity this approach brings for personal data sovereignty [32].
- **Adaptable.** The data used to train the model are from the same distribution as the data that might be collected when using the model. There is no extra labelling process. This opens the possibility for an interface capable of slowly adapting to the evolving affordances of the person who uses it, a point which we return to below.

### 2.4. An Emergent Mapping Module: The Latent Mapping

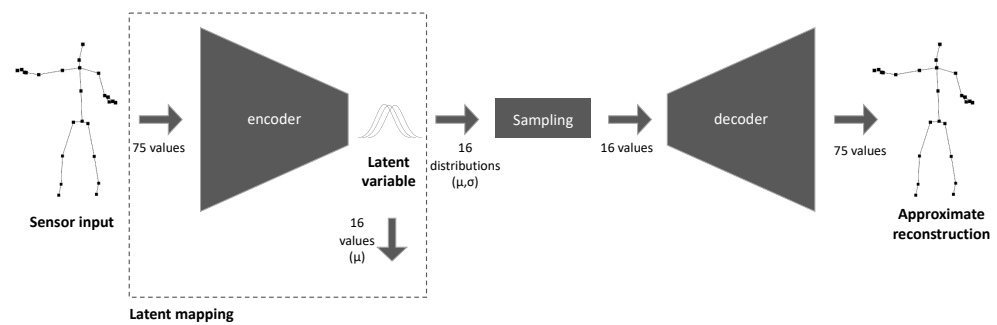
The Variational Autoencoder (VAE) can be seen as a pair of neural networks—an *encoder* and a *decoder*—that are trained together to define its latent representation in the form of a fixed number of Gaussian distributions. The encoder and decoder smoothly map samples to and from this representation respectively. Typically, the latent representation is of lower dimensionality than the input, so the VAE defines a type of lossy compression.

In a recent paper [33], we argue that the encoder of a VAE makes an effective module of a multi-layered mapping within a musical interface, allowing a creator to quickly generate complex and expressive mappings. We described this technique as a *latent mapping* (Figure 2), arguing that it brings qualities beneficial to musical expressivity:

**Consistency.** Similar inputs give similar outputs.

**Diversity.** Dissimilar inputs give dissimilar outputs.

**Range.** The entire output range can be reached by plausible inputs.



**Figure 2.** Overview of a Variational Autoencoder (VAE) with a box showing the subsection re-appropriated as the latent mapping.

Our explorations below demonstrate the feasibility of a single individual recording enough data to train it. In such a case, the mapping will be bespoke. The VAE is trained to define a latent representation that best captures what qualities makes one gesture distinctive *in the context of that individual's gestural vocabulary*.

One cannot predict the latent mapping that will emerge from training a VAE. In fact, training tends to be stochastic, and so retraining tends to lead to a completely different mapping. This makes it appropriate for an open-ended creative process, where we explore possibilities without a specific vision of what we seek, much like the process of choreographic devising we described above, where material is encountered rather than invented.

However, loss of control over the detail of the mapping does not necessarily imply loss of authorship, because the mapping emerges directly from the corpus of recorded gestures used to train the VAE. A bespoke VAE emerges from a person's gestural individuality.

In our own explorations, I recorded 16 h of free improvised movement. This would be a big task for one person if we then needed to segment and label the data for a classifier, or if I was performing specific movements for a preselected label, as in an IML process. However, as we simply need to record gesture representative of my movement vocabulary, I collected the data through a 30 min daily improvisation practice, which is a natural and pleasant addition to my mornings.

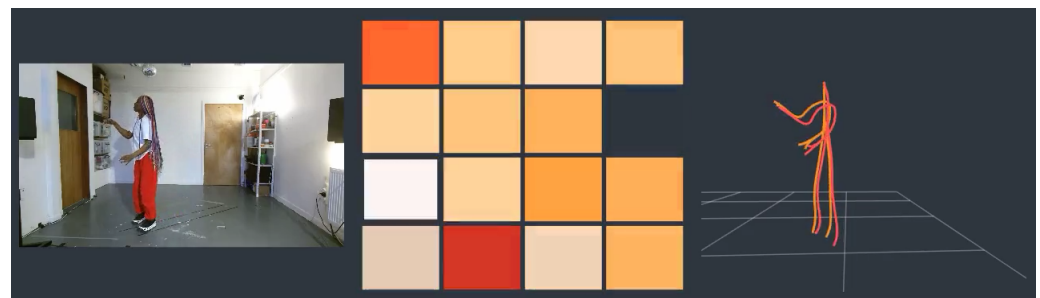
The movement was recorded in a skeletal representation as captured by a Kinect v2 sensor, which represents the body as 25 joints in 3 dimensions, giving frames of 75 components, at 30 Hz. We explored around 12 VAE models, varying the model structure, the dimensionality of the latent space and the loss function. Using an Nvidia GTX 1070, training typically approached convergence after around 20 min. The VAE's loss function captures both the model's ability to accurately reconstruct the skeleton from its latent representation, and regularisation parameters that ensure that the encoder and decoder define smooth functions between the skeletal and latent representations of the body.

We evaluated models by both the technical metrics of the loss function and through practical exploration of how the model feels when hooked up to a synthesiser, the setup of which we explain in more detail below.

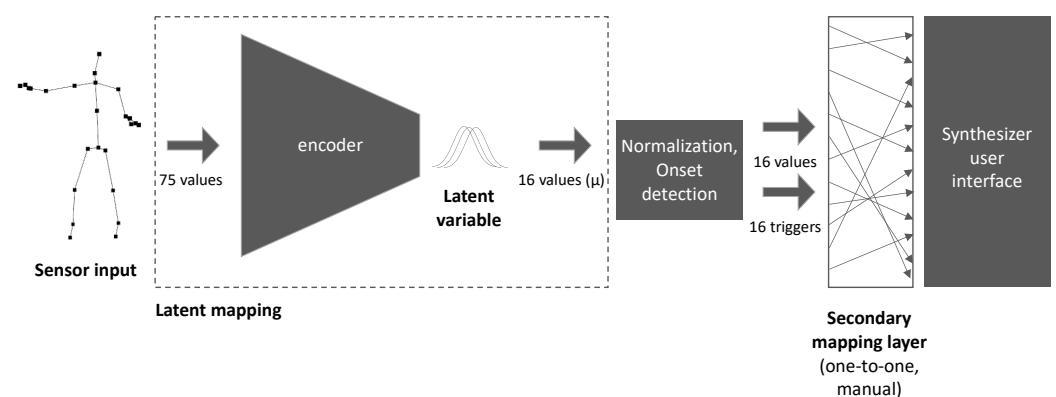
We ended up with two VAEs, both trained with an individual frame as input to define a latent mapping from these 75 dimensions into 16 gaussians.

### 2.5. Sonified Body: Transforming Dance into Sound

For our first artistic exploration with this interface, entitled *Sonified Body*, we mapped the output of the latent mapping (Figure 3) to a range of software synthesisers in Ableton Live by sending Open Sound Control [34] messages to Max for Live patches that controlled the parameters exposed by the synthesiser interfaces (Figure 4). As our focus here was more on uncovering the expressive capabilities of movement were translated, rather than sonic aesthetic, we created this intermediate mapping arbitrarily. By this, I mean we mapped parameters quickly without thinking except to mitigate situations where a single parameter, such as an amplitude control, would effectively negate the effect of all other parameters by silencing the synthesiser.



**Figure 3.** User interface of custom software implementing the latent mapping. The left image shows the RGB camera feed. The right image shows the tracked skeleton in orange, and the skeleton as reconstructed by the VAE decoder in red. The central matrix visualises the current latent mapping (16 values).



**Figure 4.** Data flow schematic of the *Sonified Body* system.

We introduced a number of postprocessing transformations to assist in this process. As there is no innate meaning to the ordering of the latent components, we applied a transformation of Principle Component Analysis (PCA) to the output of the latent mapping. This effectively reorders the components based on their variance, which helped inform our intuition in defining the intermediate mapping.

The components of the latent mapping are unbounded, whereas the inputs to Ableton Live’s software synthesisers are all uniformly bounded. If the range of the latent component is much larger than the input domain of the parameter, then the entire sonic variance may be trapped within a small subset of the total range of possible inputs, so we applied a normalising transformation. However, if we completely match the ranges through a min-max normalisation, then more subtle gestures will have less effect. We tried three approaches, each giving a different character to the interface: unnormalised, min-max and scaling to fit  $\pm 2$  standard deviations.

Finally, the latent mapping implementation outputs continuous data, which maps neatly to drone-like synthesisers with continuous input parameters. However, discrete events such as the start and end of notes are excluded by this. To introduce more percussive events, we introduced a simple onset detection module based loosely on Dahl that detects crossovers between two different low-pass filters [35]. We mapped the onset of each component to trigger a different pitched MIDI ‘note on’ message, which activated different drums in a drum kit instrument. This provided a rudimentary mapping from sudden shifts in velocity of the moving body to note on events, while keeping the representation of velocity within the learnt parameter space of the latent mapping.

Overall, our system still broadly follows Hunt et al.’s three layer mapping model from Figure 1:

- Layer 1, *controller to meaningful*, is now performed by the latent mapping combined with the postprocessing transformations.

- Layer 2, *meaningful to abstract* is our arbitrary attachment to the synthesiser interface parameters.
- Layer 3, *abstract to synthesis* is defined internally by the software synthesiser.

As we feed the latent mapping into the designed abstract parameters of the synthesiser, *Sonified Body* is not a fully emergent interface by our definition above. However, the latent mapping has replaced the preconceived abstractions of ‘meaningful’ parameters with a set of emergent qualities. This allows us to avoid imposing these abstractions onto the body. Instead of thinking of individual limbs modulating cartesian coordinates, our representation of the body is defined by how my body moves. For example, in one of our mappings, the first parameter seems to capture the quality of opening the arms wide versus hugging them against the body. Other parameters are less easily described: one captures movement across the space along the x-axis of the camera, but it is entangled with a lowering of the hips towards the floor. This may sound complicated but this is not an interface you learn about through explanation. It reveals itself through embodied exploration.

### 3. Results

#### *Explorations with Dancers*

To explore the system’s artistic potential, we ran a five day residency working individually with three dancers—Adilso, Divine and Catriona—from a range of improvisatory movement backgrounds (Figure 5). Adilso worked with us before we introduced the onset detection module (due to restrictions imposed by the covid pandemic, there was a two month interruption between Adilso and the other two dancers). We prepared around six software synthesisers and corresponding mappings prior to the residency, and devised a setup letting us quickly swap both model and synthesiser.



Figure 5. *Cont.*



**Figure 5.** Dancers Divine Tasinda, Adilso Machado and Catriona Robertson improvising with the *Sonified Body* system during the exploratory residency.

Once in the studio with each dancer, we gave a brief overview of the tracking capabilities of the Kinect but primarily allowed the dancers to discover the system through exploration. The public outputs we shared from the residency were three recordings of improvisations by the dancers, which you can see at <https://timmb.com/sonified-body>, (accessed on 13 September 2021).

#### 4. Discussion

Following a methodology rooted in artistic practice, we guided ourselves with the above criteria while allowing artistic need and intuition to take over as necessary. As we did so, we kept a tally of any contradictions that seemed to be emerging, as these can reveal nuances to the principles we set out to explore.

The most obvious contradiction is that for all the non-abstract purity of the latent mapping, we introduce a near-random mapping into the carefully designed abstractions of the synthesiser controls. In part, we explain this as being the simplest way to get a feel for the latent mapping. Emergent synthesis will come in the next stage of the project. However, the arbitrariness of this process revealed a desire to minimise introducing our own designs into this process. Where decisions needed to be made, we worked quickly and intuitively to help us avoid thinking through the abstractions of the synthesiser interface. Our efforts to prize emergent qualities made us particularly sensitive to where we were fitting the system around our own assumptions, such as when tuning the filters of the onset detector.

##### 4.1. Generalisability of the Bespoke Interface

The second contradiction is that our latent mapping was trained exclusively on recordings of myself (not a professional dancer), yet we invited professional dancers to explore it as an expressive tool. One of the motivating ideas running through this research is how technology is embedded with the assumptions of its designers. The bespoke emergent interface, created and trained by me, is both a solution to this, if I use it, and an extreme example of this, if others use it. Accordingly, we entered the residency aiming to investigate how generalisable this interface was. If the model is biased towards my vocabulary of movement, would there be a ‘funneling’ effect where the dancers ended up needing to move like me in order to use the system? As it turned out, I believe any such effect was outshone by the funneling effect of the Kinect’s skeleton tracking. With one dancer, for example, we observed a shift from subtle movements of the torso or face towards more obvious and extended limb movements. In spite of this, in our discussions with dancers, each made a point of describing the sense of freedom they felt using the system, particularly in contrast to dancing to recorded music. From our observations, each dancer retained

their own movement identity while seeking out expressive possibilities within the sensing capabilities of the Kinect.

#### 4.2. Generalised Sensing, Bespoke Interpretation

The Kinect skeleton is our next contradiction. Its representation of the body as 25 joints in cartesian coordinates is, of course, a designed abstraction. Our observation of a dancer emphasising extended limb movement suggests their gesture being constrained by an internalisation of this abstraction.

We might see the whole flow from body to sound as follows:

*Body (raw) → Kinect (abstract) → latent representation (emergent) → synthesiser parameters (abstract) → sound (raw).*

Evidently, we are far from a fully emergent interface; yet, the latent mapping does help us 'escape' the representation of the Kinect skeleton by representing the body within a space of plausible poses rather than the full cartesian representation in which any joint could be anywhere. However, information lost through that abstraction cannot be regained.

It is tempting to consider the solution of abandoning the skeletal representation and instead working with a more raw representation such as a 3D point cloud. However, this would require a significantly more complex VAE model, which would require more training data, likely more than one individual could provide.

Our ability to create a bespoke mapping trained on one individual's data depends on the generalising capacity of the Kinect, which is likely trained on data recorded from many individuals. However, a possible resolution to this contradiction is to distinguish between *sensing* and *interpreting* gesture. Sensing determines what our model can see. Interpretation determines how meaning is derived from that. Dividing between sensing and interpreting lets us delegate sensing to a joint enterprise, which will then need to address issues of dataset bias and generalisability. The emergent interface explored here would then serve to make such a general representation applicable to the needs of the individual. If we consider by way of analogy the perception of language, then we might consider the priming of human ears towards hearing speech as sensing but the phonemes of a particular language as interpretation. Sensing can be evaluated objectively over a population of people. Interpretation remains subjective to the individual. Therefore, we want sensing to be generalisable and interpretation to be bespoke.

#### 4.3. Generative Processes in the Studio

Generative algorithms are often used to create open-ended artworks. However, randomness can play a role in many creative processes, even those seeking a fixed outcome. The latent mapping fits well with an open-ended creative process where we respond to the affordances and constraints that we encounter in the interface. This has value for collaborations between creative technology and performative practices like dance because it allows the technologist to respond at the same pace of instant composition that often characterises the creative process of performers (see Section 1.3).

#### 4.4. Emergent Representations Are Complex and Imprecise

Embodied modes of thinking excel at concurrency, context-sensitivity and the complexity of dynamic systems with many interacting elements. They lack explainability and precision, opening the door to ambiguity. Ambiguity can be valuable. A smile can change the meaning of many events in many different ways. It lets us communicate a thousand possibilities without committing to any. Its meaning is situated within its context, which everyone will always experience differently anyway; yet, the action of smiling is far from imprecise. A smile requires the coordination of many muscles, each of which could transform its overall effect. The imprecision lies in how its meaning might generalise from one context to another. The more that meaning remains situated in context, the less precise we can be about it in the abstract.

Consider an interface such as a theremin or Kinect that allows non-tactile gestural control without the person touching the interface. While both of great interest to the experimental music community, they have struggled to get past the gimmick phase. The Kinect has been a great success as a tool for makers and artists but is widely considered a failure as a gaming controller and has been discontinued in this role. The theremin is notoriously difficult to play because it requires the hand to be positioned exactly in space relative to something it cannot touch; yet, in human-to-human interaction, non-tactile gesture is widespread, from hand-waving to conducting an orchestra. One potential reason for this is that untethered gesture is inherently imprecise. Instead of indicating specifics, such gestures are complex (typically non-linear and multidimensional), ambiguous and reliant on contextual clues. However, rather than limiting expression, this ambiguity is fundamental to many forms of human expression and thought. As McGilchrist points out, jokes stop being funny once explained. Poems do not become more meaningful by being made more precise [10]. Perhaps untethered gestural interfaces are simply inappropriate as inputs to a system of abstractions. All the effort of playing the theremin is to control but two parameters: pitch and amplitude.

Emergent interfaces trade precision for complexity. When deployed to control precise abstractions, they will likely be less useful than simple and precise interfaces because the loss of precise control is not compensated for by complexity. The activities that will benefit from this change are those that thrive on complexity and ambiguity of control, such as acts of real-time artistic creation [36] or social connection.

#### 4.5. Beyond Design

The loss of designed abstraction means the loss of our capacity to explicitly design interaction. However, as with the trade-off between precision and complexity, whether or not this is a disadvantage depends on our paradigm of interaction. If we consider the analogy of an interface as a language, then we can observe that some forms of language are designed explicitly on top of abstractions, such as mathematical notation, semaphores and programming languages. Others are emergent, such as natural language, body language and (stretching the analogy) musical melody. Formalist systems such as grammar and music theory are inferred from observations of language and music in the wild, whereas abstract languages are defined in terms of them. This comparison is not exact. For example, English can be used both poetically and with extreme formalism (sometimes described as 'legalese'), and there are cases where previously observed patterns of syntax become rules, such as the standardisation of English spelling in the 18th century. Nonetheless, the analogy illustrates that a lack of explicit design does not hinder our capacity to express ourselves, particularly in activities more dependent upon embodied thinking such as poetry, dance and music. In fact, the hand of a designer in the tools we use to express ourselves can be problematic. Those creating music with software synthesisers are familiar with the trap of presets, which offer a quick path to creating music but can reduce the creator's sense of authorship over their creation. An interface funnels us towards that which is easy to do, what Rokeby describes as its 'operational clichés' [37].

A designed interface is built for a non-specific user, i.e., an abstraction of the person who will eventually interact with it. However benevolent the intentions of the designer, their abstractions will be infused with their assumptions, beliefs and values. The capabilities of such a system is ultimately limited by their imagination and competence. These limits may in turn have a homogenising effect on the user. If individuals cannot perform as the user conceived of by the designer then they are excluded from the opportunities such a system affords. The impact of this has been observed in terms of demographic: the designers of many of our everyday digital systems are not representative of the populations that use them in terms of race or gender, and there are currently movements to increase demographic diversity in the tech industry. While addressing technological design that exacerbates social inequalities may be most urgent, each person thinks and sees the world

differently. The more we rely on designed interfaces to do so, the more our lives are mediated through abstractions based on the assumptions of their designers.

Emergent models do not escape these issues as they can replicate and reinforce prejudicial biases in the data they are trained on. For example, Zou et al. give an example where an automatic translation system needs to choose a gender for a pronoun and resorts to gender stereotypes [38]. However, emergent interfaces also suggest the possibility for a fundamentally different answer to this problem because it becomes possible to tailor the interface to the unique needs and expressive capabilities of an individual. This may be more valuable in tasks where individuality is a priority. Again, this suggests human–human connection and artistic creation are more natural domains for such an interface than activities rooted in formalist thinking such as software development or bureaucracy.

To illustrate this, one dancer from the *Sonified Body* residency told us they found expressive possibilities within the system open up through an exploration of vulnerability. This makes sense as a concept to introduce in a performative act such as dancing or singing. It makes less sense when organising files, drafting contracts, coding algorithms or landing a space rocket. However, it makes crucial sense in a social contexts where we want to build empathy and intimacy.

## 5. Conclusions

*We're in the process of likening ourselves to the machines with which we need to interact.*  
—McGilchrist [39]

The technological feat of computers is made possible through the stack of many layers of designed abstraction, from modulating electrical voltages up to the components of a user interface. Accordingly, the dominant paradigm in HCI is of manipulating abstractions which ultimately are designed to represent contextually situated entities. Emergent interaction offers an alternative approach where unsupervised machine learning allows representations to emerge directly from the situation in which interaction happens.

Exploring this concept through methodology of artistic practice, we have created a system transforming movement into sound that incorporates an emergent representation of the body's pose. We explored this in a residency with three dancers.

Our next step is to extend this system to utilise emergent representations in the output to create a fully emergent process from pose to audio. We plan to combine the latent mapping with an unsupervised generative model as output. This would let us directly connect the two latent spaces, although exactly how we connect them remains an open question. Such a system would be a candidate for a fully emergent system to interpret sensed gesture.

We are also interested in introducing online adaptation to the emergent mapping. Right now our system has adapted to the gestural affordances of the individual at training time. As one might expect, at performance time we observed the individual adapting to the affordances of the system's interpretation. If the system were to respond in turn by updating its model based on what it observes in the individual, then a process of multi-agential conversation emerges.

As we are likely to remain reliant on designed hardware for a while, a fully emergent system is perhaps a theoretical ideal we can move towards rather than implement outright. However, where a computer interprets human action is where computation meets the contextually situated world of human meaning and intention. Here, emergent interaction may help us reduce how far those using computer systems are bound to the assumptions of those who design them and allow us more bespoke interaction with computers. It is likely to hold most value when applied to activities reliant on non-formalist modes of cognition such as physical performance and social connection. To me, a defining quality of such activities is highlighted by McGilchrist: to merge and seek out resonance rather than to abstract and manipulate. In cases where computers are mediators of social interaction then interfaces constructed around resonance, sensing and belonging may open more humane



opportunities for social connection than interfaces designed around the manipulation of digital objects.

**Author Contributions:** Conceptualisation, T.M.-B. and P.T.; methodology, T.M.-B. and P.T.; software, T.M.-B. and P.T.; validation, T.M.-B. and P.T.; formal analysis, T.M.-B. and P.T.; investigation, T.M.-B. and P.T.; resources, T.M.-B. and P.T.; data curation, T.M.-B.; writing—original draft preparation, T.M.-B.; writing—review and editing, T.M.-B. and P.T.; visualisation, T.M.-B. and P.T.; project administration, T.M.-B.; funding acquisition, T.M.-B. All authors have read and agreed to the published version of the manuscript.

**Funding:** *Sonified Body* was funded by Creative Scotland and Present Futures festival with support from Preverbal Studio.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** The dancers invited to the residency provided informed consent for their involvement and inclusion in the published outputs. They were paid for their time at rates above the minimum set by the Independent Theatre Council.

**Data Availability Statement:** The recorded data of the first author, which was used to train the discussed models, is considered to be personal data and the author has chosen not to publish it at this point in time.

**Acknowledgments:** *Sonified Body* features the dancers Adilso Machado, Divine Tasinda and Catriona Robertson. It was mentored by Ghislaine Boddington and produced by Feral. We are grateful to the detailed and considered feedback from the anonymous reviewers.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Gillies, M. Understanding the Role of Interactive Machine Learning in Movement Interaction Design. *ACM Trans. Comput.-Hum. Interact.* **2019**, *26*. [CrossRef]
- Gillies, M.; Kleinsmith, A. Non-representational Interaction Design. In *Contemporary Sensorimotor Theory*; Bishop, J.M.; Martin, A.O., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 201–208. [CrossRef]
- Norman, D.A. Natural user interfaces are not natural. *Interactions* **2010**, *17*, 6–10. [CrossRef]
- Paszkiew, S. Data Acquisition Methods for Human Brain Activity. In *Analysis and Classification of EEG Signals for Brain-Computer Interfaces*; Springer: Cham, Switzerland, 2020; Volume 852, *Studies in Computational Intelligence*, pp. 3–9. [CrossRef]
- Wilson, R.A.; Foglia, L. Embodied Cognition. In *The Stanford Encyclopedia of Philosophy*, Spring 2017 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2017. Available online: <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/> (accessed on 13 September 2021).
- Tanaka, S. The notion of embodied knowledge and its range. *Encyclopaedia* **2013**, *37*, 47–66.
- Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*; MIT Press: Cambridge, MA, USA, 2004. [CrossRef]
- Storkerson, P. Naturalistic Cognition: A Research Paradigm for Human-Centered Design. *J. Res. Pract.* **2010**, *6*, 1–21. Available online: <http://jrp.icaap.org/index.php/jrp/article/view/230> (accessed on 13 September 2021).
- Rajko, J. A call to action: Embodied thinking and human-computer interaction design. In *The Routledge Companion to Media Studies and Digital Humanities*; Routledge: Abingdon-on-Thames, UK, 2018; pp. 195–203. [CrossRef]
- McGilchrist, I. *The Master and His Emissary*; Yale University Press: New Haven, CT, USA, 2009. [CrossRef]
- Kahneman, D. *Thinking, Fast and Slow*; Macmillan: New York City, NY, USA, 2011.
- Adams, A.T.; Costa, J.; Jung, M.F.; Choudhury, T. Mindless Computing: Designing Technologies to Subtly Influence Behavior. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 719–730; UbiComp'15. [CrossRef]
- Hayles, N.K. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*; University of Chicago Press: Chicago, IL, USA, 2000.
- Neuhaus, B. Instant Composition. 2009. Available online: <http://www.bettinaneuhaus.com/english/teacher-instant.htm> (accessed on 30 May 2021).
- Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*, 2nd ed.; Harper & Row: New York, NY, USA, 2002.
- Seevinck, J. *Emergence in Interactive Art*; Springer: Cham, Switzerland, 2017. [CrossRef]
- Fiebrink, R.; Cook, P.R.; Trueman, D. Human model evaluation in interactive supervised learning. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Vancouver, Canada, 7–12 May, 2011; pp. 147–156. [CrossRef]

18. McCallum, L.; Grierson, M.S. Supporting Interactive Machine Learning Approaches to Building Musical Instruments in the Browser. In *Proceedings of the International Conference on New Interfaces for Musical Expression*; Michon, R., Schroeder, F., Eds.; Birmingham City University: Birmingham, UK, 2020; pp. 271–272. [CrossRef]
19. Fiebrink, R.; Trueman, D.; Cook, P.R. A Meta-Instrument for Interactive, On-the-fly Machine Learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, PA, USA, 3–6 June 2009. [CrossRef]
20. Hunt, A.; Wanderley, M.M.; Paradis, M. The importance of parameter mapping in electronic instrument design. *J. New Music Res.* **2003**, *32*, 429–440. [CrossRef]
21. Brown, D.; Nash, C.; Mitchell, T. Simple mappings, expressive movement: a qualitative investigation into the end-user mapping design of experienced mid-air musicians. *Digit. Creat.* **2018**, *29*, 129–148. [CrossRef]
22. Gurevich, M.; Fyans, A.C. Digital Musical Interactions: Performer-system relationships and their perception by spectators. *Organised Sound* **2011**, *16*, 166–175. [CrossRef]
23. Kingma, D.P.; Welling, M. Stochastic gradient VB and the variational auto-encoder. In *Proceedings of the Second International Conference on Learning Representations, ICLR, Banff, AB, Canada, 14–16 April 2014*; Volume 19. Available online: <https://arxiv.org/abs/1312.6114> (accessed on 13 September 2021).
24. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**. arXiv:1406.2661.
25. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In *Proceedings of the Twenty-eighth Conference on Neural Information Processing Systems (NIPS 2014)*, Montreal, Canada, 8–13 December 2014; pp. 3104–3112. Available online: <https://arxiv.org/abs/1409.3215> (accessed on 13 September 2021).
26. Reynolds, C.W. Flocks, Herds and Schools: A Distributed Behavioral Model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, Anaheim, CA, USA, 27–31 July 1987; Association for Computing Machinery: New York, NY, USA, 1987; pp. 25–34; SIGGRAPH '87. [CrossRef]
27. Seevinck, J. Emergence in interactive artistic visualization. *Int. J. Softw. Eng. Knowl. Eng.* **2015**, *25*, 201–230. [CrossRef]
28. Crnkovic-Friis, L.; Crnkovic-Friis, L. Generative Choreography using Deep Learning. In *Proceedings of the International Conference on Computational Creativity*, Paris, France, 27 June–1 July 2016. Available online: <http://arxiv.org/abs/1605.06921> (accessed on 13 September 2021).
29. Pettee, M.; Miret, S.; Majumdar, S.; Nassar, M. CHOREO-GRAPH: Learning latent graph representations of the dancing body. In *Proceedings of the 4th Workshop on Machine Learning for Creativity and Design at NeurIPS 2020*, Vancouver, BA, Canada, 6–12 December 2020.
30. McDonald, K. Dance x Machine Learning: First Steps. 2018. Available online: <https://kcimc.medium.com/discrete-figures-7d9e9c275c47> (accessed on 31 May 2021).
31. McGregor, W. Living Archive. 2018. Available online: <https://waynemcgregor.com/research/living-archive> (accessed on 13 September 2021).
32. Boddington, G. The Internet of Bodies—Alive, connected and collective: the virtual physical future of our bodies and our senses. *AI Soc.* **2021**. [CrossRef] [PubMed]
33. Murray-Browne, T.; Tigas, P. Latent Mappings: Generating Open-Ended Expressive Mappings Using Variational Autoencoders. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2021. Available online: <https://nime.pubpub.org/pub/latent-mappings> (accessed on 13 September 2021). [CrossRef]
34. Wright, M.; Freed, A. Open sound control: A new protocol for communicating with sound synthesizers. In *Proceedings of the International Computer Music Conference (ICMC)*, Thessaloniki, Greece, 25–30 September 1997.
35. Dahl, L. Triggering Sounds From Discrete Air Gestures: What Movement Feature Has the Best Timing? In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Goldsmiths, UK, 30 June–4 July 2014.
36. Benford, S.; Ramchurn, R.; Marshall, J.; Wilson, M.L.; Pike, M.; Martindale, S.; Hazzard, A.; Greenhalgh, C.; Kallionpää, M.; Tennent, P.; Walker, B. Contesting control: Journeys through surrender, self-awareness and looseness of control in embodied interaction. *Hum.-Comput. Interact.* **2020**, *1*–29. [CrossRef]
37. Rokeby, D. The construction of experience: Interface as content. In *Digital Illusion: Entertaining the Future with High Technology*; Dodsworth, C., Jr., Ed.; ACM Press: New York, NY, USA, 1998; Chapter 2, pp. 27–48.
38. Zou, J.; Schiebinger, L. Design AI so that it's fair. *Nature* **2018**, *559*, 324–326. [CrossRef] [PubMed]
39. McGilchrist, I. Have We Lost Our Senses. Invited Lecture at The Open Senses Symposium; Institute of Philosophy, University of London: London, UK, 2017.