Towards a Machine Somaesthete: Latent Modeling of EMG Signals in Viola Playing

Lucy Strauss
Goldsmiths, University of London
United Kingdom
University of Johannesburg
South Africa
lstra003@gold.ac.uk

Matthew Yee-King
Goldsmiths, University of London
United Kingdom

ABSTRACT

Somaesthetic experiential qualities can provide a window into process of meaning-making, both human and machinic. We draw such qualities from viola performance into the design-in-progress of a novel interactive performance system. In doing so, we introduce the concept of a *Machine Somaesthete* that senses and makes sense of these qualities from a second-person perspective. Our system comprises electromyographic (EMG) muscle sensing and a Variational Autoencoder. With a novel dataset, we aim to encode latent representations of performance movement that are meaningful from a somaesthetic perspective. We present our model and our design process, then analyse latent trajectories to interrogate how our system can be considered a Machine Somaesthete, and the nature of its sensitivity to bodily experiences of viola playing. At the intersection of artificial intelligence, music performance and intra-action design, we take a sympoietic (together-making) view of knowledge creation. We and our practices are transformed as we design - and design with - machine learning systems.

CCS CONCEPTS

• **Applied computing → Performing arts**; • **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies → Machine learning**.

KEYWORDS

Soma-design, Latent Modelling, Machine Learning, Sympoiesis, Machine Somaesthete, Interactive Music Systems

ACM Reference Format:
Lucy Strauss and Matthew Yee-King. 2024. Towards a Machine Somaesthete: Latent Modeling of EMG Signals in Viola Playing. In *9th International Conference on Movement and Computing (MOCO '24), May 30–June 02, 2024, Utrecht, Netherlands.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3658852.3659074

MOCO '24, May 30–June 02, 2024, Utrecht, Netherlands

ACM ISBN 979-8-4007-0994-4/24/05
https://doi.org/10.1145/3658852.3659074

1 INTRODUCTION

In both the Sciences [43] [26] [76] [60] [39] and the Humanities [31] [66] [5], the lines between agents as separate entities are increasingly fuzzy. Concepts like 'holobiont' [76] and 'symborg' have become useful to refer to entities that are simultaneously one and multiple[1].

In this light, the separate *body* | *interface* | *world* categories of traditional phenomenology [47], embodied cognition [69] and embodied interaction design [15] have become slippery concepts. We find use in the notion of 'intra-action', where agency and creation are collective phenomena:

> *"We do not obtain knowledge by standing outside of the world; we know because 'we' are of the world. We are part of the world in its differential becoming."* [5]

In an inter- or intra-active performance system (IPS) context, we are part of the system in its differential becoming. Whereas embodied cognition brought forth notions of autopoietic (self-creating) systems [69], our intra-action design approach is more aligned with notions of sympoiesis (together-creation).[2] Through this lens, we acknowledge that we both create and create *with* our design. We and our practices emerge through design, just as our designs are shaped and created [14] [8].

This work echoes similar lines of inquiry regarding co-creativity in sound music and computing [70] and HCI [23]. We seek to answer how intra-action design can be a sympoietic process whereby the designer, the user and the design create and emerge together. To this model, we introduce a concept: *The Machine Somaesthete* that senses and makes sense of aesthetic bodily experiences within an intra-active system. Through this paper, we demonstrate how the use of the Machine Somaesthete ideal can deepen knowledge and understanding through sympoietic design of and with machine learning (ML) systems.

We use a soma-design [32] [34] approach within the Research through Design (RtD) framework [77]. Through first-person methods, our design process is informed by the aesthetic bodily experience of the designer. We share our process in the hope that it can be extensible to other designers.

In tackling these questions and concepts, we zoom into the sense-making processes of acoustic instrument performers. We invite a

[1] In terms of biological organisms, lichen are a fantastic example of holobionts as coupled organisms comprising fungi and algae.

[2] Enactivists [45] first proposed enactivism as an embodied cognition that is compatable with computing. However, recent embodied cognition discourse acknowledges that the individual-centred approach to 4E cognition no longer aligns with sympoiesis and symbiosis in biology and ecology research. There are efforts find compatibility between these two views [56] [51] [22] [13] [6], yet there is still work to be done.

ML model into the first-person somaesthetic experiences of a violist, with the aim that the model itself becomes a 'somaesthete' with a second-person perspective of the performer's sensory information. This Machine Somaesthete comprises a data-capture system and a Variational Autoencoder (VAE) ML architecture. We captured our own training dataset of electromyographic (EMG) data, recording the bioelectric signals of muscle tension. We capture these signals from the body of the first author, a musician with advanced training in viola performance. This human-machine sympoiesis lays the foundation for future work on an IPS.

This paper also documents our attempts to answer questions including: RQ.1) Can we create a VAE architecture with latent space organisation that is meaningful in terms of somaesthetics? RQ.2) Is our trained model's machinic sense-making any more meaningful than an untrained network? RQ.3) How will we know if the latent space meets this ideal? RQ.4) What bodily aspects of viola playing is our Machine Somaesthete sensitive to?

In tackling these questions, we explore the model's inferencing capacities with time-series analysis in latent space. We discuss our results through a somaesthetic lens, grounded in viola performance practice.

2 BACKGROUND

Our term *Machine Somaesthete* is a nod to seminal interactive music system designs with two components: a machine listener that makes sense of incoming music data, and a machine performer that outputs MIDI or audio signals based on the computational sense-making (information retrieval) process [40][54][55][10]. While movement can be considered in music information retrieval [30], our performance system (in progress) is not sensitive to audio or semantic representations of music. We term our system's information retrieval component a Machine Somaesthete. We present this term with the view that it can be used as a design ideal, not as a measure by which to evaluate designs. Likewise, we make no claims that our model can actually experience or perceive; it is only a model. We call our system a Machine Somaesthete because we *intend* for it to be sensitive to and make sense of somaesthetic experiences within a sympoietic system that comprises human performer, viola and machine.

Presented in this paper, our Machine Somaesthete is the information retrieval component of our IPS. The insights gained here will inform the next design stage of our IPS 'performer' component, which will eventually output sound. This foundational work deepens our understanding of our model's particular machinic interpretation of a particular somaesthetic experience (introduced in 2.1).

2.1 Somaesthetics of Viola Performance

A core notion behind the soma-design approach (designing with the body) is that "meaning-making processes start in movement and bodily realities" [32] with reference to [59]. This movement-centred view of meaning-making is central to how we draw embodied knowledge from performance practice into our design process.

Movement information from bowed string playing can be captured using accelerometers, gyroscopes [62][16][58] and even strain sensors [74] on the instrument (or bow). With these sensors, we can retrieve information about the mechanics of sound production on the instrument. There have also been efforts to retrieve mechanical information from audio signals [50][49]. We can also retrieval biomechanical information of string players' bodies with camera sensors [62][68] and mocap systems [16][36][72][27].

Our aim to capture data representations of somaesthetic qualities is distinct from the capture of mechanical and biomechanical information in movement computing. Experiential qualities of movement are not movement itself. To engage with viola practice in this way, we turn inward. Our design centers on a particular space of experiential qualities within the somaesthetic possibilities of viola playing.

In bowed-string instrument performance, players have continuous fine-grain connections with sound production. Even when playing through seemingly sharp and angular actions on the instrument (such as a change of bow direction), we (violists)[3] experience and intend these actions as rounded movement trajectories in the body. This helps us to use the inertia of our past movement in the present moment, without expending unnecessary effort. The clear lines between semantic music theoretical elements like notes and chords are fuzzy experiences in the body. Likewise, rounded, springy bodily movements and sensations help us to connect phrases of musical tension and release across notes and sub-phrases. To move in this way requires a great deal of mental imagery and somaesthetic awareness. It is as much a movement practice as a somaesthetic sensation. The Machine Somaesthete senses and makes sense of this somaesthetic quality within the sympoietic system.

Informed by this movement practice of viola performance, we (the researchers) conceptualise performance movement as a continuous space of possibilities. This framing of movement echoes sensorimotor enactivist ideas [48], the approach to movement by other string players in movement computing [65], and is further applicable to co-creative music composition practices [28]. This line of thinking speaks to our previous research, where we consider how we can inscribe IPS with embodied knowledge of music [61]. Now, our computational model zooms in on the performer's embodied experience of their own continuous space of movement possibilities.

2.2 Sensing Muscles

Bioelectricity is essential to bodily sense-making processes in music performing, during which performers interpret sensory information towards meaningful understanding of their bodily movement. This information travels through their nervous system in the form of bioelectric signals. We focus on sensory information relating to muscular tension, inviting our Machine Somaesthete to sense from the performer's bioelectric network using surface EMG sensors.

We consider our computational model to have a second-person perspective of the violist's somaesthetic experience. The machinic sense-making process acts in parallel to the performer's own bodily process, towards sympoiesis. Our choice of EMG to explore bodily awareness in music builds on related human computer interaction (HCI) work of others. The use of EMG in HCI studies exploring the second-person perspective is of interest [24]. Furthermore, the

[3]The first author is an insider to the viola performance practice community, so it is appropriate to use 'we' when writing about violists as a group.

second-person perspectives can be particularly well-suited in collaborative design, connecting shared knowledge and experience [33][21]. The joining of first-person and second-person perspectives is an applicable conceptual and practical apparatus for sympoietic system design.

There are many usages of EMG in sound and movement computing. Our work builds from similar usages of EMG with sound-producing movement of acoustic and electric[4] music instrument performance practices. Most closely related usages of EMG in terms of bowed string instrument performance practice are IPS. EMG has been used in IPS with string instruments to track EMG signals of the left forearm [64]. Such data can provide information about finger movements of the left hand and wrist. Unlike our Machine Somaesthete, this IPS [64] does not utilize ML.

EMG has also been used in an IPS to track signals on the right side of a cellist's body [41], with a similar aim to ground the design process in embodied knowledge from instrument performance [41]. Also related are systems where instrument playing, EMG sensing and ML intersect. Although not involving bowed string playing, there are examples with guitar that use these three elements to predict sound from action [19][17][18].

EMG sensing and bowed string playing most notably intersect in studies of musicians' health [3] [42], or in pedagogical research [53]. Although our IPS application is rather different, this literature has informed our design choices regarding electrode placement, as we describe in 3.1.

2.3 Organisation in Machinic Poiesis

With a sympoietic view of system design, we aim to model poietic processes computationally with various shades of autonomy, towards distribution of information and control between human and machinic intra-actors.

Sympoietic systems have been defined in contrast to autopoietic systems [12]. In our research, we do not follow these ideas and distinctions as a rigid framework; our purpose with sympoiesis is not taxonomical. Rather, we apply these definitions and distinctions somewhat metaphorically, using sympoiesis as a lens to enrich practice and research involving humans and ML in art and design. However, we do take particular inspiration from one characteristic distinction: whereas autopoietic systems are organisationally-closed [45], sympoietic systems are organisationally-ajar [12]. This means that a system's internal organisation is determined by both internal and external sources.

In computing, idea of autonomous organisation is inscribed into the architecture design of unsupervised ML model architectures like autoencoders. These models organise compact representations of training data into a latent space of reduced dimensions using feature extraction, rather than training labels. This enables us to use statistically-driven, as opposed to hand-derived features in training. As computational models, autoencoders do not actually fit the criteria of autopoietic systems; they are created and maintained by developers and trained on external data. Considering that their structural coupling is both internal and external in this sense, sympoiesis is a more fitting conceptual apparatus.

[4]Electric guitars and violins are not acoustic instruments, but also not digital musical instruments.

Already with a standard autoencoder, we can model elements of ajar organisation into our Machine Somaesthete. Yet classic autoencoders are limited in their generative capacity because training samples are mapped to single points in latent space. The spaces between mapped points in latent space do not always make sense when we sample and decode them in generative tasks. This is at odds with our conceptualisation of a continuous space of possibilities of movement in viola performance 2.1. Variational Autoencoders (VAEs) solve this problem by encoding training data as a probabilistic distribution. Rather than than encoding points in latent space, we encode vectors [38]. This probabilistic latent space can be continuous[5]. Just as semantic music theoretical elements like notes and chords are fuzzy experiences in the soma of a violist, latent encodings are vectors in latent space.

2.4 Variational Autoencoders

Given the interdisciplinary of this paper, we provide a description of VAE training. VAEs are trained in two parts: a forward training pass and a backward validation pass. In the forward pass, we pass training data through the encoder to attain latent embeddings that describe a probable location in latent space. We combine these embeddings with a noise source (often Gaussian) to sample a single point in latent space. We feed this sample through the decoder to generate a reconstruction of the original input data. In the backward training pass, we use a loss function to measure the error between the original input and the reconstructions, as well as the similarity between the latent space shape and a Gaussian (standard normal) distribution.

To achieve a probabalistic distribution of latent space, the loss function of a VAE is a sum of two components. Firstly, we aim for VAE latent spaces to have as close to multivariate Gaussian probability distribution as possible.[6] Kullback–Leibler divergence (KLD) measures the similarity between the current latent space distribution and a desirable, Gaussian distribution. Secondly, we use measure the difference between the sample reconstructed by the model and the input sample. For this, we typically use either mean square error (MSE) or BCE loss functions, depending on whether our data is valued continuous or discrete. Based on sum of these two metrics (KLD + reconstruction loss), the model weights are updated for the next forward training pass.

VAEs were first introduced a decade ago [38] and are still used within a great variety of generative ML architectures in a variety of creative domains due to their potential to generate meaningful content not found in the training set. In creative domains, VAEs enable us to generate "fantastical images and unheard sounds" [73]. Given this potential for modularity, a this model is a good starting point for our IPS.

In sound and music computing, VAEs have been trained on semantic music representations of musical structure [52], rhythm [71] and harmony [9][67]. Although we do not train our model on audio data, both EMG and sound signals are complex signals. Given this similarity, many of the signal processing techniques for audio are applicable to EMG. Pre-processing and audio feature

[5]Or quantized as is the case with VQ-VAEs.

[6]This is what enables us to sample from the latent space to generate new content that the VAE was not trained on. Gaussian distribution is also essential for generating smooth interpolations.

extraction techniques of audio training data for VAEs have included Constant-Q Transform in the wavelet domain [63] and multiband decomposition of the signal [7]. Both of these VAEs address timbral control of sound synthesis.

3 MAKING A MACHINE SOMAESTHETE

With the RtD framework in mind [77], we document our design and analysis processes with the aim that these procedures are extensible to other designers. In particular, our methods are relevant to soma-designers using quantitative measures of somaesthetic experience through data and ML. Our methods provide one possible solution for how we can "shape these tools to allow for interactions harmonizing with our somas" [32].

Our Machine Somaesthete comprises a sensing system and a machine learning model. We use the sensing system to capture a training dataset for our model of EMG data from the performer's body. Our Machine Somaesthete is one component of an intra-active music performance system, as we described in 2.

3.1 Slowstorming with Muscles of Viola Playing

Towards future live performance, we aim for an economical (non-cumbersome) technical set-up that can capture somaesthetically-relevant information from the violist's body. We sought to use a relatively number of EMG sensors, with intentional and informed sensor placement appropriate to our aims. This is to minimise a cumbersome set-up in performance settings.

Our sensor electrode placement is informed by a combination of literature on EMG studies with string players [3][42][1], string instrument pedagogy [35], performance psychology techniques for effective practice [20], as well as the violist's advanced training. With a selection of 'starting-point' electrode placements drawn from literature and prior experience, we tested these options with and away from the EMG data-capture system.

In viola performance practice, we (violists) attempt to produce a rich sound with little tension in the body. We experience 'hanging' with rather than 'pushing' the bow into the strings. We work with with gravity and force in this way not only to preserve stamina in performance and avoid injury, but also for a breadth of musical expression. We aim to be efficient with overall bodily effort by initiating movements from the larger muscle groups, rather than holding tension in the extremities. For this reason, we (the researchers) were interested from the outset to focus on the back, side and arm muscles of the right side of the body. This is the side of the body primarily responsible for bowing movements.

The violist made use of soma-slowstorming, acting and enacting performance movements with their attention drawn to specific areas of their soma. These sessions served two purposes: 1) to inform design choices relating to sensor placement; 2) to hone the violist/designer's attunement to the particular experiential qualities of this soma-design. The violist made sure to play freely as a warm-up exercise and during slowstorming sessions. In this way they practiced falling in and out of the familiar, and exploring somaesthetic experiential qualities of viola playing on multiple timescales. They also pressed against their own muscles with their hands to feel the muscular activity beneath their skin, and watched the movements of their back muscles using parallel mirrors and video for a second-person perspective of their own body. This also helped to disrupt the familiar. The violist moved back and forth between their mind, their body, their instrument, and the data capture system in various combinations. Through this iterative process, we tried and tested sensor placements.

For example, we had initially thought to capture a particular movement called 'repull' with electrode placements on the upper back around the scapula (shoulder blade). Repull denotes a subtle downward action, thought to originate in the core of the body to subtly darken the timbre of sound. Expert violists describe the action as involving rotation of the scapula [35]. We were unable to see any noticeable amplitude changes in EMG signals with electrode placements near the scapula when the violist performed repull. However, we did seem to reliably track repull with electrodes lower than the scapula and more on the side of the body than the back. It is possible that the violist plays with imperfect repull technique; or that the muscles around the scapula associated with repull are too deep in the body for surface electrodes to access.

Through this iterative process informed by practice and using soma-based techniques, we selected the four electrode placements for our dataset: the forearm, deltoid, upper back and side, all on the right side of the body.

3.2 EMG Data Capture

We captured four time-aligned channels of EMG data at a sample rate of 22050Hz using Plux BITalino EMG sensors[7] with a Bela board[8] [46]. In standard electromyography practice, signals typically undergo bandpass filtering with cutoff frequencies of 20Hz and 400-600Hz [44], making 1200Hz the maximum Nyquist rate needed to capture detail in EMG signals. Although the signal rate capabilities of Bela are unnecessary for capturing EMG data alone, it is ideal for capturing time-aligned datasets of audio and EMG data in hard real-time, as well as minimizing latency in intra-active music performance. Bela can record audio at 44100Hz, exactly double the sample rate of analog recording for easy multi-modal data alignment. These benefits serve the longevity and relevance of our dataset and model towards future work on this project.

The violist improvised freely over two recording sessions, making use of both 'standard' and 'extended' playing techniques[9]. These 37 minutes of recorded material formed our training dataset of raw EMG signals in viola playing. We left the dataset unedited, including the short moments where the violist lowered their instrument (for example, to make adjustments).

3.3 Signal Processing

We prepared the dataset for training by applying signal processing window-by-window, with a window size of 22050 samples (1 second of EMG data). Given the small size of our dataset, we used a window overlap of 80%. To each channel within each window, we applied

[7]https://www.pluxbiosignals.com/products/electromyography-emg-sensor
[8]https://learn.bela.io/products/bela-boards/bela/
[9]'Standard' techniques typically include plucking the strings with one finger from the right hand and moving the bow hair across the strings. 'Extended' techniques might include tapping on the body of the instrument with the hands, or bouncing the wood of the bow on the strings. We use inverted commas because these distinctions from western classical music have become rather conservative in a free improvisation context.

a 50Hz Butterworth notch filter to remove power line noise as is common practice [11][44]. We then applied a fast continuous wavelet transform (fCWT) [4].

Wavelet transforms are effective at both denoising bioelectric signals like electroencephalographic and EMG signals, and extracting features for predictive tasks [2][75]. EMG signals detected with surface electrodes[10] comprise a sum of electrical activities within a collection of muscle fibers that lie below the electrode. The wavelet domain gives us rich insight into non-stationary features from this sum of multiple signals. Likewise, the fCWT algorithm has applications in movement computing [25]. We chose specifically to use the fCWT algorithm because it is supported in C++ and Python languages. This allows us flexibility to perform signal processing in C++ on Bela or in Python on a laptop in future design iterations and performance settings. We use the absolute fCWT coefficient values rather than complex values to discard phase information, then resize and stack fCWT channels for each window.

3.4 VAE Design & Training

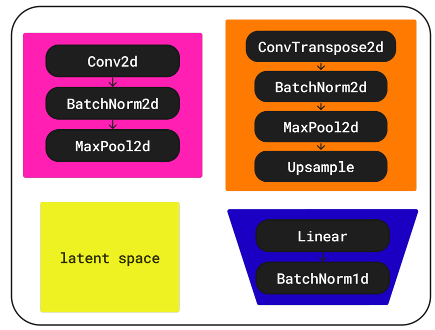
Our model architecture emerged over numerous iterations, starting with a simple VAE architecture [38]. We made incremental adjustments towards a model that could learn effectively from our novel EMG dataset. For each design iteration, we plotted training and validation losses during training runs and tested the model's generative capabilities through reconstructions and interpolations[11]. This visual feedback provided insight into latent space organisation during training. We arrived at our current model after it could provide accurate reconstructions and smooth interpolations. In future work, we will address disentanglement to make our latent space as small as possible (in terms of the number of dimensions, while maintaining its overall performance capabilities.

Our Machine Somaesthete VAE architecture (depicted in 1) includes an encoder with three convolutional layers, each followed by pooling and batch normalization layers. The encoder further reduces the input dimensions (400x400) to 50 latent dimensions through fully-connected layers representing mean, log variance, and latent space. The decoder mirrors the encoder with transpose convolutions and upsampling layers to reconstruct the original input (400x400) from the 50-dimensional latent space. We use two-dimensional convolutional layers throughout the to encourage the model to learn spatial dependencies between channels.

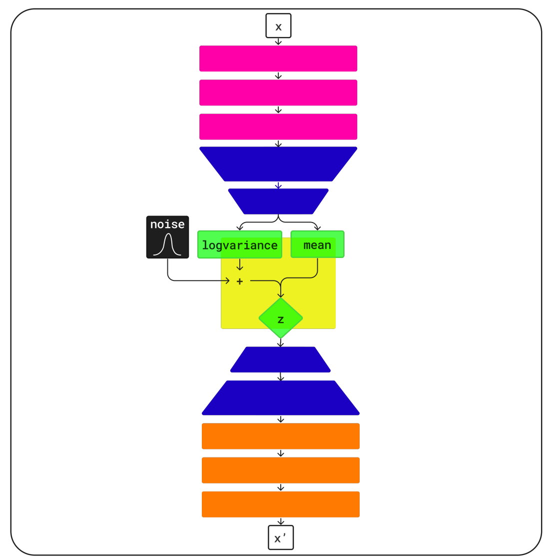
In training, we use a batch size of 24 and Xavier Glorot weights initialisation [29]. We use Adam optimization [37] with a learning rate of 1×10^{-4}. We used a loss function summing KL divergence and MSE loss. As we described in 2.4, KL divergence encourages regularises the model to learn a Gaussian distribution of latent space and MSE is an appropriate reconstruction loss function for continuous-valued training data. Our data is windowed in the time domain, but fCWT coefficients are on a continuous-valued scale.

4 MAKING SENSE OF LATENT SENSE-MAKING

Our analysis aims to understand how we can know if the latent space is somaesthetically-meaningful (RQ.3). We seek to deepen understanding of machinic sense-making by studying meaningfullness of latent space sampling in relation to application-specific somaesthetic experience. Towards the creation of an IPS, we aim for our Machine Somaesthete to eventually generate similar outputs for EMG inputs relating to similar somaesthetic experiences (and different outputs relating to different experiences) within the space of possibilities of described in 2.1. We are most curious to deepen understanding of the model's internal sense-making processes. For



(a) Key for understanding network architecture (b).



(b) Network architecture with colour-coded elements as shown in the key (a).

Figure 1: A graphical representation of our Machine Somaesthete model architecture. Each block in (b) contains all the layers depicted in its colour-corresponding block in the key (a). Not depicted, we use ReLu activations throughout the network, except for the connection between the last layer of the decoder and the reconstruction.

[10] In the medical field, practitioners use needle electrodes that can detect action potentials (electrical impulses that make up EMG signals) from individual muscle fibers.

[11] We provide a video of interpolations generated with our network architecture anonymized

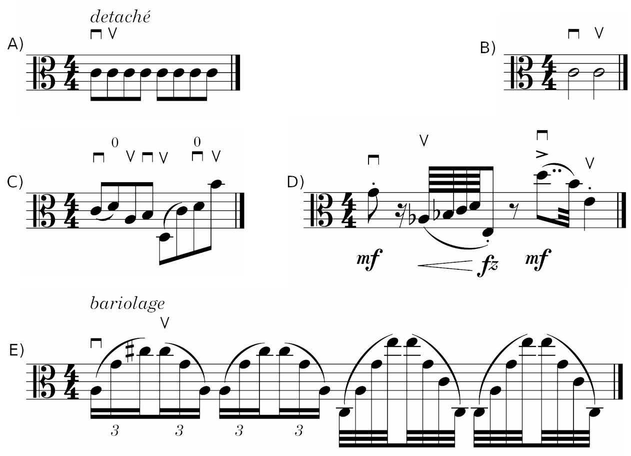


Figure 2: The five phrase categories in music staff-notation, labeled alphabetically. The violist played each of these phrases five times at 60 quaver beats per minute, with as little difference in expression between repetitions as possible.

this reason, we explore the model's generative capacities in latent space.

First, we made EMG data recordings from the violist's body as they played various musical phrases of our own devising 2. We then encoded these phrases into the latent space of our trained VAE to produce *latent trajectories*: concatenations of latent samples (z) generated from input samples of our EMG recordings.

We use dynamic time warp (DTW) to, comparing each latent trajectory to all others, including itself. We obtain a DTW a distance measure for each pair comparison and present a distance matrix of all of these measures through a heat-map 4. In the event that our model has successfully, we expect to observe clusters of lower dissimilarity measures where we compare latent trajectories from the same phrase category. Likewise, we expect to observe higher dissimilarity where we compare latent trajectories generated from different phrase categories. We show hierarchical clusters in the DTW distance matrix using a dendrogram 4.

To confirm that our observations are the result of training. we run the same analysis on latent trajectories generated using an untrained model (with the same network architecture as the trained model). We also show the DTW distance matrix resulting from our untrained model with a heatmap and a dendrogram.

Our choice to use time-series analysis in latent space is unusual, given that our VAE has no sequential connections. In the simplest technical terms, our VAE is trained to reconstruct multichannel EMG data at the window level. We choose this approach because our model will be used in music performance settings, where time is an intrinsic element. Our analysis is particular to our specific model, within a specific application, for use with a specific human body. Through this approach, we acknowledge the individuality and diversity of bodies and instruments. We discuss 5 observations from this analysis from a somaesthetic perspective, informed by viola performance practice.

4.1 Generating Latent Trajectories

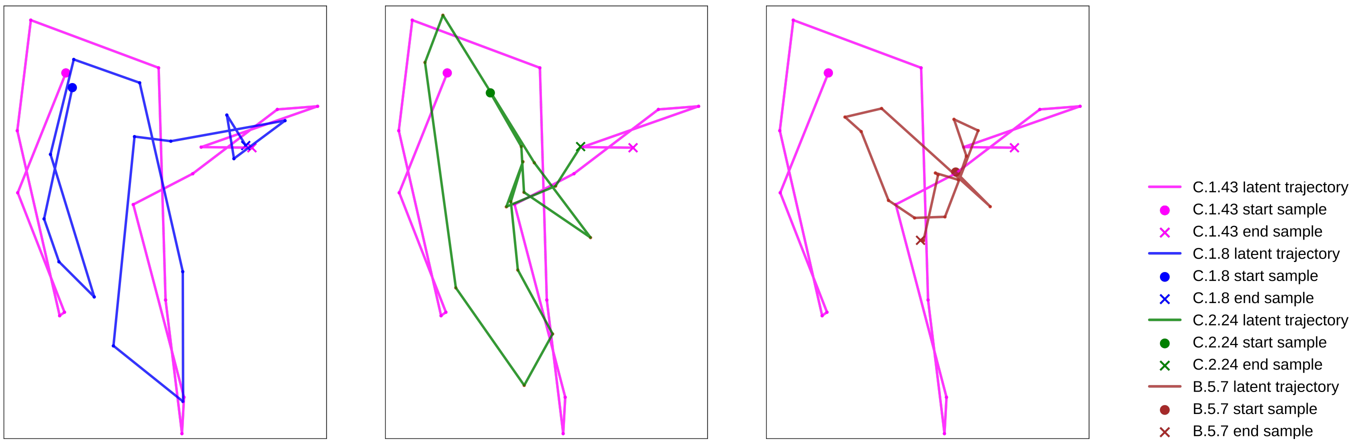
We explore the latent space organisation of our trained Machine Somaesthete through its ability to generate new samples in latent space, based on a series of EMG data recordings (inputs)[12]. We devised a collection of five musical phrases of equal lengths as shown in 2. These phrases are our five input categories. Each phrase is somaesthetically-distinct from the others (from the performer's perspective).

The violist performed each phrase five times, under the instruction that they should try to repeat each phrase as precisely as possible, with as few expressive differences between repetitions as possible. They played in time to a metronome and were able to listen back to previous repetitions. We used the same sensor placement as with our training dataset recording, and likewise recorded four channels of time-aligned EMG data 3.3. Each EMG data recording begins at the start of the first metronome beat of each phrase and ends at the end of the last beat of each phrase, to include the entire duration of the first and last beats. With five repetitions of each of the five phrase categories, we ultimately attained twenty-five biodata recordings for encoding and analysis[13].

We processed the raw 4-channel EMG recordings just as we did the training data to produce stacked 1-second-long fCWT windows with 80% overlap. These stacked windows are our input samples. We encoded each recording sample-by-sample to attain latent embedding pairs (mean and standard deviation) from each input recording. From these latent embeddings, we use reparamatization to obtain a sample (z) in latent space, as is standard practice in VAE generative tasks. We then concatenated these latent samples to form (latent) time-series. We refer to these latent time-series as latent trajectories. For each recording, we generated 5 different latent trajectories. In this way, we are able to incorporate both human and computational variations of somaesthetic qualities in our analysis. Both are relevant to our sympoietic system. We will refer to the entire collection of latent trajectories as our analysis dataset.[14] We categorise this set of EMG recordings by phrase (A to E) and take/repetition of each phrase (1 - 5) and generation (1 - 5).[15] In total, we generated 125 latent trajectories for analysis.

In preparation for the DTW analysis step, we visually familiarised ourselves with the analysis dataset. We used principal component analysis (PCA) to reduce dimensions from 50 (the number of dimensions in our model's latent space) to 2. We can already make some promising observations from the latent trajectory plots made using the trained model 3. We can see that the latent trajectory shapes generated from the same EMG input recording look

[12]While our model's learning is unsupervised in that we do not impose gesture labels during training, we do use semantic categories of movement (inputs) during analysis. The model never has access to category labels; these are for use to organise and refer to during our analysis.
[13]We have made these EMG recordings available for download. The link to the dataset can be found through our GitHub repository: https://github.com/lucystrauss/machine-somaesthete
[14]Not to be confused with the dataset we made to train our VAE.
[15]For example, we refer to the first generated trajectory from the first recording of phrase category A as: A.1.1. Likewise, we refer to the forth generated trajectory from the fifth recording of phrase category D as: D.5.4.



(a) Latent trajectories generated from the same EMG recording.

(b) Latent trajectories generated from different EMG recordings of the violist playing the same phrase.

(c) Latent trajectories generated from EMG recordings of the violist playing different phrases.

Figure 3: Latent trajectories (concatenations of latent samples over time) in a 2D plot after PCA dimensionality reduction. We show three sets of comparisons with latent trajectory C.1.43, the forty-third latent trajectory that we generated by inputting the first EMG recording of phrase category C into our trained VAE. We can observe that the pairs of latent trajectories look most similar to each other in (a) and most different to each other in (c).

the most similar to each other, and less similar when generated from different EMG input recordings from the same phrase category. Latent trajectory shapes look most different from each other when they are generated from different phrase categories. Drawing our attention to latent trajectory plots made using an untrained model, we do not make any meaningful observations about latent trajectory shape comparisons.

4.2 Time-series Analysis of Latent Trajectories

We use the fast dynamic time warp (FastDTW) algorithm [57] to compare each latent trajectory to all the others.[16] A popular choice in movement computing, DTW allows for temporal misalignment between latent trajectories. Specifically, the optimization technique of FastDTW makes efficient calculations with our multidimensional time-series of latent samples. For each latent sample, there are 50 dimensions. Within DTW framework, we use the correlation distance metric[17] to place more emphasis on temporal patterns and less on magnitude differences between sequences compared to alternative metrics (eg. Euclidean). FastDTW calculates distance measures for each dimension at a time, then aggregates these into one final distance measure for each comparison. With 125 latent trajectories, we produce a total of 7750 final distance measures for each distance matrix. As we stated in 4, we run the same analysis twice, once using our trained model and once using an untrained model.

5 DISCUSSION OF OBSERVATIONS

The heat-maps and dendrograms 4 result from the DTW analysis of latent trajectories. We include all 125 latent trajectories in the DTW calculations for the heat-maps. For the dendrograms, our DTW distance matrix is made using one sample from each phrase category recording so that we can see all the labels in the figure clearly.

For 4 (c) & (d), the latent trajectories were generated using an untrained model. We observe no meaningful groupings in these visualisations, aside from the black squares in 4 (c). These simply show where a latent trajectory is compared to itself, resulting in a distance measure of 0. Given this lack of meaningful groupings in an untrained model, we are confident that the patterns and clusters we see in 4 (a) & (b) are a result of successful model training. Given that we can see patterns and groupings in 4 (a) & (b), we are confident that the latent space organisation of our trained network is more meaningful than the latent space of our untrained network (RQ.2).

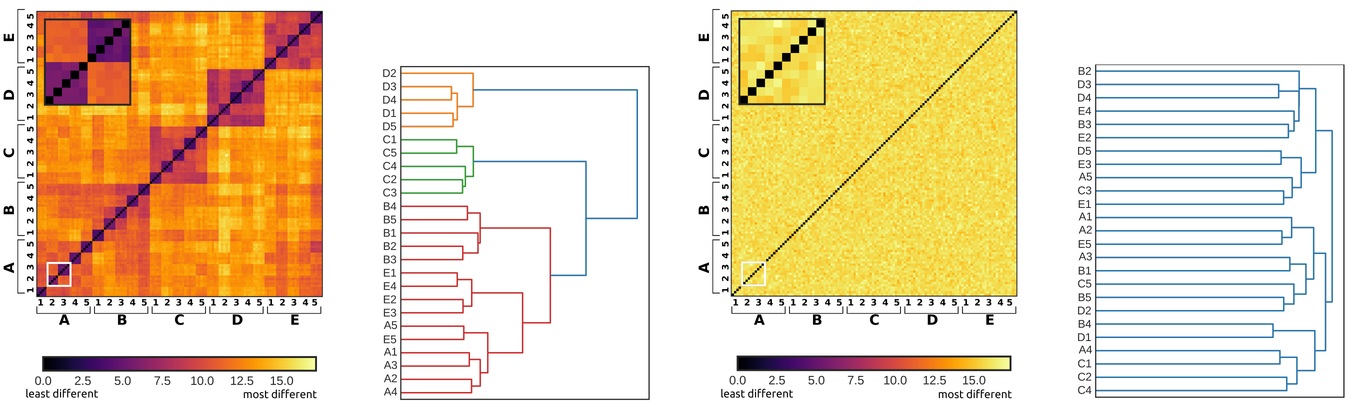
We discuss the visualisations produced with our trained model 4 (a) & (b) henceforth. From a somaesthetic perspective and informed by viola performance practice, we share our observations of these graphical representations and discuss how they can reveal elements of the model's sense-making process. Through our discussion, we make sense of machinic sense-making - in sympoiesis with the Machine Somaesthete, we make sense.

5.1 Ambiguity Between Phrases A & E

Most strikingly, we notice a relatively-low DTW distance between categories A and E. We can see this in both the heat-map 4(a) and the dendrogram 4(b). In the heat-map, this is particularly apparent between latent trajectories A5 and E5. In the dendrogram, E5 is grouped more closely with other examples from phrase category A than with the other E samples.

If we were to evaluate our model's performance without a somaesthetic perspective, we might say that it has generated incorrect latent trajectories for inputs A5 and E5. Yet our model is not

[16] We used the Python implementation of FastDTW: https://pypi.org/project/fastdtw/
[17] We used the correlation distance implementation from the SciPy library: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate.html



(a) Heat-map from trained model. **(b) Dendrogram from trained model.** **(c) Heat-map from untrained model.** **(d) Dendrogram from untrained model**

Figure 4: Heat-maps showing DTW distance matrix between 125 latent trajectories produced with our tra.ined model (a) and an untrained model (c). Labeled are 5 latent trajectories (concatenations of z) from each of the 5 recordings (played by the violist) of each of the 5 phrase categories (A to E). We see black squares (difference measure of zero) a latent trajectory is compared to itself. Medium-sized dark purple squares occur at comparisons of different latent trajectories generated from the same input. Larger, lighter purple squares have occurred where we compare latent trajectories generated using different EMG inputs to the model from the same phrase category.

intended as a gesture classifier. Its ambiguous understanding of categories A and E tells us something about how the VAE makes sense of experiential qualities of viola playing.

Phrase categories A and E sound and look very different in staff notation. In 2 we can see that phrase A is multiple repetitions of the same note with repetitive rhythm values, all with the detaché[18] bow stroke. Phrase E contains much smaller note values and each quaver (eighth note) beat contains multiple pitches under a single bow stroke[19].

However, the violist's bodily experience of playing these two phrases is similar. Even though the two phrases require different bowing techniques, the bow directions of A and E are the same for each quaver beat. The player does not experience changes in bow stroke direction as *up, down, up, down.* As we wrote in 2.1, we experience even seemingly angular movements as rounded and circular in the body. To change bow directions smoothly (without a break in sound-production), the violist extends the circular experience throughout their soma, even nodding their head as they reach the ends of the bow between strokes.

With the similarity of bow direction in mind, it is also important to note that the model has mostly generated the 'correctly'-grouped latent trajectories for A and E recordings. We have described in a fair amount of detail how similar A and E feel to play, but it is worth reminding that the model usually does not 'confuse' these two phrase categories. There is a distinction to be made between A and E. At least on the right side of the body, the distinction is the magnitude of the circular experiences of bow-change. The space of possibilities of movement is continuous and sometimes, somaesthetic qualities overlap as experiential circles spin into other places of the soma.

Our DTW analysis has revealed to us something about how our model makes sense of the somaesthetic space of possibilities of viola performance. The model is sensitive to the circular movement patterns of bow changes, even when they occur in different magnitudes or on different strings. The violist was aware of the similar shades of experiential qualities between these two phrases. We intentionally devised these two phrases because they are in some ways very similar and in other ways rather different. Yet we were unaware of the extent of the similarities before, or that the similarities are expressed so strongly in the particular parts of the body that the Machine Somaesthete can sense. These are machinic insights that the violist will feed back into their viola practice.

5.2 Similarity Between A & B

There is a closeness between phrase categories A and B. After the connection between A & E phrase groupings on the dendrogram 4 (b), B is the very next group to be linked. The relatively similarity between A and B is most apparent among the latent trajectories generated using the last few recordings of each category (A3, A4, A5 & B3, B4, B5) in the heat-map 4 (a).

From a sound standpoint, we might say that phrases A and B are most similar to each other. They both entail playing only one note - the same note on the same string - for their entire durations, albeit with different rhythms 2. Biomechanically, this requires the violist to hold their right arm at the same height to reach the second-lowest string of the viola without touching adjacent strings with the bow. In terms of muscular sensation, the violist needs to feel a similar balance of tensions in their body to be able to hold their arm at the same level for the two phrases.

[18]Playing on the string in the upper-mid half section of the bow without lifting the bow between bow strokes.
[19]This is the bariolage bowing technique.

Yet the difference in rhythm and bow direction moves these phrases further apart in the space of somaesthetic possibilities. Given the similarity between phrase categories A and E 5.1, we now know that the Machine Somaesthete is sensitive to the circular qualities of bow direction changes. With insights drawn from 4 (b) about the hierarchy of distance amongst A, B and E, we can see that the model is usually slightly more sensitive to which string the player plays on than to bow changes. This heirarchy is only inverted for E5. These observations provide insights about the bodily aspects that our Machine Somaesthete is sensitive to (RQ.4)

5.3 Distinctness of Phrase D

Category D looks the most distinct from other categories, as shown through the darkest purple shading of category comparisons on the heat-map 4 (a) and the right-most tree connection on the dendrogram (b). Even from a semantic point of view, this is unsurprising. As we notate in 2, D is the only phrase category for which there is not a cyclical repetition of rhythm. Musically, it is the most tempestuous out of all the phrases, with flurries of tension and release both rhythmically and harmonically. These flurries are reflected in the violist's soma as they play.

5.4 Organisation and Poiesis

Given the compatibility between the observed groupings (from the DTW analysis of latent trajectories) and bodily experiential qualities of viola playing, we can understand the latent space of the trained model as somaesthetically-meaningful (RQ.1).

In 4 we can see the impact of training on the model by comparing (a) with (c). All we can see is noise in the heat-map of latent trajectories generated by an untrained model 4 (c). In making these trajectories, the untrained model still makes what sense it can from inputs, retrieving mean and standard deviation from input samples, reparameterizing them with a standard normal noise source to produce new latent samples. The organisational machinic poietic process has begun, even though latent sampling is meaningless to us at this stage. Through training on our dataset, we and the model begin to make sense together.

Towards a sympoietic view of ML, we consider the historical trajectory of the dynamic organising process before our data reaches the VAE. Our data is situated in viola performance practice and pedagogy. These practices emerge over hundreds of years, as knowledge is passed down and transformed from generation to generation. Over decades of training and practice, our particular violist learns the coordinated movement of viola performance. They hold this embodied knowledge of pattern and coordination in their body. These patterns inherently exist in the dataset before the model begins to 'auto'-encode. Our sympoietic system is organisationally ajar and structurally coupled [12] to the EMG signals of viola playing from a specific body. With this view, the dynamic organising process of sympoiesis comprises at least two streams of sense-making connected through electromyography: 1) performance training of the violist, and 2) the unsupervised training of the VAE.

6 CONCLUSION

We have shown that a simple VAE model with convolutional layers can encode a somaesthetically-meaningful latent space in terms of viola performance. In line with our practice-oriented goals, this model is trained on a dataset small enough to capture from one musician in under an hour, using only four EMG sensors. We have also demonstrated that we can encode this meaningful latent space with data from the larger muscle groups further from the body's extremities. To complement these contributions, the code for our VAE model is available on GitHub[20]. Although our training dataset itself is not shared, we do provide our analysis dataset and its raw recordings on Zenodo. In our next design iteration, we will investigate disentanglement of our latent space and approaches to multimodality of movement and sound. These technical focuses will build towards generative performer component of our IPS.

As a foundation for our technical contributions, we introduced a concept of particular relevance to movement computing, sound and music computing, soma-design and HCI communities: *The Machine Somaesthete*. This RtD serves as an exemplar, showing how the use of this concept can deepen knowledge and understanding across multiple areas through the RtD process. In addition to inspiring our design process, the somaesthetic lens has lead to our unusual approach to latent space analysis. We have used latent listening trajectories to peer into the sense-making process of a Machine Somaesthete. In future work, our methods could be used to analyse the musical performance of a human musical agent who plays within the system, by analysing how they explore the latent space.

Our research questions do not articulate problems to solve, but rather open points of entry to creative work with Machine Somaesthetes. Although we addressed each question in the paper, they still hold potential to deepen our understanding of future design iterations. Interweaving between all of this work is the notion of sympoiesis, with the idea that we are both one with and separate to our creations. This ambiguity enriches our design process as we create and emerge together with our intra-active systems.

ACKNOWLEDGMENTS

This research is supported by the Arts and Humanities Research Council of the United Kingdom and the Consortium for the Humanities and the Arts South-East England.

REFERENCES

[1] Babak Afsharipour, Francesco Petracca, Mauro Gasparini, and Roberto Merletti. 2016. Spatial distribution of surface EMG on trapezius and lumbar muscles of violin and cello players in single note playing. *Journal of Electromyography and Kinesiology* 31 (Dec. 2016), 144–153. https://doi.org/10.1016/j.jelekin.2016.10.003

[2] Majid Aljalal, Saeed A. Aldosari, Marta Molinas, Khalil AlSharabi, and Fahd A. Alturki. 2022. Detection of Parkinson's disease from EEG signals using discrete wavelet transform, different entropy measures, and machine learning techniques. *Scientific Reports* 12, 1 (Dec. 2022), 22547. https://doi.org/10.1038/s41598-022-26644-7 Number: 1 Publisher: Nature Publishing Group.

[3] Emma Allingham, Birgitta Burger, and Clemens Wöllner. 2021. Motor performance in violin bowing: Effects of attentional focus on acoustical, physiological and physical parameters of a sound-producing action. *Journal of New Music Research* 50, 5 (Oct. 2021), 428–446. https://doi.org/10.1080/09298215.2021.1978506 Publisher: Routledge _eprint: https://doi.org/10.1080/09298215.2021.1978506.

[4] Lukas P. A. Arts and Egon L. van den Broek. 2022. The fast continuous wavelet transformation (fCWT) for real-time, high-quality, noise-resistant time–frequency analysis. *Nature Computational Science* 2, 1 (Jan. 2022), 47–58. https://doi.org/10.1038/s43588-021-00183-z Number: 1 Publisher: Nature Publishing Group.

[5] Karen Barad. 2003. Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs* 28, 3 (2003), 801–831. https://doi.org/10.1086/345321 Publisher: The University of Chicago Press.

[20] https://github.com/lucystrauss/machine-somaesthete

[6] Nick Brancazio. 2023. Why Not Both? Interaction and the Individual in Enactivism. *Constructivist Foundations* 18, 3 (2023), 380–382. https://constructivist.info/18/3/380 Number: 3 Publisher: Alexander Riegler.

[7] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. https://doi.org/10.48550/arXiv.2111.05011 arXiv:2111.05011 [cs, eess].

[8] Yves Candau, Jules Françoise, Sarah Fdili Alaoui, and Thecla Schiphorst. 2017. Cultivating kinaesthetic awareness through interaction: Perspectives from somatic practices and embodied cognition. In *Proceedings of the 4th International Conference on Movement Computing*. ACM, London United Kingdom, 1–8. https://doi.org/10.1145/3077981.3078042

[9] Nádia Carvalho and Gilberto Bernardes. 2023. Exploring Latent Spaces of Tonal Music using Variational Autoencoders. *AIMC 2023* (Aug. 2023). https://aimc2023.pubpub.org/pub/latent-spaces-tonal-music/release/1

[10] Nick Collins. 2011. LL: LISTENING AND LEARNING IN AN INTERACTIVE IMPROVISATION SYSTEM. (2011), 7.

[11] Carlo J. De Luca, L. Donald Gilmore, Mikhail Kuznetsov, and Serge H. Roy. 2010. Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *Journal of Biomechanics* 43, 8 (May 2010), 1573–1579. https://doi.org/10.1016/j.jbiomech.2010.01.027

[12] Beth Dempster. 2007. Boundarylessness: Introducing a Systems Heuristic for Conceptualizing Complexity. In *Nature's Edge: Boundary Explorations in Ecological Theory and Practice*, Charles S. Brown and Ted Toadvine (Eds.). State University of New York Press, 93–107.

[13] Mads Julian Dengsø and Michael David Kirchhoff. 2023. Beyond Individual-Centred 4E Cognition: Systems Biology and Sympoiesis. *Constructivist Foundations* 18, 3 (2023), 351–364. https://constructivist.info/18/3/351 Number: 3 Publisher: Alexander Riegler.

[14] Audrey Desjardins and Ron Wakkary. 2016. Living In A Prototype. In *Proceedings of the 2016 CHI*. https://dl.acm.org/doi/abs/10.1145/2858036.2858261?casa_token=g3Hqgzuch_IAAAAA:zG8rgfYg3ZKkFIF741W3ieXTr6dq3bQNujDgmdY0_CDjfRED9dOIVDrUv5u9GJZXzertkn2jEA

[15] Paul Dourish. 2001. *Where the Action is: The Foundations of Embodied Interaction*. MIT Press. Google-Books-ID: DCIy2zxrCqcC.

[16] Vincenzo D'Amato, Erica Volta, Luca Oneto, Gualtiero Volpe, Antonio Camurri, and Davide Anguita. 2021. Accuracy and Intrusiveness in Data-Driven Violin Players Skill Levels Prediction: MOCAP Against MYO Against KINECT. In *Advances in Computational Intelligence (Lecture Notes in Computer Science)*, Ignacio Rojas, Gonzalo Joya, and Andreu Català (Eds.). Springer International Publishing, Cham, 367–379. https://doi.org/10.1007/978-3-030-85099-9_30

[17] Cagri Erdem, Qichao Lan, Julian Fuhrer, Charles Patrick Martin, Jim Tørresen, and Alexander Refsum Jensenius. 2020. Towards Playing in the 'Air': Modeling Motion-Sound Energy Relationships in Electric Guitar Performance Using Deep Neural Networks. *978-88-945415-0-2* (2020), 177–184. https://www.duo.uio.no/handle/10852/79392 Accepted: 2020-09-15T18:05:18Z Publisher: Axea sas/SMC Network.

[18] Cagri Erdem, Benedikte Wallace, and Alexander Refsum Jensenius. 2022. CAVI: A Coadaptive Audiovisual Instrument–Composition. In *NIME 2022*. PubPub, The University of Auckland, New Zealand. https://doi.org/10.21428/92fbeb44.803c24dd

[19] Çağrı Erdem, Qichao Lan, and Alexander Refsum Jensenius. 2020. Exploring relationships between effort, motion, and sound in new musical instruments. *Human Technology* 16, 3 (Nov. 2020), 310–347. https://doi.org/10.17011/ht/urn.202011256767 Num Pages: 310-347 Place: Jyväskylä, Poland Publisher: Centre of Sociological Research (NGO) Section: Articles.

[20] K. Anders Ericsson. 2020. Towards a science of the acquisition of expert performance in sports: Clarifying the differences between deliberate practice and other types of practice. *Journal of Sports Sciences* 38, 2 (Jan. 2020), 159–176. https://doi.org/10.1080/02640414.2019.1688618 Publisher: Routledge _eprint: https://doi.org/10.1080/02640414.2019.1688618.

[21] Sarah Fdili Alaoui, Thecla Schiphorst, Shannon Cuykendall, Kristin Carlson, Karen Studd, and Karen Bradley. 2015. Strategies for Embodied Design: The Value and Challenges of Observing Movement. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (C&C '15)*. Association for Computing Machinery, New York, NY, USA, 121–130. https://doi.org/10.1145/2757226.2757238

[22] James Rupert Fletcher. 2023. Cognitivism ageing: The Alzheimer conundrum as switched ontology & the potential for a new materialist dementia. *Journal of Aging Studies* 66 (Sept. 2023), 101155. https://doi.org/10.1016/j.jaging.2023.101155

[23] Raune Frankjaer and Peter Dalsgaard. 2020. Knowledge-creation Processes in Crafts-based HCI Research: Introducing a Sympoietic Framework. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3419249.3420114

[24] Jules Françoise, Yves Candau, Sarah Fdili Alaoui, and Thecla Schiphorst. 2017. Designing for Kinesthetic Awareness: Revealing User Experiences through Second-Person Inquiry. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5171–5183. https://doi.org/10.1145/3025453.3025714

[25] Jules Françoise, Gabriel Meseguer-Brocal, and Frédéric Bevilacqua. 2022. Movement Analysis and Decomposition with the Continuous Wavelet Transform. In *Proceedings of the 8th International Conference on Movement and Computing*. ACM, Chicago IL USA, 1–13. https://doi.org/10.1145/3537972.3537998

[26] Scott F. Gilbert, Jan Sapp, and Alfred I. Tauber. 2012. A Symbiotic View of Life: We Have Never Been Individuals. *The Quarterly Review of Biology* 87, 4 (Dec. 2012), 325–341. https://doi.org/10.1086/668166 Publisher: The University of Chicago Press.

[27] Marco Gillies, Harry Brenton, Matthew Yee-King, Andreu Grimalt-Reynes, and Mark d'Inverno. 2015. Sketches vs skeletons: video annotation can capture what motion capture cannot. In *Proceedings of the 2nd International Workshop on Movement and Computing (MOCO '15)*. Association for Computing Machinery, New York, NY, USA, 104–111. https://doi.org/10.1145/2790994.2790995

[28] Artemi-Maria Gioti. 2020. From Artificial to Extended Intelligence in Music Composition. *Organised Sound* 25, 1 (April 2020), 25–32. https://doi.org/10.1017/S1355771819000438 Publisher: Cambridge University Press.

[29] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 249–256. https://proceedings.mlr.press/v9/glorot10a.html ISSN: 1938-7228.

[30] Rolf Inge Godøy and Alexander Refsum Jensenius. 2009. Body Movement in Music Information Retrieval. https://doi.org/10.5281/zenodo.1416668 ISBN: 9780981353708 Pages: 45-50 Publication Title: Proceedings of the 10th International Society for Music Information Retrieval Conference Publisher: ISMIR.

[31] Donna J. Haraway. 2016. *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press. https://doi.org/10.2307/j.ctv11cw25q

[32] Kristina Höök. 2018. *Designing with the Body: Somaesthetic Interaction Design*. MIT Press. Google-Books-ID: 9oZ0DwAAQBAJ.

[33] Kristina Höök, Steve Benford, Paul Tennent, Vasiliki Tsaknaki, Miquel Alfaras, Juan Martinez Avila, Christine Li, Joseph Marshall, Claudia Daudén Roquet, Pedro Sanches, Anna Ståhl, Muhammad Umair, Charles Windlin, and Feng Zhou. 2021. Unpacking Non-Dualistic Design: The Soma Design Case. *ACM Transactions on Computer-Human Interaction* 28, 6 (Nov. 2021), 40:1–40:36. https://doi.org/10.1145/3462448

[34] Kristina Höök, Baptiste Caramiaux, Cumhur Erkut, Jodi Forlizzi, Nassrin Hajinejad, Michael Haller, Caroline C. M. Hummels, Katherine Isbister, Martin Jonsson, George Khut, Lian Loke, Danielle Lottridge, Patrizia Marti, Edward Melcer, Florian Floyd Müller, Marianne Graves Petersen, Thecla Schiphorst, Elena Márquez Segura, Anna Ståhl, Dag Svanæs, Jakob Tholander, and Helena Tobiasson. 2018. Embracing First-Person Perspectives in Soma-Based Design. *Informatics* 5, 1 (March 2018), 8. https://doi.org/10.3390/informatics5010008 Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[35] J Irvine, K Kashkashian, M LaCourse, L Ramsey, K Ritscher, and C Rodland. 2020. *The Karen Tuttle Legacy*. Carl Fischer Music. https://www.carlfischer.com/bf116-the+karen+tuttle+legacy.html

[36] Kelly Jakubowski, Tuomas Eerola, Paolo Alborno, Gualtiero Volpe, Antonio Camurri, and Martin Clayton. 2017. Extracting Coarse Body Movements from Video in Music Performance: A Comparison of Automated Computer Vision Techniques with Motion Capture Data. *Frontiers in Digital Humanities* 4 (2017). https://www.frontiersin.org/articles/10.3389/fdigh.2017.00009

[37] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arXiv.1412.6980 arXiv:1412.6980 [cs].

[38] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. https://doi.org/10.48550/arXiv.1312.6114 arXiv:1312.6114 [cs, stat].

[39] Myungin Lee. 2021. Entangled: A Multi-Modal, Multi-User Interactive Instrument in Virtual 3D Space Using the Smartphone for Gesture Control. In *International Conference on New Interfaces for Musical Expression*. https://doi.org/10.21428/92fbeb44.eae7c23f

[40] George E. Lewis. 2000. Too Many Notes: Computers, Complexity and Culture in "Voyager". *Leonardo Music Journal* 10 (2000), 33–39. https://www.jstor.org/stable/1513376 Publisher: The MIT Press.

[41] Rikard Lindell and Tomas Kumlin. 2017. Augmented Embodied Performance – Extended Artistic Room, Enacted Teacher, and Humanisation of Technology. https://doi.org/10.5281/zenodo.1176177 Pages: 77-82 Publication Title: Proceedings of the International Conference on New Interfaces for Musical Expression Publisher: Zenodo.

[42] Stephanie Mann, Mads Bruun Panduro, Helene M. Paarup, Lars Brandt, and Karen Søgaard. 2021. Surface electromyography of forearm and shoulder muscles during violin playing. *Journal of Electromyography and Kinesiology* 56 (Feb. 2021), 102491. https://doi.org/10.1016/j.jelekin.2020.102491

[43] Lynn Margulis. 2008. *Symbiotic planet: a new look at evolution*. Basic books.

[44] Radek Martinek, Martina Ladrova, Michaela Sidikova, Rene Jaros, Khosrow Behbehani, Radana Kahankova, and Aleksandra Kawala-Sterniuk. 2021. Advanced Bioelectrical Signal Processing Methods: Past, Present, and Future Approach—Part III: Other Biosignals. *Sensors* 21, 18 (Jan. 2021), 6064. https://doi.org/10.3390/s21186064 Number: 18 Publisher: Multidisciplinary Digital

Publishing Institute.

[45] H. R. Maturana and F. J. Varela. 1991. *Autopoiesis and Cognition: The Realization of the Living*. Springer Science & Business Media. Google-Books-ID: nVmcN9Ja68kC.

[46] Andrew McPherson and Victor Zappi. 2015. An Environment for Submillisecond-Latency Audio and Sensor Processing on BeagleBone Black. Audio Engineering Society. https://www.aes.org/e-lib/browse.cfm?elib=17755

[47] Maurice Merleau-Ponty, Donald Landes, Taylor Carman, and Claude Lefort. 2011. *Phenomenology of Perception*. Routledge, London. https://doi.org/10.4324/9780203720714

[48] Alva Noë. 2004. *Action in Perception*. MIT Press. Google-Books-ID: kFKvU2hPhxEC.

[49] Alfonso Perez-Carrillo and Hendrik Purwins. 2017. Estimation of violin bowing features from Audio recordings with Convolutional Networks. In *Conference and Workshop on Neural Information Processing Systems (NIPS): Machine Learning for Audio Signal Processing*. https://scholar.archive.org/work/etlr4xno7zdsdb4mq6hwicqsca/access/wayback/http://media.aau.dk/smc/wp-content/uploads/2018/01/ML4AudioNIPS17_paper_38.pdf

[50] Alfonso Perez-Carrillo and Marcelo M. Wanderley. 2015. Indirect Acquisition of Violin Instrumental Controls from Audio Signal with Hidden Markov Models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 5 (May 2015), 932–940. https://doi.org/10.1109/TASLP.2015.2410140 Conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[51] Timotej Prosen. 2023. Life-Mind Continuity Beyond Individual-Centeredness: The Issue of Adaptivity. *Constructivist Foundations* 18, 3 (2023), 367–370. https://constructivist.info/18/3/367 Number: 3 Publisher: Alexander Riegler.

[52] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. 2018. A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. In *Proceedings of the 35th International Conference on Machine Learning*. PMLR, 4364–4373. https://proceedings.mlr.press/v80/roberts18a.html ISSN: 2640-3498.

[53] Carl Peter Robinson, Baihua Li, Qinggang Meng, and Matthew T.G. Pain. 2017. Pattern Classification of Hand Movements using Time Domain Features of Electromyography. In *Proceedings of the 4th International Conference on Movement Computing (MOCO '17)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3077981.3078031

[54] Robert Rowe. 1992. Machine Listening and Composing with Cypher. *Computer Music Journal* 16, 1 (1992), 43–63. https://doi.org/10.2307/3680494 Publisher: The MIT Press.

[55] Robert Rowe. 2004. *Machine Musicianship*. MIT Press. Google-Books-ID: kSKC1QQarvwC.

[56] Carl B. Sachs. 2023. Is Sympoiesis Compatible with Phenomenology? *Constructivist Foundations* 18, 3 (2023), 376–378. https://constructivist.info/18/3/376 Number: 3 Publisher: Alexander Riegler.

[57] Stan Salvador and Philip Chan. 2007. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* 11, 5 (Jan. 2007), 561–580. https://doi.org/10.3233/IDA-2007-11508 Publisher: IOS Press.

[58] Margaret Schedel and Rebecca Fiebrink. 2011. A Demonstration of Bow Articulation Recognition with Wekinator and K-Bow. *Proceedings of the 2011 International Computer Music Conference* (2011), 5.

[59] Maxine Sheets-Johnstone. 2011. *The Primacy of Movement*. John Benjamins Publishing. Google-Books-ID: 2EDgXzWMfuwC.

[60] Merlin Sheldrake. 2021. *Entangled life: How fungi make our worlds, change our minds & shape our futures*. Random House Trade Paperbacks.

[61] Lucy Strauss and Matthew Yee-King. 2023. Extensible Embodied Knowledge: Bridging Performance Practice and Intelligent Performance System Design. In *AIMC 2023*. https://research.gold.ac.uk/id/eprint/34056/

[62] Shih-Wei Sun, Bao-Yun Liu, and Pao-Chi Chang. 2020. Deep Learning-Based Violin Bowing Action Recognition. *Sensors* 20, 20 (Jan. 2020), 5732. https://doi.org/10.3390/s20205732 Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[63] Kıvanç Tatar, Daniel Bisig, and Philippe Pasquier. 2021. Latent Timbre Synthesis. *Neural Computing and Applications* 33, 1 (Jan. 2021), 67–84. https://doi.org/10.1007/s00521-020-05424-2

[64] Seth Dominicus Thorn. 2019. Transference: A Hybrid Computational System for Improvised Violin Performance. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '19)*. Association for Computing Machinery, New York, NY, USA, 541–546. https://doi.org/10.1145/3294109.3301254

[65] Seth Dominicus Thorn. 2020. Hybrid Violin Performance: Model-Free, Abductive Experiment. In *Proceedings of the 7th International Conference on Movement and Computing (MOCO '20)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3401956.3404251

[66] Anna Lowenhaupt Tsing. 2021. *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Princeton University Press. Google-Books-ID: MR4OEAAAQBAJ.

[67] Aiko Uemura and Tetsuro Kitahara. 2022. Morphing-based Reharmonization with VAE: Reducing Dissonance with Consonance-based Loss Function. https://doi.org/10.5281/zenodo.7088426 Publication Title: Proceedings of the 3rd Conference on AI Music Creativity Publisher: AIMC.

[68] Zacharias Vamvakousis, Alfonso Perez, and Rafael Ramírez. 2018. Acquisition of violin instrumental gestures using an infrared depth camera. In *Georgaki A, Andreopoulou A, editors. 15th Sound and Music Computing Conference Sonic Crossings; 2018 Jul 4-7; Limassol, Cyprus. Limassol: Cyprus University of Technology, 2018*. Cyprus University of Technology. https://repositori.upf.edu/handle/10230/43227

[69] F Varela, E Thompson, and E Rosch. 1991. *The Embodied Mind*. MIT Press, Cambridge, MA. http://mitpress.mit.edu/9780262720212/the-embodied-mind/

[70] Craig Vear. 2022. Embodied AI and Musicking Robotics. In *The Language of Creative AI: Practices, Aesthetics and Structures*, Craig Vear and Fabrizio Poltronieri (Eds.). Springer International Publishing, Cham, 113–135. https://doi.org/10.1007/978-3-031-10960-7_7

[71] Gabriel Vigliensoni, Louis McCallum, Esteban Maestre, and Rebecca Fiebrink. 2022. R-VAE: Live latent space drum rhythm generation from minimal-size datasets. *Journal of Creative Music Systems* 1, 1 (Aug. 2022). https://doi.org/10.5920/jcms.902 Number: 1 Publisher: Huddersfield University Press.

[72] Erica Volta, Maurizio Mancini, Giovanna Varni, and Gualtiero Volpe. 2018. Automatically measuring biomechanical skills of violin performance: an exploratory study. In *Proceedings of the 5th International Conference on Movement and Computing (MOCO '18)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3212721.3212840

[73] Matthew Yee-King. 2022. Latent Spaces: A Creative Approach. In *The Language of Creative AI: Practices, Aesthetics and Structures*, Craig Vear and Fabrizio Poltronieri (Eds.). Springer International Publishing, Cham, 137–154. https://doi.org/10.1007/978-3-031-10960-7_8

[74] Diana Young and Stefania Serafin. 2003. Playability Evaluation of a Virtual Bowed String Instrument. https://doi.org/10.5281/zenodo.1176579 Pages: 104-108 Publication Title: Proceedings of the International Conference on New Interfaces for Musical Expression Publisher: Zenodo.

[75] Xu Zhang, Yu Wang, and Ray P.S. Han. 2010. Wavelet transform theory and its application in EMG signal processing. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Vol. 5. 2234–2238. https://doi.org/10.1109/FSKD.2010.5569532

[76] Ilana Zilber-Rosenberg and Eugene Rosenberg. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* 32, 5 (Aug. 2008), 723–735. https://doi.org/10.1111/j.1574-6976.2008.00123.x

[77] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704