Building an Embodied Musicking Dataset for co-creative music-making

Craig Vear[1], Fabrizio Poltronieri[1], Balandino DiDonato[2], Yawen Zhang[1], Johann Benerradi[1], Simon Hutchinson[3], Paul Turowski[4], Jethro Shell[5], and Hossein Malekmohamadi[5]

[1] University of Nottingham UK
[2] Edinburgh Napier University, UK
[3] New Haven University, USA
[4] Liverpool University, UK
[5] De Montfort University, UK

Abstract. In this paper, we present our findings of the design, development and deployment of a proof-of-concept dataset that captures some of the physiological, musicological, and psychological aspects of embodied musicking. After outlining the conceptual elements of this research, we explain the design of the dataset and the process of capturing the data. We then introduce two tests we used to evaluate the dataset: a) using data science techniques and b) a practice-based application in an AI-robot digital score. The results from these tests are conflicting: from a data science perspective the dataset could be considered questionable, but when applied to a real-world musicking situation performers reported it was transformative and felt to be 'co-creative'. We discuss this duality and pose some important questions for future study. However, we feel that the datatset contains a set of relationships that are useful to explore in the creation of music.

Keywords: dataset · music performance · embodied AI.

1 Introduction

To a musician, making music is about more than organising sound in space. Depending on the musician and the context, music-making can be a deeply felt embodied experience, one that shifts their understanding of self and their relationships with others. These bonds built on trust are augmented through creative engagement and play. Crucially, these relationships occur within music-making. Similar bonds are created when we listen to music, but, from the perspective of the performer, there is a two-way interplay that can lead to moments of creativity and novelty which are surprising, unexplainable, and meaningful to the individual.

This paper outlines a research project called the *Embodied Musicking Dataset*. Its aim was to build a dataset useful in training artificial intelligence (AI) for co-creative and meaningful real-time music-making. It is important to stress that this is a proof of concept, as no precedent existed to guide the team. The concept of musicking [24] was used as a lodestone with which to build the foundational concept of embodied musicking. Introduced in more detail below, the premise is that 'musicking establishes relationships at its location, and therein lies its meaning.' [24]. From this foundational concept, we designed a dataset to capture certain musicological, physiological and psychological aspects of a musician's performance. In an attempt to build a gold-standard dataset, we formalised the process and used a single backing track that would act as a strict matrix for the dataset. We also designed a self-labelling process. 10 musicians each performing 2-5 solos, were employed, and the dataset was built.

In this paper, the authors evaluate the dataset from two perspectives: 1) a data science 'deep dive' seeking evidence of correlations or causation between the features across participants, and 2) a practice-based musical experiment that tested meaning-making in the co-creation of music with AI trained on the dataset. Our findings show that the data science techniques unearthed very little evidence of correlation or causation. The authors of this paper would even admit that, from this perspective, it is questionable. And yet, when applied to the practice-based experiments *inside* music-making, the experience for the musicians was transformational and the AI was perceived to be operating on a level of co-creativity expected from human musicians.

2 Related Work

2.1 Embodied Musicking

When musicians perform, they do not simply output sound into the world but engage in an embodied experience of *becoming* the sound they create in the flow of music-making [26], [24], [13], [4]. Relationships between musicians and music flow are particularly evident in improvisational music like jazz, free, indeterminacy, or live electronics. Performers interact with represented ideas (scores and charts) while negotiating a constantly evolving state (real-time idea generation and sound invention). They act as autonomous agents (turn-taking) or simultaneous co-creators (joint playing). (e.g., [17] [12]). Social and musical interactions create meaning in a participatory way inside the flow. Participation requires expressive alignment with these elements at deeper psychological and physiological levels than simply making a sound. [17] state, 'In order to share the act of producing and perceiving sound and movement, we need to examine the *embodied-inter(en)acted phenomenological experience* of music-making'. For the purposes of this research, we define the concept of *embodied- inter(en)acted phenomenological experience* [17] of music-making as simply *embodied musicking* using the following existing concepts:

Musicking The composer and music theorist Christopher Small describes the embodiment of music as musicking. He defines it as:

> to music is to take part, which can happen 'in any capacity, in a musical performance, whether by performing, by listening, by rehearsing or practising, by providing material for performance (what we call composing) [24].

Small stresses that 'the act of musicking establishes in the place where it is happening a set of relationships, and it is in those relationships that the meaning of the act lies'. Simon Emmerson clarified Small's principle of 'meaning' to infer the 'what you mean to me', (this subtle shift circumvents the significant issues of value and who is doing the evaluation of meaning) [4]. Therefore, meaning (or the what-you-mean-to-me) is to be found by examining the relationships within, across, through and emergent of the creative acts of musicking and the materials of these acts, e.g. *people, sound, space, and time.*

Embodied Music Cognition As defined by Nijs et al [21], within music performance the 'embodied interaction with the music implies the corporeal attunement of the musician to the sonic event that results from the performance' [21]. They further depict the embodied experience of participating in musicking as a direct, engaged approach, relying on perceiving the musical environment and skillfully coping with challenges (affordances and constraints) arising from the complex musical interactions [21]. For Nijs et al [21], and as adopted by this research project, the optimal embodied experience (flow) occurs when the:

> musician is completely immersed in the created musical reality (presence) and enjoys himself through the playfulness of the performance. Therefore, direct perception of the musical environment, skill-based playing and flow experience can be conceived of as the basic components of embodied interaction and communication pattern. [21]

Flow Theory Csíkszentmihályi's Flow Theory (1975) supports Small's and Nijs's et al. argument that the acts of doing in music are to be considered an immersive and embodied experience. Csíkszentmihályi defined flow as 'the state in which people are so involved in an activity that nothing else seems to matter'. [2] He discusses how this deep engrossment in the here-and-now of action can occur in physical pursuits (athletes entering the zone) and in 'interactions with symbolic systems such as mathematics and computer languages' (such as concentration in computer game puzzles and video game immersion), [23] both of which describe parameters of immersion within live music performance with AI. Whilst Privette & Bundrick (1991)[22] focus on the 'intrinsically enjoyable experience' of flow (emotion being another quantifiable parameter of music performance), Csíkszentmihályi & LeFevre (1989)[3] and Massimini & Carli (1988)[18] argue for characterisation of flow as a balance of 'challenges and skills' proportionately beyond normal levels found in wakefulness. In short, through the act of musicking, musicians become embodied in the music through a sense of incorporation within their environment (the soundworld), shared effort, and a loss of awareness of their day-to-day wakefulness and corporeal self-consciousness.

2.2 Musicking Datasets

Most music datasets focus on two aspects of musicking: 1) the physical properties of sound (pitch, onset, timbral construction, dynamic warp from strict count etc.) or 2) the mechanics of music organisation (harmonic progression, melodic line prediction, long and short-term time-based feature extraction) [15], [1], [14], [20], [25], [11], [9], [7], [6]. However, there is no precedent for the design and development of a music dataset of embodied musicking. Recent works which built high-quality datasets that did have a direct influence on our project include the University of Rochester Multimodal Music Performance (URMP) dataset by Li et al. (2018)[15] [1]. This is a gold-standard dataset for multi-modal musician analysis. It is primarily aimed at audio-visual analysis of music performance. However, it does not dig deep into the embodied nature of musicking. Also, the recording process of this dataset prioritised the quality of the recorded media, which, if adopted here, would have had a negative influence on the quality of the embodied performance of the musicians. *GrooVAE* (Google Magenta project) *Learning to Groove with Inverse Sequence Transformations* [8] is another gold-standard music performance dataset that influenced this project heavily. The focus is on extracting the features of drummers' groove in order to train a neural net to introduce a sense of humanisation to matrix-composed music such as midi-players and loop generators.

3 Methodology (for building the dataset)

3.1 Visual Model

The design model for our embodied musicking dataset (EMD), indeed our whole proposition, is based on capturing the multi-dimensional inter-relationships of Embodied Musicking (Figure 1). We have conceptualised these relationships as an interconnected matrix of the main components extracted from the theoretical proposition above. This concept was then used to identify the most significant human parameters to capture to create a dataset that adhered to the governing objectives of the project. Subsequently, we determined which sensors to use and the most efficient way of designing the dataset-capturing environment (described below).

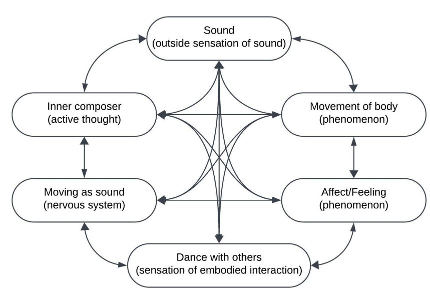

Fig. 1. Visual Model for Embodied Musicking

3.2 Primary Considerations

We needed to merge individual data from each of the 10 musicians, into a single dataset, focusing on a single instrument, style, source (score), high musical proficiency, and linking performances with a single musical backing as the ground-truth temporal track, all within the funding scope. The properties we arrived at were: *Instrument*: Upright piano; *Style*: Jazz improvisation; *Proficiency*: professional jazz musicians; *Source*: "How Deep is the Ocean" by Irving Berlin; *Temporal ground-truth (backing)*: 2 different versions of a pre-recorded human bass player.

The backing track was a crucial element in binding the dataset into a consistent temporal framework and ensuring that elements across the different recordings and musicians were bound into a cohesive matrix. To this end, a bass player recorded two different versions as if he were working in a piano-bass duet. Both recordings were five choruses of the tune in the format: head, improv 1, improv 2, improv 3, head recap. There was a two-bar count in, an option for a click track (2 & 4) and no outro (i.e. the final chorus simply stopped at the start of the first bar of the next chorus).

The choice of 'How Deep is the Ocean' as the source score was determined through consultation with several professional pianists in advance of the recording. It was agreed that this track is a well-known standard and would, therefore, not require a great deal of rehearsal. Furthermore, its harmonic construction offered the consulted musicians enough interest to sustain repeated improvisations without the risk of repeated themes and material and facilitating in-the-flow engagement.

3.3 Dataset Design, Apparatus and Setup

The dataset was designed to encapsulate the parameters outlined in *the visual model* (Figure 1). The sub-division consisted of:

Part 1: Physical-world music: Backing track audio and associated score organisation (mono). Audio recording of the piano (mono). Video of hands and fingers (embedded with the audio track). A Windows laptop was used to capture all physical-world music data. An HD webcam was used to capture the hands and fingers of the pianist. The audio was captured using a USB microphone plugged into the laptop.

Part 2: Embodied musicking: Electroencephalogram EEG [from BrainBit]. Electro-Dermal Activity EDA (arousal from Bitalino). Body tracking (using the Cubemos Skeleton SDK from the Intel Real Sense depth tracking camera). We used a Bitalino (r)evolution Bluetooth board [2] to capture EDA data from each musician. This streamed directly into the laptop for synchronous capture. This produced a single value. The BrainBit Headband [3] is a smart 4 channel EEG sensors associated with T3 and T4 temporal lobe regions and O1 and O2 occipital lobe regions. With an Intel RealSense Depth Camera D455 we captured 12 body skeleton points. The camera is positioned in front of the musician and close enough to record both facial expressions and upper torso movement. After each recording, the musicians sit down and listen back to their recording and log their sense of depth of flow through that performance.

Part 3: Flow protocol analysis (post-recording): Self-flow-evaluation as a sliding scale. A single axis slider was used to log the depth of subjective/ perceived musicking. This produced a single value and was used as a label.

A bespoke ecosystem was designed using mostly open-source technologies to support the data capture, ensure the data sync across all the sensors and preserve the data integrity. A custom software named "Blue Haze" was developed in Python captures data with a 10Hz sample rate and stores them in MongoDB[4] database using JSON-like documents. Blue Haze is an open-source project available on GitHub and customisable to other dataset-building projects[5].

3.4 Flow baseline.

To gather an understanding of musicians' flow baseline in piano playing, we ask them to complete a *Flow Short Scale* questionnaire before they attend the recording session. This *Flow Short Scale* method has been previously used in music studies. For example, Haug et al. (2020)[10] and Martin et al. (2008)[16] to a sample of musicians. We used the *Flow Short Scale* of 13 items published by Engeser (2012)[5]. These items give us information about flow experience, perceived importance, performance fluency and activity absorption. Additional items look at demand, skills, and the perceived fit of demands and skills.

3.5 Dataset contributors

We employed 10 solo improvising jazz musicians to contribute to the dataset: 5 from the UK and 5 from the US in the area of New Haven, Connecticut. We aimed for a mixture of flavours of jazz styles from commercial, through fusion, to modern. The recordings took place during the first 2020 COVID lockdown, with each musician contributing between 3 and 5 recordings to the dataset. We were strict about how we conducted these sessions and ensured that all preventative measures were implemented. The full data set can be found here https://rdmc.nottingham.ac.uk/handle/internal/10518

4 Analysis of the dataset. Test 1: Data Science

Using data science techniques, we sought to discover hidden relationships, correlations, and causation in the dataset. However, our conclusion is that there is limited evidence of these from this classic perspective. This analysis presented here, is adapted from a larger Master's project by Yawen Zhang at the University of Nottingham.

4.1 Data Preprocessing

In the Data Preprocessing phase, our focus was on ensuring the quality and completeness of the data collected from various musicians. To address missing values, we applied interpolation-based methods, carefully considering each scenario to avoid introducing biases — especially when a missing value might indicate an actual physiological or psychological state, it was retained as N/A. For physiological data like EDA and EEG, susceptible to external disturbances and equipment errors, we meticulously used statistical techniques and box plots for outlier detection, particularly scrutinizing anomalies at the start and end of performances. Data synchronization was a critical step, aligning different data types (physiological, psychological, and audio) accurately over time. This was achieved using a common backing track audio file for each performance, which provided a consistent time framework and enabled precise synchronization of data across multiple dimensions.

4.2 Data Analysis

In our project, the comprehensive exploration of varied datasets, encompassing EDA, EEG, and skeletal data, was critical. The initial phase of exploratory data analysis (EDA) was aimed at grasping the dataset's fundamental characteristics, identifying key patterns, anomalies, and relationships. This involved employing statistical methods like mean, median, and variance calculations to discern central tendencies and dispersions, complemented by visual tools such as histograms, box plots, and scatter plots to visualize data distributions and relationships between variables.

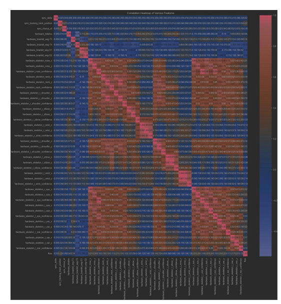

Fig. 2. Correlation Heatmap

This foundational work was integral for assuring data integrity and setting the stage for advanced analyses. It involved scrutinizing data distribution and addressing minimal missing values, particularly in essential metrics like 'flow', using median imputation to avoid biases and preserve statistical test validity. Subsequent stages of analysis ventured beyond basic descriptions, engaging in advanced statistical techniques and multifaceted data evaluations. These included deep dives into individual cases, cross-file comparative analyses using methods like ANOVA, and focused analyses employing clustering techniques. Each of these steps contributed to a holistic understanding of the dataset, revealing intricate details and complex interdependencies within the data.

A key part of our analysis was the use of a heatmap to illustrate correlations between variables (Figure 2). We found strong inter-correlations among EEG channels (T3, T4, O1, O2), which is consistent with expectations for measures of brain electrical activity. Interestingly, the 'flow' state did not show strong correlations with physiological data like EDA or EEG, hinting that musician self-reports might be less reliable or influenced by complex factors beyond basic physiology. The skeletal data, specifically the X and Y coordinates for body parts, demonstrated significant correlations, suggesting patterns of synchronous movement. The limited correlation between 'flow' and physiological data underscored the need for a deeper, more intricate analysis to unravel the underlying relationships. Nevertheless, the heatmap was instrumental in laying the groundwork for comprehending the dataset's complexity.

Deep Dive into a Single File In this phase, we explored a single performer's data, laying the foundation for future multi-file analyses. Our choice to analyze musician "Jn3VvBWcnDESzN9gUTh3bN" is significant. This musician has high data integrity, a "flow" score close to the dataset average, and noticeable variations in performances. Data is mostly complete, with just 1 and 3 missing values in sync-delta and flow attributes, respectively. We identified outliers using the Interquartile Range (IQR) method, mainly concentrated in skeletal data.

Analysis of Physiological Data and Flow The trends of EDA and flow exhibit some parallels; at certain points where EDA peaks are evident, as visualized in the provided graph (Figure 4) fluctuations in the flow state are also apparent, and occasionally, there seems to be a temporal lag between the peaks of the two. Acknowledging that EDA is influenced by factors like ambient temperature, humidity, and individual skin conductivity, we note that its peaks do not always correlate directly with "flow" states, suggesting that while EDA may indicate transient physiological changes during a performance, it may not fully capture the musician's psychological state. This complexity underlines the plausible but intricate link between EDA and the musician's psychological state, a connection we aim to explore further in subsequent analyses.

Fig. 3. Moving Window Average Time Series for the Four EEG Channels

EEG data was constrained by the sampling frequency of 7.68 Hz, calculated as the inverse of the average sampling interval (130.153 ms), which only enables study frequencies up to 3.9 Hz (Nyquist–Shannon sampling theorem) and makes it challenging to conduct meaningful frequency domain analysis of EEG. Furthermore, EEG data is often susceptible to noise and may also be influenced by non-cognitive factors such as muscle movements, eye blinks, or even electrical interference. A moving window average of 10 seconds was employed during the trend analysis assisting us in observing the primary signal trends more clearly. Upon observation, we discern that the different EEG channels (T3, T4, O1, O2) exhibit highly similar patterns throughout the entire time span. The high correlation among the four channels is evident even without a formal correlation analysis (Figure 3).

In our analysis of the motion data, significant issues were encountered due to a substantial number of data points being zeros, negative values, or having low confidence levels, leading to a large portion of data being deemed unusable. To confront these data irregularities, we explored various ARIMA (Autoregressive Integrated Moving Average) models, adjusting the parameters multiple times in an attempt to find an optimal fit for our non-stationary and volatile dataset.

The graph reflects the state of the motion data with one of the parameter sets we tested, where (p,d,q) = (1,1,10). This particular model configuration was an effort to mitigate the impact of unreliable data points while seeking to uncover any latent trends. However, even after processing with the ARIMA model, the data did not yield distinct "flow" patterns, underscoring the need for advanced analytical methods. Our choice of ARIMA was driven by its potential to model non-stationary time series data, a characteristic of the complex and noisy signals we encountered.

Fig. 4. Time Series Plot with Marked EDA Peaks

From the comparison, the musician's overall movement, as represented by the nose's position, doesn't show a clear correlation with their flow state. The dynamics of the nose's position might be influenced by various factors, and its interplay with the musician's psychological state (as represented by flow) isn't straightforward. The implications and potential reasons for these observations would warrant deeper exploration in further analysis.

Cross-file Comparisons In our cross-file comparisons, we analyzed the data from multiple sessions to discern any overarching patterns across different musicians and performances. This involved a comparative analysis of EDA and 'flow' data, looking for consistencies and discrepancies that could inform our understanding of physiological responses during musical performances. The analysis found that EDA and 'flow' trends across different sessions and chorus IDs. The visualization captures the EDA and 'flow' data points for each chorus, scaled and overlaid to facilitate comparison. Through this graphical analysis, we can observe the variability and potential correlations between EDA responses and the reported 'flow' states across multiple performance sessions. The image underscores the diversity of physiological reactions and psychological states experienced by musicians, highlighting the need for a nuanced interpretation of these complex datasets.

Focused Single-File Analysis After initial explorations and analyses, we delved deeper into specific file analysis to understand musicians' physiological reactions during specific tasks. Continuing with the second-stage analysis, our goal was to analyze their physiological reactions (EDA data) across five performances (i.e., five choruses). We used KMeans clustering method, an unsupervised machine learning method that groups similar data points together with different cluster numbers. In the analysis, three clusters emerged as the optimal choice.

Their respective sizes were 38, 30, and 13, leading us to hypothesize Cluster 0: Encompassing the majority of bars, this suggests these bars share statistical similarities in EDA. Cluster 1: Bars in this cluster have a different EDA reaction than those in Cluster 0. Cluster 2: The smallest cluster bars here might represent sections where the musician showcases specific emotions or technical prowess. These bars, being fewer, might represent special or challenging parts of the track. Based on the clustering results, bars in Cluster 0 might predominantly belong to one or two choruses, representing sections where the musician showcases technical and emotional peaks. In contrast, bars in Cluster 2 might be spread across all choruses, acting as the "baseline" throughout.

4.3 Conclusion of analysis 1

Through this analysis, we have come to recognize that there might be some form of correlation between music and physiological activity. Although no explicit patterns were discerned, potential connections and trends were observable from the data. The flow state is a unique psychological and physiological state linked to heightened focus, skill-challenge balance, and time perception distortion. Our findings suggest that when musicians enter a flow state, their physiological indicators might manifest patterns consistent with this psychological state.

5 Analysis of the dataset. Test 2: Inside Musicking

To question the dataset's relevancy of purpose, we needed to evaluate it within the domain that it originated: namely, inside musicking. To this end, we designed *Jess+* an intelligent digital score system that uses AI and a robotic arm to amplify and communicate the creativity of an inclusive ensemble. This project was conducted in the real-world with professional musicians and in a collaboration between Orchestras Live (a national producer creating inspiring orchestral experiences for communities across England) and Sinfonia Viva (a British orchestra based in Derby, England).

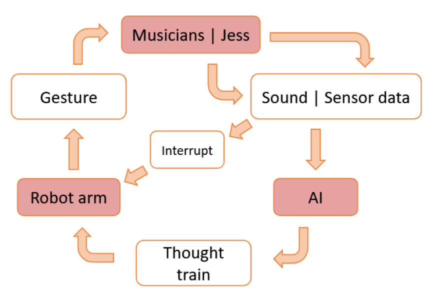

Fig. 5. Interaction Design

The digital score consisted of a robot arm, controlled by creative AI, which moved in a way much like a conductor or dancer might respond to music, in some experiments it drew on paper. The system consisted of a realtime feedback loop: live sound -> AI response -> movement -> human interpretation. The human ensemble consisted of a trio of musicians: Jess (disabled musician playing a digital instrument), Clare (violin) and Deidre (cello). The main goal was for the digital score to empower disabled musicians to engage in live music conversations despite disability-related barriers. Using a user-centered design and agile workflow, we created an AI-robot digital score with musician collaboration, following a closed-loop, realtime interaction design illustrated in Figure 5. The design of *Jess+* is discussed in 2 separate papers, which are currently in production at the time of writing.

The Embodied Musicking Dataset was central in training a core part of the *Jess+* system. The AI Factory consisted of 7 convolutional encoder-decoder deep learning models with an hourglass-shaped architecture[19], trained to predict features from other features of the dataset including audio envelope, *core* position, EEG, EDA and flow, with the goal of this AI-stack to learn a representation of features with relation to each other in the context of musicking.

In realtime deployment, predicted data from deep learning networks was either fed to neighboring networks or used as raw data for deciding robot movements. These were predefined in advance and in consultation with the musicians. However, the AI was allowed to make critical decisions about speed, tempo, velocity, duration and interruption of each of these movements so as to surprise the human musicians. The team spent 5 months on the project, meeting the 3 musicians 5 times to test and develop the digital score. This culminated in a formal sharing with the partners using two types of mode: pens on paper (top photo in Figure 6), and "dancing feather" (bottom photo in Figure 6)[6].

5.1 Results

Through an extensive and iterative research process, we conducted a comprehensive qualitative investigation into musicians' reflections on using the digital score. Our findings revealed that the design decisions implemented in this case study enriched the musicians' experiences. They discussed how the AI and the robot were working with them inside musicking. And that its behaviours and interactions inspired relationships and bonding that they perceived to be co-creative. Though each musician had a unique connection with the robot, the disabled musician, Jess (who was also playing a digital instrument), felt a strong bond with the system during music-making, seeing *Jess+* as an extension of herself. She thought the extension's purpose was to visually represent the music, referring to it as a "friend" and "story-teller." Non-disabled musicians (playing cello and violin) recognized the Jess-system connection, labelling it a "creative accompanist".

Fig. 6. Final Performance

All musicians highlighted how they felt being in-the-loop with *Jess+* and how this transformed their own practices. Jess felt that the system allowed her to "express the emotions that she is sometimes not able to express through her current digital setup". She felt "extended through the system", which meant that she could "feel like she was able to express her feelings directly onto a score". This led to her wanting to "explore that part of me and I wanted – you know, I want my emotions that are in here to get expressed outwardly through that". For the non-disabled musicians, they reflected how

> The robot arm was liberating to improvise with as it was non-judgemental. At times, it united the three musicians' music, and at other times, it could also be independent from us (as we knew it would return to respond to what we later did). This, in turn, influenced the musicians to start or stop, to 'gel' together harmonically or feel the freedom to play outside harmonic or rhythmic frameworks. In my opinion, improvising with a human (especially someone new to you) carries psychological elements that could interfere with making music together, so the robot arm provided the opportunity for freedom of musical and emotional expression that would take much more time to establish and develop between humans".

5.2 Conclusion of analysis 2

From an embodied and subjective perspective, our findings revealed that the musicians formed unexpected and distinct relationships with the creative-AI robot arm (*Jess+*). Although they viewed its role differently, they all acknowledged that it cultivated a set of relationships between the ensemble that a) augmented their creativity, and b) stimulated a sense of an inclusive ensemble without a hierarchy. They felt that an inclusive ensemble was formed involving disabled and non-disabled musicians *and* non-human musicians (*Jess+*). They also felt that they were in-the-loop with it *inside* musicking and that it had its own "voice". They were particularly taken by they way they felt that it listened to them and co-created with them, and significantly that it operated in a non-judgemental way. Discussed in much more detail in a forth-coming paper, they "viewed the system as an additional layer of creativity and felt empowered by its inclusive potential".

On the one hand, these insights shed new light on human-AI co-creativity and collaboration. On the other hand, however, the team does not know why the central AI, and its AI-factory design of deep learning models trained on the Embodied Musicking Dataset worked or was able to conjure such intense relationships as those experienced by these professional musicians. Simple good will or a lack of transparency were factors that were mitigated against through the research design and methodology. We suspect that it was to do with the musicking-behaviours embedded into the training data that were collected through improvised performance, and the musicking-focus behaviours embedded into the AI and symbolic algorithms.

6 Discussion

The experiences described by the professional musicians provide convincing evidence that something *musicking* was captured in this dataset. Classic data science techniques for extracting correlation and causation within the dataset, though, have been unable to point to why or how. This conflict brings forth more questions:

1. Can embodied musicking be represented in a dataset? We think this dataset provides some convincing evidence that it can. We are not certain that we were able to capture it in its entirety, but the approach and philosophy that we used to frame this proof-of-concept indicates a positive move forward. Musicking is more than just emitting notes into the air through time. There are emotions and affectual responses involved. There is also some brain activity involved, although we are not convinced that that is purely logical planning and reasoning (e.g. "I need to play this next", or "the next logical note to play is this"). We suspect that these decisions are more inherited and embedded into the whole mind-body system.

2. Should we be designing embodied musicking datasets using scientific principles? The EMD was built using a "gold-standard" approach. We had a single backing track, and focused on a single instrument, and invited professional musicians to contribute data using a standard system of instruments to capture their embodied musicking. Why? Surely, creativity and humanness are messy. The deep dive into the dataset through the classic data science techniques revealed very little to suggest that a "gold standard" approach was appropriate. There are some critical questions about the sample-rate at which the dataset was built (brain waves captured at 10Hz are really only going to show limited information about mental activity), whether the EDA tracking of arousal was not standardised, and can the individual musician really be trusted to label their performance objectively (interesting to note that only recording they perceived to be "good" were allowed into the dataset, as such we have not data of a poor embodied performance). And yet, given these problematic concerns (from the perspective of data science), the AI that was implemented in *Jess+* transformed musicianship and creativity because the humans trusted, recognised and enjoyed the sensation of the AI.

This leads to our final question *3. what is really going on here?* Clearly, something is happening when we train neural nets using the EMD to co-create with human musicians, but what is it? Might it be the messy-ness of it that we recognise somehow as human? We mentioned in the introduction that, from a data science perspective, the EMD is questionable, and we are happy with that description. Perhaps there is some quality in its questionable-ness that we relate to: perhaps there is something in the way that the system was designed from an embodied perspective, and the AI-stack operated as a poetic random-number generator. Or perhaps the level of trust amongst the team and professional musicians convinced them that the AI was to be trusted, and they went along with it (although given the amount of discussion, challenge and debate in the practical sessions, we doubt this). Regardless, the next stages of our research need to lean into these questions.

7 Conclusion

In this paper, we shared findings from designing, developing, and deploying a proof-of-concept dataset aiming to capture physiological, musicological, and psychological aspects of embodied musicking. This dataset was then tested using a) classical data science techniques and b) practice-based techniques.

Data science revealed potential correlations between music and physiological activity. While no clear patterns emerged, we observed possible connections and trends in the data. Flow state has been verified as a unique psychological and physiological state associated with heightened focus, a balance between skills and challenges, and a distortion in the perception of time. Our findings suggest that when musicians enter a flow state, their physiological indicators might manifest patterns consistent with this psychological state. This offers a new view on understanding and promoting the flow state, implying that monitoring physiological indicators can identify and improve flow experiences. However, this testing also unveiled challenges. The low database sampling rate and skeletal data confidence issues hindered EEG and high-frequency data analysis. This calls for a more holistic approach in future research, incorporating a range of physiological, cognitive, and emotional measurement techniques.

In practice-based performance, the dataset trained a deep learning model and powered an intelligent digital score that extended the creativity of disabled and non-disabled musicians within an inclusive music ensemble. The professional musicians involved all acknowledged that the AI/robot was co-creative, that they felt in-the-loop with it and that it transformed their creativity. They recognised it as a co-creator, and a member of the ensemble, even a "friend". Their improvisations were open, enhancing and for all three, extending their techniques and confidence. In summary, we face a conflict: the dataset may seem flawed, yet we believe it holds something vital to musicking, a set of relationships that are somehow perceived to be from within musicking itself.

Acknowledgements This dataset project was funded with a generous grants from the Human-Data Interaction Network funded through the Engineering and Physical Sciences Research Council grant number EP/R045178/1. This work on Jess+ funded by the ERC under the European Union's Horizon 2020 research and innovation programme (ERC-2020-COG – 101002086). Additional funding was received from TAS Hub and the Faculty of Arts at the University of Nottingham. We would like to address our sincere thanks for the support we received from Orchestras Live[7] and Sinfonia Viva[8], and of course our musicians Jess, Clare and Deirdre

References

1. Cancino-Chacón, C.E., Grachten, M., Goebl, W., Widmer, G.: Computational Models of Expressive Music Performance: A Comprehensive and Critical Review. Frontiers in Digital Humanities **5** (2018), https://www.frontiersin.org/articles/10.3389/fdigh.2018.00025

2. Csikszentmihalyi, M.: Beyond boredom and anxiety: Experiencing flow in work and play.-san fransisco, ca: Jossey-bass p. 4 (1975)
3. Csikszentmihalyi, M., LeFevre, J.: Optimal experience in work and leisure. Journal of personality and social psychology **56**(5), 815 (1989)
4. Emmerson, S.: Living electronic music. Routledge (2017)
5. Engeser, S.: Theoretical integration and future lines of flow research. Advances in flow research pp. 187–199 (2012)
6. Friberg, A., Bresin, R., Sundberg, J.: Overview of the KTH rule system for musical performance. Advances in Cognitive Psychology **2**, 145–161 (2006). https://doi.org/10.2478/v10053-008-0052-x
7. Friberg, A., Colombo, V., Frydén, L., Sundberg, J.: Generating Musical Performances with Director Musices. Computer Music J. **24**(3), 23–29 (Sep 2000). https://doi.org/10.1162/014892600559407, https://direct.mit.edu/comj/article/24/3/23-29/93448
8. Gillick, J., Roberts, A., Engel, J., Eck, D., Bamman, D.: Learning to groove with inverse sequence transformations. In: International Conference on Machine Learning. pp. 2269–2279. PMLR (2019)
9. Giraldo, S., Ramírez, R.: A machine learning approach to ornamentation modeling and synthesis in jazz guitar. J. of Mathematics and Music **10**(2), 107–126 (May 2016). https://doi.org/10.1080/17459737.2016.1207814, https://www.tandfonline.com/doi/full/10.1080/17459737.2016.1207814
10. Haug, M., Camps, P., Umland, T., Voigt-Antons, J.N.: Assessing differences in flow state induced by an adaptive music learning software. In: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). pp. 1–4. IEEE (2020)
11. Huang, C.Z.A., Vaswani, A., Uszkoreit, J., Shazeer, N., Hawthorne, C., Dai, A.M., Hoffman, M.D., Eck, D.: Music transformer: Generating music with long-term structure. arXiv preprint arXiv:1809.04281 (2018)
12. Hytonen-Ng, E.: Experiencing'flow'in jazz performance. Routledge (2016)
13. Leman, M.: Embodied music cognition and mediation technology. MIT press (2007)
14. Lerch, A., Arthur, C., Pati, A., Gururani, S.: An interdisciplinary review of music performance analysis. arXiv preprint arXiv:2104.09018 (2021)
15. Li, B., Liu, X., Dinesh, K., Duan, Z., Sharma, G.: Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. IEEE Transactions on Multimedia **21**(2), 522–535 (2018)
16. Martin, A.J., Jackson, S.A.: Brief approaches to assessing task absorption and enhanced subjective experience: Examining 'short'and 'core'flow in diverse performance domains. Motivation and Emotion **32**, 141–157 (2008)
17. Martínez, I.C., Damesón, J., Pérez, J.B., Pereira Ghiena, A., Tanco, M.G., Alimenti Bel, D.: Participatory sense making in jazz performance: Agents' expressive alignment. In: 25th Anniversary Conference of the European Society for the Cognitive Sciences of Music (Ghent, Bélgica, 31 de julio al 4 de agosto de 2017) (2017)
18. Massimini, F., Carli, M.: 16. the systematic assessment of flow in daily experience (1988)
19. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
20. Müller, M., Grosche, P., Wiering, F.: Automated analysis of performance variations in folk song recordings. In: Proc. Int. Conf. on Multimedia Information Retrieval. pp. 247–256 (Mar 2010)

21. Nijs, L., Lesaffre, M., Leman, M.: The musical instrument as a natural extension of the musician. In: the 5th Conference of Interdisciplinary Musicology. pp. 132–133. LAM-Institut jean Le Rond d'Alembert (2009)
22. Privette, G., Brundrick, C.M.: Peak experience, peak performance, and flow: Correspondence of personal descriptions and theoretical constructs. Journal of social behavior and personality **6**(5), 169 (1991)
23. Siekpe, J.S.: An examination of the multidimensionality of flow construct in a computer-mediated environment. Journal of Electronic Commerce Research **6**(1), 31 (2005)
24. Small, C.: Musicking: The meanings of performing and listening. Wesleyan University Press (1998)
25. Todd, N.P.M.: The dynamics of dynamics: A model of musical expression. The J. of the Acoustical Society of America **91**(6), 3540–3550 (Jun 1992). https://doi.org/10.1121/1.402843, https://pubs.aip.org/asa/jasa/article/91/6/3540-3550/968369
26. Vear, C.: The Digital Score: Musicianship, Creativity and Innovation. Routledge (2019)

Notes

1. http://www2.ece.rochester.edu/projects/air/publications/li2018creating.pdf
2. https://plux.info/kits/36-bitalino-revolution-board-ble-810121002.html
3. https://brainbit.com/
4. https://www.mongodb.com/
5. https://github.com/Creative-AI-Research-Group/embodiedMusickingDataset/tree/master/blue%20haze
6. Here are some of the recordings from this sharing session: https://youtu.be/MBPQNmAXvXk https://youtu.be/7dQKIpjKJu4 https://youtu.be/sK4KAmv3ikw
7. https://www.orchestraslive.org.uk/
8. https://www.sinfoniaviva.co.uk/