



## OPEN ACCESS

## EDITED BY

Anna Xambó,  
Queen Mary University of London,  
United Kingdom

## REVIEWED BY

Rodrigo Constanzo,  
Royal Northern College of Music,  
United Kingdom  
Koray Tahiroglu,  
Aalto University, Finland

## \*CORRESPONDENCE

Torin Hopkins  
✉ Torin.Hopkins@colorado.edu

RECEIVED 03 February 2025

ACCEPTED 04 July 2025

PUBLISHED 04 August 2025

## CITATION

Hopkins T, Sun R, Chuan Weng SC, Ma S-Y,  
Crum J, Hirshfield L and Do E-L (2025) BrAln  
Jam: neural signal-informed adaptive system  
for drumming collaboration with an AI-driven  
virtual musician.

*Front. Comput. Sci.* 7:1570249.  
doi: 10.3389/fcomp.2025.1570249

## COPYRIGHT

© 2025 Hopkins, Sun, Chuan Weng, Ma,  
Crum, Hirshfield and Do. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# BrAln Jam: neural signal-informed adaptive system for drumming collaboration with an AI-driven virtual musician

Torin Hopkins<sup>1,2\*</sup>, Ruojia Sun<sup>1</sup>, Suibi Che Chuan Weng<sup>1</sup>,  
Shih-Yu Ma<sup>1,2</sup>, James Crum<sup>2</sup>, Leanne Hirshfield<sup>2</sup> and  
Ellen Yi-Luen Do<sup>1,2</sup>

<sup>1</sup>ATLAS Institute, University of Colorado Boulder, Boulder, CO, United States, <sup>2</sup>Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO, United States

Collaboration between improvising musicians requires a dynamic exchange of subtleties in human musical communication. Many musicians can intuit this information, however, translating this knowledge to embodied computer-driven musicianship systems—be they robotic or virtual musicians—remains an ongoing challenge. Methods of communicating musical information to computer-driven musicianship systems have traditionally been accomplished using an array of sensing techniques such as MIDI, audio, and video. However, utilizing musical information from the human brain has only been explored in limited social and musical contexts. This paper presents “BrAln Jam,” utilizing functional near-infrared spectroscopy to monitor human drummers’ brain states during musical collaboration with an AI-driven virtual musician. Our system includes a real-time algorithm for preprocessing and classifying brain data, enabling dynamic AI rhythm adjustments based on neural signal processing. Our formative study is conducted in two phases: (1) training individualized machine learning models using data collected during a controlled experiment, and (2) using these models to inform an embodied AI-driven virtual musician in a real-time improvised drumming collaboration. In this paper, we discuss our experimental approach to isolating a network of brain areas involved in music improvisation with embodied AI-driven musicians, a comparative analysis of several machine learning models, and *post hoc* analysis of brain activation to corroborate our findings. We then synthesize findings from interviews with our participants and report on the challenges and opportunities for designing music systems with functional near-infrared spectroscopy, as well as the applicability of other physiological sensing techniques for human and AI-driven musician communication.

## KEYWORDS

functional near-infrared spectroscopy (fNIRS), brain-computer interfaces, embodied AI, music, neuroscience, machine learning, music improvisation, human-computer interaction

## 1 Introduction

Music improvisation is an incredibly demanding team task that necessitates instantaneous synchronization among two or more skilled musicians, each possessing a shared lexicon of musical knowledge. Many actions must be executed within milliseconds, and even minor deviations in timing can have adverse consequences for musical performance (Chafe et al., 2004; Schuett, 2002; Chew et al., 2004). To meet these demands,

collaborating musicians often communicate using their bodies, using eye contact, and other means to make musical information more predictable over several time frames (Hopkins et al., 2022b). In the context of collaborative music improvisation (also known as jamming), the integration of embodied AI collaborators opens new creative avenues, as well as challenges for human-computer musical interaction. A central issue being the AI's capability to extract meaningful musical information from human musicians during a jam session (Bretan and Weinberg, 2016; Miranda, 2021; McCormack et al., 2019).

Traditionally, computer systems in real-time musical interaction have used music information retrieval techniques (Weinberg et al., 2005; Bretan and Weinberg, 2016; Rowe, 2004; Hopkins et al., 2021, 2020, 2023d). These methods enable AI-driven musicians (AIMs) to “listen and react” to the output from human musicians. However, this approach is limited by several factors including processing speed, computational demands, introduction of latency, and eschewing other communication cues that human musicians often rely on—including gesture and other forms of embodied communication (Hopkins et al., 2022b, 2024, 2023c; Bretan and Weinberg, 2016). Drawing inspiration from strategies that improvising musicians use to navigate these challenges (Leman, 2007; Lesaffre et al., 2017; Hallam et al., 2009; Lewis and Piekut, 2016; Midgellow, 2019), alternative approaches have been developed (Bretan and Weinberg, 2016; Lesaffre et al., 2017). These include gesture analysis through computer vision and motion capture (Bretan and Weinberg, 2016; Hopkins et al., 2022b), brain-computer interfaces (Miranda, 2021; Yuksel et al., 2016, 2019), and the incorporation of electrodermal activity sensors to gauge the confidence levels of humans collaborating with AIMs (McCormack et al., 2019).

In this paper, we focus on a metric we call “rhythmic predictability,” which has been broadly explored in the neuroscience of rhythm and improvisation (Vuust et al., 2018; Berkowitz and Ansari, 2008). We enable an embodied AIM to adjust its playing based on the neural signals of a human drummer and their perception of rhythmic predictability throughout the improvised drumming session. The AIM differentiates neural signals associated with incoherent rhythms from musical ones and refines its drumming as a result.

We draw on several areas of research in neuroscience of music improvisation and rhythm perception, as well as AI collaboration to inform our metric. First, the perception of rhythm has been associated with activation in the supplementary motor area and pre-motor cortex, both of which are located within the frontal lobe (Patel and Iversen, 2014; Vuust and Witek, 2014; Grahn and Rowe, 2009). Secondly, generative music activities such as those exhibited in music improvisation have been associated with activation patterns in the dorsolateral prefrontal cortex and supplementary motor area (Bengtsson et al., 2007), as well as the temporoparietal junction de Manzano and Ullén (2012). Other literature in creative music cognition suggests activation in the default mode network may play an important role (Beaty et al., 2023). However, these regions are too deep within the brain to measure reliably with neuroimaging techniques used in ecological settings, such as functional near-infrared spectroscopy (fNIRS) or electroencephalography (EEG).

Additionally, inspired by the broader metric of “trust and reliability,” which is often used for quantifying collaboration effectiveness in embodied AI-human interaction (Bobko et al., 2022; Eloy et al., 2022; Glikson and Woolley, 2020; McCormack et al., 2020), activation in these same areas are also implicated with an additional emphasis on the temporoparietal junction and frontal lobe (Eloy et al., 2022; Bobko et al., 2022).

While the dynamics of trust have been explored in musical settings (Hopkins et al., 2023a; McCormack et al., 2019, 2020), much of the relevant literature comes from a long history of human-automation interaction studies (Eloy et al., 2022; Bobko et al., 2022; Howell-Munson et al., 2022; Eloy et al., 2019). These studies often manipulate agent characteristics such as *reliability* to assess trust. In particular, when agents provide unreliable information, their human collaborators are likely to ignore the information and rely solely on their own experience, resulting in a lack of trust in the agent (Hoff and Bashir, 2015; Glikson and Woolley, 2020; Eloy et al., 2022).

To measure real-time brain data, choosing a neuroimaging technology often requires assessing trade-offs between spatial and temporal resolution. The two most popular neuroimaging techniques used in ecological settings are fNIRS (the more spatially resolute option) and EEG (the more temporally resolute option). Because the supporting neuroscientific findings in music perception are region-specific and studies related to trust in human-AI collaboration utilize fNIRS for real-time neural signal detection hardware, we opt for fNIRS in the design of our system. fNIRS is also more robust for use in real-time drumming and is becoming increasingly popular in real-time neural signal detection generally (Seidel-Marzi et al., 2021). Though there is more latency associated with fNIRS than EEG, the desire for spatial resolution and activity-resistant neural signal detection hardware led us to design our system using fNIRS.

Throughout the article, we explore the concept of rhythmic predictability as a quantifiable metric, enabling AIMs to evaluate the predictability of their generated rhythms. Grounded in both our observations and existing research, including jam sessions and musician interviews (Hopkins et al., 2022b,a), this notion addresses the inherent capacity of musicians to assess their own rhythmic predictability, as well as that of the ensemble. Such capabilities become apparent when musicians spontaneously adjust their performance to enhance synchronization, particularly when the musical harmony appears to be misaligned (Weng et al., 2023).

To operationalize the concept of rhythmic predictability, we implement a two-step approach. First, we define rhythmic predictability as a neurologically measurable attribute, capturing it through fNIRS. Second, we incorporate this metric into “BrAIIn Jam,” an innovative real-time system that monitors the neural activity of human drummers during interactive musical collaboration with an embodied AI-driven virtual musician (AIVM) (Hopkins et al., 2023b) (see Figure 1). The system employs a real-time algorithm specifically designed for preprocessing and classifying brain data. This enables the AI to dynamically adapt

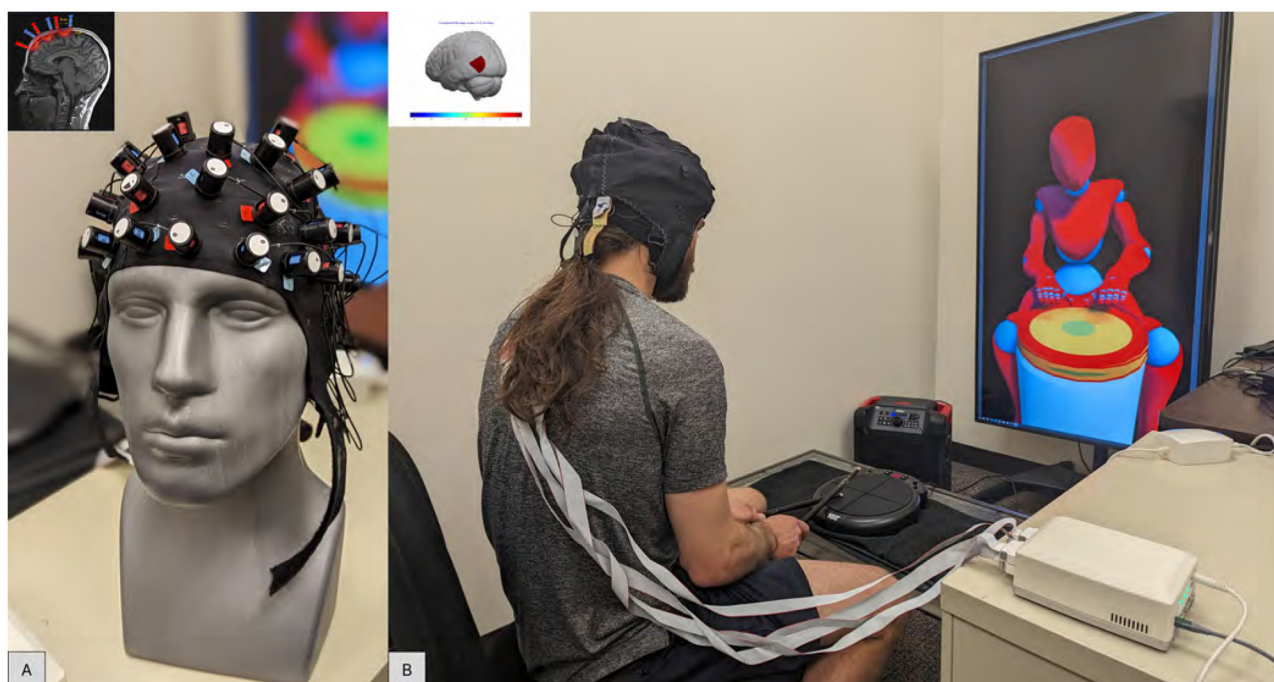


FIGURE 1

(A) fNIRS signal and equipment for reading neural signals. (B) Drummer playing with the BrAIn jam system using fNIRS to read brain activation while playing with a AI-Driven Virtual Musician.

its rhythmic patterns based on the fNIRS readings (with 3–5 second delay due to hemodynamic response time), which serve as indicators of the human player's affective state. Lastly, we explore how this affective information can be utilized in an open-ended creative setting to enhance the level of synchronization and overall musical performance between human and AI musicians.

Building on the foundation of rhythmic predictability, we make several key contributions that extend the current understanding of AI-human musical interaction and offer practical tools for real-time, affective state-informed musical collaboration:

- (1) We formalize the concept of rhythmic predictability as a metric and how to reliably measure it during collaboration with an embodied AI musician. We substantiate our claims through an empirical investigation that integrates an analysis of both neural and machine learning data. This provides a nuanced understanding of how musicians inherently gauge rhythmic predictability with embodied AI musicians and how this can be quantified for use in adaptive AI systems.
- (2) We introduce “BrAIn Jam,” an adaptive system that employs fNIRS to capture real-time neural activity of human musicians. The system features a real-time preprocessing algorithm for fNIRS data, which we make available for future explorations in brain-computer interfaces using fNIRS technology.
- (3) We offer valuable insights into the practicalities and limitations of incorporating affective state data into real-time musical systems. These insights are derived from a player test of the BrAIn Jam system and address the specific constraints posed by fNIRS technology.

## 2 Related work

### 2.1 Embodied AI-driven musicianship

Recent advancements in music technology have led to the development of AI-driven systems that can assist in musical performance and composition. Musical and automated robots have a long history of use in musical performance (Kapur, 2005; Singer et al., 2003, 2004; Rowe, 2004; Dannenberg et al., 2011; Kajitani, 1989; Raman, 1920; Roads, 1986; Takanishi et al., 1996; Williamson, 1999). However, only within the last few decades have researchers focused on creating adaptive, improvisational robots that display musicianship. One of the earliest works in the field of robotic musicianship was conducted by (Weinberg et al., 2005) and Eigenfeldt and Kapur (2008), who independently introduced robotic drumming systems that used sensors and actuators to control a robotic drummer. Though the system was able to play complex rhythms, it was limited in its ability to generate music and was only able to improvise with human input in a call and response manner. Hoffman later proposed a robotic marimba player which was capable of call and response, overlay improvisation, and phrase-matching (Weinberg and Driscoll, 2007; Miranda, 2021; Bretan and Weinberg, 2016). The authors showed that the system was able to generate accompaniments that were consistent with the human player and had a high degree of musical coherence.

Nonetheless, the field faces inherent challenges, notably the “concatenation cost” and “embodiment cost,” which refer to the computational and physical time required to assemble and execute musical phrases, respectively (Bretan and Weinberg, 2016; Miranda, 2021). While efforts are underway to enhance these

systems by enabling them to anticipate musicians' decisions, this is primarily achieved through the integration of extra-musical information to inform musical predictions (Bretan and Weinberg, 2016; Lesaffre et al., 2017). Despite these advancements, considerable work remains to enhance the responsiveness and social adaptability of these systems in the context of improvised music (Bretan and Weinberg, 2016).

## 2.2 Use of extra-musical information in interactive musical systems

While existing work has attempted to bridge the communication gap between human musicians and interactive systems, they often rely on overt cues that may not accurately represent the internal emotional or cognitive state of the human participant. Commonly used strategies such as gesture analysis, electrodermal activity, and brain-computer interfaces are described in the following sections.

### 2.2.1 Gesture analysis

Communicating via gesture is vital for musicians in improvised and live performance. Players often coordinate the movement of different sections, whether to take a solo, or even to end a song synchronously (Bretan and Weinberg, 2016; Hopkins et al., 2022b). Additionally, gesture-based anticipation in human-robot interaction has been shown to increase anthropomorphic inferences and acceptance of the robots by humans (Eyssel et al., 2011). Gesture analysis has been explored in the robotic marimba player, Shimon, through the use of computer vision (Bretan and Weinberg, 2016; Miranda, 2021). By analyzing head movement, Shimon can make inferences about tempo and rhythm that can support the analysis of musical information. OpenCV has been employed to enhance the accuracy and efficiency of this gesture-based analysis, contributing to more nuanced and responsive musical interactions (Miranda, 2021).

### 2.2.2 Electrodermal activity

The use of skin conductance as a measure of engagement has been explored but lacks the specificity to decode complex cognitive states. Skin conductance has been positively correlated with sympathetic nervous system responses (also known as fight or flight responses), and is particularly useful at discerning levels of stress in adaptive computational systems (McCormack et al., 2019; Williams et al., 2019). McCormack et al. presents a collaborative improvising AI drummer that communicates its confidence through an emoticon-based visualization. The AI system was trained on both musical performance data and real-time skin conductance of musicians improvising with professional drummers (McCormack et al., 2019). The electrodermal activity data served as extra-musical cues to inform the AI's generative process. Temporal Convolutional Networks were used to analyze this skin conductance data, adding a layer of complexity to the AI's understanding of the musician's engagement levels (McCormack et al., 2019).

### 2.2.3 Brain-computer interfaces

EEG has also been employed to explore collaborative sonification and the emotional states of musicians (Leslie et al., 2014; Mullen et al., 2015; Leslie and Mullen, 2011). Leslie and Mullen (2011) utilized EEG data to create a collaborative sonification system called MoodMixer, which aimed to enhance musical collaboration by translating EEG signals into auditory feedback. This approach opened up new avenues for understanding the emotional and cognitive states of musicians during live performances inspiring several interactive musical systems using EEG Mullen et al. (2015).

Music systems that adapt to the player based on affective state have been exhibited by Yuksel et al. (2016) to inform music training systems and support creativity. In this project, fNIRS devices were used to measure workload and adapt the learning or creative environment accordingly. The team proposes a training mode and a creativity mode, whereby separate affective states and brain regions are recorded and act as a proxy to discern level of cognitive workload. A Support Vector Machine (SVM) with a linear kernel to adapt music training systems based on fNIRS-measured workload, was used to provide a method for real-time adaptation (Yuksel et al., 2016).

## 2.3 Neuroscience of rhythmic predictability

Rhythmic predictability is a quantifiable metric inspired by the neuroscience of rhythm, music improvisation, and social reliability.

### 2.3.1 Neural underpinnings of rhythm

Rhythm is a complex cognitive construct that entails a dynamic interplay between several regions of the brain. There is substantial evidence to suggest that auditory-motor coupling plays a strong role in how we perceive and interact with our environments. The predictive coding framework—though difficult to measure at the neuronal level—provides a model for explaining how the brain minimizes prediction errors in the processing of rhythmic auditory stimuli and has been used to explain how the brain processes incongruous rhythms (Vuust and Witek, 2014; Vuust et al., 2018). This framework predicts the existence of a tightly connected network in the brain that processes auditory stimuli and plans motor movement as a result. This is validated through electrical signals sensed from the brain that correspond to unexpected deviations in rhythmic patterns, otherwise known as mismatch negativity (Vuust and Witek, 2014; Patel and Iversen, 2014; Chen et al., 2006; Hallam et al., 2009).

Additionally, according to the action simulation for auditory prediction hypothesis, the motor system simulates periodic movement to assist in predicting beat timing (Patel and Iversen, 2014). This theory suggests that sound is processed and relayed to areas closely associated with the latter stages of planning motor output, such as the premotor cortex and supplementary motor area of the frontal lobe (Grahn and Rowe, 2009; Hallam et al., 2009; Bengtsson et al., 2007). Areas of the brain associated with planning and coordinating more complex motor patterns such as the areas in



the prefrontal cortex (Thaut et al., 2014; Fukuie et al., 2022; Chen et al., 2006) and cerebellum (Chen et al., 2008; Hallam et al., 2009) have also been shown to be recruited to assist in the processing and response to more complex rhythms.

Other theories focus more on attention, such as the dynamic attending theory, implicate similar regions to explain the processing and execution of rhythms. The dynamic attending theory proposes that attentional oscillations synchronize with external rhythmic stimuli to form temporal expectations (Hallam et al., 2009; Vuust and Witek, 2014; Vuust et al., 2018). These theories implicate further contribution from the basal ganglia [playing a role in filtering motor output (Vuust and Witek, 2014; Kung et al., 2013)], as well as frontal and parietal lobes in dynamically attending to and predicting rhythms (Vuust and Witek, 2014).

These models as well as neuroimaging evidence suggests that brain area networks are able to respond to incongruous rhythmic patterns. Given the location of these areas in the neocortex, we postulate that researchers can utilize these signals to build adaptable music tools.

### 2.3.2 Neural underpinnings of music improvisation

Music improvisation is a complex musical activity that requires musicians to make extemporaneous musical decisions while communicating with each other. Thus, there are several components of music improvisation that elicit well-known patterns of neural activation. However, research that has sought to isolate neural activation in association with music improvisation, several areas emerge only in conditions involving improvisation (Berkowitz and Ansari, 2008; Berkowitz, 2010; Hallam et al., 2009).

From a social cognition perspective, some of these areas have also been shown to be active while speaking—demonstrating the connection between music improvisation and speech and language processing (Berkowitz, 2010; Berkowitz and Ansari, 2008; Hallam et al., 2009). While other areas such as the temporoparietal junction and dorsolateral prefrontal cortex have been implicated in social situational processing generally (Berkowitz and Ansari, 2010; Carter et al., 2012; Carter and Huettel, 2013; Schurz et al., 2017; de Manzano and Ullén, 2012; Bengtsson et al., 2007).

From a music cognition perspective, improvised music activities have also been associated with activity in regions that support the perception of rhythm (Hallam et al., 2009), complex musical structure (Fukuie et al., 2022), movement in association with music (Levitin, 2002), and motor planning (Vuust and Witek, 2014). Activation patterns in the dorsolateral prefrontal cortex and supplementary motor area have been implicated in improvising music as well as planning and executing musical motor patterns (Bengtsson et al., 2007). Other literature in creative cognition suggests activation in the default mode network also plays a role in the generation of new creative ideas (Beaty et al., 2023).

These areas have been shown to work together to enable musicians to make extemporaneous creative musical decisions, while executing fine motor control over complex musical patterns. Thus, recognizing these areas as contributors to the signal likely present in rhythmic predictability is paramount.

## 2.4 Neural correlates of trust and reliability in human-agent teaming

Trust is a multifaceted construct that has garnered considerable attention across various domains, including interpersonal relationships and human-automation interactions (Fett et al., 2014; Howell-Munson et al., 2022; Carter and Huettel, 2013; Eloy et al., 2019). Emerging insights from social neuroscience have pinpointed specific brain regions involved in interpersonal trust and the ability to anticipate complex behaviors in social scenarios, notably the temporoparietal junction (TPJ) (Carter and Huettel, 2013; Carter et al., 2012). This area has been associated with socially predictive behavior, a component of the trust areas measured in human-agent teaming studies.

Building on this neuroscientific foundation, our concept of rhythmic predictability was inspired by the overlap between trust and reliability. Observations of drummers demonstrate their ability to correct for both their own and collective rhythmic predictability or unpredictability, highlighting the need for adaptive systems to better anticipate musical decisions (Bretan and Weinberg, 2016; Hopkins et al., 2023a). In human-agent teaming studies, reliability when tailored for musical contexts, serves as a cornerstone for establishing trust (Glikson and Woolley, 2020; Hoff and Bashir, 2015; Howell-Munson et al., 2022; Yuksel et al., 2017; Eloy et al., 2022, 2019).

In the musical interaction domain, this form of reliability becomes crucial. Unpredictable or unreliable musical performance can erode the trust that is essential for successful musical collaborations (McCormack et al., 2019, 2020). Musicians, therefore, aim to establish a high level of rhythmic reliability to build trust with their fellow musicians. This involves a delicate balance: being predictable enough to maintain the structural integrity of the musical collaboration, while also introducing elements of unpredictability and improvisation to sustain engagement and interest among the participants.

## 2.5 Functional near-infrared spectroscopy

fNIRS systems can employ different techniques of illumination (Gervain, 2015), with continuous-wave (CW) fNIRS representing the most frequently adopted approach in cognitive neuroscience. In this approach, near-infrared light at two or three different wavelengths is constantly emitted from sources into the scalp, and the light collected by detectors indexes changes in concentrations of oxygenated ( $\Delta\text{HbO}_2$ ) and deoxygenated ( $\Delta\text{HbR}$ ) hemoglobin in the brain.

These local changes in concentrations of  $\text{HbO}_2$  and  $\text{HbR}$  are computed from light intensities at these wavelengths using the modified Beer-Lambert Law Delpy et al. (1988); Kocsis et al. (2006), which accounts for light absorption and scattering in biological tissue:

$$\Delta A = \Delta c \cdot \varepsilon \cdot d \cdot B,$$

where  $A$  is light attenuation,  $\varepsilon$  is the absorption coefficient of the chromophore (e.g.,  $\text{HbR}$ ),  $c$  is the concentration of the

chromophore,  $d$  is the distance between the points where light enters and exits the tissue (approximately 3 cm), and  $B$  is the differential pathlength factor (DPF), which accounts for the effect of scattering on pathlength. Pairs of emitters (i.e., sources) and detectors form multiple NIRS channels.

Cross-validation of fNIRS as a neuroimaging method: fNIRS has been cross-validated not only in terms of the temporal characteristics of hemodynamic changes but also the spatial localizations of these changes. For example, several combined fNIRS-fMRI studies have shown that these methodologies are highly comparable in this respect (e.g., Cui et al., 2011; Heinzel et al., 2013; Noah et al., 2015; Okamoto et al., 2004; Sato et al., 2013). For example, Noah and colleagues (2015) developed a protocol for conducting multi-modal experiments with fNIRS and fMRI to ensure signal comparability, testing it using a complex yet naturalistic motor task, namely a dancing video game (Noah et al., 2015). Thus, fNIRS is a valid neuroimaging method that has temporal and spatial resolutions which represent an adequate compromise between that of fMRI and EEG, respectively; that is, it has greater temporal resolution than fMRI, but not EEG, and greater spatial resolution than EEG, but not fMRI (Ferrari and Quaresima, 2012).

### 2.5.1 “Real-world” neuroimaging

There has been a considerable and rapid rise in technological advancements to the wearability and portability of fNIRS in recent years (see Pinti et al., 2018a for review). These systems enable participants to freely perform tasks without constraints on the body and researchers to investigate situations that are difficult to contrive in laboratory settings (e.g., interpersonal interactions), providing an unprecedented opportunity to study complex cognition more naturalistically (e.g., Hirsch et al., 2018; Stuart et al., 2019). For example, researchers have recently investigated prefrontal hemodynamics whilst participants walked around a real-world street environment (Burgess et al., 2022; Pinti et al., 2015) and others have investigated alterations in hemodynamics during slacklining (Seidel-Marzi et al., 2021). With respect to interpersonal interaction, Pinti et al. (2021) assessed face-to-face deception in interacting dyads, and Kelley et al. (2021) recently compared brain activity during eye-contact interactions with a humanoid robot. In other words, the neuroscientific questions for which fNIRS is particularly well-suited are those predicting the involvement of brain regions in ecological tasks requiring unrestricted movement, human-to-human interaction, and so forth (Pinti et al., 2018b).

### 2.5.2 Hemodynamic considerations in naturalistic situations

Because greater ecological validity increases situational complexity, tasks that represent real-world activities will often require behaviors such as free-movement actions and language. Behaviors such as speech production (i.e., verbal communication) during interpersonal interactions and intentional body movements during complex motor tasks (e.g., playing an instrument) create marked changes in respiration compared to resting states (e.g., Scholkmann et al., 2013b,a). Since respiration (e.g., changes

in arterial CO<sub>2</sub>) is closely linked to cerebral oxygenation and hemodynamics, task-related changes in physiological systems such as respiration and heart rate represent sources of noise (e.g., system confounds) in fNIRS signals (Tachtsidis and Scholkmann, 2016). Filtering techniques during pre-processing, such as the use of band-pass filters together with multi-modal monitoring, typically address this issue in normal, laboratory-based experiments (see Pinti et al., 2019 for review); however, researchers have found that HbR signals are typically less affected by systemic confounds (Dravida et al., 2017) and, therefore, have used this type of signal for interpreting results in ecologically valid neuroimaging paradigms (e.g., Crum et al., 2022; Hirsch et al., 2021, 2017).

## 2.6 Affective state classification in machine learning

fNIRS has emerged as a non-invasive, lightweight neuroimaging modality that offers significant advantages over traditional methods like electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) (Ayaz et al., 2022). Its portability and wearability have been significantly enhanced in recent years, enabling naturalistic studies in complex cognitive and social settings (Pinti et al., 2018a; Hirsch et al., 2018; Stuart et al., 2019). These advancements have found applications in various domains, including human-agent teaming and musical performance (Eloy et al., 2022; Howell-Munson et al., 2022; Vanzella et al., 2019; Hopkins et al., 2023a).

Concurrently, machine learning techniques, particularly deep learning architectures, have revolutionized affective state classification (Bandara et al., 2019). Convolutional Neural Networks (CNNs) excel at capturing spatial characteristics of neuroimaging data, while Long Short-Term Memory (LSTM) networks are adept at learning temporal patterns (Rabbani and Islam, 2023; Wang et al., 2023). Support Vector Machines (SVMs), known for their user-friendliness and wide availability, are also commonly employed (Yuksel et al., 2016, 2019; Wang et al., 2023). The synergy between fNIRS and machine learning has been particularly impactful. For instance, CNNs and LSTMs have been combined to improve the accuracy of time-series classification of fNIRS data (Bandara et al., 2019), while SVMs have been utilized for workload adaptation in fNIRS-based systems for music (Yuksel et al., 2016, 2019; Miranda, 2021).

However, the generalizability of these machine learning models across different settings and populations remains a concern (Maleki et al., 2022; Krois et al., 2021). To address this, there is a growing call for more rigorous validation methods and diverse datasets. One promising avenue is to focus on intra-individual variability by collecting data from fewer participants but under varied conditions (Jankowsky and Schroeders, 2022; Rybner et al., 2022). This integrated approach not only leverages the strengths of both fNIRS and machine learning but also addresses the challenges of model generalizability, thereby offering a comprehensive solution for affective state classification in complex cognitive and social interactions.

## 3 Materials and methods

### 3.1 Research methods

To address the challenge of measuring and implementing rhythmic predictability in real-time musical improvisation, we adopted a two-phase approach informed by a 40-hour field study of observed jam sessions. In phase I, we experimentally identified correlates of rhythmic predictability using fNIRS, formalizing it as a measurable attribute both neurologically and computationally.

We trained individual models for each participant following a rigorous standard fNIRS experiment protocol. We established three experimental conditions in which the participant was asked to improvise on a drum pad with an embodied virtual AI musician: 1) improvise with musical beats, 2) improvise with a steady rhythm at 90 beats per minute [a tempo that can easily be followed (Hopkins et al., 2022a; Leman, 2007)], and 3) improvise with random rhythms generated computationally randomly. Additionally, we established a control condition whereby the participant played along with a steady pulse at 90bpm. Between conditions participants stared at a fixation cross on screen for 10 seconds.

Throughout the experiment, we split the incoming fNIRS signal and captured it both in CSV files using Python and in Aurora 2021.9 software for *post hoc* analysis. The CSV files were later used to train and compare several machine learning models, with the best performing models chosen to be used in the system in phase II. The fNIRS data captured in Aurora was used to verify that our signal was properly acquired and provide preliminary evidence for activation in specific regions of the brain during experimentation.

In phase II, we developed “BrAI Jam,” a system that integrates this metric into real-time interactions between human drummers and an AIVM using individualized machine learning models trained on data from the participant players. Participants played drums with three versions of the AIVM. The first version used dynamic switching of drum beats based on the brain activity of the participant. The second used continuous switching of the drum beat with no neural feedback. Lastly, as a control condition, participants mimicked an AIVM that played a consistent steady pulse at 90bpm. We then conducted semi-structured interviews and statements to gain insights into their experience with the system.

Further description of methodologies and hypotheses for phase I and phase II of the experiment are discussed in Section 5 (Experimental Design). In the remainder of this section, we describe a formative field study, the BrAI Jam system design, and equipment used.

### 3.2 Field study: observations of collaboration between human musicians

#### 3.2.1 Field study methods

To inform the design of the BrAI Jam system, we recorded over 40 hours of jam sessions with 5 musicians (ranging from 25 to 31 years of age) over the course of several months to better understand how musicians communicate with each other in an improvised setting. The recordings were made using several web

cameras recording at 1080p and using OBS studio to combine angles into a single video for *post hoc* analysis.

The group consisted of two guitarists, a bassist, a keyboardist, and a drummer. All members of the group had more than five years of experience playing their respective instruments. The group regularly engaged in jamming sessions, typically meeting once or twice a week. When jamming, the group played music extemporaneously. That is, it was never rehearsed, and known songs were rarely—if ever—played.

On some occasions, the first author joined some of the jam sessions to better understand the group dynamics through ethnography. Notes were taken after the sessions and video was later analyzed by two other members of the research team. The video analysis focused on how players communicated with each other and how they responded when recognizing collective musical incohesion.

#### 3.2.2 Observations

The dynamics of communication among improvising musicians has been well explored. Notably, musicians tend to use body language, eye contact, head motion, and facial expression (Hopkins et al., 2022b; Bretan and Weinberg, 2016) to communicate with one another during a jam session.

Our observations surfaced a collective behavior pattern the musicians used when recognizing the music was suffering (i.e. melodically, rhythmically) due to incohesive musical actions between the players. They tended to use three strategies to overcome this issue. First, they would make eye contact or visibly demonstrate their confusion. Secondly, they would simplify the music playing only necessary notes/chords and rhythms to convey the main musical motif or rhythm. Lastly, they would exaggerate gestural cues such as rhythm with the body and simplify visual cues (such as chord shapes on guitar) to better synchronize the group.

We concluded that playing extemporaneously inevitably led to moments of miscommunication or incohesion between members. Some members even noted this as a desirable trait of an improvised jam session. Whereby players could “find where to meet musically”.

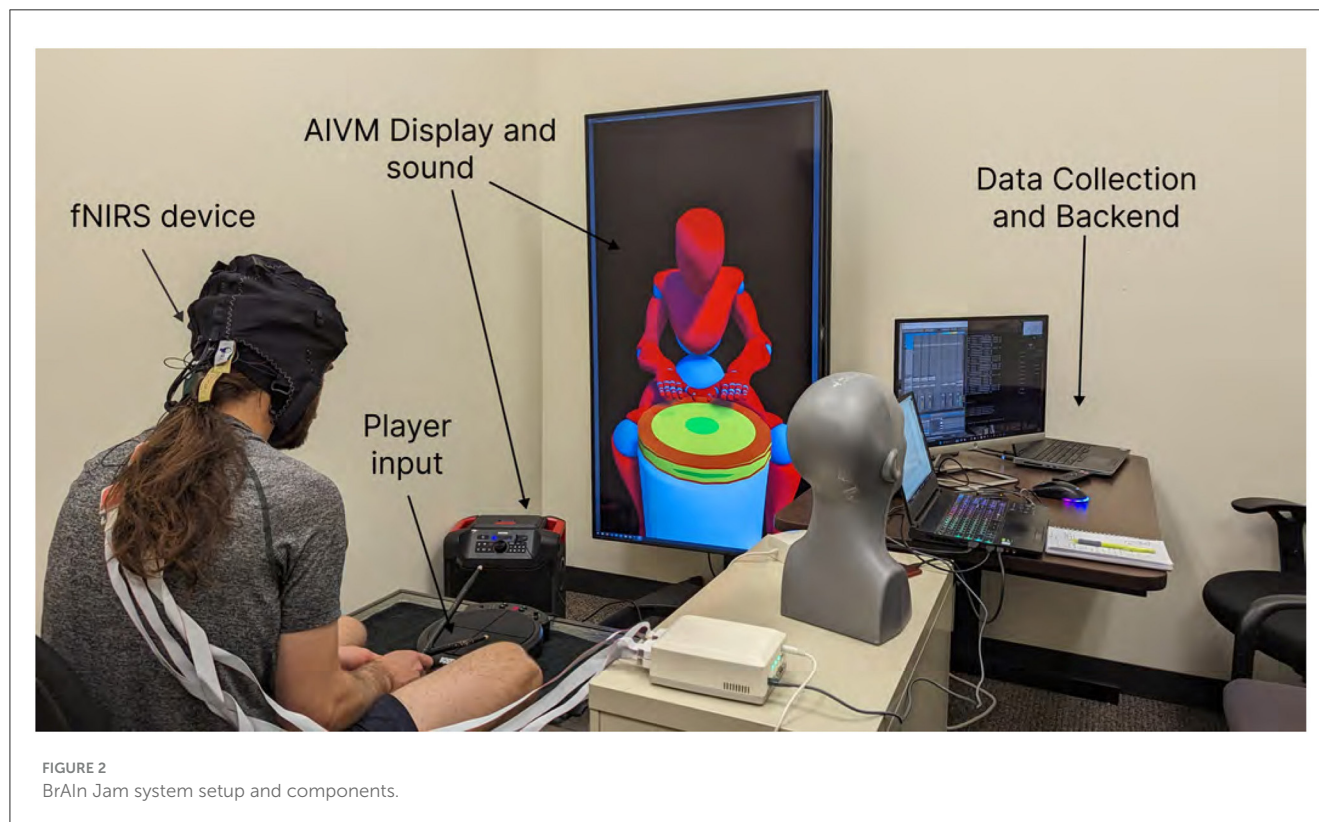
We decided to focus on musical incohesion and unpredictability due to its observed relevance in jam sessions, supporting frameworks from previous work in the neuroscience of rhythm, and because of an opportunity to better inform AI-driven musicians of human dynamics during musical play.

### 3.3 Equipment and system design

#### 3.3.1 AI-Driven Virtual Musician Equipment

We received musical input from the player using a KAT MIDI drum that was directly connected to the AIVM. The AIVM system we used was derived from an open-source AIVM system (Hopkins et al., 2023b,d) incorporating the Unity Game Engine and Ableton Live 11. A Python backend running Google Magenta’s LSTM RNN was used to generate new drum beats for the system, trained on Google Magenta’s Groove dataset (Gillick et al., 2019). The AIVM was displayed on a vertically oriented TV screen in front of the drummer while they played (see Figure 2).





### 3.3.2 fNIRS equipment

The present study adopted a 50-channel layout that was configured using FOLD Toolbox (Zimeo Morais et al., 2018) to collect data from subregions of the prefrontal cortex and of more posterior, temporoparietal regions that were identified from previous literature discussed above (see Figure 3). More specifically, data was collected from two participants using a portable NIRxSport2 system (NIRx, Berlin, Germany) (see Figure 4). This is a CW-fNIRS device that uses two wavelengths (760 and 850nm) with a sampling rate of 5.0863 Hz.

### 3.3.3 fNIRS data

Real-world tasks often involve free-movement actions and language, introducing changes in respiration and heart rate (Scholkmann et al., 2013b,a). These physiological changes can introduce noise in fNIRS signals (Pinti et al., 2019). While filtering techniques are used in laboratory experiments (Pinti et al., 2019), HbR signals are typically less affected by confounds (Dravida et al., 2017), making them suitable for ecologically valid neuroimaging (Crum et al., 2022; Hirsch et al., 2021, 2017). However, real-time adaptive systems have mainly focused on HbO2 (Howell-Munson et al., 2022; Yuksel et al., 2016, 2019).

To address this, we developed a real-time preprocessing algorithm applying a linear bandpass filter (0.01 - 0.5 Hz) and the modified Beer-Lambert law. Our analysis focused on HbR data for a cleaner signal, collected for training the real-time model. Both *post hoc* fNIRS analysis and real-time data were collected.

NirsLAB software (Version 2019.04, NIRx) was used for data processing. Time series for each participant were adjusted to

include 5 s before the first round and 30 seconds after the final round, capturing the full hemodynamic response. Data quality was assessed using the coefficient of variation (CV), with channels having elevated CV values inspected for motion artifacts or poor optode-scalp contact. A CV threshold of 15% was employed, aligning with standard practices (Piper et al., 2014; Pfeifer et al., 2018), following guidelines by Yücel et al. (2021). A pre-whitening autoregressive model-based algorithm was applied to correct for motion and serially correlated errors (Barker et al., 2013; Yücel et al., 2021), without prior traditional filtering.

Subsequent analysis used a General Linear Model on round-level fNIRS data for HbO2 and HbR, resulting in 'beta' values quantifying the fit between observed brain activity and the expected hemodynamic response function. Each NIRS channel was individually fitted for each participant, yielding unique beta values. Real-time HbO2 data produced unusable results due to significant extra-cortical physiological activity caused by vigorous drumming. Contrast analysis in NirsLAB identified brain regions with significant variations in activation across conditions. For a comprehensive understanding of fNIRS analyses using the general linear model, refer to Barker et al. (2013); Tak and Ye (2014); Yücel et al. (2021).

### 3.3.4 System diagram

The BrAln Jam system is comprised of an fNIRS device and associated data processing programs, Python scripts, and Unity Engine. Figure 5 shows how these components are connected and send information to each other. The system diagram is separated



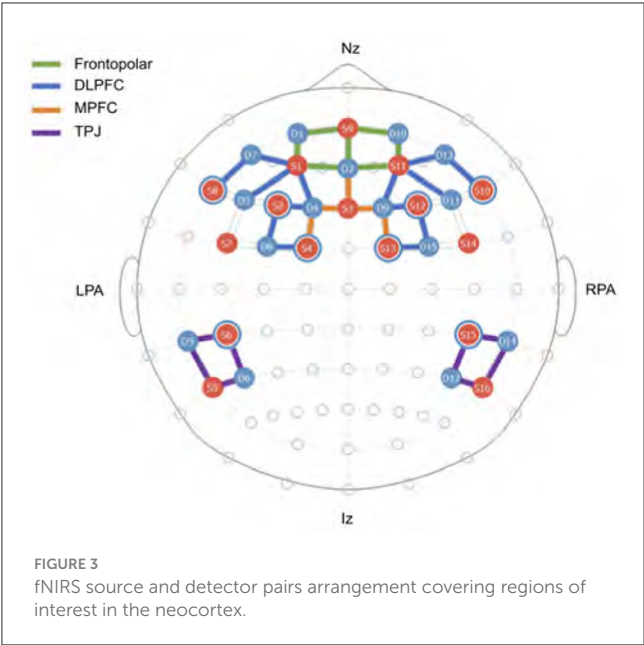
into two components: (1) the real-time system, and (2) *post hoc* analysis software.

The real-time system comprises two inputs—an fNIRS device and a MIDI drum pad. The MIDI drum pad is used to capture drumming information from the player, which is in turn used to inform the generation of new drum beats played by the virtual drummer. The fNIRS device is used to stream information that is passed to a machine learning classifier which determines whether the music is sufficiently rhythmically musical. If the classifier determines that the music is not, then it manipulates the drum beat at a musically relevant interval by launching a new clip on the downbeat using Ableton’s built-in launching function. All

generated clips are instantiated in Ableton to control for tempo, sound quality, and timing of musical changes.

Tempo control is accomplished using the global tempo setting in Ableton. This ensures that the notes for all generated drum beats are not faster or slower than any other. This avoids random speeding or slowing of the tempo between beats, which could distract the drummer and render the conditions incomparable. Sound quality and timbre are controlled using the same instrument pack for all conditions in Ableton, ensuring a consistent and realistic drum sound is produced for all drum beats. Ableton then sends OSC messages to Unity which controls the animation of an embodied virtual AI drummer.

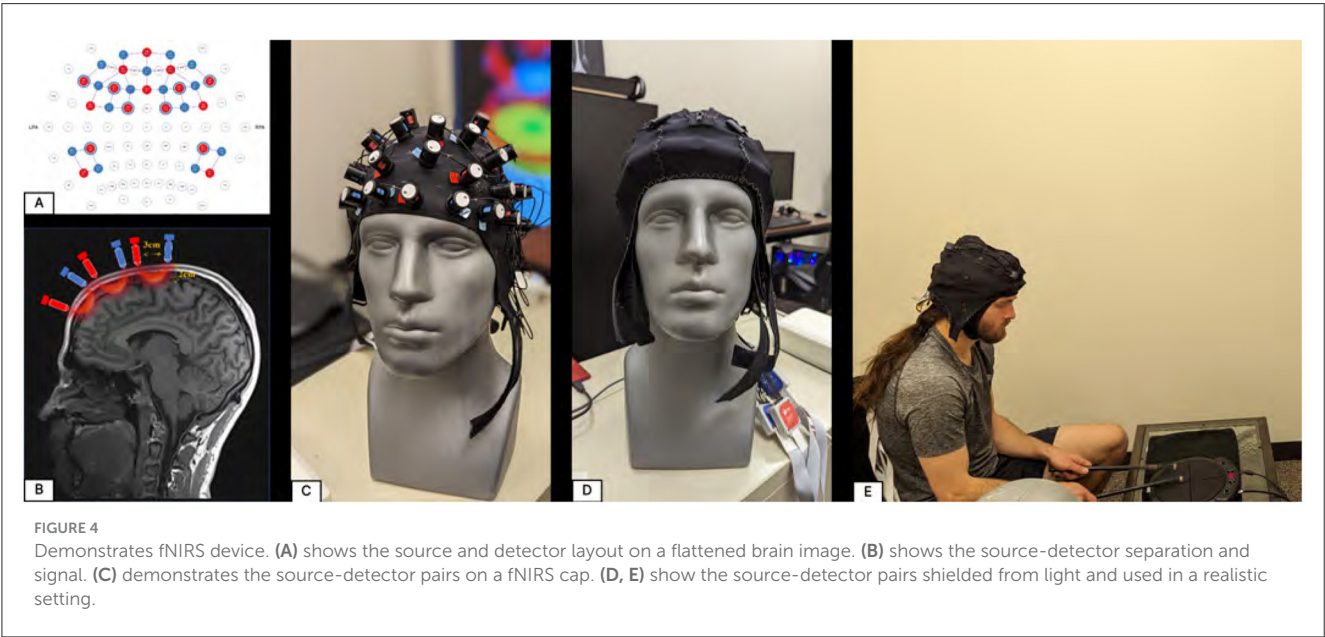
The *post hoc* system comprises nirsLAB and Python code which converts incoming open sound control (OSC) messages to arrays that populate a CSV file. These files can be manually verified, as well as used for model training in post. The nirsLAB software is used to supply evidence that the acquired signals were clean and to determine what areas of the brain were most active during the training periods.

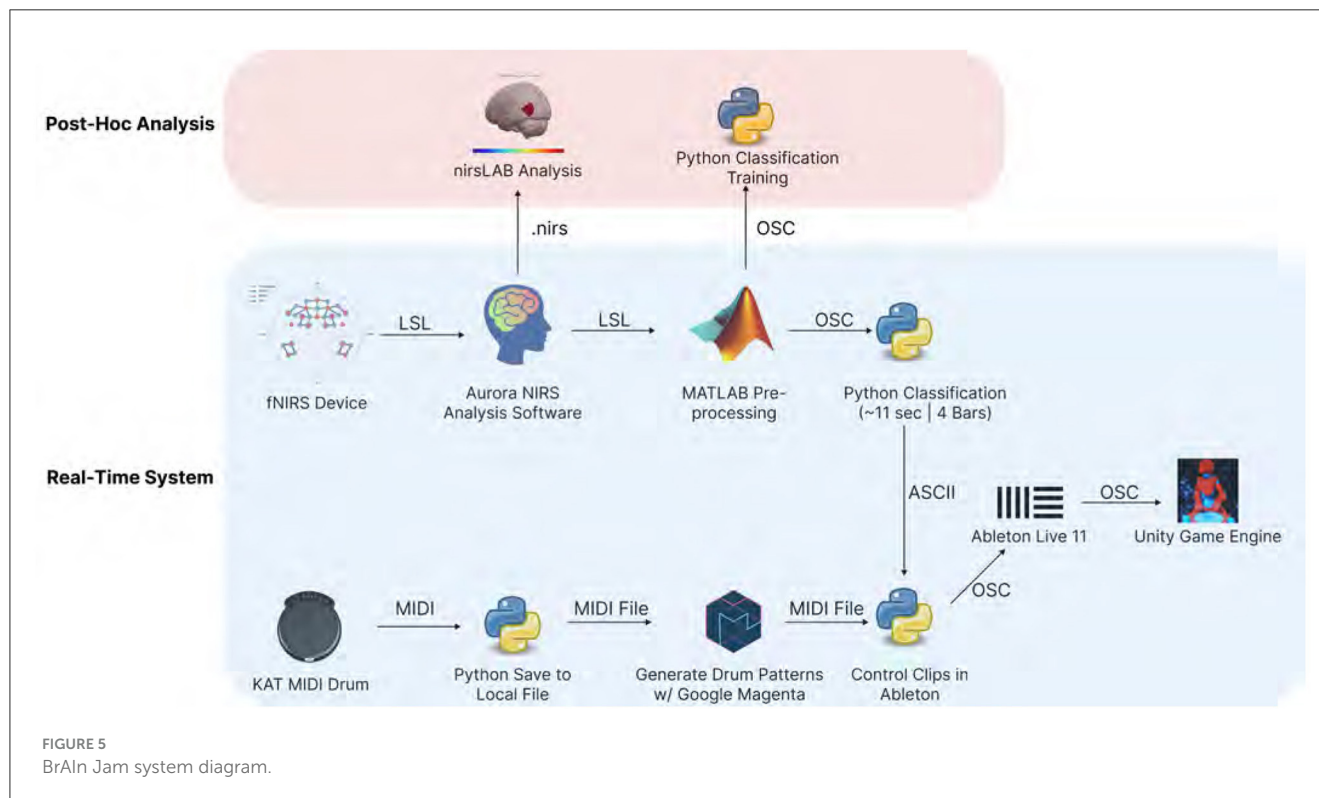


## 4 Experimental design

### 4.1 Hypotheses and research objectives

The experimental framework of this study is divided into two distinct phases, each designed to rigorously investigate specific aspects of human-AI musical interaction and the neural correlates of rhythmic predictability. In phase I, we explore the neural correlates of rhythmic predictability by presenting musicians with three distinct auditory conditions to improvise with: musical phrases, random phrases, and a steady pulse. In phase II, we demonstrate the dynamic neurofeedback system and a non-dynamic system to musicians and gather feedback on how this





technology can be used in practice. The research is guided by three primary hypotheses:

#### 4.1.1 Hypotheses

1. **The first hypothesis** posits that the neural responses to musical phrases and the steady pulse will differ in a discernible manner from those elicited by random phrases. This is premised on the idea that musical phrases, being inherently structured, will engage different cognitive and affective processes compared to random, unstructured phrases.
2. The second hypothesis predicts that the best performing machine learning model will be the RNN LSTM hybrid model used to previously predict fNIRS activation in a musical task (Bandara et al., 2019).
3. The third hypothesis postulates that detecting low rhythmic predictability and dynamically changing the drumbeat will result in greater perceived rhythmic cohesion with an embodied virtual drummer as compared to continuous switching of a drumbeat without feedback.

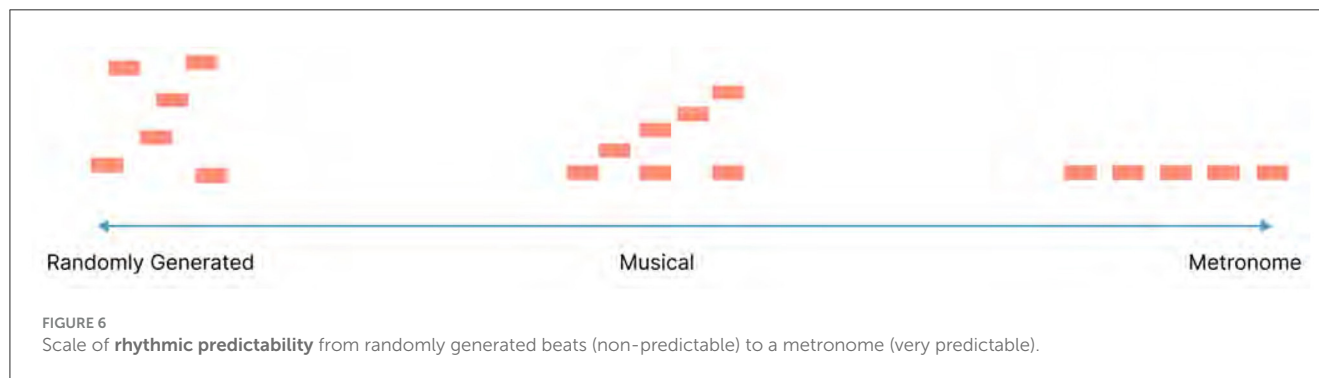
## 4.2 Phase I: neural correlates and model training

In the initial phase, participants engaged in musical improvisation using an electronic drum pad, accompanied by an AIVM whose actions were triggered by pre-recorded hand drum beats. This phase incorporated three experimental conditions along with a control condition.

A virtual musician system (Hopkins et al., 2023b) was implemented to provide an embodied representation of the AI. This was important to ensure elements of human-agent teaming studies remained consistent in the investigation of trust (Yuksel et al., 2017; Bobko et al., 2022; Eloy et al., 2022). A virtual musician system was used to accomplish anthropomorphic embodiment without the embodiment cost, latency, and resources associated with robotics (Bretan and Weinberg, 2016).

Participants were exposed to four conditions interleaved with a 10 sec fixation cross to enable hemoglobin signals to return to baseline between trials (Hopkins et al., 2023a; Crum et al., 2022). Participants engaged in the following conditions and were asked to improvise on the electronic hand drum, each corresponding to several levels of rhythmic predictability (see Figure 6):

1. **Condition 1: Randomly generated drum beats.** Randomly generated beats were programatically randomized for note and time step. This made drumbeats unpredictable to the musicians.
2. **Condition 2: Pre-selected musical drum beats.** These drum beats were generated by an Long Short-Term Memory Recurrent Neural Network (LSTM RNN) trained on 1,200 drum samples in the Magenta Groove data set (Gillick et al., 2019). After generation, 200 drum beats were hand selected by a music educator for their predictability.
3. **Condition 3: Pulse-based metronome beat.** A regular drum beat at 90bpm was used to simulate a steady and reliable drum beat.
4. **Control Condition: Mimic a metronome with the AIVM.** This control condition enabled us to compare the other conditions against a physically similar, rhythmic exercise with the AIVM, to ensure brain activation in the regions of interest were confined to aspects of rhythmic reliability rather than the control factors.



#### 4.2.1 Machine learning model comparison

A primary objective of this experiment was to scrutinize the neural correlates associated with rhythmic predictability and to address how affective state can be incorporated into a real-time improvisation system with AIVMs. To address the bias-variance trade-off in machine learning, we opted for a data-rich approach. Rather than recruiting a large number of musicians and collecting limited data from each, we selected two experienced improvising musicians (Age:  $M = 28$ ,  $SD = 1$ ; Years playing:  $M = 16.5$ ,  $SD = 6.5$ ) and exposed them to the conditions for three hours each, yielding a cumulative six hours of training data. This amount of data is consistent with averages in capture time by many fNIRS experiments as surveyed by Eastmond et al. (2022).

We employed this strategy to mitigate the limitations associated with cross-validation, which often compromises the development of individualized models (Krois et al., 2021; Rybner et al., 2022; Jankowsky and Schroeders, 2022; Maleki et al., 2022; Rocks and Mehta, 2022; Zhang et al., 2022). While a larger participant pool would enhance the statistical generalizability across subjects, our focus was on obtaining the most reliable data possible—not cross validation. Therefore, we maximized the data collection time per participant to ensure a more nuanced understanding of the neural correlates of rhythmic predictability.

The primary objective of the implemented model is to classify time-series data of musicians improvising with one of three specified conditions *Random*, *Musical*, and *Metronomic* music. The time-series data is chunked into overlapping segments, each of size approximately 1,900 data points with an overlap of 680 data points. This corresponded to 34 channels, 11 s (4 bars of music at 90bpm, time signature 4/4), and 4 s (hemodynamic response time) respectively at a sampling rate of 5 Hz. Label encoding is employed to convert these string labels into integers, facilitating machine learning model training. Following this, the data is split into training and test sets, with 80% of the data being used for training and the remaining 20% for testing. Several models were then implemented for a comparative analysis:

- An SVM model [a commonly used model for fNIRS classification (Yuksel et al., 2016, 2019)] with a linear kernel and regularization parameter  $C = 0.1$  is trained on the segmented and scaled data.
- A Random Forest classifier with 50 estimators and a maximum depth of 10 is also trained on the same data.
- A hybrid CNN and LSTM model is employed [based on a recent classification strategy for fNIRS data (Bandara et al.,

2019)], consisting of a 1D convolutional layer with 64 filters and a kernel size of 3, followed by max-pooling and an LSTM layer with 64 units. Dropout layers are inserted for regularization. The model employs sparse categorical cross-entropy loss and is optimized using the Adam optimizer. Early stopping is employed during training to prevent overfitting.

- A simple dense neural network model consisting only of dense layers is trained. This model serves as an additional baseline for comparison.
- A standalone LSTM model with 64 units is also trained, serving as another variant for comparison.

Performance metrics, including accuracy, precision, recall, and F1-score, are computed for each model post-training. This comprehensive approach gives an in-depth comparison of different machine learning algorithms, offering valuable insights into their respective efficacies for fNIRS classification.

#### 4.3 Phase II: interaction with AI-driven virtual musician

In the second phase, participants engaged with an AIVM under two experimental conditions, complemented by a control condition. The aim of this phase was twofold: to assess the model's capability to accurately predict brain states and to explore the integration of rhythmic predictability classification in real-time systems, such as AIVMs.

1. **First Condition: dynamic switching with fNIRS**—This condition utilized functional near-infrared spectroscopy (fNIRS) to dynamically control the switching of AI-generated drum beats.
2. **Second condition: continuous alteration**—In contrast to the first, this condition did not employ intelligent switching but instead featured a continuous alteration of AI-generated drum beats.
3. **Control condition: metronome mimicry with AIVM**—This control condition served as a baseline, mimicking a metronome with the AIVM to isolate brain activation related to rhythmic reliability from other factors.

To assess the effectiveness of the different conditions, each participant was interviewed individually after interacting with the AIVM. The interviews were semi-structured and focused on the



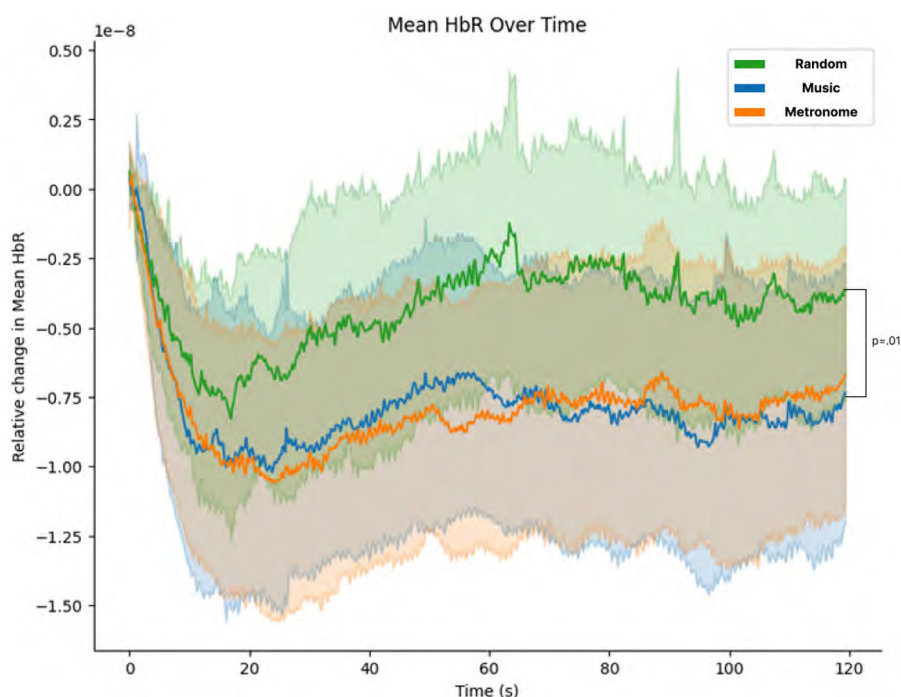


FIGURE 7

Mean HbR pre-processed signal overtime using our real-time algorithm. Blue and orange lines indicate musical and metronome conditions respectively, while the green line represents playing to randomly generated drum beats.

differences between the two conditions and overall impressions of the system. These one-on-one discussions enabled a more in-depth exploration of the participants' experiences than could be achieved through surveys or other quantitative methods. Interviews were recorded, transcribed, and analyzed separately by two members of the research team using inductive thematic analysis.

## 5 Results

### 5.1 Phase I results

#### 5.1.1 Results for hypothesis 1

The first hypothesis posited that the neural responses to musical phrases and a steady rhythmic pulse will differ discernibly from those elicited by random phrases. This is premised on the idea that musical phrases—being inherently structured—will engage different cognitive and affective processes compared to random, unstructured phrases. Our results support this hypothesis as demonstrated in Figures 7 and 8. The areas largely associated with differences in each subject are the TPJ, which are also known to be associated with social cognition, theory of mind, and predictive social planning (Carter and Huettel, 2013; Carter et al., 2012; Schurz et al., 2017). Our results also support this hypothesis as demonstrated in Figure 9. Dorsolateral prefrontal cortex activation is also shared in the musical and metronome improvising conditions, while not present in the random condition. This suggests that the DLPFC is in part responsible for discerning musical and regular patterns in music.

#### 5.1.2 Results for hypothesis 2

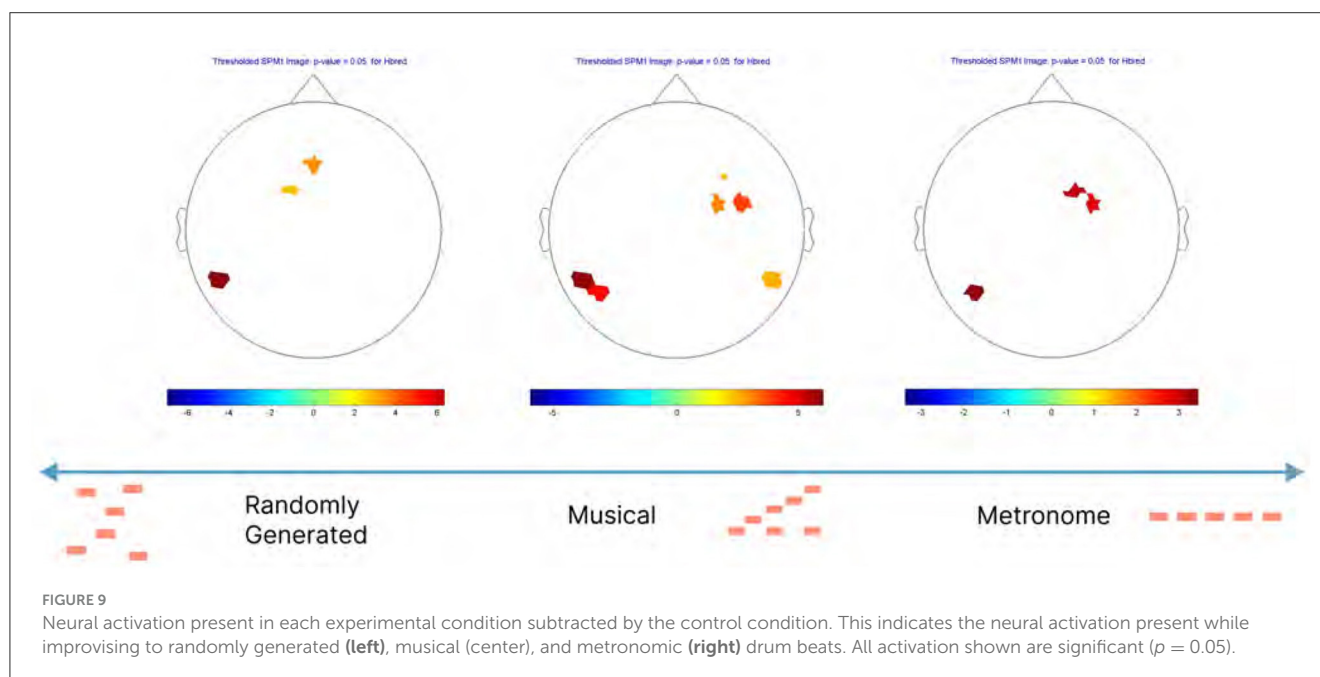
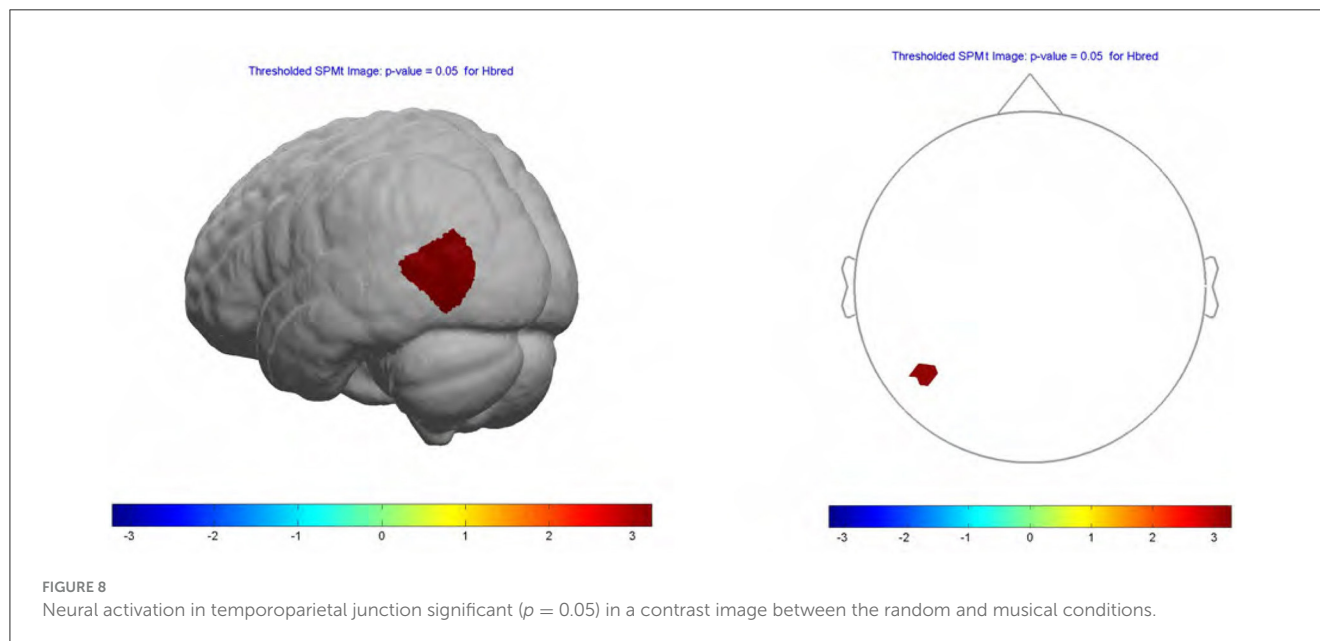
The second hypothesis predicted that the CNN+LSTM RNN model would perform the best based on prior work. This was not supported by our results. Our machine learning model comparison demonstrates that accuracy, precision, recall, and F1 score all scored higher for the random forest (RF) model (Table 1). The random forest model performed better on all scores and also demonstrated less variance between participants.

#### 5.1.3 Measuring rhythmic predictability in code

The HbR concentration data collected from the real-time data stream was averaged for both of the players over all of the channels. Mean standard error was applied to each of the conditions over the duration of each condition and were shown to be distinct. A t-test was performed between each of the conditions, and as hypothesized there was a statistically significant ( $p = 0.01$ ) difference between the randomly generated beats and the musical ones (see Figure 7). This suggests that “rhythmic predictability” can be measured while improvising on the drum.

#### 5.1.4 Measuring rhythmic predictability in nirsLAB

To further specify the regions associated with the difference in signal a thresholded  $t$ -test ( $p = 0.05$ ) was performed between each of the conditions and the control condition (tapping a drum to a metronome). This suggests that the difference in signal between improvisation and tapping to a metronome with an AIVM significantly raises activation in the regions of the



brain shown in Figure 9. To summarize, improvisation elicited signal in the temporoparietal junction for all conditions as compared to controls. Also, areas of the brain associated with workload (frontopolar area) became significantly more active when improvising to random drum beats as compared to the other conditions. Activation in the right dorsolateral prefrontal cortex and medial prefrontal cortex in addition to temporoparietal junction activation were active in the metronome condition as compared to the control. Interestingly, the musical condition elicited the most activation in the temporoparietal junction of all of the conditions with additional activation in the right dorsolateral prefrontal cortex.

When comparing random generated beats vs musical beats there is even more evidence of temporoparietal junction activation.

Figure 8 demonstrates statistical significance ( $p = 0.05$ ) in activation of the left temporoparietal junction (TPJ). Figure 8 also demonstrates this on a 3 dimensional representation of the brain area. This suggests that when musicians are improvising the TPJ is very active compared to playing with a metronome, and when beats become perceivably musical, activation in this region increases again significantly. We discuss the implications of these findings in the discussion section.

### 5.1.5 Machine learning results

In the evaluation of various machine learning models (see Table 1), the Random Forest (RF) algorithm consistently outperformed other models across all metrics for both Player 1

TABLE 1 Performance metrics for various models.

Metric	Model				
	Vanilla LSTM	Dense Neural Network	CNN + LSTM	SVM	RF
<b>Player 1</b>					
Accuracy	0.9	0.898	0.9	0.9	<b>0.932</b>
Precision	0.7222	0.9235	0.924	0.8913	<b>0.936</b>
Recall	0.7396	0.9240	0.9	0.8894	<b>0.932</b>
F1 Score	0.7282	0.9235	0.903	0.8854	<b>0.932</b>
<b>Player 2</b>					
Accuracy	0.906	0.935	0.809	0.889	<b>0.949</b>
Precision	0.9240	0.9134	0.861	0.9240	<b>0.951</b>
Recall	0.9000	0.9021	0.809	0.9000	<b>0.949</b>
F1 Score	0.9029	0.9040	0.814	0.9029	<b>0.950</b>

Bold figures indicate the highest value per row of data.

and Player 2. Specifically, for Player 1, RF achieved the highest scores in Accuracy (0.932), Precision (0.936), Recall (0.932), and F1 Score (0.932), thus exhibiting the most robust performance on this dataset. The high Precision (0.936) indicates that the Random Forest model reliably avoided false positives, meaning that nearly all instances classified as positive were genuinely positive. The similarly high Recall (0.932) shows that RF effectively captured most actual positive cases, missing very few. Similarly, for Player 2, RF again led in Accuracy (0.949), Precision (0.951), Recall (0.949), and F1 Score (0.950), reinforcing its effectiveness. The consistently high Precision (0.951) and Recall (0.949) for Player 2 further underscore RFs capability to accurately identify true positives while minimizing incorrect classifications.

While generalization across multiple individuals was not the primary focus of this study, the results do indicate some intriguing findings regarding model performance and variability across different datasets. The Vanilla LSTM model exhibited significantly lower Precision (0.7222) and Recall (0.7396) for Player 1, indicating frequent incorrect positive predictions (false positives) and simultaneous misses of true positives (false negatives). These results suggest potential underfitting, where the model inadequately represented the complexity of the data, failing to capture critical patterns.

Conversely, the CNN/LSTM hybrid model demonstrated a notable performance decrease in Precision (from 0.924 for Player 1 to 0.861 for Player 2) and Recall (from 0.900 for Player 1 to 0.809 for Player 2). Such drops imply generalization issues, indicative of overfitting. Although not initially aimed at exploring generalization, these results highlight the variability in model effectiveness across different individuals, offering insights into the possible limitations of generalized models.

Importantly, obtaining a substantial amount of data for each individual proved beneficial in addressing the bias-variance tradeoff inherent in model training (Zhang et al., 2022). The individualized datasets enabled the Random Forest model to achieve a balance between bias and variance, demonstrating

both high Precision and Recall consistently. These findings underscore some advantages of individualized data collection and suggest that personalized or individualized models might offer performance benefits by capturing unique patterns within specific datasets. Consequently, the Random Forest model emerges as the most reliable and effective choice for both players.

## 5.2 Phase II results

### 5.2.1 Hypothesis 3

The third hypothesis predicted that dynamic switching to simplified metronomic beats would result in greater cohesion between the AI and human player. We also received mixed feedback on the system and the implementation of the adaptive signal. Though our hypothesis was unsupported, our qualitative results led to a more nuanced view of how brain signals can be used for real-time adaptive musical improvisation systems.

### 5.2.2 Qualitative insights

Both participants reported enjoying musical collaboration using BrAIn Jam in both conditions, with and without fNIRs-informed dynamic switching. Furthermore, both participants experienced differences with the system in the two conditions. Two researchers independently extracted themes from the transcribed interviews using grounded theory. These were thoroughly discussed and agreed upon between the two of them. Phrases that supported the themes were taken directly from the transcription for clarity. We extracted the following themes from the qualitative surveys and interviews:

- **Synchronization.** Both participants emphasized the ability to synchronize with the AIVM as a major factor in the experience. Participant 1 reported an easier time staying in sync with the AIVM in the first condition because when it changed beats to the simpler pattern, “it was easier to find the rhythm again and get back in sync again.” Participant 2 reported that it was tough to get in sync with the AIVM in the affective classification condition because when they were adjusting to the AIVM, it would change again. These experiences support that dynamic switching can facilitate synchronization between the human musician and AIVM for a more seamless musical collaboration. However, if the dynamic switching were not applied appropriately, for example, if it occurred too frequently or with poor timing, it could be more demanding of the player to work to synchronize to the musical change.
- **Responsiveness.** Both participants reported experiencing that BrAIn Jam with affective classification was responsive whereas the one without classification was not. According to participant 1, the AIVM in condition 1 “seemed like it was adapting to what I played.” On the other hand, the AIVM in condition 2 “wasn’t dependent on what I was doing at all.” Participant 2 reported that while BrAIn Jam in condition 1 was more responsive, it did not always respond in a way



that supported the music making, “like it was listening to you but wasn’t making the right decision.” These results suggest that using fNIRS classification, we were able to make our system more responsive, which is important in a musical collaboration context; however, the responses can be further refined through improving the musical fidelity and affective classification accuracy.

- **Consistency.** Both participants mentioned that BrAIn Jam was not very consistent in its musical patterns in both conditions, as in, it didn’t sustain the same beat pattern for long periods of time. However, participant 2 reported the system without affective classification was more consistent. This could be a result of frequent beat switches due to detection of low trust states. Although the possible difference in consistency between 2 conditions was not intentionally designed, these insights reveal that this is an implicit factor in designing adaptive systems where the system behavior changes depending on the player’s state, and can contribute to experiences of reliability. This element may be particularly noticeable depending on the interaction task, such as in musical collaboration. Furthermore, both participants mentioned that they would like BrAIn Jam to “play the same thing for longer” overall. Participant 1 comments, “even if it may be a bit more repetitive, it may be less cognitively demanding than changing every 20-30 seconds.”

Interestingly, the two participants had different preferences in terms of interaction with BrAIn Jam, which was largely based on which system they *synchronized* more easily with. Participant 1 preferred the condition with classification because “when it would change things up, it would go in a more simple direction, and it was easier to match playing [with it].” In this case, the musical *responsiveness*, specifically, the dynamic switching based on affective state, supported the player’s experience of drumming with the AIVM. On the other hand, participant 2 preferred the system without fNIRS-informed dynamic switching because it was more *consistent*. Participant 2 experienced that the AIVM with dynamic switching “wasn’t very solid, it didn’t keep something going for long enough to pick up on it.” For them, the “main difference and factor in enjoyment is consistency.” This outcome reveals that personal preferences with adaptive behavior are critical for shaping the interaction experience. Participant 1 considered the BrAIn Jam’s *responsiveness* helpful for getting back in sync musically; on the other hand, participant 2 prioritized *consistency* over adaptiveness when it came to supporting the musical collaboration experience.

In terms of the overall interaction with the AIVM, while participant 1 was more focused on their own instrument, participant 2 was “definitely focused on the [virtual musician].” Participant 1 reported that the visual element “could be helpful at times, seeing the hands hit the drums makes it a bit easier to match or play along with.” Overall, the participants reported that they “liked the beats it was playing” and it was “fun all around.” They both mentioned that they would use a system like BrAIn Jam in everyday contexts, either in a jam session with multiple players, or as a practice tool.

## 6 Discussion

Based on the results two of our three hypotheses were confirmed. The first two and their implications are addressed in the discussion of phase 1. The last hypothesis is addressed in phase II and raises larger questions about including biological signals in adaptive systems for music.

### 6.1 Phase I discussion

It’s crucial to interpret our findings in the context of a small participant pool paradigm. While the results showed statistical significance, the limited participant pool is a factor that must be considered. The decision to focus on a high-data, low-participant pool was made to better understand the bias-variance tradeoff in machine learning models applied to complex neuroscientific data.

#### 6.1.1 Neuroscientific findings

Our study identified the Temporoparietal Junction (TPJ) and the Dorsolateral Prefrontal Cortex (DLPFC) as key brain regions for processing musical and rhythmic patterns. Specifically, we observed significant differences in TPJ activation between random and musical beats, suggesting that the TPJ, potentially in conjunction with the DLPFC, is instrumental in rhythmic predictability. These neuroscientific findings are particularly promising for the development of future adaptive musical systems. They provide a neurologically substantiated metric for rhythmic predictability that could be augmented by other technologies such as EEG, EDA, computer vision, or Music Information Retrieval methods.

The fNIRS technology used in our study is well-suited for “in-the-wild” recordings, offering greater ecological validity. However, it’s important to note that the technology has an inherent 3-5 second lag due to the delay in hemodynamic responses. This constraint limits its application to capturing longer-form changes in musical material rather than real-time adjustments. Despite this limitation, the technology remains valuable for understanding trends in musical improvisation over extended periods, thereby enriching our understanding of rhythmic predictability and other similar metrics in real-world settings.

By incorporating these nuanced neuroscientific insights, we not only deepen our understanding of the neural mechanisms involved in musical interactions but also lay the groundwork for more sophisticated and effective adaptive musical systems.

#### 6.1.2 Machine learning insights

In the realm of machine learning, the RF model demonstrated a superior balance in the bias-variance tradeoff, effectively capturing the complexity of the data while avoiding overfitting. This is in contrast to the Vanilla LSTM model, which exhibited high bias and underfit the data, failing to capture its inherent complexity. Similarly, the state-of-the-art CNN/LSTM hybrid model showed high variance, overfitting the data for one player and underperforming on new data for another.

The bias-variance tradeoff is crucial in machine learning, especially in neuroscience where the data often contains complex, nuanced signals. A high-bias model like Vanilla LSTM would be too simplistic to capture the intricate neural activation patterns, while a high-variance model like the CNN/LSTM model would risk overfitting to noise, making it less generalizable to new, unseen data.

The RF model's ability to effectively balance bias and variance is particularly important for tasks that require distinguishing clearly defined neural signals. Given that our study identified specific brain regions the TPJ and DLPFC as key in processing musical and rhythmic patterns, the RF model's robustness and generalizability make it highly applicable. Its balanced performance allows for more accurate classification or prediction tasks related to neural activation patterns, offering the potential for deeper insights into the neural mechanisms underlying musical perception and cognition.

By integrating these neuroscientific and machine learning findings, we not only advance the understanding of neural mechanisms in musical interactions but also pave the way for more nuanced and effective adaptive musical systems.

## 6.2 Phase II discussion

From our system implementation and user experiences with BrAIIn Jam, we highlight the following insights for designing adaptive systems with affective models.

### 6.2.1 From experiment to real interaction scenario

Many human-human interaction tasks, including musical collaboration, are complex and open-ended. However, to integrate affective information from brain imaging data, it is typically necessary to identify a specific neural correlate we are interested in and design a highly controlled experiment to model this metric. There is an inherent complexity and tension in then applying this affective model, which relies on simplifications and assumptions, to a real interaction context. As such, additional steps should be taken to understand how to apply the model to meet the realities of the task and support the holistic experience. During our phase 2 testing, we learned many lessons about how to best apply our affective model in a realistic jam scenario. For example, while we expected dynamic switching to a metronomic drum pattern would increase rhythmic predictability, for participant 2, this switch made the AIVM's drumming "tough to lean on", indicating that we have to further refine how and when BrAIIn Jam adapts or introduce user-specific customization. An iterative testing and design process can help reveal the role that the affective information plays and how it interacts with other aspects in ways that one might not anticipate.

Additionally, better understanding the underlying mechanisms for how musicians change musical elements in an extemporaneous jam session is important for building a dynamic AI-driven music system. Tolerable levels of rhythmic predictability likely vary between participants, may exist on a spectrum, and may change over time. Simply detecting whether a beat is rhythmically

predictable may not fully expose how musicians communicate about this or enact change based on predictability. Further investigating how musicians deal with unpredictability with more musicians and gradations of rhythmic predictability could lead to a better understanding of how musicians communicate during a jam session between each other, as well as with an AIVM.

### 6.2.2 Affective information integration for longer form adaptation

In a musical jam context, musical information is often the most dominant and immediate factor in shaping the interaction, as musicians must react nearly instantaneously to achieve musical synchronization and expression. However, affect is always underlying this music co-creation and it can interact with the music making in subtle and complex ways over varying time scales. Thus, having access to affective information such as rhythmic predictability provides new opportunities for designing dynamic adaptive musical systems.

In BrAIIn Jam, we used rhythmic predictability to dynamically switch the complexity of the AIVMs drum beats in real-time. Although both participants experienced that BrAIIn Jam was responsive, they reported that the system made musical changes that were too frequent. Thus, depending on the demands of the situation, affective classification can be more suitable for shaping changes in interaction that happen over a longer period of time. Changes at this time scale align with the natural progression of musical jams. As participant 2 describes, in a jam session, they'll "jam on the same loop for minutes at a time." This application of affective information also makes sense when considering what it represents. A metric such as rhythmic predictability is likely not subject to rapid fluctuation, but rather, builds or decays over stretches of time. While many interactive systems are designed for immediate feedback on the primary task at hand (in this case, making music that sounds good), we can create richer interactions by taking an interest in how affective experiences unfold over time.

### 6.2.3 Implications of findings for musical activities

Our findings suggest that the perception of rhythmic predictability can be measured and classified. These findings have the potential to impact collaborative improvisation with virtual musicians, human-human improvisation, musical practice, play, and composition.

Specifically, an AIVM drummer that interacts similarly to a human in a jam session has the potential to transform one's ability to practice musicianship. Practice tools in music today are not commonly dynamic. For example, the most common practice tool—a metronome—plays one tone at regular intervals. True, human-like, musical collaboration requires a deep knowledge of music—and more importantly, human dynamics. Musicians acquire this knowledge over the course of a lifetime playing with other musicians; however, many are too embarrassed, lack time, or ability to play with others. AIVMs may provide a number of those musicians with the ability to gain and hone these skills on their own, giving them confidence to seek out jam sessions with other humans.

TABLE 2 Comparison of different technologies for musical control.

Technology	Temporal Res.	Spatial Res.	Information	Prediction Types	Speed of Control	Advantage	Limitation
fNIRS	3-5s	3 cm × 2 cm	Cognitive	Affective	Multiphrase	Spatial res.	Temporal res., rest times
EEG	< 1ms	3-4 cm	Cognitive	Affective	Phrase	Temporal res.	Sweat, motion, spatial artifacts
EDA	ms-seconds	N/A	Arousal	Affective	Phrase	Arousal reliable	1 channel limited
EMG	Immediate	N/A	Muscular	Motion	Phrase	Immediacy	Limited to muscle movement
Eyetracking	20-40ms	N/A	Vision	Attention, indirect	Phrase	Temporal res.	Limited to eyegaze
CV	- (10-100 ms)	Camera view	Movement	Motion, affective	Note-level	Temporal res.	Recognition unreliable, camera-dependent

An improvising AIVM may also be a welcomed addition to a jam session with a limited number of people. Such an AIVM, capable of sensing rhythmic predictability, could autonomously adjust its rhythms in real-time if perceived as unpredictable, thus supporting cohesive group dynamics. Also, if human musicians within the session become increasingly rhythmically unpredictable, an AIVM could detect these changes and proactively contribute to shifts in musical direction, maintaining engaging and balanced interactions.

Additionally, gathering data on what is perceived as rhythmically predictable could further enable databases to be labeled with this information. This insight could lead to a better understanding of how predictability in music is observed. These insights could enable the exploration of critical questions such as: is unpredictability desirable in some cases? How does it differ from musician to musician? Does musical expertise influence perceived rhythmic predictability? How does the ebb and flow of predictability in music contribute to how musicians communicate in a jam session? Can rhythmic predictability be optimized for, and what effects would this have on the resultant music?

### 6.2.4 Choosing the right technology for the adaptive outcome

The insights gained from our BrAIn Jam experiment prompted us to take a closer look at the capabilities and limitations of various non-scanner bound technologies that could be harnessed for musical prediction and real-time interaction. To facilitate this examination, we created a comparative table summarizing these technologies along with their respective advantages and disadvantages (see Table 2).

One notable recurring theme across these technologies is the inherent limitation in the type and speed of information they can acquire. For instance, electrodermal activity emerges as a reliable indicator of sympathetic arousal, providing valuable insights into the affective dimension. However, it falls short in capturing more nuanced cognitive states or specific neural network activities.

Electroencephalography, with its rapid data acquisition, excels in providing timely information. However, it may be less suitable for interactive and performative environments due to its limitations in pinpointing precise brain regions' activations (Asadzadeh et al., 2020; Liu et al., 2020). This drawback can impede its capacity

to make accurate decisions about neural networks and cognitive processes, relying more on statistical correlations.

Computer vision and eyetracking, while showing promise in recording high-density passive affective information quickly and with fewer technological constraints, have their own set of limitations. They may not approximate cognitive states as effectively as direct brain access technologies and are prone to artifacts stemming from hardware and software constraints. However, of the mentioned technologies computer vision is the only that can make note-level predictions before they happen (Bretan and Weinberg, 2016).

fNIRS emerges as a technology poised to make reliable predictions about cognitive states. However, its disadvantages include its limited temporal resolution, which introduces a delay of 3 to 5 seconds in rendering useful information. Additionally, it demands rest periods, significantly constraining its applicability in long-form, uncontrolled settings.

In light of these insights, we recognize that the choice of technology for musical prediction and interaction should be thoughtfully tailored to the specific demands of the musical context, taking into account the temporal and informational requirements for achieving desired outcomes. Each technology brings its unique strengths and limitations to the table, and a judicious selection or combination can lead to more seamless and meaningful musical interactions.

### 6.3 Limitations

One of the most significant limitations of this study is the limited sample size, consisting of only two players. While the study justified this approach by collecting an extensive dataset of 6 hours per player, thereby allowing for a more nuanced understanding of the bias-variance tradeoff, it does raise questions about the generalizability of the findings. The Random Forest model, although robust in its performance metrics for these two players, has not been tested on a broader population, which limits the external validity of the results.

In the context of neuroscience, our study concentrated on specific brain regions FPA, TPJ, MPFC, and DLPFC thereby potentially narrowing the scope of our findings. While a replication using fMRI could offer more precise insights into the brain regions involved in rhythmic predictability, it would compromise



ecological validity. It's important to acknowledge the inherent limitations of fNIRS technology and adaptivity based on this metric. For instance, fNIRS has a 3-5 second delay in sensing blood flow changes in the brain (Pfeifer et al., 2018). Moreover, long-term activity can lead to the oversaturation of oxygenated and deoxygenated hemoglobin, necessitating activity to be limited to less than five minutes followed by a 10-second rest phase to return levels to baseline (Crum et al., 2022; Hopkins et al., 2023a).

These constraints primarily affect the technology's applicability for real-time improvisational systems, limiting its impact to longer-form interactions ranging from approximately 5 seconds to 5 minutes or more. These timing constraints lend themselves well to interactions such as making musical adjustments in a jam; however, fNIRS may not be a good choice of technology for interactions that require more immediate musical responses such as note-level instrument control. Despite these challenges, the baseline differentiation in activation levels supports the integration of multiple types of signals, each with their own trade-offs, in the design of adaptive systems based on biological signals (such as combining fNIRS for spatial resolution and EEG for temporal resolution).

## 6.4 Future works

Given the limited sample size, future work should aim to validate the Random Forest model on a more diverse and larger sample while maintaining the depth of data collection. This would provide a more comprehensive understanding of the model's generalizability and effectiveness. Real-time applications of the model should also be explored, possibly through optimization techniques that reduce computational load without significantly sacrificing accuracy.

The present study serves as an initial exploration into the complex interplay between neuroscience and machine learning in the context of musical systems. Each biosignal technology, including fNIRS, comes with its own set of advantages and limitations. Future research could focus on finding correlates of the signals measured by fNIRS to understand how this metric generalizes to other technologies such as EEG, GSR, or computer vision. Establishing fNIRS as a baseline could yield valuable insights into the adaptability and reliability of these other technologies.

These findings provide a foundation for designing future systems that optimize musical outputs based on perceived rhythmic predictability. Specifically, if the generative musical output from a model lacks rhythmic predictability, real-time feedback from an fNIRS device could be utilized to enhance this predictability dynamically. This enhancement could occur by using classified brain states to modulate model logits, guiding the generative process toward producing music that aligns with the player's desired level of rhythmic predictability.

Moreover, the study opens avenues for considering longer-form adaptive signals in musical systems. Current adaptive mechanisms, although predictive to some extent, are constrained in their ability to influence musical form over extended periods. Future systems could be designed to navigate through a series of smooth musical transitions, leading to varying levels of rhythmic, melodic, or harmonic complexity. Such a system would not only enhance the

musical experience, but also offer a more nuanced understanding of how musical elements evolve over time.

Other aspects of the metric could also be explored. For instance, can we quantify more specific levels of predictability versus randomness, and perhaps even quantify musical complexity in terms of neural signatures? Could we similarly quantify rhythmic, melodic, or harmonic complexity? Answers to these questions would not only corroborate data obtained through Music Information Retrieval techniques but could also inform musical choices in real-time adaptive systems. These insights would likely align more closely with the decision-making strategies of human players (Bretan and Weinberg, 2016), thereby enriching the interactive musical experience.

By addressing these areas, future work can build upon the foundational insights gained in this study, offering a more comprehensive understanding of the role of neuroscience and machine learning in the development of adaptive musical systems.

## 6.5 Conclusion

In this study, we have made significant strides in understanding the complexities of embodied AI-human musical interaction through several key contributions. Firstly, we formalized rhythmic predictability as a quantifiable metric, integrating neural and machine learning data to provide a nuanced understanding of how musicians gauge this element. Secondly, we introduced "BrAIIn Jam," an innovative adaptive system that employs fNIRS technology to capture real-time neural activity. Thirdly, we provided valuable insights into the practical constraints and opportunities of incorporating affective state data into real-time musical systems.

Our findings reveal the neural underpinnings of musical improvisation with embodied AI, particularly the significant activation of the TPJ and brain regions associated with cognitive workload. These neural activations varied depending on the type of beats-random or musical-indicating the TPJ's critical role in musical improvisation. Additionally, we observed that machine learning models like Random Forests outperformed other models in capturing the complexity of the data, suggesting their potential utility in real-time adaptive systems.

Interestingly, the study also highlighted the importance of individual preferences in AI-human interactions. While some musicians found the BrAIIn Jam's adaptiveness beneficial, others prioritized consistency, indicating that personal preferences play a crucial role in shaping the interaction experience. These findings not only extend our understanding of AI-human musical collaboration but also offer practical tools and insights for future research in this burgeoning field.

Looking ahead, our future work should aim to validate the Random Forest model on a more diverse and larger sample while maintaining the depth of data collection. This would provide a more comprehensive understanding of the model's generalizability and effectiveness. Real-time applications of the model should also be explored, possibly through optimization techniques that reduce computational load without significantly sacrificing accuracy. Moreover, we can consider longer-form adaptive signals in musical systems, navigating through a series of smooth musical transitions, leading to varying levels of rhythmic, melodic, or harmonic

complexity. Further exploration of the metric, such as quantifying specific levels of predictability versus randomness and musical complexity in terms of neural signatures, could also enrich the interactive musical experience.

By addressing these areas, future work can build upon the foundational insights gained in this study, offering a more comprehensive understanding of the role of neuroscience and machine learning in the development of adaptive musical systems with embodied AI.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by University of Colorado Boulder Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

TH: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. RS: Conceptualization, Formal analysis, Writing – review & editing, Data curation, Validation, Visualization, Writing – original draft. SC: Conceptualization, Data curation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. S-YM: Data curation, Investigation, Software, Writing – review & editing. JC: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. LH: Funding acquisition, Resources, Supervision, Validation, Writing – review & editing. ED: Funding acquisition,

Project administration, Resources, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Funding for this project was provided by the Institute of Cognitive Science at the University of Colorado Boulder. Additionally, authors six and seven shared funding for personnel and equipment use.

## Acknowledgments

We (the authors) would like to acknowledge the Institute of Cognitive Science, ATLAS Institute, and Department of Computer Science for their support of the resources necessary to complete this work. Additionally, we would like to thank our participants for taking the time to participate in our study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Asadzadeh, S., Rezaii, T. Y., Beheshti, S., Delpak, A., and Meshgini, S. (2020). A systematic review of eeg source localization techniques and their applications on diagnosis of brain abnormalities. *J. Neurosci. Methods* 339:108740. doi: 10.1016/j.jneumeth.2020.108740
- Ayaz, H., Baker, W. B., Blaney, G., Boas, D. A., Bortfeld, H., Brady, K., et al. (2022). Optical imaging and spectroscopy for the study of the human brain: status report. *Neurophotonics* 9:S24001. doi: 10.1117/1.NPh.9.S2.S24001
- Bandara, D., Hirshfield, L., and Velipasalar, S. (2019). Classification of affect using deep learning on brain blood flow data. *J. Near Infrar. Spectrosc.* 27, 206–219. doi: 10.1177/0967033519837986
- Barker, J. W., Aarabi, A., and Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fmris. *Biomed. Opt. Express* 4, 1366–1379. doi: 10.1364/BOE.4.001366
- Beaty, R. E., Merseal, H. M., and Zeitlen, D. C. (2023). “Network neuroscience of domain-general and domain-specific creativity,” in *The Routledge International Handbook of Creative Cognition* (London: Routledge), 433–452.
- Bengtsson, S. L., Csikszentmihályi, M., and Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *J. Cogn. Neurosci.* 19, 830–842. doi: 10.1162/jocn.2007.19.5.830
- Berkowitz, A. (2010). *The Improvising Mind: Cognition and Creativity in the Musical Moment*. Oxford: Oxford University Press.

- Berkowitz, A. L., and Ansari, D. (2008). Generation of novel motor sequences: the neural correlates of musical improvisation. *Neuroimage* 41, 535–543. doi: 10.1016/j.neuroimage.2008.02.028
- Berkowitz, A. L., and Ansari, D. (2010). Expertise-related deactivation of the right temporoparietal junction during musical improvisation. *Neuroimage* 49, 712–719. doi: 10.1016/j.neuroimage.2009.08.042
- Bobko, P., Hirshfield, L., Eloy, L., Spencer, C., Doherty, E., Driscoll, J., et al. (2022). Human-agent teaming and trust calibration: a theoretical framework, configurable testbed, empirical illustration, and implications for the development of adaptive systems. *Theoret. Issues Ergon. Sci.* 2022, 1–25. doi: 10.1080/1463922X.2022.2086644
- Bretan, M., and Weinberg, G. (2016). A survey of robotic musicianship. *Commun. ACM* 59, 100–109. doi: 10.1145/2818994
- Burgess, P. W., Crum, J., Pinti, P., Aichelburg, C., Oliver, D., Lind, F., et al. (2022). Prefrontal cortical activation associated with prospective memory while walking around a real-world street environment. *Neuroimage* 258:119392. doi: 10.1016/j.neuroimage.2022.119392
- Carter, R. M., Bowling, D. L., Reeck, C., and Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* 337, 109–111. doi: 10.1126/science.1219681
- Carter, R. M., and Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends Cogn. Sci.* 17, 328–336. doi: 10.1016/j.tics.2013.05.007
- Chafe, C., Gurevich, M., Leslie, G., and Tyan, S. (2004). “Effect of time delay on ensemble accuracy,” in *Proceedings of the International Symposium on Musical Acoustics* (Nara: ISMA), 46.
- Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* 20, 226–239. doi: 10.1162/jocn.2008.20018
- Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage* 32, 1771–1781. doi: 10.1016/j.neuroimage.2006.04.207
- Chew, E., Zimmermann, R., Sawchuk, A. A., Kyriakakis, C., Papadopoulos, C., François, A., et al. (2004). “Musical interaction at a distance: distributed immersive performance,” in *Proceedings of the MusicNetwork Fourth Open Workshop on Integration of Music in Multimedia Applications* (Barcelona: MusicNetwork Barcelona), 15–16.
- Crum, J., Zhang, X., Noah, A., Hamilton, A., Tachtsidis, I., Burgess, P., et al. (2022). An approach to neuroimaging interpersonal interactions in mental health interventions. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 7, 669–679. doi: 10.1016/j.bpsc.2022.01.008
- Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of fNIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069
- Dannenberg, R. B., Brown, H. B., and Lupish, R. (2011). “McBlare: a robotic bagpipe player,” in *Musical Robots and Interactive Multimodal Systems* (Springer), 165–178.
- de Manzano, Ö., and Ullén, F. (2012). Goal-independent mechanisms for free response generation: Creative and pseudo-random performance share neural substrates. *Neuroimage* 59, 772–780. doi: 10.1016/j.neuroimage.2011.07.016
- Delpy, D. T., Cope, M., Van Der Zee, P., Arridge, S., Wray, S., and Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Phys. Med. Biol.* 33, 1433–1442. doi: 10.1088/0031-9155/33/12/008
- Dravida, S., Noah, J. A., Zhang, X., and Hirsch, J. (2017). Comparison of oxyhemoglobin and deoxyhemoglobin signal reliability with and without global mean removal for digit manipulation motor tasks. *Neurophotonics* 5:1. doi: 10.1117/1.NPh.5.1.011006
- Eastmond, C., Subedi, A., De, S., and Intes, X. (2022). Deep learning in fNIRS: a review. *Neurophotonics* 9, 041411–041411. doi: 10.1117/1.NPh.9.4.041411
- Eigenfeldt, A., and Kapur, A. (2008). “An agent-based system for robotic musical performance,” in *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME), 144–149.
- Eloy, L., Doherty, E. J., Spencer, C. A., Bobko, P., and Hirshfield, L. (2022). Using fNIRS to identify transparency- and reliability-sensitive markers of trust across multiple timescales in collaborative human-human-agent triads. *Front. Neuroergonom.* 10:838625. doi: 10.3389/fnrgo.2022.838625
- Eloy, L., Stewart, A. E. B., Jean Amon, M., Reinhardt, C., Michaels, A., Sun, C., et al. (2019). “Modeling team-level multimodal dynamics during multiparty collaboration,” in *2019 International Conference on Multimodal Interaction*, 244–258.
- Eyssel, F., Kuchenbrandt, D., and Bobinger, S. (2011). “Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism,” in *Proceedings of the 6th International Conference on Human-Robot Interaction* (Human-Robot Interaction), 61–68.
- Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage* 63, 921–35. doi: 10.1016/j.neuroimage.2012.03.049
- Fett, A.-K. J., Gromann, P. M., Giampietro, V., Shergill, S. S., and Krabbendam, L. (2014). Default distrust? An fMRI investigation of the neural development of trust and cooperation. *Social Cognit. Affect. Neurosci.* 9, 395–402. doi: 10.1093/scan/nss144
- Fukuie, T., Suwabe, K., Kawase, S., Shimizu, T., Ochi, G., Kuwamizu, R., et al. (2022). Groove rhythm stimulates prefrontal cortex function in groove enjoyers. *Sci. Rep.* 12:7377. doi: 10.1038/s41598-022-11324-3
- Gervain, J. (2015). “Near-infrared spectroscopy,” in *International Encyclopedia of the Social & Behavioral Sciences*, 2.
- Gillick, J., Roberts, A., Engel, J., Eck, D., and Bamman, D. (2019). “Learning to groove with inverse sequence transformations,” in *International Conference on Machine Learning* (ICML).
- Glikson, E., and Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Acad. Managem. Annals* 14, 627–660. doi: 10.5465/annals.2018.0057
- Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009
- Hallam, S., Cross, I., and Thaut, M. (2009). *Oxford Handbook of Music Psychology*. Oxford: Oxford University Press.
- Heinzel, S., Haueßinger, F. B., Hahn, T., Ehli, A. C., Plichta, M. M., and Fallgatter, A. J. (2013). Variability of (functional) hemodynamics as measured with simultaneous fNIRS and fMRI during intertemporal choice. *Neuroimage* 71, 125–134. doi: 10.1016/j.neuroimage.2012.12.074
- Hirsch, J., Noah, J. A., Zhang, X., Dravida, S., and Ono, Y. (2018). A cross-brain neural mechanism for human-to-human verbal communication. *Soc. Cogn. Affect. Neurosci.* 13, 907–920. doi: 10.1093/scan/nsy070
- Hirsch, J., Tiede, M., Zhang, X., Noah, J. A., Salama-Manteau, A., and Biriotti, M. (2021). Interpersonal agreement and disagreement during face-to-face dialogue: an fNIRS investigation. *Front. Hum. Neurosci.* 14:601. doi: 10.3389/fnhum.2020.606397
- Hirsch, J., Zhang, X., Noah, J. A., and Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *Neuroimage* 157, 314–330. doi: 10.1016/j.neuroimage.2017.06.018
- Hoff, K. A., and Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Factors* 57, 407–434. doi: 10.1177/0018720814547570
- Hopkins, T., Bae, S. S., Uhr, J., Banić, A., Zheng, C., and Do, E. Y.-L. (2020). “User interfaces in smart cities,” in *Handbook of Smart Cities* (TENOR), 1–33.
- Hopkins, T., Doherty, E., Ofer, N., Weng, S. C. C., Gyro, P., Tobin, C., et al. (2023a). Stringesthesia: dynamically shifting musical agency between audience and performer based on trust in an interactive and improvised performance. *arXiv [preprint] arXiv:2309.05786v1*. doi: 10.1145/3616195.3616209
- Hopkins, T., Jude, A., Phillips, G., and Do, E. Y.-L. (2023b). *Virtual AI Jam: AI-Driven Virtual Musicians for Human-in-the-Loop Musical Improvisation*. Available online at: <https://aimc2023.pubpub.org/pub/9064kmvw> (Accessed October 1, 2024).
- Hopkins, T., Pascente, P., Seltzer, W., Masterson, K., and Do, E. Y.-L. (2021). “The jam station: Gamifying collaborative musical experiences through algorithmic assessment,” in *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '21* (New York, NY: Association for Computing Machinery).
- Hopkins, T., Vanukuru, R., Weng, S. C. C., Tobin, C., Banic, A., Gross, M. D., et al. (2023c). “Networking AI-driven virtual musicians in extended reality,” in *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (Sydney: IEEE), 752–756.
- Hopkins, T., Weng, S. C. C., and Do, E. Y.-L. (2023d). *Rhythmic Conversations: Using AI-Driven Virtual Musicians to Enhance Audience Participation in Musical Performance*. Available online at: <https://aimc2023.pubpub.org/pub/qot7sw4x> (Accessed October 1, 2024).
- Hopkins, T., Weng, S. C.-C., Vanukuru, R., Wenzel, E., Banic, A., and Do, E. Y.-L. (2022a). “How late is too late? Effects of network latency on audio-visual perception during remote musical collaboration,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (Christchurch: IEEE), 686–687.
- Hopkins, T., Weng, S. C. C., Vanukuru, R., Novack, S., Tobin, C., Wenzel, E., et al. (2024). “XR jam: Design considerations for music networking with AI in XR,” in *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)* (Los Angeles, CA: IEEE), 97–101.
- Hopkins, T., Weng, S. C. C., Vanukuru, R., Wenzel, E. A., Banic, A., Gross, M. D., et al. (2022b). AR drum circle: real-time collaborative drumming in AR. *Front. Virtual Real.* 3:847284. doi: 10.3389/fvrv.2022.847284
- Howell-Munson, A., Doherty, E., Gavriel, P., Nicolas, C., Norton, A., Neamtu, R., et al. (2022). “Towards brain metrics for improving multi-agent adaptive human-robot collaboration: a preliminary study,” in *2022 Symposium on Human-Computer Interaction for Work*, 1–10.



- Jankowsky, K., and Schroeders, U. (2022). Validation and generalizability of machine learning prediction models on attrition in longitudinal studies. *Int. J. Behav. Dev.* 46, 169–176. doi: 10.1177/01650254221075034
- Kajitani, M. (1989). Development of musician robots. *J. Rob. Mech.* 1, 254–255. doi: 10.20965/jrm.1989.p0254
- Kapur, A. (2005). “A history of robotic musical instruments,” in *ICMC*, 4599.
- Kelley, M. S., Noah, J. A., Zhang, X., Scassellati, B., and Hirsch, J. (2021). Comparison of human social brain activity during eye-contact with another human and a humanoid robot. *Front. Robotics AI* 7:599581. doi: 10.3389/frobt.2020.599581
- Kocsis, L., Herman, P., and Eke, A. (2006). The modified beer-lambert law revisited. *Phys. Med. Biol.* 51:5. doi: 10.1088/0031-9155/51/5/N02
- Krois, J., Garcia Cantu, A., Chaurasia, A., Patil, R., Chaudhari, P. K., Gaudin, R., et al. (2021). Generalizability of deep learning models for dental image analysis. *Sci. Rep.* 11:6102. doi: 10.1038/s41598-021-85454-5
- Kung, S.-J., Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* 25, 401–420. doi: 10.1162/jocn\_a\_00325
- Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.
- Lesaffre, M., Maes, P.-J., and Leman, M. (2017). *The Routledge Companion to Embodied Music Interaction*. Oxfordshire: Taylor & Francis.
- Leslie, G., and Mullen, T. (2011). “Moodmixer: EEG-based collaborative sonification,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 296–299.
- Leslie, G., Ojeda, A., and Makeig, S. (2014). Measuring musical engagement using expressive movement and eeg brain dynamics. *Psychomusicol.: Music, Mind Brain* 24:75. doi: 10.1037/pmu0000031
- Levitin, D. J. (2002). *Foundations of Cognitive Psychology: Core Readings*. Cambridge, MA: MIT Press.
- Lewis, G., and Piekut, B. (2016). *The Oxford Handbook of Critical Improvisation Studies*. Oxford: Oxford University Press.
- Liu, X., Makeyev, O., and Besio, W. (2020). Improved spatial resolution of electroencephalogram using tripolar concentric ring electrode sensors. *J. Sensors* 2020:6269394. doi: 10.1155/2020/6269394
- Maleki, F., Ovens, K., Gupta, R., Reinhold, C., Spatz, A., and Forghani, R. (2022). Generalizability of machine learning models: quantitative evaluation of three methodological pitfalls. *Radial. Artif. Intellig.* 5:e220028. doi: 10.1148/ryai.220028
- McCormack, J., Gifford, T., Hutchings, P., Llano Rodriguez, M. T., Yee-King, M., and d’Inverno, M. (2019). “In a silent way: Communication between ai and improvising musicians beyond sound,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (ACM)*, 1–11.
- McCormack, J., Hutchings, P., Gifford, T., Yee-King, M., Llano, M. T., and Dinverno, M. (2020). Design considerations for real-time collaboration with creative artificial intelligence. *Organised Sound* 25, 41–52. doi: 10.1017/S1355771819000451
- Midgelow, V. L. (2019). *The Oxford Handbook of Improvisation in Dance*. Oxford: Oxford University Press.
- Miranda, E. R. (2021). *Handbook of Artificial Intelligence for Music*. Cham: Springer.
- Mullen, T., Khalil, A., Ward, T., Iversen, J., Leslie, G., Warp, R., et al. (2015). “Mindmusic: playful and social installations at the interface between music and the brain,” in *More Playful User Interfaces: Interfaces that Invite Social and Physical Interaction* (Springer), 197–229.
- Noah, J. A., Ono, Y., Nomoto, Y., Shimada, S., Tachibana, A., Zhang, X., et al. (2015). fmri validation of firs measurements during a naturalistic task. *J. Visual. Exp.* 2015:100. doi: 10.3791/52116
- Okamoto, M., Dan, H., Shimizu, K., Takeo, K., Amita, T., Oda, I., et al. (2004). Multimodal assessment of cortical activation during apple peeling by nirs and fmri. *Neuroimage* 21, 1275–1288. doi: 10.1016/j.neuroimage.2003.12.003
- Patel, A. D., and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the action simulation for auditory prediction (asap) hypothesis. *Front. Syst. Neurosci.* 8:57. doi: 10.3389/fnsys.2014.00057
- Pfeifer, M. D., Scholkman, F., and Labryère, R. (2018). Signal processing in functional near-infrared spectroscopy (fNIRS): methodological differences lead to different statistical results. *Front. Hum. Neurosci.* 11:641. doi: 10.3389/fnhum.2017.00641
- Pinti, P., Aichelburg, C., Gilbert, S., Hamilton, A., Hirsch, B., Burgess, W. P., and Tachtsidis, I. (2018a). A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Jap. Psychol. Res.* 60, 347–373. doi: 10.1111/jpr.12206
- Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., et al. (2015). Using fiberless, wearable fnirs to monitor brain activity in real-world cognitive tasks. *J. Visual. Exp.* 2015, e53336–e53336. doi: 10.3791/53336-v
- Pinti, P., Devoto, A., Greenhalgh, I., Tachtsidis, I., Burgess, P. W., and de C Hamilton, A. F. (2021). The role of anterior prefrontal cortex (area 10) in face-to-face deception measured with fNIRS. *Soc. Cogn. Affect. Neurosci.* 16, 129–142. doi: 10.1093/scan/nsaa086
- Pinti, P., Scholkman, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current status and issues regarding pre-processing of fnirs neuroimaging data: An investigation of diverse signal filtering methods within a general linear model framework. *Front. Hum. Neurosci.* 12:505. doi: 10.3389/fnhum.2018.00505
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2018b). The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience. *Ann. New York Acad. Sci.* 1464, 5–29. doi: 10.1111/nyas.13948
- Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., et al. (2014). A wearable multi-channel fnirs system for brain imaging in freely moving subjects. *Neuroimage* 85, 64–71. doi: 10.1016/j.neuroimage.2013.06.062
- Rabbani, M. H. R., and Islam, S. M. R. (2023). Deep learning networks based decision fusion model of eeg and fnirs for classification of cognitive tasks. *Cognitive Neurodynam.* 4, 1–18. doi: 10.1007/s11571-023-09986-4
- Raman, C. V. (1920). *Experiments with Mechanically-Played Violins*.
- Roads, C. (1986). The tsukuba musical robot. *Comp. Music J.* 10, 39–43. doi: 10.2307/3679483
- Rocks, J. W., and Mehta, P. (2022). Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models. *Phys. Rev. Res.* 4:013201. doi: 10.1103/PhysRevResearch.4.013201
- Rowe, R. (2004). *Machine Musicianship*. Cambridge, MA: MIT Press.
- Rybner, A., Jessen, E. T., Mortensen, M. D., Larsen, S. N., Grossman, R., Bilenberg, N., et al. (2022). Vocal markers of autism: Assessing the generalizability of machine learning models. *Autism Res.* 15, 1018–1030. doi: 10.1002/aur.2721
- Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., Atsumori, H., et al. (2013). A nirs-fmri investigation of prefrontal cortex activity during a working memory task. *Neuroimage* 83, 158–173. doi: 10.1016/j.neuroimage.2013.06.043
- Scholkman, F., Gerber, U., Wolf, M., and Wolf, U. (2013a). End-tidal co2: An important parameter for a correct interpretation in functional brain studies using speech tasks. *Neuroimage* 66, 71–79. doi: 10.1016/j.neuroimage.2012.10.025
- Scholkman, F., Wolf, M., and Wolf, U. (2013b). The effect of inner speech on arterial CO2 and cerebral hemodynamics and oxygenation: A functional NIRS study. *Adv. Exp. Med. Biol.* 789, 81–87. doi: 10.1007/978-1-4614-7411-1\_12
- Schuetz, N. (2002). *The Effects of Latency on Ensemble Performance* (Bachelor Thesis). CCRMA Department of Music, Stanford University, Stanford, CA, United States.
- Schurz, M., Tholen, M. G., Perner, J., Mars, R. B., and Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: a review using probabilistic atlases from different imaging modalities. *Hum. Brain Mapp.* 38, 4788–4805. doi: 10.1002/hbm.23675
- Seidel-Marzi, O., Hähner, S., Ragert, P., and Carius, D. (2021). Task-related hemodynamic response alterations during slacklining: an fnirs study in advanced slackliners. *Front. Neuroergonom.* 2:644490. doi: 10.3389/fnrgo.2021.644490
- Singer, E., Feddersen, J., Redmon, C., and Bowen, B. (2004). “Lemur’s musical robots,” in *Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME)*, 181–184. doi: 10.1145/1027527.1027569
- Singer, E., Larke, K., and Bianciardi, D. (2003). “Lemur guitarbot: Midi robotic string instrument,” in *NIME*, 188–191.
- Stuart, S., Belluscio, V., Quinn, J. F., and Mancini, M. (2019). Pre-frontal cortical activity during walking and turning is reliable and differentiates across young, older adults and people with parkinson’s disease. *Front. Neurol.* 10:536. doi: 10.3389/fneur.2019.00536
- Tachtsidis, I., and Scholkman, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics* 3, 031405–031405. doi: 10.1117/1.NPh.3.3.031405
- Tak, S., and Ye, J. C. (2014). Statistical analysis of fnirs data: a comprehensive review. *Neuroimage* 85, 72–91. doi: 10.1016/j.neuroimage.2013.06.016
- Takanishi, A., Sonehara, M., and Kondo, H. (1996). “Development of an anthropomorphic flutist robot WF-3RII,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS’96* (Osaka: IEEE), 37–43.
- Thaut, M. H., Trimarchi, P. D., and Parsons, L. M. (2014). Human brain basis of musical rhythm perception: common and distinct neural substrates for meter, tempo, and pattern. *Brain Sci.* 4, 428–452. doi: 10.3390/brainsci4020428
- Vanzella, P., Balardin, J. B., Furucho, R. A., Zimeo Morais, G. A., Braun Janzen, T., Sammler, D., et al. (2019). fnirs responses in professional violinists while playing duets:

evidence for distinct leader and follower roles at the brain level. *Front. Psychol.* 10:164. doi: 10.3389/fpsyg.2019.00164

Vuust, P., Dietz, M. J., Witek, M., and Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Ann. N. Y. Acad. Sci.* 1423, 19–29. doi: 10.1111/nyas.13622

Vuust, P., and Witek, M. A. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* 5:1111. doi: 10.3389/fpsyg.2014.01111

Wang, X., Belkacem, A. N., Li, P., Zhang, Z., Liang, J., Du, D., et al. (2023). “The classification method of EEG motor imagery based on info-LSSVM,” in *Proceedings of the 2022 11th International Conference on Computing and Pattern Recognition, ICCPR '22* (New York, NY: Association for Computing Machinery), 471–477.

Weinberg, G., and Driscoll, S. (2007). “The design of a robotic marimba player: introducing pitch into robotic musicianship,” in *Proceedings of the 7th International Conference on New Interfaces for Musical Expression (NIME)*, 228–233.

Weinberg, G., Driscoll, S., and Parry, R. M. (2005). “Haile-an interactive robotic percussionist,” in *ICMC*.

Weng, S. C., Hopkins, T., Vanukuru, R., Tobin, C., Banic, A., Leithinger, D., et al. (2023). “How field of view affects awareness of an avatar during a musical task in augmented reality,” in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (IEEEVR)*, 633–634.

Williams, D., Hodge, V., Gega, L., Murphy, D., Cowling, P., and Drachen, A. (2019). “AI and automatic music generation for mindfulness,” in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio* (Audio Engineering Society).

Williamson, M. M. (1999). *Robot Arm Control Exploiting Natural Dynamics* (PhD thesis). Massachusetts Institute of Technology, Cambridge, MA, United States.

Yücel, M. A., Lühmann, A., Scholkmann, F., Gervain, J., Dan, I., Ayaz, H., et al. (2021). Best practices for fnirs publications. *Neurophotonics* 8, 012101–012101. doi: 10.1117/1.NPh.8.1.019802

Yuksel, B. F., Collisson, P., and Czerwinski, M. (2017). Brains or beauty: How to engender trust in user-agent interactions. *ACM Trans. Internet Technol.* 17:2998572. doi: 10.1145/2998572

Yuksel, B. F., Oleson, K. B., Chang, R., and Jacob, R. J. K. (2019). *Detecting and Adapting to Users' Cognitive and Affective State to Develop Intelligent Musical Interfaces*, pages 163–177. Cham: Springer International Publishing.

Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., et al. (2016). “Learn piano with bach: An adaptive learning interface that adjusts task difficulty based on brain state,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (ACM)*, 5372–5384.

Zhang, P., Gao, H., Hu, Z., Yang, M., Song, D., Wang, J., et al. (2022). A bias-variance evaluation framework for information retrieval systems. *Inform. Proc. Managem.* 59:102747. doi: 10.1016/j.ipm.2021.102747

Zimeo Morais, G. A., Balardin, J. B., and Sato, J. R. (2018). fnirs optodes location decider (fold): a toolbox for probe arrangement guided by brain regions-of-interest. *Sci. Rep.* 8:3341. doi: 10.1038/s41598-018-21716-z