Can You Feel My Brain? Investigating Attentional Engagement through EEG in an Immersive Musical Environment

Rômulo Vieira
MídiaCom Lab, Institute of Computing
Fluminense Federal University
Niterói, Brazil
romulo_vieira@midiacom.uff.br

Carla Estefany Caetano Silva
MídiaCom Lab, Institute of Computing
Fluminense Federal University
Niterói, Brazil
carlaestefany@id.uff.br

Daniela Gorski Trevisan
Institute of Computing
Fluminense Federal University
Niterói, Brazil
daniela@ic.uff.br

Débora C. Muchaluat-Saade
MídiaCom Lab, Institute of Computing
Fluminense Federal University
Niterói, Brazil
debora@midiacom.uff.br

Pablo César
Centrum Wiskunde & Informatica
TU Delft
Amsterdam, The Netherlands
garcia@cwi.nl

***Abstract*—Bioinspired technologies have expanded the use of physiological signals in music, ranging from recommendation systems to biomusic as a form of artistic expression. However, studies analyzing attentional engagement in immersive musical environments are still scarce, and they also fail to consider the impact of haptic elements on this metric. Moreover, traditional measurement methods are costly and require specialized personnel, limiting their adoption in artistic contexts. This work proposes an analysis of attention levels in an immersive musical experience using PhysioDrum, a multisensory virtual reality (VR) application. The aim is to analyze how immersive experiences, vibrotactile elements, and design choices impact attention and engagement levels in musical contexts. The results obtained through a mixed methods approach, combining nonparametric statistical tests and exploratory analyses, indicate that attention levels reach their peak during the initial phase of interaction. In subsequent sessions, attentional focus is maintained, accompanied by a decrease in the cognitive effort required. Haptic feedback was found to enhance system adaptation and encourage sustained engagement, while dynamic tasks with increasing difficulty levels foster more stable attentional states.**

***Index Terms*—Io3MT, Musical Metaverse, EEG, Haptic Feedback**

I. Introduction

Bioinspired technologies emerge as a significant advance in musical interactions, offering new possibilities for applications that use physiological information in communication, detection of affective states of music listeners, and the sonification of biological data [1].

In the communication field, researchers have explored the structural and temporal similarities between music and natural language, developing neural decoding models capable of reconstructing music from brain activity. These models pave the way for the development of brain-computer interfaces (BCI) that enable neural signal-mediated communication, especially for people with verbal communication difficulties, such as individuals on the autism spectrum, patients undergoing rehabilitation processes, and people with anxiety disorders. These interfaces expand interaction possibilities, promoting social inclusion and emotional well-being for people who traditionally face barriers to conventional communication [1].

Emotion detection, in turn, is based on the premise that physiological signals, such as electrocardiographic (ECG), electroencephalographic (EEG), Galvanic Skin Response (GSR), and respiratory rate activity, reflect human emotional states implicitly and continuously, without the need for self-reporting or interruption of the user experience. Several studies have already demonstrated the effectiveness of these signals in predicting emotional responses, such as identifying songs that induce specific affective states, personalizing music recommendations (emotion-aware music discovery), and assessing the level of engagement and relevance perceived by users during music listening [2]. These advances are particularly relevant in affective interaction systems, in which the ability to interpret emotions in real time can enrich the user experience (UX), promoting greater personalization and immersion [2]–[4].

Additionally, sonification research seeks to map physiological data onto musical parameters such as pitch, intensity, and timbre, establishing a bridge between biological activity and its sonic representation. This approach allows neuroelectrical information to be transformed into interpretable musical structures, enabling both new forms of artistic expression and tools for monitoring physiological states in real-time [5], [6].

Despite the potential of these applications, many of the current solutions rely on invasive techniques, such as electrocorticography (ECoG), or require sophisticated and expensive equipment, such as functional magnetic resonance imaging (fMRI), limiting their use to controlled environments and highly specialized professionals. Furthermore, the state of the art still neglects aspects related to attentional engagement in immersive musical scenarios, preventing the development of a comprehensive understanding of the mental processes involved in such experiences [1], [7].

To address these gaps, this study proposes the capture, analysis, and interpretation of users' attentional focus in an immersive musical environment through an accessibl approach. The investigation is structured around two central research questions (RQ):

- How can users' attentional focus be effectively captured, analyzed, and interpreted in an immersive musical environment using a low-cost, and non-invasive method?
- What is the impact of haptic feedback on users' attention during immersive musical interactions?

As an experimental setup, this research employs PhysioDrum [8], a system grounded in the theoretical frameworks of the Musical Metaverse (MM) [9] and the Internet of Multisensory, Multimedia, and Musical Things (Io3MT) [10]. The PhysioDrum environment comprises a virtual drum kit accessed via a virtual reality (VR) headset, enabling immersive interaction through an smart musical instrument (SMI) [11] called RemixDrum [12], complemented by two foot pedals. Thmotion systotion embodies the concept of phygital (physical plus digital), wherein actions performed in the physical domain directly affect the virtual environment and vice versa. Further, the platform supports real-time control of Pure Data patches, the dynamic manipulation of visual elements using Processing programming language, and the delivery of haptic feedback in response to user interaction, thereby fostering a multisensory and interactive musical experience.

During the experiment, participants engaged in two rhythmic performance sessions of increasing difficulty — one without and one with haptic feedback. Throughout these activities, peripheral biosignals associated with the Autonomic Nervous System (ANS) were continuously monitored. EEG signals, indicative of brain activity, were discreetly recorded using electrodes placed on participants' foreheads and processed using the Bitalino[1] platform. The acquired data were subjected to preprocessing, followed by extraction of features in both the time and frequency domains. Statistical analyzes were then performed to explore the potential relationships between the physiological states of the participants and their immersive musical experiences. The integration of descriptive and inferential approaches in these analyzes enhances the robustness and validity of the findings.

As one of the first studies to explore peripheral physiological signals in MM and Immersive Io3MT-based environments, this research has the potential to provide new evidence on the feasibility of predicting attentional responses from physiological data, as well as provide insights into how design choices, like the presence of haptic feedback and the physicality of the system impact attention levels and consequently the immersion, engagement and enjoyment of participants. The results can contribute both to the advancement of the design of immersive musical systems and the development of physiologically responsive artistic experiences.

The remainder of this paper is organized as follows. Section II reviews related work concerning the use of biological signals in various musical applications. Section III describes the experimental system used in this study, along with the protocol adopted to capture and analyze brain activity of participants. Section IV presents the experimental procedures and discusses the statistical and exploratory analyzes conducted to assess the impact of PhysioDrum on the attentional focus of the participants. Section V examines the results in greater depth, highlighting patterns of attentional modulation across sessions, the role of haptic feedback in facilitating system adaptation and sustaining focus, and the broader implications of these findings for the design of immersive musical environments. Finally, Section VI summarizes the main findings and outlines the conclusions drawn from this research.

II. Related Work

Based on an extensive review of academic literature, industry white papers, and publicly available corporate documentation, there is a growing consensus around the use of physiological signals as reliable indicators of emotional states [13]. Within this context, numerous musical applications have adopted such methods for a variety of purposes. These include, for example, inferring the musical preferences of users through the analysis of EEG responses and detecting emotional reactions triggered by specific phonograms [2], [14].

The use of physiological signals as indicators of emotional states has also garnered significant interest, particularly among researchers and healthcare professionals. Vital signs such as heart rate, blood pressure variability, and ventilatory asynchrony are routinely used to infer patient needs and responses, especially in contexts involving pain assessment and management [13], [15], [16].

Within this framework, the study conducted by Rahman et al. [17] proposes the transformation of ANS signals from individuals diagnosed with Dissociative Identity Disorder (DID) into musical structures, with the aim of enriching interpersonal communication. The authors monitor a range of physiological signals, including electrodermal activity (EDA); skin temperature at the fingertips — which typically exhibits gradual changes approximately 15 seconds after emotional or physiological stimuli; blood volume pulse (BVP) — which reflects fluctuations in blood flow and amplitude; and respiratory rate, known to be sensitive to stress and relaxation states.

These signals are then algorithmically mapped to corresponding musical parameters: EDA is linked to melodic variation, fingertip temperature to musical tonality, heartbeat to percussive rhythm, and respiration to a continuous auditory effect resembling exhalation. This musical biofeedback system enables more nuanced and accessible monitoring of users' physical and emotional conditions, particularly by caregivers without specialized medical training [13], [16].

[1] https://www.pluxbiosignals.com/collections/bitalino

More recent work has used physiological signals to promote greater connection or intimacy between individuals. Janssen and colleagues [18] demonstrated that the auditory perception of one person's heartbeat by another influences social behavior in a way that is analogous to conventional intimacy signals, such as eye contact and physical proximity.

In a complementary line of inquiry, the research conducted by Bootsma [19] investigates music generation based on biofeedback with the objective of enhancing empathy and emotional connection among loved ones. In this approach, physiological signals are translated into musical parameters. Heart rate is mapped onto rhythmic structures; electrodermal activity informs background harmonies; and EDA peaks trigger specific sound effects. The resulting music is further contextualized by associating it with textual and voice messages received from family members, thereby fostering a heightened sense of presence and emotional proximity.

Despite the numerous immersive applications utilizing physiological data [20]–[22], particularly EEG, no studies have specifically focused on using biosignals to evaluate attentional engagement during immersive musical experiences. Understanding this psychophysiological dimension is essential not only for refining the design of interactive systems, but also for expanding the potential of musical and recreational technologies in educational and therapeutic settings. In this context, the primary contribution of the present study lies in the simultaneous measurement of attentional states and the investigation of the impact of haptic feedback on these parameters. Additionally, this work advances the state of the art by using a low-cost and non-invasive approach to brain signal acquisition, thereby promoting greater accessibility for future research and practical applications.

III. Methodology

This section outlines the methodological procedures of the study, beginning with the theoretical foundations guiding the design of immersive musical environments. It then describes PhysioDrum, the multisensory and immersive musical platform developed based on these principles and used as the main experimental tool. Subsequently, it details the acquisition of physiological signals, focusing on instrumentation and data collection protocols.

A. Design Guidelines for Immersive Musical Systems

Some points are recurrent in the development of VR applications, such as the possibility of exploring the environment from different perspectives and freedom of navigation. Furthermore, this work is grounded in a set of specific requirements for immersive Io3MT services [8], which will be outlined and explained below.

*1) **Design for Functionality**:* Although environments designed for artistic creation are inherently hedonic in nature, such systems must be purposefully developed to support meaningful expression, enabling artists and users to convey their ideas and emotions effectively. To this end, a range of technical considerations must be addressed, including reliable synchronization of information exchange, minimal latency, and realistic interactions that accurately reflect users' physical actions. Moreover, the design must ensure good ergonomic conditions and minimize the risk of cybersickness, thereby fostering a comfortable and immersive user experience.

*2) **Design for Immersiveness**:* The level of technological immersion provided by an XR system is directly related to the number of sensorimotor contingencies it supports, that is the set of different actions that a user can employ to perceive different actions in the system, such as moving the head and eyes to change the line of sight, kneeling to look at the floor more closely, or turning the head to locate an audiovisual source. An immersive music system, therefore, must support as many features as possible. At the same time, it must communicate its limitations to users so that they do not attempt to perform actions not supported by the application.

Other factors that contribute to immersion are virtual body ownership, where users can see realistic (via volumetric video) or digital (avatars and cartoons) representations of their body, as well as the use of natural actions, such as blowing, plucking, and bowing, which increase the tangibility and physicality of the environment, leading to new levels of abstraction, immersion, and imagination. High sound quality and fidelity are also recommended.

*3) **Design for Feedback**:* In addition to providing an increased sense of presence and interactivity in the environment, feedback improves the usability of the application, making it more intuitive and easy to learn. In addition, it enhances the narrative understanding of the artistic work and amplifies the user experience.

From a technical point of view, auditory feedback must be related to the position of the virtual devices, whereas multimodal feedback can be achieved through precise synchronization between visual, auditory, and tactile elements.

*4) **Design for Social Connection**:* One of the fundamental characteristics of music lies in its capacity to foster shared social experiences. Accordingly, incorporating connectivity into immersive musical environments is essential to enable users to interact with one another and/or with objects within the virtual scene. Such environments should support non-verbal communication and facilitate realistic interactions, whether the participants are co-located or engaging remotely, thereby enriching the sense of presence and collaboration.

*5) **Design for Creativity**:* The last guideline for the development of immersive music systems advocates for extending their functionality beyond artistic creation and performance, encompassing additional domains such as education, training, experimentation, and leisure. To achieve this, such systems must incorporate features that assist and guide users throughout the learning process, offering adaptive and personalized support based on individual skill levels. In this way, immersive music environments can serve as powerful tools for entertainment and creative engagement, fostering the development of artistic expression, critical thinking, and sustained concentration.

B. Technical Implementation of PhysioDrum

In accordance with the aforementioned design guidelines, PhysioDrum[2][3] [8], [23] was developed. The system was implemented using Unity 2022.3.30, in conjunction with the Meta Quest 3 headset and the Meta Interaction SDK version 1.3.2. The virtual environment features a drum kit that replicates the components of a traditional acoustic drum set, including a bass drum, snare drum, floor tom, two rack toms, ride cymbal, and hi-hat.

Interaction with the virtual drum components is achieved through physical actions performed using the RemixDrum [12] and two foot pedals. The RemixDrum is an SMI that consists of a traditional drumstick augmented with embedded sensors. These include touch sensors, which control the activation and deactivation of audio tracks developed in the Pure Data programming language, and accelerometers, which detect motion along the X, Y, and Z axes to dynamically modify the RGB color patterns of a multimedia artwork created in Processing.

In the third version of RemixDrum, designed specifically for this study, colored spheres were integrated into the drumstick. Their movements are captured and processed using a computer vision algorithm implemented in Python 3.8. This algorithm analyzes the displacement of the spheres and translates their motion into digital commands, enabling real-time interaction between the physical and virtual environments. Furthermore, a coin-type vibration motor was incorporated into the instrument's electronic circuit to deliver haptic feedback, thereby enhancing users' sensory perception of collisions within the virtual drum set.

The foot pedals, in turn, are connected to an electrical circuit capable of detecting their activation and transmitting this information to the virtual application via Open Sound Control (OSC) messages. These devices are designed to emulate the behavior of traditional drum pedals, triggering the sounds of the bass drum and hi-hat within the virtual environment. Although their physical design and ergonomics differ from those of conventional pedals, their functional replication is consistent, ensuring a coherent and responsive user experience.

This operational mode, which leverages physical movements to control and trigger properties in the digital world, is rooted in the concept of phygital [24]. As the name implies, this approach integrates physical and digital processes, establishing connections and networks that bridge these domains to enable novel functionalities and forms of interaction. As a result, the system provides a fluid and intuitive interaction experience, effectively bridging physical actions and their corresponding virtual responses.

The system was run on an ASUS TUF Gaming F15 laptop, equipped with an NVIDIA GeForce RTX 4050 graphics card, Intel i7 processor, and 16GB of RAM. Figure 1 illustrates this architecture, also highlighting the 3D drum visualized by users, while Figure 2 shows the equipment used in this interaction.

[2] https://github.com/romulovieira-me/version2-physiodrum
[3] Demo video: https://youtu.be/EGDFz3pzZWg

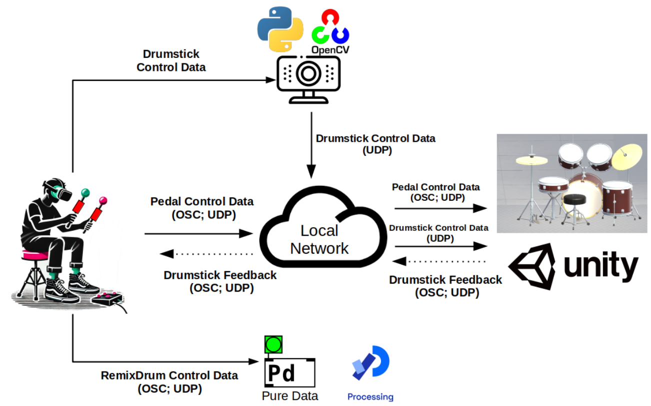
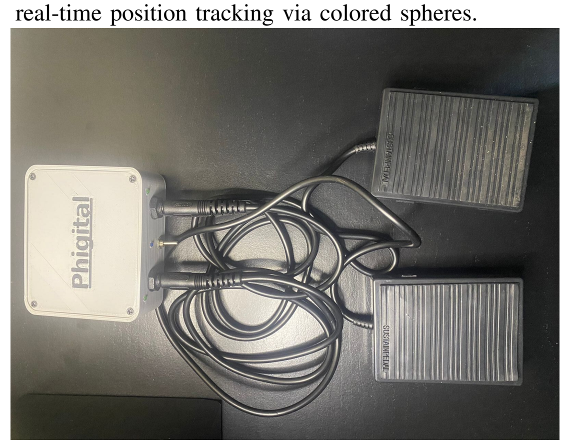


Fig. 1: PhysioDrum Architecture [23].

(a) RemixDrum 3.0 equipped with haptic feedback and real-time position tracking via colored spheres.



(b) Electronic foot pedals used for triggering virtual drum components. The pedals are connected to a circuit that transmits activation signals to the application via OSC.

Fig. 2: PhysioDrum equipment [23].

C. Protocol for Collection and Analysis of Physiological Data

The Bitalino modular device, known for its low cost and widespread use in academic research, was used to capture EEG signals. Three electrodes were placed on the participants' left prefrontal region at positions Fp1, F3, and F7 of the international 10–20 system. These sites were chosen to capture neural activity from the prefrontal cortex, a region associated with decision-making, planning, problem-solving, emotional regulation, attention, and memory. Data were transmitted via Bluetooth, the device's default communication protocol, to a Python script responsible for acquisition and storage. The signals were saved in CSV files for later inspection.

For data evaluation, statistical methods were employed for data cleaning, transformation, normalization, and segmentation, as detailed below [25]:

- **Channel Selection:** Only the frontal channel Fp1 from the 10–20 system [26] was selected for analysis, in order to facilitate potential comparisons with other datasets and previously collected data;
- **Segmentation:** The data were analyzed in 200-millisecond intervals. This granularity allows capturing quick variations in attention state without losing temporal resolution;
- **Filters:** High-pass filters with a cutoff frequency of 0.5 Hz and low-pass filters with a cutoff frequency of 50 Hz were applied to attenuate undesired frequency components;
- **Independent Component Analysis (ICA):** This technique was applied to eliminate noise resulting from ocular movement, muscle activity, and cardiac pulsation;
- **Wavelet Transform:** It was employed using the Daubechies method[4] [27] to decompose the signals into their respective time and frequency components [28]. This spectral decomposition allows identifying activation patterns across frequency bands commonly associated with distinct cognitive states;
- **Continuous Wavelet Transform (CWT):** Applied to further decompose the EEG signals, enabling the identification of specific temporal and frequency characteristics [29];
- **Spectrograms:** To visualize the temporal distribution of signal power, spectrograms were generated using coefficients from both the Daubechies wavelet and the CWT. The resulting values were normalized to a scale from 0 to 256 and categorized into five distinct bands: 1 = [0–31], 2 = [32–76], 3 = [77–153], 4 = [154–204], and 5 = [205–256]. Each spectrogram was produced with a resolution of 256 × 256 pixels [30].

Ultimately, the generated data are analyzed to assess participants' ability to sustain attentional focus across different sessions and to examine the impact of haptic feedback on these measured indicators.

Based on these analyses, it becomes possible to infer which system functionalities contributed to maintaining user engagement, as well as to derive insights into how immersive musical applications should be designed to foster sustained immersion and engagement.

IV. Experimental Setup & Data Analysis

The following section describes the experimental setup employed to investigate participants' attentional responses within an immersive musical environment. It also presents a comparative analysis of attention levels between sessions and the impact of haptic feedback on these measures.

A. Participants

A controlled laboratory study was conducted between May 5 and 27, 2025, at Fluminense Federal University, employing a between-subjects design. 30 participants were randomly and evenly assigned to two groups: Group A received uniform haptic feedback, whereas Group B experienced varying haptic responses. Participants ranged in age from 18 to 54 years (M = 26.67, SD = 10.23), with 24 identifying as male and 6 as female. The sample included individuals from diverse educational backgrounds, particularly in Computer Science, Engineering, and Psychology. Among them, two participants reported intermediate or advanced knowledge in both music and technology, six indicated expertise in VR systems, and ten had prior experience in music. The remaining participants were beginners in both areas.

The experimental procedures were conducted as part of the SenseGames Project and received approval from the Research Ethics Committee (REC) of Fluminense Federal University (protocol number: 88638025.0.0000.8160). All activities adhered to the institution's ethical standards, including the acquisition of voluntary participation through a duly signed Informed Consent Form. Participants were assured of the confidentiality and anonymity of the data provided and were explicitly informed of their right to withdraw from the study at any time without incurring any penalty.

B. Procedure

The testing process was organized into three successive stages. The first consisted of a free exploration phase, during which each participant had three minutes to interact freely with the PhysioDrum system. This initial period aimed to allow users to become familiar with the required motion dynamics, such as speed and range of movement necessary to interact with the virtual drum kit, as well as to adapt to the auditory and visual feedback provided in response to their actions.

Subsequently, participants were instructed to perform four rhythmic tasks corresponding to standard musical tempo markings: largo (40 BPM), adagio (66 BPM), andante (76 BPM), and moderato (108 BPM) [31]. In the *largo* task, participants alternated between the hi-hat and the snare drum, striking one instrument per measure at a very slow tempo. The *adagio* task introduced a more complex sequence, consisting of a high tom on the first beat, a combined hit on the snare drum and floor tom on the second and third beat, and a low tom on the fourth beat. In the *andante* task, participants performed a repeating three-beat cycle, striking the bass drum twice followed by the snare drum, thereby increasing both tempo and motor demand. Finally, the *moderato* task required a continuous four-beat pattern comprising bass drum, cymbal, snare drum, and cymbal, executed in sequence at a faster pace. Each participant completed the task under two conditions: one with vibrotactile feedback enabled and the other without it, in which the response was delivered immediately after striking a

[4]Daubechies wavelets are a class of orthogonal wavelets defined by Ingrid Daubechies, widely used in signal processing due to their compact support and high number of vanishing moments, which allow for accurate representation of both smooth and transient features in the signal.

drum component. The sequence of conditions was randomized to minimize potential order effects and prevent systematic bias. This game-like activity expanded the usability of the system, ensuring its appeal to both novice and experienced users. On average, each session lasted approximately 15 minutes.

C. Statistical Comparison Between Sessions

Upon completion of the testing sessions, a statistical comparison was conducted between different levels of attentional engagement, categorized into three ranges: low (0%–45%), medium (46%–74%), and high (75%–100%), for each of the two experimental sessions [32]. In this study, attention is defined as a continuous index derived from EEG signals. Increased power in the beta (13–30 Hz) and low gamma (30–45 Hz) frequency bands is consistently linked to attentional focus, active engagement, and sensorimotor coordination. In contrast, elevated theta (4–8 Hz) or delta (0.5–4 Hz) activity are often associated with fatigue, cognitive overload, or reduced vigilance [33]. Therefore, higher attention levels are considered desirable, as they indicate greater immersion, sustained focus, and responsiveness to task demands.

The primary objective of the analysis was to assess whether attention levels varied between sessions, and whether task repetition and familiarity contributed to more stable or elevated attentional states. To this end, a within-subject design was adopted, using each participant's performance as their own baseline. This approach allows for isolating the effects of experimental conditions while controlling for individual differences [34].

For each participant *p*, the time spent within each attention range was estimated for Session 1 ($T_{p,1}$) and Session 2 ($T_{p,2}$) by multiplying the proportion of EEG samples falling into each attention category by the total session duration.

This approach relies on relative proportions of EEG samples to produce normalized time estimates that are robust to sampling rate variations [35]–[37]. Rather than modeling temporal dependencies or state transitions, we focus on aggregate time spent in each attention range across sessions, a method widely used in psychophysiological research to assess cognitive states such as attention or engagement [38], [39]. Temporal modeling would require high-resolution annotations and specialized tools (e.g., Hidden Markov Models), which are beyond the scope of this exploratory study [35].

The difference between these time estimates was then calculated for each participant, as defined in Equation 1:

$$D_p = T_{p,2} - T_{p,1} \tag{1}$$

The resulting set of differences D_p, representing within-subject changes in time spent per attention category between sessions, was subsequently tested for normality. Specifically, it was examined whether the distribution of these differences approximated a Gaussian (normal) distribution. To assess the normality of the data, the Shapiro–Wilk test [40] was applied.

Across all attention ranges, the results indicated a violation of the normality assumption, with *p-values* below 0.001. Practically, this suggests that the distribution of D_p was asymmetric, potentially exhibiting outliers or a concentration of values away from the curve center. Given the violation of the normality assumption, the Wilcoxon signed-rank test [41] was selected as a nonparametric alternative suitable for paired samples.

Table I presents the statistical results. The Wilcoxon statistic W measures the symmetry in the differences between the two sessions, specifically how the time spent in each attention band during Session 1 compares to that in Session 2. A lower value of W suggests a greater difference between the analyzed conditions.

The associated *p-value* indicates the probability that the observed differences between sessions could have occurred due to random variation, assuming that there is no significant effect (null hypothesis). If the value is below the significance threshold of $\alpha = 0.05$, it indicates a statistical difference.

Medians are reported as descriptive statistics to summarize the distribution of time spent within each attention range in the two sessions. In the context of non-normally distributed data, the median serves as a more appropriate measure of central tendency than the mean, as it is less sensitive to skewness and outliers. Moreover, since the Wilcoxon signed rank test evaluates whether the median of the paired differences is significantly different from zero, reporting medians aligns directly with the underlying statistical hypothesis of the test.

The analysis revealed a statistically significant difference in the low attention range, with a consistent reduction in the time spent in this category during Session 2. This result is supported not only by the magnitude of the difference, but primarily by the directionality and consistency of the changes across participants (nearly all showed decreased time in the low range, with few ties or reversals). Such uniformity increases the rank-based sensitivity of the Wilcoxon test, resulting in a lower *p-value*. This pattern may reflect reduced inattention due to system adaptation, implicit learning of task dynamics, or growing familiarity with the interactive environment.

In contrast, although the high attention range exhibited a similar magnitude of mean increase, the effect was more variable between individuals: some participants improved while others regressed or remained unchanged. This lack of directional consistency, combined with greater dispersion and the presence of ties, weakens the statistical signal, resulting only in a trend toward significance (*p-value* = 0.052). The attention range did not show significant changes between sessions, reinforcing the notion that most attentional adaptation occurred at the extremes of the engagement scale.

Taken together, these results indicate that attentional modulation occurred primarily at the lower and upper bounds of the attention spectrum. The observed pattern, a statistically supported reduction in low attention and a non-significant upward trend in high attention, suggests that repeated exposure to the task led to measurable improvements in attentional engagement, even in the absence of external instructional or sensory cues. This dynamic appears to reflect a process of familiarization and implicit learning, whereby participants

TABLE I: Results of the Wilcoxon test comparing the time spent within each attention range between sessions.

Attention Range	Median – Session 1 (s)	Median – Session 2 (s)	W Statistic	*p-value*
Low	149.0 s	135.0 s	89.0	0.048
Medium	111.4 s	112.9 s	130.0	0.396
High	65.9 s	74.2 s	90.0	0.052

progressively attune to the demands of the interactive environment.

D. Overall Distribution of Attention Levels

To complement the statistical findings, the overall distribution of attentional levels was also analyzed using histograms. Figure 3 presents these values in Sessions 1 and 2, revealing consistent overall patterns, while also highlighting relevant individual variations between participants.

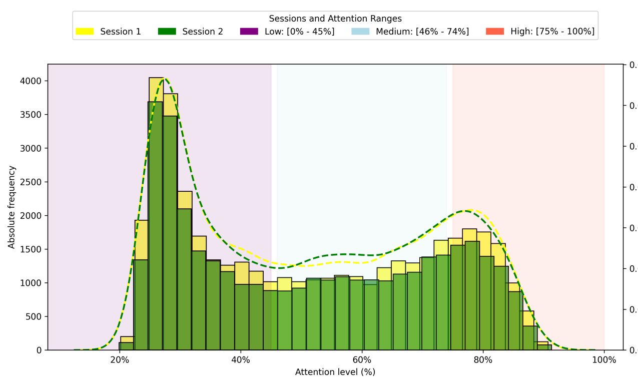


Fig. 3: General Attention Distribution by Session.

It was observed that most attention-related occurrences were concentrated within the low attention range, particularly during the initial moments of the experiment, with frequency peaks around 30% in both sessions. This pattern is consistent with findings reported in the literature [42], which describe a cognitive latency period of approximately 1 to 2 seconds before the attentional system reaches higher levels of engagement, as users gradually adapt to the environment and interactive stimuli.

Despite the initial predominance of low attention levels, a gradual increase in attentional engagement was also observed throughout the task, with substantial occurrences in the medium and high ranges, particularly toward the end of the sessions. This pattern suggests a progressive increase in user engagement, potentially associated with task complexity and growing familiarity with the environment. In summary, participants spent less time in a low attention state during the second session, indicating greater stability and focus at this stage.

Comparatively, the results suggest that Session 1 elicited a higher absolute frequency of attentional events across all ranges, possibly due to a novelty effect. Session 2 showed indications of longer episodes of sustained high attention, in alignment with the statistical analysis. This adjusted attentional profile is also consistent with findings from studies on implicit learning and cognitive automation in immersive interactive systems, where task repetition enables greater cognitive efficiency and sustained engagement over time [43], [44].

E. Statistical Evaluation of Haptic Feedback on Attentional Engagement

A second statistical analysis was conducted to evaluate the impact of haptic feedback on participants' attention levels. This analysis compared the time spent in different attention ranges under conditions with and without tactile stimulation. Before making this comparison, the data distribution was reassessed using the Shapiro–Wilk test to confirm the normality assumption. The results revealed significant deviations from normality across all attention ranges ($p < 0.05$), thereby precluding the use of parametric statistical methods.

Given the non-normal distribution of the data and the fact that participants experienced both conditions, a within-subjects design was used, and the Wilcoxon signed-rank test was applied again. Table II summarizes the descriptive statistics and test results.

In the low attention range, the median time decreased in the haptic condition compared to the non-haptic condition, suggesting a potential reduction in periods of low attentional engagement when tactile stimulation was present. Although the result was not statistically significant, the directional trend observed in the data points to a modest effect. The corresponding Wilcoxon signed-rank statistic reflects a predominance decrimental shift across participants, reinforcing the hypothesis of reduced distraction under multisensory stimulation.

For the medium attention range, the medians across conditions were nearly identical, and the statistical test indicated no evidence of systematic differences. This suggests that haptic feedback had no meaningful influence on moderate levels of attentional engagement, and participants maintained similar attentional performance regardless of tactile stimulation.

In contrast, the high attention range showed an increase in median time during the haptic condition, suggesting an enhancement in sustained attentional states when tactile feedback was available. Whereas the result did not reach statistical significance, the consistent directionality of change across participants and the descriptive pattern observed support the interpretation of a possible trend toward deeper and more stable attentional engagement induced by multisensory cues.

Although no statistically significant effects were observed, the analysis shows a consistent pattern indicating that haptic feedback may decrease instances of inattentiveness and promote sustained focus. These findings highlight the potential value of multisensory stimulation in enhancing attentional dynamics during interactive tasks.

TABLE II: Wilcoxon test results comparing the time spent in each attention range between sessions with and without haptic feedback.

Attention Range	Median – Non-Haptic (s)	Median – Haptic (s)	W Statistic	*p-value*
Low	144.2 s	135.3 s	102.0	0.087
Medium	112.6 s	113.1 s	117.0	0.365
High	66.8 s	75.1 s	104.0	0.095

F. Exploratory Analysis of Haptic Feedback Based on Session Order

Figure 4 illustrates the temporal evolution of participants' attention throughout the experimental sessions, divided into four conditions: Session 1 with haptic feedback, Session 1 without haptic feedback, Session 2 with haptic feedback, and Session 2 without haptic feedback. Each plot represents a bivariate density map, where the horizontal axis corresponds to time (in seconds) and the vertical axis to attention level (0–100). Color intensity ranges from purple (low density) to yellow (high density of occurrence). The red dashed line represents the attention threshold of 55%, which is used as a reference to identify moderate to high levels of attentional engagement. This value was empirically defined as a baseline, as it corresponds to the nearest integer to the upper bound average of the confidence intervals calculated during the participants' resting state. Values below this threshold are associated with lower spectral band amplitudes, which are consistent with physiological patterns typically observed during rest. The adoption of this threshold aligns with common practices in EEG studies that do not employ supervised learning, enabling the analysis of relative fluctuations in attention based on a standardized and data-driven reference point [29], [38], [39], [45].

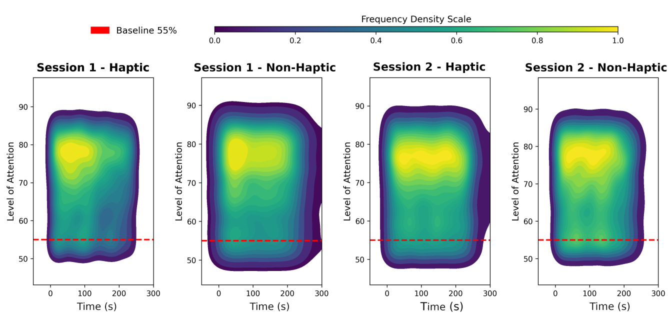


Fig. 4: Distribution of attention focus density across experimental sessions.

Overall, the results indicate that participants maintained attention levels above the baseline for most of the interaction period, with a significant concentration of values between 70% and 85%. In the conditions involving haptic feedback, particularly in Session 2, a more pronounced density peak is observed in the upper region of the graph, suggesting a higher frequency of elevated attention throughout the task. This more concentrated and earlier distribution (between 50 and 150 seconds) suggests that tactile feedback contributed to a more effective elevation and maintenance of attentional levels.

In Session 1 with haptic feedback, a more concentrated and sustained density at higher attention levels can be observed at the beginning of the experiment, indicating that tactile feedback contributed positively to system adaptation and the execution of initial tasks. In contrast, in Session 2 without haptic feedback, attention levels remained stable, albeit at slightly lower values compared to the previous session. This stability may be attributed to participants' growing familiarity with the task and adaptation to the experimental environment, making haptic feedback less essential during a second exposure.

These findings reinforce the hypothesis that haptic feedback may serve as a catalyst for initial engagement, particularly in contexts where users are still acclimating to the interface and task dynamics, and subsequently helps to smooth attentional fluctuations, contributing to a more stable concentration profile throughout the experience. Likewise, this feature appears to support a gradual transition toward autonomous engagement.

V. Discussion

The results suggest that the PhysioDrum experience supports a progressive stabilization and elevation of attentional focus over time. When comparing the two experimental sessions, a significant reduction in time spent in low attention states was observed, along with a trend toward increased high attention during the second exposure to the task. While the first session exhibited a broader distribution of attentional events, the second was characterized by longer and more consistent episodes of sustained attention. This transition suggests that task repetition contributed to greater cognitive stability and continuous focus during interaction, likely due to increasing familiarity with the immersive environment.

The analysis of haptic feedback effects complements these findings by revealing a systematic, though not statistically significant, trend of improvement at both ends of the attentional spectrum. Tactile stimulation was associated with a reduction in inattentive episodes and a prolongation of high-attention states. These effects were particularly evident in the first session with feedback, where attention levels rose rapidly in the early stages, indicating that multisensory input may facilitate initial adaptation to the task and environment. In contrast, during the second session without feedback, attention levels remained stable, albeit slightly lower, suggesting that participants developed attentional self-regulation mechanisms based on previous experience.

The findings suggest a hybrid adaptive mechanism in which attentional focus is initially driven by external sensory input, such as haptic feedback, and progressively maintained through internal cognitive regulation shaped by task repetition and environmental predictability. This dynamic is particularly significant in immersive musical performance contexts, where achieving a balance between sensory engagement and cognitive stability is critical. The findings suggest that PhysioDrum can support elevated attention levels without relying on explicit instruction or additional external prompts, indicating its potential to promote immersion, fluid motor execution, and expressive interaction. In this process, haptic feedback plays a dual role: facilitating rapid initial engagement and subsequently supporting the refinement of attentional control, thereby enhancing both consistency and quality of performance.

These results carry important implications for the design of immersive musical systems and other applications that demand sustained attentional engagement. By demonstrating that attentional states can be progressively stabilized through task repetition and selectively enhanced via haptic feedback, this study provides empirical support for integrating adaptive multisensory strategies into interactive environments. In musical contexts, such mechanisms can improve performers' concentration, reduce cognitive load during complex tasks, and foster deeper expressive involvement. Beyond musical applications, the approach may benefit domains such as cognitive training, neurorehabilitation, and immersive education, where maintaining focus over extended periods is essential. PhysioDrum may serve as an example of how multimodal feedback and iterative interaction design can be leveraged to support attentional regulation and potentially enhance user experience in real-time, sensor-driven systems.

Despite the promising outcomes, the study presents certain limitations concerning the participant profile. The sample was predominantly composed of male individuals (24 out of 30), which reduces the representativeness of diverse user demographics and constrains the generalizability of the findings. In addition, fewer than half of the participants reported prior experience with musical practice or XR systems, which may have influenced their initial attention levels and interaction patterns with the multisensory stimuli. These factors contribute to the variability of the response and may have amplified the effects of novelty and adaptation between sessions. Lastly, as the study was conducted in a controlled laboratory environment, it may not fully capture the dynamics of real-world or longitudinal use, highlighting the importance of conducting follow-up studies in more ecologically valid settings.

VI. Conclusion

This paper presented an analysis of users' attentional focus in immersive musical and multisensory environments, employing non-invasive methods and involving participants with varying levels of musical and technological proficiency. It also examined in detail the role of tactile feedback in influencing users' concentration and engagement with the system. The goal of this study was not to examine how the reported attention values relate to applications in broader HCI or XR domains. Instead, the analysis was deliberately focused on the role of attention within immersive musical applications. To the best of our knowledge, this is the first research effort to explore these dimensions within this specific context.

The findings revealed individualized responses, with significant differences in physiological characteristics depending on the order of exposure to experimental conditions. These variations were influenced by whether participants received vibrotactile feedback in the initial session or by their increased familiarity and engagement in the second session, factors that directly impacted levels of attention and focus. Such results are consistent with the inherently subjective and emotional nature of musical practice, where both biological responses and system perception vary not only between users but also within the same user over time, depending on contextual and temporal factors.

Participants showed higher attentional focus during the initial adaptation phase, with varied tasks and difficulty levels helping sustain engagement. Haptic feedback supported both adaptation and long-term focus. These results deepen understanding of how the brain processes musical information across interaction stages and highlight the value of physiological signals as personalized input for developing emotionally responsive immersive musical systems—the core contribution of this research.

Future work will explore how attention interacts with immersive environments across diverse user profiles—including musicians, VR experts, and neurodivergent individuals to inform personalized design strategies for expressive and therapeutic applications. We also aim to develop an open, multimodal dataset with physiological signals, interaction logs, and annotations to support reproducibility and adaptive system training. Finally, upcoming studies will triangulate physiological, quantitative, and qualitative data to enhance user experience and guide the evolution of PhysioDrum.

Acknowledgment

The authors would like to thank the support of the following funding agencies CAPES, CAPES Print, CNPq, FINEP, FAPERJ, INCT-ICONIoT and INCT-MACC.

References

[1] I. Daly, "Neural decoding of music from the eeg," *Scientific Reports*, vol. 13, no. 1, p. 624, 2023.

[2] X. Hu, F. Li, and T.-D. J. Ng, "On the relationships between music-induced emotion and physiological signals." in *ISMIR*, 2018, pp. 362–369.

[3] I. Arapakis, I. Konstas, and J. M. Jose, "Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 461–470.

[4] O. Barral, M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, N. Ravaja, S. Kaski, and G. Jacucci, "Exploring peripheral physiology as a predictor of perceived relevance in information retrieval," in *Proceedings of the 20th international conference on intelligent user interfaces*, 2015, pp. 389–399.

[5] S. Sanyal, S. Nag, A. Banerjee, R. Sengupta, and D. Ghosh, "Music of brain and music on brain: a novel eeg sonification approach," *Cognitive neurodynamics*, vol. 13, pp. 13–31, 2019.

[6] S. Cheung, E. Han, A. Kushki, E. Anagnostou, and E. Biddiss, "Biomusic: An auditory interface for detecting physiological indicators of anxiety in children," *Frontiers in neuroscience*, vol. 10, p. 401, 2016.

[7] I. Daly, N. Nicolaou, D. Williams, F. Hwang, A. Kirke, E. Miranda, and S. J. Nasuto, "Neural and physiological data from participants listening to affective music," *Scientific data*, vol. 7, no. 1, p. 177, 2020.

[8] R. Vieira, S. Wei, T. Röggla, D. C. Muchaluat-Saade, and P. César, "Immersive io3mt environments: Design guidelines, use cases and future directions," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 2024, pp. 1–10.

[9] L. Turchet, "Musical metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1811–1827, 2023.

[10] R. Vieira, D. C. Muchaluat-Saade, and P. César, "Towards an internet of multisensory, multimedia and musical things (io3mt) environment," in *2023 4th International Symposium on the Internet of Sounds*. IEEE, 2023, pp. 1–10.

[11] L. Turchet, "Smart musical instruments: vision, design principles, and future directions," *IEEE Access*, vol. 7, pp. 8944–8963, 2018.

[12] R. Vieira, M. Rocha, C. Albuquerque, D. C. Muchaluat-Saade, and P. César, "Remixdrum: A smart musical instrument for music and visual art remix," in *2023 IEEE 9th World Forum on Internet of Things (WF-IoT)*. IEEE, 2023, pp. 1–7.

[13] S. Blain-Moraes, S. Chesser, S. Kingsnorth, P. McKeever, and E. Biddiss, "Biomusic: A novel technology for revealing the personhood of people with profound multiple disabilities," *Augmentative and Alternative Communication*, vol. 29, no. 2, pp. 159–173, 2013.

[14] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an eeg-based recognition of music liking using time-frequency analysis," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3498–3510, 2012.

[15] M. B. Happ, "Interpretation of nonvocal behavior and the meaning of voicelessness in critical care," *Social science & medicine*, vol. 50, no. 9, pp. 1247–1255, 2000.

[16] S. Blain and P. McKeever, "Revealing personhood through biomusic of individuals without communicative interaction ability," *Augmentative and Alternative Communication*, vol. 27, no. 1, pp. 1–4, 2011.

[17] J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, and Z. Jin, "Towards effective music therapy for mental health care using machine learning tools: human affective reasoning and music genres," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 11, no. 1, pp. 5–20, 2021.

[18] J. H. Janssen, J. N. Bailenson, W. A. IJsselsteijn, and J. H. Westerink, "Intimate heartbeats: Opportunities for affective communication technology," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 72–80, 2010.

[19] R. Bootsma, "Listening to your loved one's biomusic: Auralizing biosignals to enhance social connection," Master's thesis, Utrecht University, 07 2024.

[20] B. Wan, Q. Wang, K. Su, C. Dong, W. Song, and M. Pang, "Measuring the impacts of virtual reality games on cognitive ability using eeg signals and game performance data," *IEEE Access*, vol. 9, pp. 18 326–18 344, 2021.

[21] I. V. Petukhov, A. E. Glazyrin, A. V. Gorokhov, L. A. Steshina, and I. O. Tanryverdiev, "Being present in a real or virtual world: A eeg study," *International journal of medical informatics*, vol. 136, p. 103977, 2020.

[22] R. S. Calabrò, A. Naro, M. Russo, A. Leo, R. De Luca, T. Balletta, A. Buda, G. La Rosa, A. Bramanti, and P. Bramanti, "The role of virtual reality in improving motor performance as revealed by eeg: a randomized clinical trial," *Journal of neuroengineering and rehabilitation*, vol. 14, pp. 1–16, 2017.

[23] R. Vieira, D. C. Muchaluat-Saade, and P. Cesar, "Physiodrum: Bridging physical and digital realms in immersive musical interaction," in *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, 2025, pp. 356–358.

[24] C. Mele, T. R. Spena, M. Marzullo, and I. Di Bernardo, "The phygital transformation: a systematic review and a research agenda," *Italian Journal of Marketing*, vol. 2023, no. 3, pp. 323–349, 2023.

[25] C. E. C. Silva, R. Vieira, D. G. Trevisan, and D. C. Muchaluat-Saade, "Towards analysing user attention using electroencephalography in immersive multisensory virtual environments," in *ACM International Conference on Interactive Media Experiences Workshops (IMXw)*. SBC, 2025, pp. 91–95.

[26] G. H. Klem, "The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, pp. 3–6, 1999.

[27] C. Vonesch, T. Blu, and M. Unser, "Generalized daubechies wavelet families," *IEEE transactions on signal processing*, vol. 55, no. 9, pp. 4415–4429, 2007.

[28] R. Ramos, B. Valdez-Salas, R. Zlatev, M. Schorr Wiener, and J. M. Bastidas Rull, "The discrete wavelet transform and its application for noise removal in localized corrosion measurements," *International Journal of Corrosion*, vol. 2017, no. 1, p. 7925404, 2017.

[29] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[30] M. Stephane, "A wavelet tour of signal processing," 1999.

[31] A. Fernández-Sotos, A. Fernández-Caballero, and J. M. Latorre, "Influence of tempo and rhythmic unit in musical emotion regulation," *Frontiers in computational neuroscience*, vol. 10, p. 80, 2016.

[32] C. Silva, R. Vieira, D. Trevisan, and D. Muchaluat-Saade, "Towards analysing user attention using electroencephalography in immersive multisensory virtual environments," in *Proceedings of the ACM International Conference on Interactive Media Experiences Workshops*. Porto Alegre, RS, Brasil: SBC, 2025, pp. 91–95. [Online]. Available: https://sol.sbc.org.br/index.php/imxw/article/view/35234

[33] A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using eeg spectral power and erps in the n-back task," *Journal of neural engineering*, vol. 9, no. 4, p. 045008, 2012.

[34] V. A. Thompson and J. I. Campbell, "A power struggle: Between-vs. within-subjects designs in deductive reasoning research," *Psychologia*, vol. 47, no. 4, pp. 277–296, 2004.

[35] F. Lotte, "A tutorial on eeg signal-processing techniques for mental-state recognition in brain–computer interfaces," *Guide to brain-computer music interfacing*, pp. 133–161, 2014.

[36] M. X. Cohen, *Analyzing neural time series data: theory and practice*. MIT press, 2014.

[37] X. Gu, Z. Cao, A. Jolfaei, P. Xu, D. Wu, T.-P. Jung, and C.-T. Lin, "Eeg-based brain-computer interfaces (bcis): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 5, pp. 1645–1666, 2021.

[38] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using eeg's and peripheral physiological signals," in *International workshop on multimedia content representation, classification and security*. Springer, 2006, pp. 530–537.

[39] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, 2019.

[40] N. M. Razali, Y. B. Wah *et al.*, "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests," *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[41] G. Divine, H. J. Norton, R. Hunt, and J. Dienemann, "A review of analysis and sample size calculation considerations for wilcoxon tests," *Anesthesia & Analgesia*, vol. 117, no. 3, pp. 699–710, 2013.

[42] R. Hassan, S. Hasan, M. J. Hasan, M. R. Jamader, D. Eisenberg, and T. Pias, "Human attention recognition with machine learning from brain-eeg signals," in *2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 2020, pp. 16–19.

[43] I. Miguel-Alonso, D. Checa, H. Guillen-Sanz, and A. Bustillo, "Evaluation of the novelty effect in immersive virtual reality learning experiences," *Virtual Reality*, vol. 28, no. 1, p. 27, 2024.

[44] P. Bondesan, A. Canal, S. Fleury, A. Boisadan, and S. Richir, "Implicit learning of professional skills through immersive virtual reality: a media comparison study," in *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2025, pp. 442–449.

[45] J. Xu, F. Ren, and Y. Bao, "Eeg emotion classification based on baseline strategy," in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, 2018, pp. 43–46.