Immersive Io3MT Environments: Design Guidelines, Use Cases and Future Directions

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

Shu Wei
Department of Psychiatry
University of Oxford
Oxford, UK
shu.wei@psych.ox.ac.uk

Thomas Röggla
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
t.roggla@cwi.nl

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*—The Internet of Multisensory, Multimedia, and Musical Things (Io3MT) is an emerging field at the intersection of computer science, arts, and humanities. It focuses on integrating technologies and datasets to explore human sensory perception, multimedia elements, and musical compositions for artistic and entertainment purposes. This position paper advocates for merging Io3MT with Extended Reality (XR) in creative contexts. Through literature review and expert focus group discussions, we have identified five key design guidelines that aid in the development of immersive environments for artistic creation within the Io3MT framework. We developed PhysioDrum as a practical demonstration of these guidelines. This work provides insights into the infrastructure, tools, opportunities, and research challenges arising from integrating Io3MT and XR.**

***Index Terms*—Io3MT, Extended Reality, Immersive Environment, Artistic Creation**

I. Introduction

The Internet of Multisensory, Multimedia, and Musical Things (Io3MT) [1] represents a domain within the Internet of Sounds (IoS) [2] that arises from the confluence of concepts from the exact sciences, such as Networked Music Performance (NMP), Wireless Multimedia Sensor Networks (WMSNs), Multiple Sensorial Media (Mulsemedia), and Interactive Art, as well as concepts from the humanities and the arts. From a computational perspective, it aims to aggregate, in a non-hierarchical manner, all these pieces of information in one place and simultaneously, thereby allowing a specific category of data to act upon or influence the others. For instance, a chord sequence may alter a color pattern in a video, or the display of a particular image might trigger scents reminiscent of the landscape depicted in that figure. This mode of operation gives rise to a network of devices capable of detecting, acquiring, processing, and exchanging data that facilitate the connection between the digital and physical realms through information and communication technologies, thus promoting the emergence of new creative applications and services that explore multimodal, multi-platform, and transmedia relationships [1]. While Internet of Musical Things (IoMusT) [3]suggests the use of multimedia elements, the field has yet to establish protocols, data formats, architectures, and tools for integrating sound with multimedia and multisensory information. Io3MT addresses this gap, enabling interoperability between these three types of information, where changes in one can influence another, fostering new forms of musical and artistic expression. While this discussion goes beyond the scope of this work, further exploration can be found in [1].

Despite being in its early stages, practical applications of these concepts have already been explored, such as in networked music performance [1] and multimedia remixing practices [4] utilizing smart musical instruments (SMIs) [5]. Nevertheless, current literature lacks comprehensive proposals extending these capabilities to the realm of Extended Reality (XR). In parallel, the predominant focus of the current state of the art in XR technology has been directed towards enhancing audience experiences, with a strong emphasis on visual components, often at the expense of musical production and sensory effects, which are frequently treated as secondary considerations. Furthermore, there is an absence of a well-defined framework that systematically categorizes the various design dimensions of XR systems with an emphasis on artistic factors. Additionally, there is a limited understanding of musical interactions within shared and collaborative virtual environments. [6], [7].

Therefore, this study proposes a combination of Io3MT concepts and the Musical Metaverse [8] to create immersive, multimodal, and interoperable entertainment applications. Consequently, our ultimate goal is to: i) introduce five design principles for developers, artists, researchers, and students to create their own environments through a common lens, also indicating enabling technologies and affordances that facilitate the creation of content that amplifies user experiences; and ii) present a demonstration called PhysioDrum[1], which serves as a proof of concept for the proposed guidelines and helps to enhance musical expression possibilities. This paper does not aim to conduct a quantitative analysis of the prototype, such as measuring latency, user satisfaction scores, or performance accuracy.

The structure of this paper is outlined as follows. Section II reviews prior studies on frameworks for music practice mediated by technology and the utilization of music in XR environments. Section III describes the focus group and subsequent data analysis for the formulation of design guidelines. Section IV presents the identified guidelines and propose potential use cases. Section V suggests a practical implementation to demonstrate the validity of the framework. Section VI provides a discussion on the functionality of this installation and presents the lessons learned from its development. Finally, Section VII summarizes the findings derived from the development of this work.

II. Related Work

From a review of academic research, industry documents, and corporate descriptions, related work can be categorized into two groups: Frameworks for Specifying Technology Use in Musical Practice and Frameworks for Creating Musical Experiences in XR.

A. Frameworks for Specifying Technology Use in Musical Practice

One of the initial endeavors to delineate observations on design factors, artistic considerations, and human elements in musical devices development was conducted by Cook [9]. Those principles stress the importance of designing devices that are distinct from traditional computers, learning from user interactions, and avoiding mere replication of existing instruments. The guidelines advocate for accommodating diverse musical backgrounds, immediate sound generation, using appropriate digital protocols like MIDI, and maintaining alternatives like Open Sound Control (OSC). Other principles include avoiding batteries, preferring wired connections, innovating with new algorithms and controllers, enhancing existing instruments, and using everyday objects as inspiration for new musical devices.

Subsequent to the initial formulation of the design principles, the author undertook a comprehensive reevaluation to assess their continued relevance and application in practice [10]. The review confirmed its overall validity, however, several suggestions for modifications were made in light of subsequent technological advancements. The principle that originally discouraged the use of batteries has become obsolete as modern batteries are smaller, more efficient, and longer-lasting. Meanwhile, the guideline advocating for wired connections has been gradually supplanted by more flexible and lighter wireless technologies such as Bluetooth, Wi-Fi, and ZigBee. Moreover, the author extends the principles to musical interface design by proposing new guidelines, such as playful designs to enhance engagement, the inclusion of simple and recognizable forms to facilitate use, backward compatibility, the inclusion of diagnostic tools, and involving novices in the design process, as their perceptions can aid in innovation.

Following those developments, Wang [11] outlined principles for expressive visual design in computational music, emphasizing real-time audiovisual integration. Key principles include design sound and graphics together, ensuring visual attractiveness, prioritizing musical experience, introducing constraints, enhancing interaction with visual feedback, focusing on essential elements, using motion animations judiciously, embedding personality in visuals, developing unique aesthetics, iterating on architecture, and using algorithm visualizations to deepen understanding and inspire innovation.

The DIAMOUSES project [12], beyond proposing an application for real-time exchange of audiovisual streams, outlined a conceptual architecture for performance in such environments. Furthermore, notable contributions include Paine's work [13], which highlights structural and musical features of new interfaces for musical expression, and Weinberg's research [14], which defines and classifies aesthetic and technical aspects of interconnected musical networks, offering historical perspective on technological innovations in this field and discussing the paradigms underpinning these studies.

B. Frameworks for Creating Musical Experiences in XR

Although limited, there are some frameworks that propose guidelines for musical practice in virtual environments. These have been designed in the wake of rapid developments in the field of XR, as well as the increasing availability of low-cost technologies for this area of study. One of the prominent works in this domain is the proposal by Serafin et al. [15], which focuses on presenting and elucidating design principles, evaluation methods, case studies, and considering future challenges concerning virtual reality musical instruments (VRMIs). They emphasize the importance of feedback, mapping, and minimizing latency to enhance the multisensory experience and reduce cognitive load, thus preventing cybersickness. Furthermore, the framework suggests improving the ergonomics of the equipment, highlighting the importance of creating a sense of presence, representing the user's body accurately and promoting social interactions in these immersive environments.

The WAVE system [16] integrates 3D sound and XR to create immersive musical instruments, facilitating activities such as performance and composition. It features low-cost hardware, open-source software, and maintains high sound quality. Key features include scalability, mobility, visual feedback, easy environment control, and real-time sound and movement tracking.

Further, there is the work of Turchet et al. [17], which conduct a detailed analysis of 199 studies from the last decade, covering technical, artistic, perceptual, and methodological dimensions on the intersection between music and XR. This review is enriched by interviews with experts and provides insights that lead to the proposal of a research agenda for the

[1]Accessed on: August 19, 2024: https://l1nq.com/DqVMu

field as well as a reference framework. In addition, Ciciliani [18] investigates the possibilities of composition within 3D virtual environments, specifically examining how the design of virtual spaces (topologies) influences interactive sound sources and sonic events.

In addition to the framework proposals, there are some practical applications that help to understand musical capabilities in XR environments, such as Musical Metaverse Playground [8]. This project consists of two prototypes of playgrounds using WebXR and Web Audio technologies to create and test immersive sound experiences conveniently executed in web browsers integrated into commercially available standalone Head-Mounted Displays (HMD). Dziwis et al. [19] propose integrating two Bytebeat-based live coding languages into metaverse systems, enabling real-time, collaborative live coding in a virtual environment with immersive features like spatial audio rendering and XR device integration.

Analogous to these proposals and applications, this paper proposes five design principles that will form a framework for artistic creation based on the principles of Io3MT. However, the research presented here does not begin with *a posteriori* definition of the guidelines. Instead, it results from a user-centered process (focus group) to initially define the guidelines and use cases for these scenarios, and only thereafter moves to practical implementation, whereas previous frameworks were derived from an observation of the general situation. Another distinctive feature of this research is the incorporation of physical objects for interaction within the virtual environment, enhancing the physicality of XR environments driven by music.

III. Focus Group

We conducted a focus group with four specialists (three women and one man). Participant 1 (P1) is a postdoctoral researcher with experience as a developer and evaluator of XR systems, and also a drummer. Participant 2 (P2) is a master's student, proficient in VR, and a guitarist. Participant 3 (P3) holds a master's degree, is a bassist, and has experience in the development and use of XR for gaming. Participant 4 (P4) is a postdoctoral researcher, drummer, and pianist, with experience in the development of XR systems. All contributors are from the field of computer science and possess an amateur background in music, with a particular emphasis on acoustic instruments, as well as a focus on entertainment and performance.

The goal was to understand the current state of XR-driven artistic experiences and to envision their integration with Io3MT concepts. It is important to emphasize that different groups of experts, each with distinct interests and potential target audiences, will yield varying insights and opinions. Consequently, the content derived from this discussion should not be regarded as rigid or conclusive, but rather as points subject to review and refinement. Creative methods like brainstorming, group dynamics, and voting ensured balanced participation and diverse perspectives. The 1 hour and 40-minute session took place at Centrum Wiskunde & Informatica (CWI) in the Netherlands, and had the following structure.

- **Introduction (10 minutes):** The session began with the moderator and participants introducing themselves, followed by the moderator presenting an outline of the workshop flow. Subsequently, participants proceeded to sign the consent form.
- **Warm-up activity (10 minutes):** During the warm-up, participants shared their experiences in extended reality environments, not limited to artistic applications, discussing positive and negative aspects, as well as debating features that would enhance their experiences. This phase allowed for a broad range of ideas beyond artistic contexts.
- **Part 1 - Exploratory inquiries on experiences (30 minutes):** In this stage, participants explored components that might compose immersive environments based on Io3MT, discussing intended objectives and behaviors. Voting exercises prioritized key points, and interactive dynamics encouraged the exchange of ideas.
- **Part 2 - Exploratory inquiries on features (35 minutes):** This part of the workshop focused on specific resources, with participants pairing up to sketch user journeys upon entering the environment. Then, pairs were swapped to compare experiences, refine ideas, and discuss valuable moments that should be included in the environment being developed.
- **Final remarks (15 minutes):** Lastly, the moderator and participants reflected on XR-based artistic experiences, addressing implementation challenges and future research directions.

A thematic analysis [20] was performed to extract insights from the focus group data, providing a comprehensive understanding of participants' perceptions and experiences. Dovetail software was used to analyze transcripts and audio recordings. A tagging technique (commonly called coding) was applied to identify and categorize emerging themes and patterns, increasing objectivity and reducing bias. A single coder used the same eight tags applied by Lee et al. [21] to distinguish the opinions and feelings of focus group participants. Table I provides a summary of each theme and primary functionalities.

IV. Framework and Feasible Scenarios

Common points emerged from both the literature review and focus group discussions. Positive aspects of XR interactions include the ability to explore applications from various perspectives, engaging environments, freedom of navigation, and immersive experiences that prolong enjoyment. Areas for improvement include enhancing control precision, reducing HMD discomfort (weight, heat, wired connections), and addressing interaction and mobility limitations. Participants suggested incorporating progressive interaction stages and integrating multimodal resources and physicality to improve immersion.

TABLE I: Summary of Thematic Analysis Themes.

Themes	Function
Positive Moments	Highlights and categorizes positive interactions and experiences within XR environments
Negative Moments	Focuses on negative experiences to identify criticisms and sources of frustration
Could be Improved	Related to aspects of the XR system that can be improved or enhanced
Behavior/Actions	Delineates behaviors and actions that are commonly observed within the XR environments among participants
Requirements	Explores technical requirements, including necessary tools and specifications for optimal functionality
New Features/Tools	Identifies and categorizes suggested new features and tools that could enhance the quality and immersiveness
Raised Issues/Concerns	Focuses on less critical, minor issues that were mentioned during discussions
Additional Benefits	Highlights extra, non-essential benefits or positive aspects provided by the XR environments

A. *Design Principles for Immersive Io3MT Environments*

Based on this discussion, the focus group experts considered the pros and cons of XR systems and how they could be integrated with the concepts proposed in Io3MT. The outcome of this discussion led to the formulation of five design principles aimed at creating an immersive Io3MT environment. They are presented in Figure 1 and elaborated subsequently.

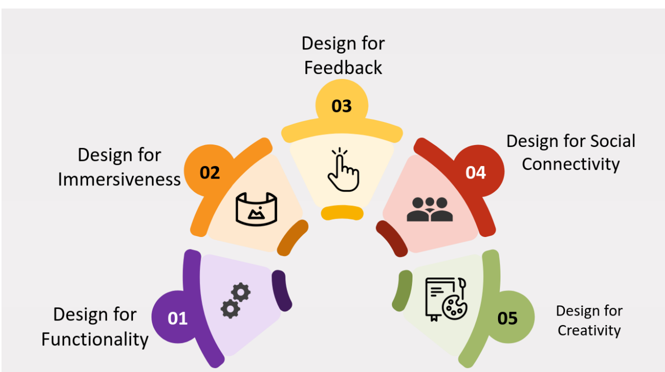


Fig. 1: Summary of proposed design principles.

Design for Functionality

Despite the hedonic nature of environments geared towards artistic creation, they must be developed to serve a specific purpose and also function as a means of expressing the ideas and emotions of the artists. In this regard, a series of technical factors must be taken into account to make these systems viable, such as synchronization in the exchange of information. Therefore, all interactions should exhibit minimal latency.

System interactions can be realistic and mirror real-world practices, avoiding reliance solely on simple button clicks. It is advisable to incorporate traditional sound creation methods like blowing, plucking, and bowing into the environment to enhance expressiveness and performance possibilities. These methods also improve audience communication by visually linking gestures to musical outcomes, aiding in the interpretation of the performer's intent. Drawing inspiration from existing instruments to expand their capabilities, rather than creating entirely new applications, can further these results.

At the same time, good ergonomics must be guaranteed not only for the controls and musical instruments used in this environment, but also for any HMD that may be used. Developers must be attentive to the tensions and discomforts that may arise from the use of such equipment, especially in long rehearsal and/or performance sessions, in addition to the limitations posed by wires and the weight of such hardware. Regarding the good usability of the system, cybersickness should be avoided.

System factors influencing this condition include the aforementioned latency, screen oscillation, calibration, and ergonomics. In addition to optimization factors such as efficient tracking and high refresh and frame rates, developers must pay attention to how the user's movement in the virtual environment is facilitated [15].

Design for Immersiveness

The degree of technological immersion offered by a particular system can be characterized by the range of normal sensorimotor contingencies supported by the system, i.e., the set of actions a user can perform to perceive something. For example, moving the head and eyes to change the line of sight, kneeling to get a closer look at the ground, or turning the head to locate the position of a sound source. Therefore, it is advisable that immersive Io3MT environments clearly communicate the system limitations, discouraging users from relying on contingencies not fully supported by the application. In addition to this factor, the sense of body ownership and embodiment also contributes to immersion. In other words, seeing a body in the virtual environment that one feels ownership of enhances the sensation of being inside that environment. This way, user representation can occur via a cartoon, avatar, or volumetric video [22].

Designers and developers of these environments should also consider the use of natural actions, which conform to the real world, and "magical" elements, which are not limited by the laws of physics, human anatomy, or the current state of technological development [15]. Importantly, both forms of interaction can be combined, such as manipulating non-isomorphic objects, creating synthetic sounds, and assigning ethereal attributes to avatars. The advantage of natural techniques is that the familiarity of such approaches can enhance usability, while magical approaches allow the artist to overcome limitations of the real world, such as financial and technical aspects. The combination of these factors with haptic elements, such as multisensory feedback and guidance, and an increase in the tangibility and physicality of the environment, can lead to new levels of abstraction, immersion, and imagination [15].

Another paramount consideration in crafting guidelines for an immersive XR environment geared towards artistic practice is the incorporation of a high-quality sound system. Elevating the auditory experience in XR not only heightens overall immersion but also contributes to the emotional resonance of the artistic content. A sophisticated sound system enables accurate spatialization, cultivating a heightened sense of realism by placing users within a three-dimensional sonic landscape. This spatial awareness, coupled with attention to detail in sound design, allows artists to convey nuanced expressions and intensify narrative impact. Furthermore, a high-quality sound system serves as a conduit for responsive feedback, enhancing user engagement and enriching the cross-modal integration of sensory elements.

Enhancing the immersiveness of Io3MT environments involves multiple factors, including multisensory interactions using controllers, joysticks, virtual musical instruments, and new interfaces for musical expression. Aesthetic elements like visual appeal and color also play crucial roles. This synthesis of sound and visuals, along with customization options such as user profiles, instrument selection, and predefined settings, creates a personalized and engaging experience. Artists' ability to combine superior audio with visual elements not only boosts immersion but also reflects their dedication to delivering a polished and professional experience.

Design for Feedback

Given that the fundamental concepts of Io3MT anticipate the creation of systems integrating musical, multimedia, and sensory elements, it is imperative to map this information to generate feedback when executed. From an auditory feedback perspective, it is crucial for virtual objects to correspond to the location and movement of real devices. Multimodal feedback can be achieved through precise synchronization among visual, auditory, and tactile elements, such as haptic vibrations. Incorporating this information, which may also relate to sonic nuances, enhances narrative comprehension and amplifies the user experience.

Simultaneously, the integration of sensory elements provides tangible physical responses, aiding in the development of musical skills. As these topics are strongly interconnected, tactile and kinesthetic cues are not only inevitable but also highly recommended for musical performance. Additionally, this type of information enhances the sense of presence and interactivity in the XR environment, improves the usability of physical equipment, and helps create more intuitive and easy-to-learn interfaces, allowing the use of such equipment without third-party items such as gloves or joysticks.

Design for Social Connection

One of the fundamental aspects of music is its ability to create shared social experiences. Conversely, current immersive applications often focus heavily on individual activities, where each person enters a unique virtual world. This is primarily due to the occlusive properties of HMD, blocking any visual communication with the external world. Recent developments in XR technologies, however, point towards the desire to create shared experiences, as seen in the concept of Social VR [22]. This principle refers to virtual environments that enable multiple users to interact, immersively and remotely, in various types of applications, ranging from group meetings to live concerts and classes, facilitated by unique social mechanics. It allows participants to organize into groups, interact with objects seen in the scene, engage in non-verbal communication, and experience realistic interactions, not with the aim of replicating reality entirely, but enhancing and expanding existing communication channels from the physical world. Hence, such an immersive environment can be utilized to encourage an appropriate blend of virtual end-users and co-present participants for public, private (individual or one-on-one), and semi-private (small group) interactions. The interest in utilizing other modal channels allows the exploration of possibilities that have not been thoroughly investigated in this field, shedding light on new social musical experiences in virtual reality.

Design for Creativity

In addition to the core functionality of playing or creating art, the proposed system should extend its purpose to serve as a comprehensive tool for teaching, training, experimentation, and leisure. The integration of assistance and guidance can help the learning process, offering adaptive and personalized support to users. By allowing the exploration of scenarios impossible in real life, the platform transforms into an experimental environment, empowering users to discover new instruments, techniques, and creative possibilities that transcend the limits of physical reality.

This training capability provides an opportunity for users to enhance their musical skills and experience a new artistic creation environment, enabling them to test equipment they are not familiar with and gain a better understanding of the acoustics or general characteristics of a space where they will perform soon. The alignment of sound and visual content, as well as the automatic generation of music and virtual band members for accompaniment, expands creative and technical possibilities. Another important dimension is the ability for musical editing. By bringing features like mixing and mastering to immersive environments, not only do new forms of music customization emerge, but it also offers a programmatic and innovative approach to this well-established practice in the musical domain.

Conversely, an exclusively functional approach would undermine artistic expression, which inherently involves elements such as creativity. Therefore, such applications ought to possess a ludic essence, thereby serving as conduits that amplify entertainment while facilitating practical and tangible activities, fostering creativity, and promoting the development of artistic language, critical thinking, and concentration.

B. Use Cases

From this discussion, several case studies were also proposed. These scenarios illustrate how different technologies interact and how each requirement can be adapted to technical circumstances in a meaningful way, becoming an important source of experience and serving as a foundation for the development of future applications. Each case study is self-contained and they are presented as follows.

Scenario 1: Virtual Tests and Building of Musical Instruments/Equipment

One of the potential applications of an immersive Io3MT environment lies in its ability to allow users to interact with virtual musical instruments through intuitive interfaces and haptic devices that replicate tactile sensations, such as plucking strings, adjusting knobs, and feeling the textures of the instruments. Optimizing the flow of information is crucial for synchronization and minimizing latency, ensuring authentic sound reproduction and effective tactile feedback, while also meeting the functional requirements of the system.

This extended reality system uses HMD units with auditory and tactile feedback to create a realistic musical environment, aiding technique improvement with real-time guidance. Users can adjust the virtual setting to simulate various performance spaces and test instruments. It supports multi-user interactions for group training and emulates band dynamics, enhancing immersion, feedback, and social connection while fostering creativity. Virtualizing instruments reduces physical resources, leading to cost savings and greater efficiency.

In the same way, developing physical prototypes often faces challenges such as slow assembly and high costs. By integrating Io3MT with XR technologies, new musical interfaces can be visualized and prototyped early, using photorealistic rendering to create detailed virtual models. This approach speeds up development, reduce costs, and supports iterative design by allowing developers to update components and simulate multisensory experiences in an XR environment, enhancing the usability and expressiveness of instruments.

Scenario 2: Providing Cultural Heritage

Integrating Io3MT with XR technologies can help preserve and access the cultural heritage of various communities, including rare musical instruments, styles and performance techniques. This combination enables the creation of immersive environments that allow users to authentically experience historical music through interactive models. Such environments offer tactile feedback, simulating the textures and resistance of ancient instruments, and can also recreate the historical contexts in which these musical styles were developed, providing a rich visual and cultural backdrop.

These immersive environments enable shared experiences, such as virtual workshops and concerts, where participants explore cultural music together. They also offer collaborative spaces for musicians, historians, and students to discuss techniques, and serve as digital archives, expanding access to and preserving cultural heritage globally.

Scenario 3: Enhancing Creative Industries

The proposed integrated framework can extend to various creative industry applications. For example, it could synchronize television content with lighting and aromatic effects or use HMD devices as secondary screens for additional content like cast details or plot summaries. Mixed reality video games also offer potential, allowing interaction with physical objects that control their digital counterparts.

This integration could enable the creation of remote-access virtual museums and galleries, offering personalized tours and interactive experiences with artworks, including audio or video commentaries. Users could also engage in discussions on artistic and cultural topics. Further possibilities include virtual reality experiences based on literary works, with synchronized audiovisual effects to enhance the narrative, and art therapy sessions set in immersive virtual landscapes with sensory elements.

These functionalities aim to enhance immersion across various scenarios, such as virtual museum visits and therapy applications, while maintaining social connectivity through audiovisual or tactile feedback. This interaction fosters creativity and the exploration of new artistic concepts.

Scenario 4: Enhancing Live Performances

The intersection of Io3MT and XR technologies significantly enhances live performances, enabling musicians and spectators to interact in novel ways. Artists can perform within dynamic virtual settings that react to music in real time, extending sensory experiences beyond visual and auditory to include other senses involved in controlling musical and multimedia information.

Functionality benefits include the use of intuitive control interfaces that replicate real instrument sensations or create new musical interactions. Virtual environments overcome geographical barriers, allowing the inclusion of remote participants and fostering online communities centered around specific musical styles or artists. These settings can also dynamically adapt to the rhythm or mood of the music.

For the audience, the experience is diversified with options to explore the virtual space and select unique viewing angles, impossible in traditional settings. During presentation, real-time tactile and auditory feedback helps artists fine-tune their performances, while multimodal audience feedback intensifies the overall experience, making spectators feel everything from vibrations to virtual atmospheric effects. This immersive approach transforms passive viewership into an interactive experience. Participants can delve into the history of the music, engage in virtual workshops, or attend masterclasses with the artists, enriching their understanding of the musical piece and the artist's creative process.

V. PhysioDrum

To illustrate the introduced concepts, a proof of concept for live performance was developed following the five proposed guidelines.

A. Implementation

This application aims to construct a virtual drum set composed of its most traditional elements, such as the bass drum, snare, two toms tuned to low and high pitches, a crash cymbal, and a hi-hat, allowing for natural and fluid interaction with the instrument. Key functional requirements include minimal latency between hitting the drum and sound generation, minimal striking errors, and appropriate sound volume and quality.

To achieve this, the application considered expressiveness, ensuring that haptic actuators provide multiple control levels (e.g., volume, intensity, frequency) with reliable timing. The system also minimizes user inconvenience by simplifying actuator placement and reducing excessive cables. Additionally, it is easily re-configurable with straightforward input-output mapping.

Physical drumsticks were chosen over HMD controllers as the input device to enhance the physicality of the experience. Drumstick movements were captured by a touch sensor and accelerometer, adapted from the RemixDrum [4]. These signals were encapsulated in OSC messages and sent via UDP packets to a computer on the same local network. Data from the right drumstick were directed to a Pure Data patch, triggering a pre-recorded track based on boolean variables from the touch sensor. Meanwhile, parameters such as volume, reverb, and flange were controlled by movements along the accelerometer's X, Y, and Z axes, creating a remix between the digital drum sound and that generated by Pure Data.

The integration with the virtual world was implemented using Unity (version 2022.3) and Meta's Interaction SDK V1.3.2, with hand tracking enabled. Quest 2's infrared cameras captured the user's hand movements, allowing seamless interaction between the physical and virtual environments.

The resulting musical track was inspired by John Cage's work "I Ching" and Aleatoric Music. Users can control only certain aspects of the composition, in this case, the drum rhythms, while the technology and randomness of the data sent by the drumsticks interact together in certain parts of the audible content. This represents a clear counterpoint to Western musical tradition, which emphasizes control over almost every aspect of composition.

As for the data from the left drumstick movements, they were sent to a digital artwork created using the Processing programming language, controlling its RGB color patterns using the accelerometer data from the X, Y, and Z axes. The artwork was titled "Origami", refers to traditional Japanese paper-folding art. This name reflects the changing colors and figure movements in different frames, visually resembling the folding process.

The application architecture and interaction between elements are illustrated in Figure 2. The digital drum, replacing the physical one as in the first version of RemixDrum [4], integrates its sound with random sounds generated in Pure Data, triggered and modified by actuators on the physical drumstick. These movements also influence the color and motion patterns of an animation in Processing, highlighting the unique integration of Io3MT and XR and the distinctiveness of the prototype.

Figure 3 illustrates the main components of PhysioDrum, featuring the RemixDrum virtual drum kit. This setup includes a drumstick equipped with actuators that map physical

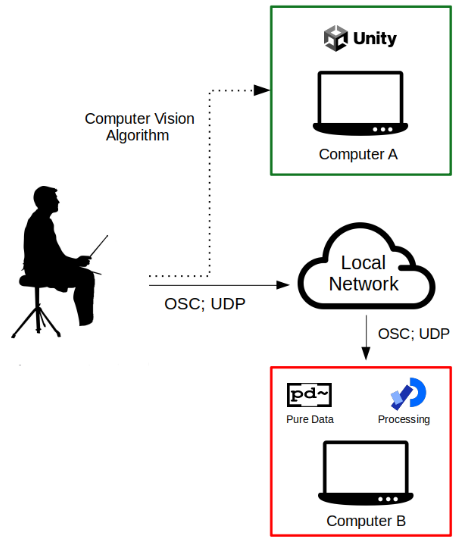


Fig. 2: Functional Architecture for PhysioDrum.

movements to the virtual environment and control audiovisual elements. Additionally, the artwork generated in Processing is displayed on the screen, along with sliders indicating the status of each sound parameter in Pure Data, offering the user visual feedback on these specific elements.

B. Design Analysis

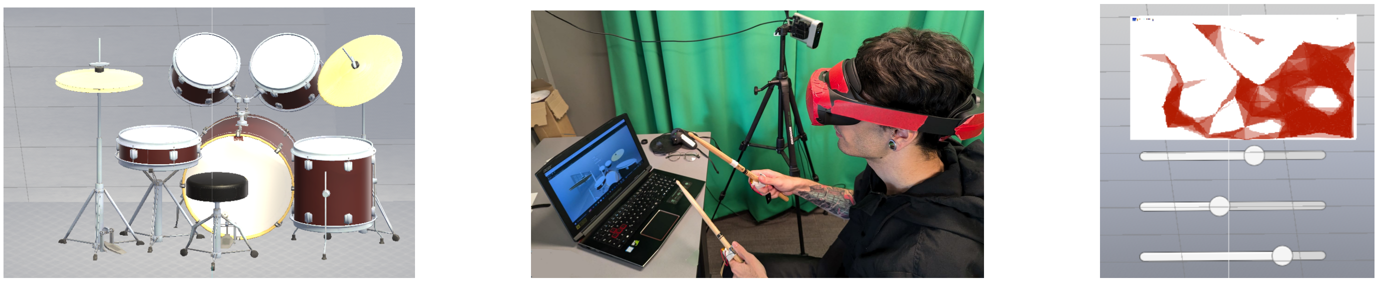
Table II summarizes the design guidelines and technologies used in creating PhysioDrum.

In terms of functionality, effective synchronization and real-time responsiveness were achieved. The virtual drum emitted sounds instantly upon detecting drumstick contact, and movements in RemixDrum 2.0 adjusted musical and artistic parameters in real time. Users could also access the drum's music signal through a screen-based interface on another computer, enabling spontaneous music editing and visual art creation. The local network infrastructure ensured low latency for real-time music editing, while off-the-shelf headsets helped minimize discomfort.

For the second guideline on immersiveness, sensing capabilities were a key focus. RemixDrum 2.0 enabled personalized cyber-physical actions, ensuring audiovisual integration by linking drumstick movements with both sound and visuals. Each interaction generated a unique multimedia experience.

Feedback was provided through auditory and visual channels. Auditory feedback reflected interactions with the virtual drum, confirming that physical movements triggered corresponding digital responses. Sliders for volume, reverb, and flange in Pure Data, along with Processing's replicated artwork in the virtual environment, allowed musicians to visualize the system's current state.

Although the fourth guideline was less explored, certain elements foster social connectivity. Network interaction and integration of Pure Data and Processing enabled real-time remote control of music and visual content. It allows users to adjust volume and effects. Physical and virtual tools could



(a) Virtual drumkit scene. (b) Real usage scenario. (c) Visual feedback.

Fig. 3: Structural composition and practical application of PhysioDrum.

be integrated to enhance cyber-physical interaction, with each participant contributing to the collective artistic experience. This framework encourages interaction among musicians and between musicians and the audience.

Finally, for creativity, we see a wide range of use cases for this application beyond virtual musical performance. Extended scenarios for this multi-user and multi-platform system include music workshops, virtual rehearsals, instrument teaching, and technical demonstrations of using immersive media for artistic creation.

TABLE II: Requirements of each guideline employed in PhysioDrum.

Design Guideline	Feature	Tools
Design for Functionallity	Specific purpose; low latency; realistic interactions; ergonomic design; real-time music editing	RemixDrum 2.0, Unity, Meta Quest 2, Pure Data
Design for Immersiveness	VR; embodiment; synthesized sounds; multisensory feedback; graphic arts	Meta Quest 2 Computer Vision algorithms, Pure Data, Unity, Processing
Design for Feedback	Visual and auditory feedback	Unity, OSC, UDP, Pure Data
Design for Social Connectivity	Network interaction; sound parameter and multimedia control	Pure Data, Processing
Design for Creativity	Extended use cases include music workshops, instrument lesson, virtual rehearsal and Tech Demo	RemixDrum 2.0, Unity, Pure Data, Processing

In summary, PhysioDrum aligns with the Musical Metaverse [17] through consistent avatars (hands) and stable object behavior, such as the drum set and environment. The three-dimensional space meets key Metaverse criteria, with integrated physical and virtual spaces allowing actions in one to influence the other. This enables seamless interaction via Pure Data or Processing, promoting interoperability and synchronized musical data. While the social aspect was less explored, the framework offers a solid foundation for future social integration.

VI. Discussion & Lessons Learned

Upon concluding the analysis on the real applicability of the proposed guidelines, it is possible to highlight several points based on the operation mode of the application. By integrating physicality into the system and fostering interaction akin to that of a traditional musical instrument, PhysioDrum aids in the transfer of familiar motor skills, making its learning process easier and more intuitive. This approach also supports spatial memory and facilitates interactions [23], [24].

Simultaneously, Cadoz [25] emphasized that effective musical interfaces in the digital world must meet three criteria: authenticity of the performer's gestures, accurate gesture detection to prevent information loss, and appropriate resistance to mimic the nature of physical movements. These parameters have been successfully incorporated into the proposed system.

Another advantage of this hybrid environment is the enhancement of immersive experiences through the integration of functionality in both physical and digital devices. In this way, the artifacts have an added value, being able to switch between their various modes of use at any time with a simple gesture. Each mode offers a distinct visual surface, representing new opportunities with the instrument and the creative process as a whole [26], [27]. The modular design of the application also ensures greater system customization, supporting various drum configurations and additional multimedia content. In this context, issues of accessibility, noise control, and ubiquity are also addressed by this system.

Cost reduction is another positive aspect. Learning a musical instrument requires substantial time and financial investment. Furthermore, beginners must purchase equipment in advance to determine if it is the right instrument for them. Specifically, for drums, there is also the need to implement a spacious, soundproof environment [28]. The PhysioDrum, on the other hand, offers a system that allows users to play anywhere, anytime, at a low cost. Moreover, the open-source and affordable nature of RemixDrum 2.0 further enhances its economic viability.

In the context of Io3MT, it is important to attain a level of sensitivity and control comparable to that of acoustic instruments [1]. By examining the connections between sound and touch, new strategies for composition and performance begin to emerge for performers using digital instruments. These involve technological implementations that utilize haptic information channels, offering insights into how tacit knowledge of the physical world can be introduced into the digital domain, reinforcing the view that sound is a "kind of touch", or further clarifying the connection between a performer and an instrument as a "multimodal participatory space" (rather than a "control space)", aligning with the Io3MT concepts. Consequently, new strategies for composition and performance may emerge [25].

However, some noticeable limitations are identified in the proposed system, such as the absence of foot interaction. Consequently, the bass drum is triggered by the drumsticks, diverging from the traditional usage paradigm of this instrument. As well, the hi-hat is restricted to the "open" position, which limits its musical capabilities. Furthermore, the design principle aimed at exploring social interactions in the virtual environment also received little attention during development.

VII. Conclusion

There is an increasing interest in music XR experience, forming a substantial research area yet lacking Io3MT standards integration. This paper, therefore, introduces the first-ever design framework and guidelines for the combined practice of musical XR and Io3MT through a focus group study. We also present a proof-of-concept prototype, PhysioDrum, a multi-sensory VR drumming experience, to demonstrate the proposed design guidelines. Future improvements of this implementation include adding multi-user interaction in the VR drumming scene and haptic feedback on the physical drumstick input device. Furthermore, we plan to incorporate quantitative metrics into the application analysis and conduct another focus group with a larger number of participants, aiming to gather new insights and obtain more generalizable results. We consider musical XR with Io3MT an emerging area for fostering musical expressions through new forms of immersive media. Our guidelines can serve as a foundational framework to aid future design and development in this innovative field.

Acknowledgment

The authors would like to thank the support of the following funding agencies CAPES, CAPES Print, CNPq, FINEP and FAPERJ. We would also like to thank Ashutosh Singla for his support in running the application.

References

[1] R. Vieira, D. Muchaluat-Saade, and P. César, "Towards an internet of multisensory, multimedia and musical things (io3mt) environment," in *Proceedings of the 4th International Symposium on the Internet of Sounds*, ser. IS2 '23. New York, NY, USA: IEEE, 2023, p. 231–238. [Online]. Available: https://doi.org/10.1145/3561212.3561246

[2] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The internet of sounds: Convergent trends, insights and future directions," *IEEE Internet of Things Journal*, 2023.

[3] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018.

[4] 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)*, 2023, pp. 1–7.

[5] L. Turchet, "Smart musical instruments: Vision, design principles, and future directions," *IEEE Access*, vol. 7, pp. 8944–8963, 2019.

[6] S. Ppali, V. Lalioti, B. Branch, C. S. Ang, A. J. Thomas, B. S. Wohl, and A. Covaci, "Keep the vrhythm going: A musician-centred study investigating how virtual reality can support creative musical practice," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–19.

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

[8] A. Boem and L. Turchet, "Musical metaverse playgrounds: exploring the design of shared virtual sonic experiences on web browsers," in *2023 4th International Symposium on the Internet of Sounds*, 2023, pp. 1–9.

[9] P. Cook, "2001: Principles for designing computer music controllers," *A NIME Reader: Fifteen years of new interfaces for musical expression*, pp. 1–13, 2017.

[10] P. R. Cook, "Re-designing principles for computer music controllers: a case study of squeezevox maggie," in *NIME*, vol. 9, 2009, pp. 218–221.

[11] G. Wang, "Principles of visual design for computer music," in *ICMC*, 2014.

[12] C. Alexandraki and D. Akoumanakis, "Exploring new perspectives in network music performance: The diamouses framework," *Computer Music Journal*, vol. 34, no. 2, pp. 66–83, 2010.

[13] G. Paine, "Towards unified design guidelines for new interfaces for musical expression," *Organised Sound*, vol. 14, no. 2, pp. 142–155, 2009.

[14] G. Weinberg, "Interconnected musical networks: Toward a theoretical framework," *Computer Music Journal*, vol. 29, no. 2, pp. 23–39, 2005.

[15] S. Serafin, C. Erkut, J. Kojs, N. C. Nilsson, and R. Nordahl, "Virtual reality musical instruments: State of the art, design principles, and future directions," *Computer Music Journal*, vol. 40, no. 3, pp. 22–40, 2016.

[16] L. Valbom and A. Marcos, "Wave: Sound and music in an immersive environment," *Computers & Graphics*, vol. 29, no. 6, pp. 871–881, 2005.

[17] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[18] M. Ciciliani, "Virtual 3d environments as composition and performance spaces," *Journal of New Music Research*, vol. 49, no. 1, pp. 104–113, 2020.

[19] D. Dziwis, H. von Coler, and C. Porschmann, "Live coding in the metaverse," in *2023 4th International Symposium on the Internet of Sounds*, 2023, pp. 1–8.

[20] V. Braun and V. Clarke, *Thematic analysis.* American Psychological Association, 2012.

[21] S. Lee, A. Striner, and P. Cesar, "Designing a vr lobby for remote opera social experiences," in *Proceedings of the 2022 ACM International Conference on Interactive Media Experiences*, 2022, pp. 293–298.

[22] J. Li and P. Cesar, "Social virtual reality (vr) applications and user experiences," in *Immersive Video Technologies*. Elsevier, 2023, pp. 609–648.

[23] S. Willemsen, A.-S. Horvath, and M. Nascimben, "Digidrum: a haptic-based virtual reality musical instrument and a case study," in *17th Sound and Music Computing Conference*. Axea sas/SMC Network, 2020, pp. 292–299.

[24] M. Feick, N. Kleer, A. Zenner, A. Tang, and A. Krüger, "Visuo-haptic illusions for linear translation and stretching using physical proxies in

virtual reality," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, 2021, pp. 1–13.

[25] L. Hayes, "Vibrotactile feedback-assisted performance." in *NIME*. Citeseer, 2011, pp. 72–75.

[26] K. Okada, F. Ishizawa, A. Kobayashi, A. Yoshii, M. Sakamoto, and T. Nakajima, "Virtual drum: Ubiquitous and playful drum playing," in *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)*. IEEE, 2014, pp. 419–421.

[27] D. Zaveri, R. Parekh, R. Lobo, S. Ardekar, V. Vora, and A. Kore, "Aero drums-augmented virtual drums," in *2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*. IEEE, 2022, pp. 516–520.

[28] R. C. L. Suen, K. T. Chang, M. P.-H. Wan, W. Y. Chua, and Y. C. Ng, "Mobile and sensor integration for increased interactivity and expandability in mobile gaming and virtual instruments," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 2015, pp. 703–708.