FLUMINENSE FEDERAL UNIVERSITY

RÔMULO AUGUSTO VIEIRA COSTA

Internet of Multisensory, Multimedia and Musical Things (Io3MT) Environments: Requirements, Use Cases, and Evaluation

NITERÓI

2025

RÔMULO AUGUSTO VIEIRA COSTA

Internet of Multisensory, Multimedia and Musical Things (Io3MT) Environments: Requirements, Use Cases, and Evaluation

Thesis presented to the Computing Graduate Program of Fluminense Federal University in partial fulfillment of the requirements for the degree of Doctor of Science.
Topic Area: Computer Science.

Advisor:
Prof. Dr. Débora Christina Muchaluat Saade

Co-Advisor:
Prof. Dr. Pablo Santiago César Garcia

NITERÓI

2025

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

C837i Costa, Rômulo Augusto Vieira
Internet of Multisensory, Multimedia and Musical Things (Io3MT) Environments: Requirements, Use Cases, and Evaluation / Rômulo Augusto Vieira Costa. - 2025.
302 f.: il.

Orientador: Débora Christina Muchaluat Saade.
Coorientador: Pablo Santiago César Garcia.
Tese (doutorado)-Universidade Federal Fluminense, Instituto de Computação, Niterói, 2025.

1. Internet das Coisas. 2. Realidade Estendida. 3. Sistemas Multimídia. 4. Computação Musical. 5. Produção intelectual. I. Saade, Débora Christina Muchaluat, orientadora. II. Garcia, Pablo Santiago César, coorientador. III. Universidade Federal Fluminense. Instituto de Computação. IV. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

RÔMULO AUGUSTO VIEIRA COSTA

Internet of Multisensory, Multimedia and Musical Things (Io3MT) Environments: Requirements, Use Cases, and Evaluation

Thesis presented to the Computing Graduate Program of Fluminense Federal University in partial fulfillment of the requirements for the degree of Doctor of Science.
Topic Area: Computer Science.

Approved on November 25, 2025.

EXAMINATION COMMITTEE

Prof. Dr. Débora Christina Muchaluat Saade - Advisor, UFF

Prof. Dr. Pablo Santiago César Garcia - Co-Advisor, CWI & TU Delft

Prof. Dr. Célio Vinicius Neves de Albuquerque, UFF

Prof. Dr. Igor Monteiro Moraes, UFF

Prof. Dr. Luca Turchet, University of Trento

Prof. Dr. Marcelo Ferreira Moreno, UFJF

Niterói
2025

Para minha irmã, Rosi, por ser maior que as muralhas.

Acknowledgements

Este trabalho não seria possível sem o auxílio da minha orientadora, Débora. Desde as primeiras conversas, Débora mostrou-se respeitosa, animada e profundamente comprometida com minha pesquisa. Já no papel de orientadora, agiu como os grandes fazem: atenta às minhas opiniões, mas convicta do melhor caminho a ser seguido. Por todo o apoio, oportunidades concedidas e pelo conhecimento generosamente compartilhado, registro aqui minha gratidão.

Agradeço a todos os integrantes do Laboratório MídiaCom. Estendo esse agradecimento a todos os servidores e professores da Universidade Federal Fluminense (UFF), em especial à secretaria do PGC, pela agilidade e presteza incondicionais.

Agradeço aos funcionários e professores do Colégio Potência, onde esta travessia começou. Em especial Maria Elizabeth Vasconcelos, Maria Gorete, Edelvando Tonholo, Joana Dornelas e principalmente Érika Cerqueira, primeira referência e incentivadora a seguir na vida acadêmica.

Aos professores Jim Marciano, Eugênio Pacelli, Jean Carlo Mendes e Alex Vitorino, por acreditarem no meu potencial e incentivarem a continuidade desta caminhada intelectual.

Deixo minha gratidão aos professores e servidores da Universidade Federal de São João del-Rei (UFSJ) e a todos os companheiros do laboratório ALICE (Arts Lab in Interfaces, Computers, and Everything Else), sobretudo ao meu antigo orientador e sempre amigo, Flávio Schiavoni, e aos amigos João Teixeira e Isadora Franco.

Em 2023, não satisfeito em me tirar *Das Gerais*, este curso de doutorado decidiu me levar para ainda mais longe: Amsterdã. Essa jornada só se tornou possível graças a Pablo César, cuja receptividade e generosidade abriram-me as portas do Centrum Wiskunde & Informatica. Mais tarde, na condição de co-orientador, agradeço-lhe pelos ensinamentos que me tornaram um pesquisador e cientista melhor.

Aos amigos do Distributed and Interactive Systems Group, que fizeram do meu tempo no CWI uma memória inesquecível: Abdallah El Ali, Xuemei Zhou, Jiahuan Pei, Haochen (Peter) Huang, Renske Bootsma, Varun Pradhan, Karolina Wylężek, Zohrab Sarabian, Karthikeya Venkatraj, Silvia Rossi, Irene Viola, Simon Gunkel, Ashutosh Singla, Ofelya Alijeva e a Karima El Bacha, por toda a hospitalidade e suporte com a burocracia de se mudar para um novo país.

Agradeço especialmente a Moonisa Ahsan, pelas dicas de carreira, produtividade e apoio na escrita desta tese. Agradeço carinhosamente à Shu Wei, minha principal parceira intelectual e artística durante este doutorado.

À minha psicóloga, Franciely Damasceno, que com seu profissionalismo e carisma, ajudou-me a conquistar a vitória mais importante desses quatro anos: a conquista de mim mesmo.

Aos que se foram e não estão mais aqui: Paulo dos Anjos (Polica) e Dona Lilinha.

Aos amigos que guardo debaixo de sete chaves dentro do coração: Iago, Matheus, Milady, William e Marcelo.

Este caminho foi percorrido com a melhor trilha sonora possível. Por isso, agradeço a Lucas Silveira, Tyler Joseph, Josh Dun, Oliver Sykes e Joe Mulherin. A Milton Nascimento, Lô Borges e a todo o Clube da Esquina por serem a minha trilha sonora favorita quando a saudade de casa apertava.

Agradeço a CAPES (incluindo o projeto CAPES Print), RNP e FAPERJ, por financiarem esta pesquisa. Reforço aqui a importância do investimento público na ciência e meu comprometimento em defender uma educação gratuita, de qualidade e universal.

Reservo os melhores agradecimentos desta tese a minha família. À minha querida irmã e melhor amiga, Rosiane, por ser minha maior inspiração e também por ser a minha maior incentivadora e sempre vibrar com as minhas conquistas. “Minha irmã, eu sou como você é. Saí do mesmo escuro e ando por aí, toda noite eu sei que amanhã tem mais, que a gente muda e continua a sonhar, aprendendo”. Obrigado por ser maior que as muralhas e por me transmitir essa estranha mania de ter fé na vida. Ao meu cunhado, Guilherme, exemplo de dedicação, lealdade e trabalho duro. Ao meu pai, Ronaldo, por me mostrar que estudar seria buscar o caminho que vai dar no sol, por manter vivo em mim meu coração de estudante e por me ensinar que sonhos não envelhecem. O que levo de ti guardo como tesouro e procuro todo dia honrar. À minha mãe, Rosângela. Mamãe, eu sou fruto das suas orações. Obrigado por não ter desistido de mim, por ser o som, a cor e o suor, e nosso exemplo de fé, trabalho duro e honestidade. Essa eu fiz por nós! Por todos os que vieram antes e todos os que virão depois.

Agradeço ao Rômulo do passado, por ainda morar no meu coração e em toda vez que o adulto balança ele vir para me dar a mão. Você é o que você queria ser quando crescesse.

A todos que contribuíram para a expansão do meu conhecimento, minha eterna gratidão.

Vencemos! Mas ainda há muito mais por vir. Caminho não tem fim.

"Vão tentar derrubar que é pra me ver crescer
E às vezes me matar que é pra eu renascer
Como uma supernova que atravessa o ar
Eu sou a maré viva, se entrar, vai se afogar".
(Fresno)

Resumo

A Internet das Coisas Multissensoriais, Multimídia e Musicais (Io3MT) pode ser compreendida como uma rede de transmissão que integra, em um mesmo ecossistema, dispositivos e dados capazes de explorar os cinco sentidos humanos (tato, audição, visão, olfato e paladar), conteúdos multimídia e informações musicais de forma intercambiável e não hierarquizada, disponibilizando aplicações e serviços de alcance global. Esta tese apresenta o primeiro modelo de referência dedicado a esse domínio, delineando sua arquitetura padrão, os tipos de dados envolvidos, os requisitos de comunicação em rede e as ferramentas apropriadas à sua implementação. O propósito é fornecer diretrizes que permitam a cientistas, artistas, *designers* e profissionais da indústria conceber e desenvolver, de modo prático, ambientes fundamentados nos princípios da Io3MT. Para validar a proposta, foram realizadas duas provas de conceito, cada uma explorando dimensões distintas do paradigma, nas quais se conduziram experimentos voltados à análise de desempenho da rede e à aferição da Qualidade da Experiência (QoE). Esses experimentos envolveram um protótipo de dispositivo multissensorial e um cenário imersivo para performance artística em tempo real. Os resultados obtidos confirmam a viabilidade técnica do modelo de referência, ao mesmo tempo em que demonstram seu potencial estético e expressivo, evidenciando a capacidade da Io3MT de sustentar experiências interativas, criativas e multissensoriais.

Palavras-chave: Internet das Coisas, Io3MT, Multissensorial, Multimídia, Arte Interativa, Performances Musicais pela Rede, Realidade Estendida, Imersividade.

Abstract

The Internet of Multisensory, Multimedia, and Musical Things (Io3MT) can be understood as a transmission network that integrates, within a unified ecosystem, devices and data capable of engaging the five human senses (touch, hearing, vision, smell, and taste), multimedia content, and music information in an interchangeable and non-hierarchical manner, thereby providing globally accessible applications and services. This thesis introduces the first reference model of this domain, outlining its standard architecture, data types, communication requirements, and the tools suitable for its implementation. The objective is to establish guidelines that enable scientists, artists, designers, and industry practitioners to conceive and develop environments grounded in the principles of Io3MT. To validate the proposed framework, two proof-of-concept implementations were conducted, each exploring distinct dimensions of the paradigm. In these prototypes, experiments were carried out to evaluate network performance and to assess the resulting Quality of Experience (QoE). The experiments comprised a device prototype and an immersive environment designed for real-time artistic performance. The results confirmed the technical feasibility of the reference model while also demonstrating its aesthetic and expressive potential, thereby highlighting Io3MT's capacity to sustain interactive, creative, and multisensory experiences.

Keywords: Internet of Things, Io3MT, Multisensory, Multimedia, Interactive Art, Networked Musical Performance, Extended Reality, Immersiveness.

List of Figures

List of Tables

List of Abbreviations and Acronyms

3MT	:	Multisensory, Multimedia and Musical Things;
5G	:	Fifth Generation;
6DoF	:	Six Degrees of Freedom;
A2D	:	Anywhere Anytime Drumming;
ACM	:	Association for Computing Machinery;
ADHD	:	Attention-Deficit / Hyperactivity Disorder
AI	:	Artificial Intelligence;
AMQP	:	Advanced Message Queuing Protocol;
API	:	Application Programming Interface;
AR	:	Augmented Reality;
AWS	:	Amazon Web Services;
AVI	:	Audio Video Interleave;
BCI	:	Brain-Computer Interface;
Bit	:	Binary Digit;
BLE	:	Bluetooth Low Energy;
BPM	:	Beats Per Minute;
CC	:	Common Core;
CMOS	:	Complementary Metal-Oxide Semiconductor;
CoAP	:	Constrained Application Protocol;
CWI	:	Centrum Wiskunde & Informatica;
DAW	:	Digital Audio Workstation;
DHCP	:	Dynamic Host Configuration Protocol;

DIY : Do-It-Yourself;

DJ : Disc Jockey;

DMI : Digital Music Instrument;

DNS : Domain Name System;

DYMO : Dynamic Music Object;

EEG : Electroencephalogram;

FLAC : Free Lossless Audio Codec;

FSR : Force-Sensitive Resistor;

FTA : Fake Time Approach;

GEQ : Game Experience Questionnaire;

GPS : Global Positioning System;

GUI : Graphical User Interface;

HAP : HomeKit Accessory Protocol;

HCI : Human-Computer Interaction;

HE : Heuristic Evaluation;

HLL : Hierarchical Live Looping;

HMD : Head-Mounted Display;

HTML : HyperText Markup Language;

HTTP : HyperText Transfer Protocol;

HQ : Haptic Questionnaire;

HX : Haptic Experience;

I2C : Inter-Integrated Circuit;

IDE : Integrated Development Environment;

IEC : International Electrotechnical Commission;

IEEE : Institute of Electrical and Electronics Engineers;

IETF : Internet Engineering Task Force;

IP	:	Internet Protocol;
Io3MT	:	Internet of Multisensory, Multimedia and Musical Things;
IoAuT	:	Internet of Audio Things;
IoMT	:	Internet of Multimedia Things;
IoMusT	:	Internet of Musical Things;
IoS	:	Internet of Sounds;
IoT	:	Internet of Things;
IPTV	:	Internet Protocol Television;
ISO	:	International Organization For Standardization;
ITU	:	International Telecommunication Union;
IVE	:	Immersive Virtual Environment;
JSON	:	JavaScript Object Notation;
JPEG	:	Joint Photographic Experts Group;
JSP	:	Jakarta Server Pages;
LAA	:	Latency Acceptance Approach;
LED	:	Light-Emitting Diode;
LEO	:	Low Earth Orbit;
M2M	:	Machine-To-Machine;
Mbps	:	Megabits per second;
MIDI	:	Musical Instrument Digital Interface;
MIDI CC	:	Musical Instrument Digital Interface Continuous Controller;
MM	:	Musical Metaverse;
MQTT	:	Message Queuing Telemetry Transport;
MR	:	Mixed Reality;
MuDI	:	Multimedia Digital Instrument;
Mulsemedia	:	Multiple Sensorial Media;

mDNS : Multicast Domain Name System;

N : Nausea;

NAT : Network Address Translation;

NASA-TLX : National Aeronautics and Space Administration Task Load Index;

NCL : Nested Context Language;

NG-RAN : Next-Generation Radio Access Network;

NIME : New Interfaces for Musical Expression;

NMP : Networked Music Performance;

NPS : Net Promoter Score;

NUI : Natural User Interface;

O : Oculomotor;

OSC : Open Sound Control;

OSGi : Open Services Gateway Initiative;

OT : Operational Technologies;

OWL : Web Ontology Language;

P2P : Peer-to-Peer;

PCM : Pulse Code Modulation;

PDF : Portable Document Format;

PNG : Portable Network Graphics;

PQ : Presence Questionnaire;

PPP : Point-to-Point Protocol;

PWM : Pulse-Width Modulation;

QoE : Quality of Experience;

QoS : Quality of Service;

RDFS : Resource Description Framework Schema;

REST : Representational State Transfer;

RFID : Radio-Frequency Identification;

RQ : Research Question;

RTT : Round-Trip Time;

RWO : Real World Objects;

SAM : Self-Assessment Manikin;

SDK : Software Development Kit;

SD : Standard Deviation;

SEDL : Sensory Effect Description Language;

SEV : Sensory Effect Vocabulary;

SMI : Smart Musical Instrument;

SoS : System of Systems;

SQL : Structured Query Language;

SSQ : Simulator Sickness Questionnaire;

STEAM : Science, Technology, Engineering, Arts and Mathematics;

SVG : Scalable Vector Graphics;

SUS : System Usability Scale;

TCP : Transmission Control Protocol;

TS : Total Score;

TXT : Text File;

UDP : User Datagram Protocol;

UML : Unified Modeling Language;

URI : Uniform Resource Identifier;

URL : Uniform Resource Locator;

USB : Universal Serial Bus;

UX : User Experience;

VO : Virtual Objects;

VR : Virtual Reality;

VRMI : Virtual Reality Musical Instrument;

WAN : Wide Area Network;

Wi-Fi : Wireless Fidelity;

WLAN : Wireless Local Area Network;

WMSNs : Wireless Multimedia Sensor Networks;

WebXR : Web Extended Reality;

XML : Extensible Markup Language;

XMPP : Extensible Messaging And Presence Protocol;

XR : Extended Reality

Contents

1 Introduction

The term Internet of Things (IoT) (ASHTON, 2009; MARSAN, 2015; GENG, 2017; LIN et al., 2017; AL-FUQAHA et al., 2015) was first introduced by Kevin Ashton in 1999 to characterize the use of Radio-Frequency Identification (RFID) technologies within supply chain management. Since its inception, the concept has been progressively expanded and reinterpreted across multiple domains. For instance, (GERSHENFELD; KRIKORIAN; COHEN, 2004) conceptualize IoT as a network of everyday objects capable of interconnection and data exchange through diverse communication protocols. In a similar vein, (ATZORI; IERA; MORABITO, 2010) describe IoT as the pervasive presence of heterogeneous devices that, by means of unique identifiers, can interact and collaborate with surrounding entities to pursue shared objectives. Despite variations in emphasis, these definitions converge on a common understanding: IoT denotes the ability of devices, sensors, actuators, and everyday objects to connect to the Internet, thereby generating, transmitting, and consuming data with minimal human intervention.

From a technological standpoint, IoT results from the integration of techniques that extend the addressability, identification, sensing, and actuation capabilities of objects. This integration enables the processing of embedded information while fostering more efficient communication and collaboration across networked systems. Furthermore, IoT places particular emphasis on scalability and interoperability, ensuring that devices can operate seamlessly in heterogeneous environments, maintain connectivity during geographic mobility, and exchange data in real time (MATTERN; FLOERKEMEIER, 2010).

Based on the technologies used in each environment, three generations of IoT can be identified. As previously mentioned, the first generation emerged with RFID tags, which are commonly utilized for monitoring logistics and tracking applications. The second generation was characterized by the introduction of sensors and actuators, which enabled data collection from the physical world and its digital representation. In contrast, the third generation is built on virtualization, establishing a connection between Real World Objects (RWO) and Virtual Objects (VO) (FLORIS; ATZORI, 2015; ATZORI; IERA; MORABITO, 2010; GUBBI et al., 2013).

The consolidation of IoT has enabled its application in multiple domains, including home

automation, smart city management, energy grid supervision, environmental monitoring, vehicular systems, and public safety (HALLER, 2010; VIEIRA; BARTHET; SCHIAVONI, 2020).

The potential for extensive interconnectivity, coupled with ease of implementation and relatively low operational costs, has positioned IoT as a central enabler in the emergence of new paradigms of communication. Within this context, the Internet of Multimedia Things (IoMT) (ALVI et al., 2015) has emerged as a prominent subfield, distinguished by the interaction and integration of heterogeneous multimedia objects with other Internet-connected devices. IoMT is a network of uniquely identifiable and addressable devices capable of cooperating to exchange multimedia content, such as text, audio, images, and video, among users (NAUMAN et al., 2020; ALVI et al., 2015; FLORIS; ATZORI, 2015, 2016b).

This subfield supports existing applications in the Internet of Things sector and facilitates the development of multisensory services (CUNNINGHAM; WEINEL, 2016). It enables real-time security and monitoring based on multimedia models, as well as advancements in e-health (CAPPELEN; ANDERSSON, 2016), environmental monitoring (GABRIELLI; TURCHET, 2022), virtual reality (SERAFIN; ERKUT; KOJS; NORDAHL, et al., 2016), and the transmission of 360-degree videos and audio applications.

Research exploring the specific use of the Internet and electronic devices for sound interaction in cyberspace has led to the emergence of another subfield known as the Internet of Sounds (IoS) (TURCHET; LAGRANGE, et al., 2023). This umbrella term integrates IoT and concepts from engineering and the humanities, such as digital audio processing, acoustic monitoring, music, and the arts. The Internet of Sounds facilitates the integration and collaboration of diverse devices with varying capabilities for detection, processing, and communication in co-located and remote environments. Its primary objective is to establish a network of devices that can detect, acquire, process, and exchange sound-related data, thereby supporting artistic activities and commercial applications.

To address the various musical and non-musical domains, IoS can be divided into two paradigms: the Internet of Audio Things (IoAuT) (TURCHET; FAZEKAS, et al., 2020) and the Internet of Musical Things (IoMusT) (TURCHET; FISCHIONE, et al., 2018; TURCHET; ANTONIAZZI, et al., 2020). The first approach, IoAuT, refers to a network of computing devices embedded in physical objects that can produce, receive, analyze, process, transmit, and understand sound information in distributed environments.

IoMusT, on the other hand, is a collection of ecosystems, networks, and musical things[1] as well as protocols and services related to music in both physical and digital environments (TURCHET, 2018c). Specifically, it targets the music industry and its diverse stakeholders,

[1] Electronic devices capable of acquiring, processing, acting on, or exchanging data that serve a musical purpose.

including artists, amateur and professional musicians, audience members, music students and teachers, studio producers, record labels, publishers, and sound engineers. This initiative aims to create new models for immersive concerts, audience participation in artistic performances, remote rehearsals, e-learning, and smart studios (TURCHET; NGO, 2022).

1.1 Multimodal Representations in Musical Systems

Musical practice is inherently a multisensory activity, as exemplified by the act of playing an instrument, which involves the transformation of mechanical energy into acoustic energy (MULDER, 1994). Moreover, musical instruments offer a range of multimodal feedback, as illustrated in Figure 1. This figure outlines both the musician's actions and the corresponding sensory responses, which may include:

- Aural (auditory) feedback, such as the reproduction of a sound corresponding to a played musical note;
- Visual feedback, like the vibration of a guitar string;
- Proprioceptive feedback, experienced when a violinist senses the bow's position or the spacing of fingers on the instrument's neck;
- Haptic feedback, like the vibration felt in a drumstick after striking a drum kit piece.

Together, these feedback modalities form a closed perceptual-motor loop, in which the musician continuously performs actions, perceives their outcomes, evaluates them, and adjusts subsequent movements accordingly. This cycle occurs rapidly and repeatedly throughout musical performance, underpinning the dynamic interaction between perception and motor control (YOUNG; MURPHY; WEETER, 2017).

Parallel to this, archaeological and anthropological research has revealed that, since ancient times, music has had a syncretic nature. Its combination with dance and poetry formed a unique set that allowed the expression of human feelings, sensations, and perceptions through melody, rhythm, words, and gestures (FILIMON, 2023). Furthermore, artistic syncretism can be associated with another structural principle: synesthesia. This phenomenon corresponds to a neurological condition in which the stimulation of one sensory modality triggers involuntary responses in another sensory domain (HARRISON, 2001; DIMOVA, 2024).

From this perspective, several connections can be identified between musical creation and other human sensory experiences. One illustrative case is the historical association between sound and color, which dates back to the pre-Aristotelian period, when early philosophers

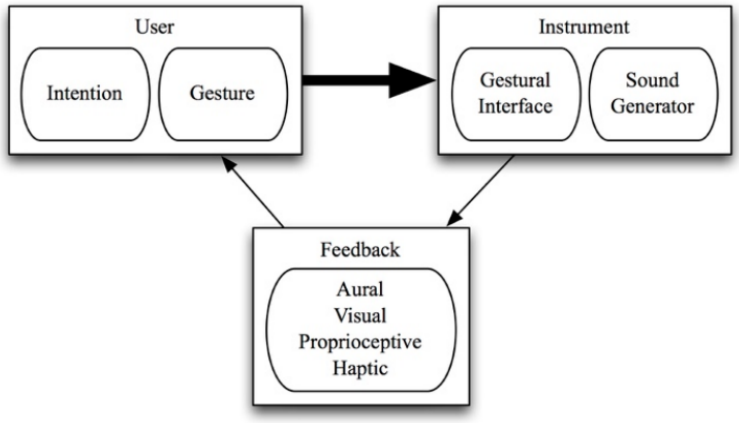


Figure 1: Physical interactions and multimodal feedback in musical practice (YOUNG; MURPHY; WEETER, 2017).

believed that musical harmony embodied the union of distinct colors. In the 20th century, this synesthetic relationship gained renewed prominence. Influenced by *avant-garde* movements such as Dadaism and Futurism, composers began to explore the notion of a sonic space beyond sound, while visual artists sought to construct a visual space beyond the image, thereby expanding the boundaries of sensory expression across artistic domains (FILIMON, 2023).

As a result, numerous proposals and studies have explored the relationship between hearing and vision. A notable example is Arthur Lange's Spectrotone Chart (Figure 2) (LANGE, 1943), a color-coded graphical representation of orchestral instruments designed to assist composers, arrangers, and sound engineers in understanding how various instruments are distributed across the audible frequency spectrum (20 Hz to 20 kHz). The chart highlights how the spectral and perceptual characteristics of each instrument contribute to musical texture, aiding in the creation of balanced orchestral arrangements, the avoidance of sound masking, and the optimization of timbral combinations[2].

In the context of audio mastering, David Gibson proposed a correlation between sound and color, using visual metaphors to understand how different sound elements, such as frequency, equalization, panning, and volume, occupy three-dimensional space (GIBSON, 2005). Figure 3 illustrates this metaphor. Each sphere represents a range of instruments or sound elements, while its size is related to the intensity of the sound. The sphere's position on the vertical axis represents the frequency, with high-pitched sounds positioned higher and low-pitched sounds positioned lower. The depth (Z-axis) of the graphic elements refers to the perceived distance of the sound source in relation to the listener. The horizontal axis indicates the stereo width; sounds positioned to the left or right are panoramically mixed in these channels, while central

[2]Timbre refers to the distinctive characteristic of a sound that enables the identification of different instruments playing the same note.

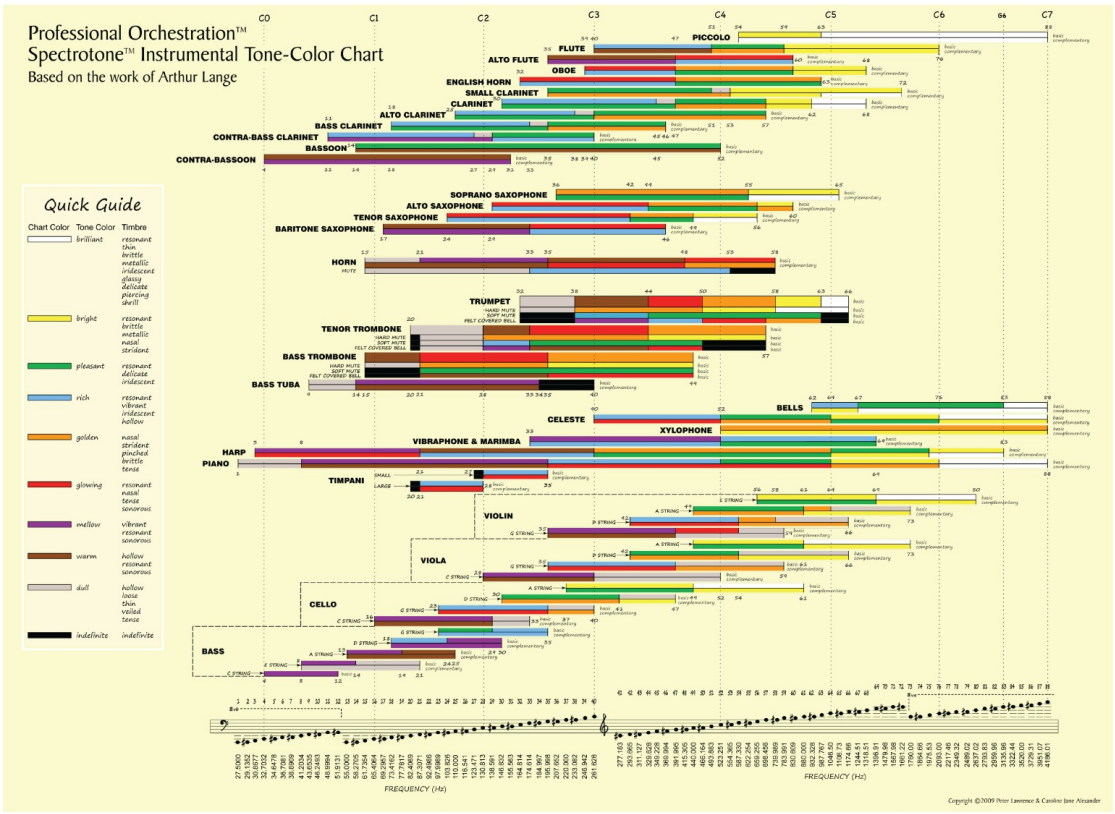


Figure 2: Lange's Spectrotone Chart presents each musical note positioned clearly on a mini musical staff. This layout benefits both musicians and non-musicians alike. Below the staff, the chart includes the frequencies in Hertz (Hz) for each note, allowing for a deeper exploration of their full potential in recordings and mixing process (LANGE, 1943).

sounds are equally present in both channels. The opacity of the spheres represents the density or presence of the sound in the mix. Finally, the visual texture (smooth, rough, shiny) is an analogy for the sound texture (metallic, soft, harsh).

Composers such as Alexander Scriabin (1872–1915) and John Lennon (1940–1980) explored cross-sensory associations by linking sound to other human senses. Scriabin, in particular, developed a system that connected musical notes to specific colors, aiming to evoke a synesthetic experience in his compositions. His approach was partially based on the circle of fifths and influenced by Isaac Newton's color theory. To materialize this vision, Scriabin created the *Clavier à Lumières* (Keyboard with Lights), an instrument designed to project colors in real time, synchronized with the musical tones being performed (BOWERS, 1996; PEACOCK, 1988).

John Lennon, in turn, frequently employed synesthetic language to convey his musical ideas in abstract and subjective terms. A notable example appears in the composition of "Being for the Benefit of Mr. Kite!", featured on the album "Sgt. Pepper's Lonely Hearts Club Band" (1967). Inspired by a Victorian-era circus poster, Lennon envisioned the song as a multisensory experience capable of evoking not only the auditory environment of a circus but also its tactile

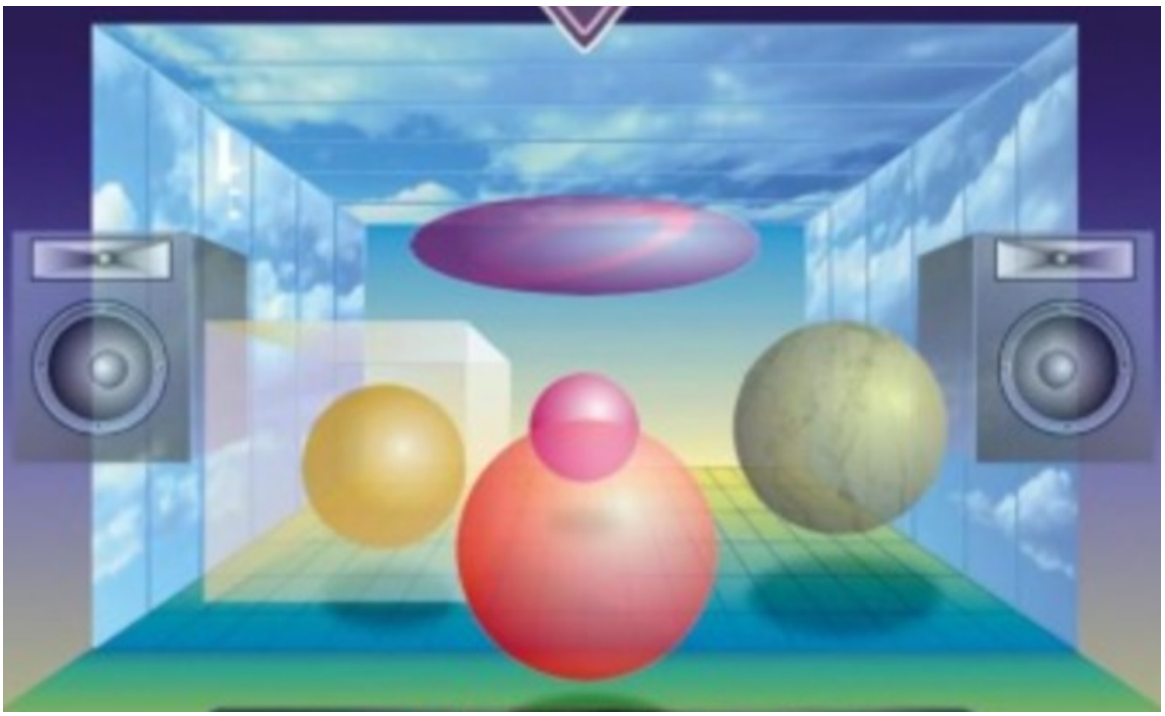


Figure 3: Visual and Spatial Representation of Sound Mixing (GIBSON, 2005).

and olfactory dimensions. In communicating this concept to producer George Martin, Lennon remarked that he wanted the song to "sound like an orange", illustrating his inclination to associate auditory elements with other sensory modalities. He further stated that the track should allow listeners to "smell the sawdust on the floor", reinforcing his intent to craft an immersive sensory narrative (MACDONALD, 2007; LEWISOHN, 2021).

All these approaches reflect a conception of music that transcends the boundaries of hearing, positioning it as a musculoskeletal, neuromotor, and symbolic practice (MULDER, 1994). However, it is important to emphasize that the perception of multisensory elements does not occur automatically. For an individual to comprehend and assimilate meaning from multiple sensory modalities, it is necessary to actively capture, interpret, and integrate information received from distinct sensory channels. Moreover, this information must be cognitively processed and aligned to form a coherent perceptual experience (GHINEA et al., 2014; MAROIS; IVANOFF, 2005; MAYER, 2003).

In this context, sensory experiences can serve as tools for translating subjective perceptions into sonic material, thereby expanding not only the creative possibilities in musical composition, but also illustrating technical concepts related to music production. This multisensory approach enables creators to express themselves more effectively, while offering audiences a heightened level of engagement and immersion (DALSGAARD; SCHNEIDER, 2025).

1.2 Motivation

Despite the importance of such multimodal and multimedia elements in musical practice, the literature review (TURCHET; LAGRANGE, et al., 2023; TURCHET; FAZEKAS, et al., 2020; TURCHET; FISCHIONE, et al., 2018; SILVA, 2025) presents IoS as a rather fragmented field lacking a representation and explanation of technology stacks, especially with regard to the means of communication, storage, analysis, interpretation, and retrieval of these types of information (TURCHET; VIOLA, et al., 2018; VIEIRA; SCHIAVONI; SAADE, 2022).

As a result, misconceptions can eventually lead to an organization lacking an understanding of how to manage and integrate different entities. This can cause stakeholders to overlook relevant design aspects such as standardization, reusability, and compatibility, just to name a few.

Likewise, the integration of multimedia and multisensory data demands a revision and extension of existing IoS models and the devices employed in these systems, particularly in light of their limited energy resources and processing capabilities. This makes it difficult to combine musical data, multimedia, and sensory stimuli in a single object. In addition, this type of transmission already has known problems, such as latency, jitter, privacy and security issues, and lack of interoperability. It also fails to consider the requirements and challenges imposed by devices and multimedia traffic over the network together with other scalar data (TURCHET; FISCHIONE, et al., 2018).

Given the above, it becomes evident that current IoS environments are not adequately equipped to address these emerging challenges. Moreover, integrating multimedia and multisensory elements demands the development of a new set of specialized analytical tools capable of processing these three categories of information in a cohesive and meaningful way.

Aiming to overcome these issues, this thesis proposes a new research domain that integrates all these concepts in the same place and at the same time, allowing them to be interchangeable and non-hierarchical. This vision is given the name of Internet of Multisensory, Multimedia, and Musical Things (Io3MT) (VIEIRA; SAADE; CÉSAR, 2023, 2024; VIEIRA; WEI, et al., 2024; VIEIRA; SAADE; CÉSAR, 2025), with its theoretical concepts based on works already existing in the literature, in order to allow the integration and crossmodal correspondence between multimedia, multisensory, and musical information. This division serves as a fundamental cornerstone, being an expanded vision of its parallel areas that foster the operationalization of computational tools and techniques that enable and support multimedia and multisensory elements in the IoS.

Consequently, Io3MT aims to unite the aforementioned research areas, musical concepts

and existing communities to foster cross-collaborations, as well as address the challenges arising from this combination within a shared perspective at the system level. As a result, a deeper understanding and connection with music are expected, using these new elements as aesthetic and artistic factors. Since the musician's unilateral perception is changed, not only is the music heard, but it also becomes possible to feel and see it while respecting the division proposed by Luca Turchet when separating IoMusT (TURCHET; FISCHIONE, et al., 2018) and IoAuT (TURCHET; FAZEKAS, et al., 2020).

Figure 4 illustrates the manner in which IoT constitutes the overarching research domain from which progressively specialized subdomains emerge. These specializations first address multimedia elements and subsequently incorporate sonic dimensions, thereby distinguishing between non-musical audio — as characterized by IoAuT — and musical audio, as defined within IoMusT. Io3MT, by encompassing both categories of sonic information, can be understood as a refinement of IoS that consolidates these elements while also introducing multimedia components, an unprecedented integration within this area. Moreover, Io3MT builds upon the multisensory affordances foregrounded in IoMusT, thereby extending the conceptual and technical scope of existing sound-centric IoT paradigms.

It is anticipated that Io3MT will not only facilitate new musical applications and services, such as augmented performances and installations but will also have a significant impact on various sectors of society, including creative industries and cultural digital experiences. The concepts behind Io3MT can enhance virtual environments and contribute to the creation of smart museums and galleries. Additionally, it can lead to the development of more realistic video games, fostering a deeper connection between players and extending their gaming experience. In the realm of home entertainment, Io3MT can make movies, videos, books, and music on streaming services more immersive. Furthermore, it has potential applications in healthcare and well-being services, as well as in educational implementations, among other areas.

1.3 Research Questions

In light of this, the following research questions (RQ) arise:

- RQ1: What are the functional requirements for Io3MT environments that enable the integration of multisensory and multimedia data within IoS-based systems?
- RQ2: Which evaluation methodologies are most suitable for assessing Quality of Experience (QoE) in different Io3MT environments that combine auditory, multimedia, and multisensory stimuli?

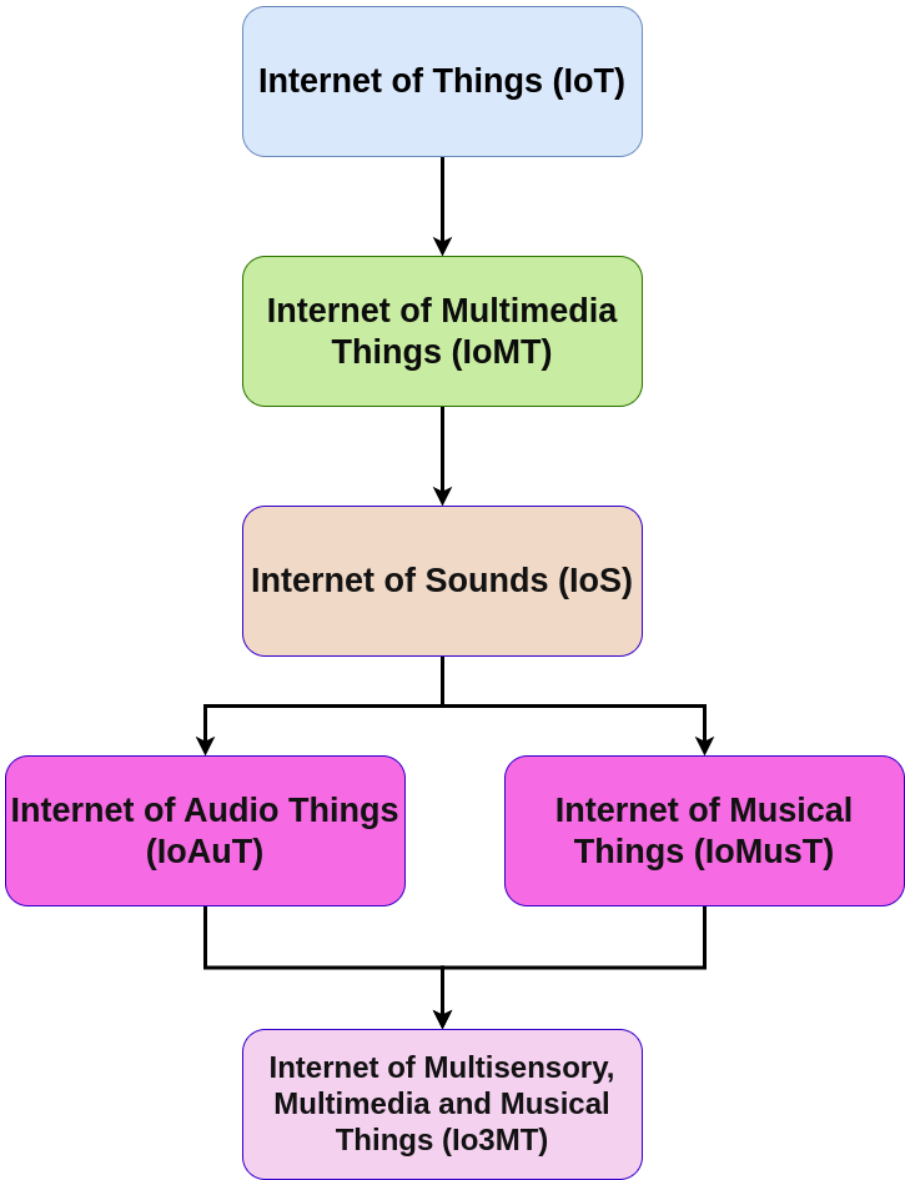


Figure 4: A schematic representation illustrating the relationship between the Internet of Multisensory, Multimedia, and Musical Things (Io3MT), the related domains of the Internet of Musical Things (IoMusT) and the Internet of Audio Things (IoAuT), and the broader foundational domains of the Internet of Multimedia Things (IoMT) and the Internet of Things (IoT).

- RQ3: In what ways can the integration of music, multimedia content, and multisensory feedback enhance artistic expressiveness and aesthetic experience within Io3MT-based systems?

1.4 Goals

Due to the multifaceted and complex nature of Io3MT research, existing models for IoT, IoS, and their related areas are insufficient to specify and implement these new environments. A scientific structure to capture, measure, quantify, and judge the user experience needs to be systematized. Therefore, the main goal of this thesis is to provide a holistic view of Io3MT,

explaining a series of guidelines and indicating a reference model. In addition, this thesis implements Io3MT concepts in different use cases and evaluates them quantitatively and qualitatively.

Furthermore, this thesis aims to meet the following specific goals:

- Conduct a systematic mapping to identify the requirements relevant to modeling IoT reference systems and explore how existing approaches can be extended to meet the needs of Io3MT;
- Provide a comprehensive overview of the core elements, components, and aspects of Io3MT to help structure and organize it effectively;
- Design and implement scenarios that integrate the proposed approach;
- Evaluate the scenarios based on network performance, QoE, and the efforts related to standardization and implementation;
- Foster collaborations among existing communities involved in creating art, music, and entertainment mediated by technology.

On the other hand, this thesis does not aim to answer how devices produce each sensory effect, media, or musical information, how this data is loaded and processed, or how the temporal aspects are managed.

1.5 Methodology

This thesis focuses mainly on the design and analysis of a reference model for the Io3MT environment. To achieve this purpose, concepts from distributed systems, IoT, computer music, Human-Computer Interaction (HCI), and artistic performances are used. This constitutes **applied research**, as it seeks to transfer and adapt practical contributions from each of these fields to the domain under investigation.

This research also advances the understanding of the behavior of Io3MT, with a particular focus on its underlying techniques. It includes a critical analysis of the theoretical foundations that justify the development of the present study. To this end, a **selective bibliographic review** was conducted, encompassing scientific publications found in books, peer-reviewed journals, and proceedings of national and international conferences. These works address both the core themes and the peripheral concepts relevant to the research, thereby providing the necessary theoretical and methodological support for its advancement

There is also an **experimental research** that assesses network performance and QoE in the proposed tests. The first measurement is crucial for ensuring satisfactory data delivery and meeting the established performance requirements. This study focuses on several key metrics (VIEIRA; SCHIAVONI; SAADE, 2022; TURCHET; CASARI, 2024): latency, which measures the time it takes for data packets to travel between the source and destination; jitter, which refers to the variation in latency; and throughput (also known as transfer rate), which indicates the amount of data transmitted from one point to another on the network within a specific timeframe (ZHU et al., 2004; GOZDECKI; JAJSZCZYK; STANKIEWICZ, 2003; ROCHA; SOUZA FILHO, 2001). These values are obtained from measurements taken using a network analyzer (Wireshark[3]) and subsequently the averages, standard deviations and minimum and maximum values are calculated for each of these metrics.

The second aspect uses techniques based on User Experience (UX) design to verify user satisfaction with the different applications presented. In the first evaluation, a semi-structured interview (WILSON, 2013) is used, a research approach combining predefined questions or topics with flexible questions to delve deeper into areas of interest during interaction with interviewees.

A specific UX protocol incorporating quantitative and qualitative metrics was developed for the second experiment, which assesses the Io3MT concepts in an immersive environment with vibrotactile feedback. Since the experiment takes place in virtual reality (VR), two questionnaires are utilized: the Simulator Sickness Questionnaire (SSQ) (KENNEDY et al., 1993) and the Presence Questionnaire (PQ) (WITMER; SINGER, 1998). The SSQ measures physical discomfort, disorientation, oculomotor symptoms, and nausea. Meanwhile, the PQ evaluates involvement, immersion, sensory fidelity, and interface quality factors.

Then, the System Usability Scale (SUS) (BROOKE et al., 1996) was used to obtain a subjective measure of perceived usability, while the National Aeronautics and Space Administration Task Load Index (NASA-TLX) (HART; STAVELAND, 1988) provided a way to quantify the cognitive and physical workload involved in the tasks.

To analyze the haptic aspects of the system, the questionnaire proposed by (SATHIYAMURTHY et al., 2021) was used. This questionnaire evaluates five key constructs: harmony (the tactile integration with other senses), expressiveness (the ability to convey nuances), autotelia (the intrinsic pleasure derived from haptic information), immersion (the level of user engagement), and realism (the credibility of the sensations experienced).

Finally, a semi-structured interview is conducted with the participants — following the same methodological approach adopted in the first study — in order to deepen the qualitative

[3] https://www.wireshark.org/

insights and contextualize the quantitative findings obtained through the questionnaires.

A detailed discussion of these methods will be presented at a later stage of the text, in a contextually appropriate section.

1.6 Publications

In terms of knowledge dissemination, 20 academic works were published, comprising journal articles, full and short papers in conference proceedings, book chapters, demonstrations, and tutorials at both national and international conferences. These outputs are detailed below:

- **Articles Published in Journals**
  - MATTOS, Douglas; VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora C.; GHINEA, Gheorghita. Assessing Mulsemedia Authoring Application Based on Events With STEVE 2.0. IEEE Access, v.13, p.100970-100986, 2025. DOI: https://doi.org/10.1109/ACCESS.2025.3576167.
- **Full Papers Published in Conference Proceedings**
  - VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora C.. A Survey on the Internet of Musical Things: Environment Challenges, Standards, Services, and Future Visions. In: IEEE 8th World Forum on Internet of Things (WF-IoT), Yokohama, Japan, 2022. p. 1-6;
  - VIEIRA, Rômulo; ROCHA, Marcelo; ALBUQUERQUE, Célio; MUCHALUAT-SAADE, Débora C.; CÉSAR, Pablo. RemixDrum: A Smart Musical Instrument for Music and Visual Art Remix. In: IEEE 9th World Forum on Internet of Things (WF-IoT), Aveiro, Portugal, 2023. p. 1-7;
  - VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora C.; CÉSAR, Pablo. Towards an Internet of Multisensory, Multimedia and Musical Things (Io3MT) Environment. In: 4th International Symposium on the Internet of Sounds, Pisa, Italy, 2023. p. 1-10;
  - VIEIRA, Rômulo; WEI, Shu; RÖGGLA, Thomas; MUCHALUAT-SAADE, Débora C.; CÉSAR, Pablo. Immersive Io3MT Environments: Design Guidelines, Use Cases and Future Directions. In: IEEE 5th International Symposium on the Internet of Sounds (IS2), 2024, Erlangen, Germany. Proceedings... Erlangen: IEEE, 2024. p. 1-10. DOI: https://doi.org/10.1109/IS262782.2024.10704141;

- VIEIRA, Rômulo; FARIAS, Flávio; MACENA, Euller; MUCHA- LUAT SAADE, Débora C. Assessing 360-degree multisensory experiences with AMUSEVR. In: Proceedings of the 1st International Workshop on Multi-Sensorial Media and Applications (MSMA '25), Ireland, 2025. New York: Association for Computing Machinery, 2025. p. 63–71. DOI:10.1145/3728485.3759238. ISBN 9798400718427;
- SILVA, Carla Estefany Caetano; VIEIRA, Rômulo; TREVISAN, Daniela Gorski; MUCHA- LUAT SAADE, Débora Christina. Análise de sinais de eletroencefalograma para medição de atenção em um ambiente musical imersivo multissensorial. In: Brazilian Symposium on Multimedia and the Web (WebMedia 2025), 2025, Rio de Janeiro, Brazil. Proceedings of the Brazilian Symposium on Multimedia and the Web. [S.l.]: SBC – Sociedade Brasileira de Computação, 2025. ISSN 2966-2753;
- VIEIRA, Rômulo; SILVA, Carla Estefany Caetano; TREVISAN, Daniela Gorski; MUCHA- LUAT SAADE, Débora C.; CÉSAR, Pablo. Can You Feel My Brain? Investigating Attentional Engagement through EEG in an Immersive Musical Environment. In: IEEE International Symposium on the Internet of Sounds (IS2 2025), 2025. Proceedings [...]. [S.l.]: IEEE, 2025. **(In press)**.

- **Short Papers Published in Conference Proceedings**

  - VIEIRA, Rômulo; SCHIAVONI, Flávio; MUCHALUAT-SAADE, Débora Christina. Sunflower: a proposal for standardization on the Internet of Musical Things environments. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), 40., 2022, Fortaleza, CE. Anais... Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 25-32. ISSN 2177-9384.
    DOI: https://doi.org/10.5753/sbrc_estendido.2022.222447;
  - VIEIRA, Rômulo; SCHIAVONI, Flávio; MUCHALUAT-SAADE, Débora C. A Proposal for Standardization of Internet of Musical Things (IoMusT) Environments. In: IEEE 8th World Forum on Internet of Things (WF-IoT), 2022, Yokohama, Japan. Proceedings... Yokohama: IEEE, 2022. p. 1-2;
  - VIEIRA, Rômulo; IVANOV, Marina; ABREU, Raphael; SANTOS, Joel A. F. dos; MATTOS, Douglas; MUCHALUAT-SAADE, Débora Christina. Autoria de aplicações multissensoriais para TV 3.0 com a ferramenta STEVE. In: Workshop Futuro da TV Digital Interativa – Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 29., 2023, Ribeirão Preto, SP. Anais... Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 143-149. ISSN 2596-1683;
  - SANTOS, Joel dos; VIEIRA, Rômulo; JOSUÉ, Marina Ivanov; OLIVEIRA, Karen Sá; MUCHALUAT-SAADE, Débora Christina. Multidevice Support in the Next

Generation of the Brazilian Terrestrial TV System. In: ACM International Conference on Interactive Media Experiences Workshops (IMXw '24), 2024. Proceedings... New York: Association for Computing Machinery, 2024. p.34-38. DOI: https://doi.org/10.1145/3672406.3672412;

- IVANOV, Marina; VIEIRA, Rômulo; SANTOS, Joel A. F. dos; MUCHALUAT-SAADE, Débora Christina. TV 3.0: integração e controle de renderizadores de efeitos sensoriais. In: Workshop Futuro da TV Digital Interativa – Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 30., 2024, Juiz de Fora, MG. Anais... Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 297-302. ISSN 2596-1683.
  DOI: https://doi.org/10.5753/webmedia_estendido.2024.244585;
- VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora Christina; CÉSAR, Pablo. Internet of Multisensory, Multimedia and Musical Things (Io3MT): Framework Design, Use Cases, and Analysis. In: ACM International Conference on Interactive Media Experiences (IMX '25), 2025. Proceedings... New York: Association for Computing Machinery, 2025. p. 484-487. DOI: https://doi.org/10.1145/3706370.3731931;
- SILVA, Carla Estefany Caetano; VIEIRA, Rômulo; TREVISAN, Daniela Gorski; MUCHA- LUAT SAADE, Débora Christina. Towards analysing user attention using electroencephalography in immersive multisensory virtual environments. In: ACM International Conference on Interactive Media Experiences Workshops (IMXW), 25., 2025, Niterói, RJ. Anais... Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 91-95.
  DOI: https://doi.org/10.5753/imxw.2025.2092;
- BARRÉRE, Eduardo; SOUSA, Li-Chang Shuen Cristina Silva; MUCHALUAT-SAADE, Débora C.; MORENO, Marcelo F.; NETO, Carlos de Salles Soares; SANTOS, Joel André Ferreira dos; JOSUÉ, Marina Ivanov P.; COSTA, Rômulo Augusto Vieira; COSTA, Iago Victor Silva; GONÇALVES, João Víctor Cruz; SOUSA, Sarah Regina Bezerra. AppEduTV3.0: Aplicativo educacional para a TV 3.0. In: Brazilian Symposium on Multimedia and the Web (WebMedia 2025), 2025, Rio de Janeiro, Brazil. Proceedings of the Brazilian Symposium on Multimedia and the Web. [S.l.]: SBC – Sociedade Brasileira de Computação, 2025. ISSN 2966-2753.

- **Book Chapters**

  - VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora Christina; CÉSAR, Pablo. Exploring Artificial Intelligence for Advancing Performance Processes and Events in Io3MT. In: International Conference on Multimedia Modeling (MMM), 30., 2024,

Amsterdam. Proceedings... Part IV. Berlin: Springer-Verlag, 2024. p. 234-248. DOI: https://doi.org/10.1007/978-3-031-53302-0_17.

- **Demonstrations**

  - VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora Christina; SCHIAVONI, Flávio Luiz. Sunflower: An Interactive Artistic Environment based on IoMusT Concepts. In: Proceedings of the 2022 ACM International Conference on Interactive Media Experiences (IMX '22). New York, NY, USA: Association for Computing Machinery, 2022, p. 245-248;
  - VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora Christina; CÉSAR, Pablo. PhysioDrum: Bridging Physical and Digital Realms in Immersive Musical Interaction. In: ACM International Conference on Interactive Media Experiences (IMX '25), 2025. Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (IMX '25) New York: Association for Computing Machinery, 2025. p. 356-358. DOI: https://doi.org/10.1145/3706370.3731698.

- **Tutorials**

  - VIEIRA, Rômulo; MUCHALUAT-SAADE, Débora C.; SCHIAVONI, Flávio. Introdução à Internet das Coisas Musicais: prática utilizando Pure Data e Processing. In: Tutoriais – Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 28., 2022, Curitiba. Anais... Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 127-131. ISSN 2596-1683.

1.7 Thesis Structure

This thesis is structured as follows. Chapter 2 introduces the terminologies employed throughout the text, as well as the main research fields that serve as the foundations of Io3MT.

Chapter 3 discusses related work concerning networked musical practice, and applications that combine traditional media with sensory factors.

Chapter 4 presents the proposed Io3MT reference model, outlining its main guidelines and specifying the technical and artistic requirements that must be addressed. It also discusses the protocols, data types, and tools that can be employed to enable the practical implementation of these concepts.

Chapter 5 addresses the construction and subsequent analysis of a use case that serves as a proof of concept for the reference model proposed. More specifically, it introduces a

smart musical instrument (SMI), named RemixDrum, which, through the integration of sensors, actuators, and wireless communication, enables new forms of artistic expression aligned with the domain discussed in this study. The QoE evaluation consists of a semi-structured interview with an expert user, whereas the network analysis encompasses observations regarding latency, jitter, and network throughput.

Chapter 6 presents the expansion of Io3MT concepts into immersive scenarios. To this end, a focus group composed of four experts was conducted to define a set of guidelines to be considered in the design of such environments. As a proof of concept, an application called PhysioDrum was developed. This system extends the RemixDrum by incorporating a power supply module and vibration motors to deliver vibrotactile feedback to users. In addition, an electronic interface was implemented to integrate physical pedals into the system, enabling users to control different components of the virtual drum kit.

Chapter 7 addresses the experimental study of PhysioDrum. For this purpose, an evaluation protocol was developed to formalize the main constructs and both quantitative and qualitative methods for assessing immersive musical experiences. Subsequently, a group of 30 participants performed four rhythmic tasks of increasing complexity, designed to evaluate the overall quality of the application and, more importantly, to investigate the impact of haptic feedback within this type of environment.

Finally, Chapter 8 presents the conclusions of this research, emphasizing its primary contributions in relation to the proposed reference model and the implemented use cases. The chapter also discusses inherent limitations and delineates potential avenues for future investigations aimed at advancing this emerging research domain.

2 Background

This chapter aims to provide a non-exhaustive overview of the key domains and concepts underlying the Internet of Multisensory, Multimedia, and Musical Things (Io3MT). It presents the main definitions, characteristics, and epistemological foundations associated with its pillars, as well as the critical factors to be considered for its adoption.

2.1 On the Relation Between Multimedia, Multisensory and Musical Elements

A comprehensive understanding of Io3MT necessitates a clear conceptual definition of its three fundamental domains. While these concepts share similar technological abstractions, they originate from distinct disciplines and display significant differences in terms of epistemology, functionality, and phenomenology.

The following section outlines the main characteristics and particularities of each domain, aiming to establish a solid theoretical foundation for the critical articulation proposed by Io3MT.

2.1.1 An Overview on Multimedia Systems

The concept of multimedia is inherently multidisciplinary, intersecting several important industries, including computing, telecommunications, content publishing, audio and video electronics, as well as the television, film, and broadcasting sectors (STEINMETZ; NAHRSTEDT, 2002). Formally, multimedia is defined as the integration of various forms of media — understood as means of representing and distributing information — such as text, images, audio, video, and animation, all within a unified digital environment (FELDMAN, 1994; FURHT, 2002).

From a functional perspective, multimedia can be divided into two main forms: linear and non-linear (MEIXNER, 2017). Linear multimedia is defined by the sequential and continuous presentation of information, featuring a predetermined beginning and end without any user

intervention. These systems can be automated to display content at fixed time intervals, thus removing any interactive control that users might have over the presentation flow.

In contrast, non-linear multimedia does not follow a chronological or predetermined order. Instead, it is inherently interactive, requiring active participation from users for navigation and decision-making while exploring the content.

Multimedia data can be created manually by humans, such as animations; captured by various sensors, like cameras, microphones, or motion capture systems; or synthetically generated by computer systems, including 3D virtual environments or sounds produced using programming languages (SALEME; SANTOS, et al., 2019).

Multimedia systems integrate the hardware and software components required to support the combination of at least two fundamental types of media: discrete media, which consist of static and time-independent elements such as text and images, and continuous media, which are time-dependent and include formats such as video, audio, and animation. These media are delivered through presentation devices, such as screens, speakers, or projectors, each characterized by specific attributes such as color pattern, intensity, resolution, and typography, all designed to stimulate human sensory perception. These presentation devices may vary in dimensionality, encompassing both two-dimensional displays (e.g., monitors) and more complex three-dimensional environments, such as holographic projection systems (FELDMAN, 1994; SALEME; SANTOS, et al., 2019).

Moreover, these systems must guarantee temporal synchronization across continuous media streams, for example by maintaining accurate alignment between audio and video in videoconferencing applications. They should also provide mechanisms for interaction, navigation, and content manipulation, thereby defining themselves as interactive systems (HALSALL, 2001).

Multimedia systems can also operate in a distributed configuration, in which the components responsible for sending (source) and receiving (sink) data streams are geographically dispersed. Ideally, the networks that support such architectures should function with minimal or no transmission errors. Nevertheless, many multimedia applications are designed to tolerate a certain degree of packet loss or data corruption, since the demands of real-time delivery and temporal synchronization often take precedence over strict error correction. In some cases, this trade-off may even involve the intentional discarding of packets, prioritizing continuity and responsiveness over completeness of data (FURHT, 2002; STEINMETZ; NAHRSTEDT, 2002).

Even so, to ensure a satisfactory user experience, it is necessary to meet specific requirements, including low latency, minimal jitter, accurate synchronization, and support for multipoint communication through multicast distribution protocols (FURHT, 2002).

2.1.2 Mulsemedia: Sensory Expansion of Multimedia

In the contemporary context, multimedia applications are widely disseminated across a diverse range of devices, including computers, smartphones, tablets, televisions and Head-Mounted Displays (HMDs). These platforms improve the user experience by offering greater immersion and interactivity. They also introduce innovative narratives and applications in fields such as communication, education, and marketing (MATTOS et al., 2025).

However, most of the content available is based on a combination of video and audio, focusing mainly on just two human senses: hearing and sight, a clear contrast with the fact that more than 60% of human communication is non-verbal and all five senses (sight, hearing, touch, taste and smell) are used to understand and interact with the real world (MATTOS et al., 2025; GHINEA et al., 2014).

Interoceptive capabilities are usually not taken into consideration, such as kinesthesia, responsible for movement; equilibrioception, which concerns the sense of motor stability; thermoception, which reflects the ability to feel cold and heat; proprioception, which is the awareness of the body's position in physical space; nociception, reflected in the sensation of pain; and interoception, which is the ability to feel internal organs (GSÖLLPOINTNER; SCHNELL; SCHULER, 2016; MATTOS et al., 2025).

To explore all these senses in interacting with digital media and create a greater fusion between physical and digital media, Multiple Sensorial Media (Mulsemedia) (GHINEA et al., 2014) emerged. Furthermore, the use of this concept can increase immersion in content and the quality of experience perceived by users. It is essential to note that this concept differs slightly from digital multisensory experiences, which do not encompass multimedia information (WALTL; TIMMERER; HELLWAGNER, 2010; RAINER et al., 2012; YUAN; GHINEA; MUNTEAN, 2014; MONKS et al., 2017).

Applications developed according to these models are typically structured into three phases: authoring (or production), distribution, and rendering process (COVACI et al., 2018). The first phase focuses on the creation and synchronization of various sensory effects with multimedia content. Several techniques can be utilized to achieve this, including the capturing and processing of sensor data, automatic video extraction, or manual authoring of effects. This process can be structured in two ways: time-based, similar to how a video editor operates, or event-based, where specific occurrences trigger different actions.

After completing this process, the second stage begins, where these sensory effects are encoded, processed, and distributed. Once done, the third and final phase renders this information for end users.

It is possible to observe that mulsemedia perception is not something trivial, being the result of steps that combine sensory processing and cognitive reasoning for a full understanding of the semantic context in which it is inserted (GHINEA et al., 2014). Therefore, specific guidelines must be followed in the design of these environments. These principles should include a range of devices and applications that can translate information and emotions between the digital domain and the physical world and vice versa (SALEME; FALBO, et al., 2018). They should also incorporate temporal relationships between different media and sensory effects, showcase expected behaviors, and allow for user customization based on individual preferences. In addition, these environments are traditionally organized into layers, enabling each layer to evolve independently, although this separation is not a strict requirement (SALEME, Estevão; SANTOS; GHINEA, 2019).

2.1.3 What is Music?

Music constitutes an ubiquitous phenomenon across human cultures, fulfilling a wide array of functions that range from celebration and entertainment to religious practices and funerary rituals. Beyond its symbolic and sociocultural significance, this form of creative expression plays a fundamental role in the modulation of emotional states and the development of cognitive processes, contributing to both the construction of subjective experience and the regulation of affective dynamics (VUUST et al., 2022; VIEIRA; GONÇALVES; SCHIAVONI, 2020; LEE, S. et al., 2024).

However, the definition of music does not lend itself to a single and definitive formulation. Traditionally, music has been described as the simultaneous and successive combination of sounds, with order, balance, and proportion within a specific time interval (MED, 1996; SCHAFER, 1992). Melody, harmony, counterpoint, dynamics, and rhythm constitute the primary structural dimensions of a musical composition (MED, 1996).

Melody is defined as the arrangement of sounds in a sequence over time, reflecting the horizontal aspect of musical structure. Harmony, on the other hand, involves the simultaneous combination of sounds, creating a vertical dimension in music. Counterpoint consists of the layering of multiple independent melodic lines that are played together, thereby incorporating both the horizontal and vertical elements of musical composition.

Dynamics refer to the changes in sound intensity within a music piece. These variations are typically represented by volume levels that range from *pianississimo* (very soft) to *fortississimo* (very loud). These variations contribute to the emotional expression and interpretative articulation of the music.

Lastly, rhythm refers to the temporal organization of sounds and silences, determining the

order and proportion in which sonic events occur, both in melody and harmony. It constitutes the temporal basis that structures music, enabling the perception of regular patterns and rhythmic contrasts.

The combination of these elements creates musical sound, characterized by pitch, duration, intensity, and timbre, which collectively influence the auditory experience (MED, 1996). Pitch is determined by the frequency of sound vibrations. Higher and faster frequencies are perceived as acute (high-pitched) sounds, while lower frequencies are perceived as deep (low-pitched) sounds. The sequential arrangement of sounds with varying pitches gives rise to melody, whereas the simultaneous combination of different pitches forms chords, which constitute the basis of harmony.

Duration refers to the temporal extension of a sound and it is an important feature to shape both rhythmic and melodic structures in music. Intensity, in turn, corresponds to the amplitude of sound vibrations and is influenced by the force exerted by the emitting actor. More intense sounds are perceived as louder and more energetic, whereas less intense sounds have a softer tone. The variation of intensity during a musical performance create dynamics, responsible for conveying expressive and emotional nuances.

Timbre, commonly described as the "color" of sound, results from the specific combination and relative intensity of harmonics that accompany the fundamental frequency generated by a sound source. This spectral composition imparts distinctive acoustic characteristics to each source, enabling perceptual differentiation. The variation and overlap of distinct timbres add diversity to a composition, a concept known as instrumentation.

This technical and formal definition, while relevant in Western contexts, does not adequately capture the plurality of musical practices around the world (NETTL, 1983). In many cultures, music is closely connected with dance, language, and rituals, reflecting its deep integration into social and symbolic dynamics. Consequently, musicologists, philosophers, sociologists, and philologists work to define what music is in different contexts (IAZZETTA, 2001).

From an anthropological perspective, (MERRIAM, 1960) suggested that music should not be viewed merely as a collection of sounds, but rather as a cultural behavior that encompasses intentions, contexts, and meanings. The various functions of music, such as emotional expression, communication, reinforcement of social norms, and social integration, underscore its complex role in human societies. Thus, music is not only the sounds created but also the system of values and practices that support it.

Still in this context, the naturalist movement suggests that music can exist independently within nature and is inherently tied to it. Proponents of this idea argue that music does not qualify as art; rather, it is the act of creating and expressing music that embodies artistic

value to this action. While listening to music can offer leisure, and opportunities for learning, it ultimately stems from mastering the science behind it (MED, 1996). Music is therefore recognized as a natural and universal phenomenon.

The theory of natural resonance supports the idea that harmonic relationships have a mathematical nature that influences on the auditory perception of consonance and dissonance. This establishes the dominance of natural practice over formal methods (MORAES, 2010; FREITAS, 2018). The theory also suggests that, because music is a natural and intuitive phenomenon, people can compose and perceive music in their minds without necessarily learning or fully understanding it. Composing, improvising, and performing are art forms that draw on this musical phenomenon. From this perspective, music can exist independently of communication or even perception, as it arises from physical interactions that do not require human involvement.

In contrast, functionalist theorists (MEYER, 2008; MASSI, 1992) argue that music does not exist independently of perception. According to this view, music is only realized when a musical artifact mediates the relationship between creator and listener. This dialogue is mediated by a formative musical gesture, as represented in musical notation, or by a formalized gesture that emerges through performance and interpretation.

From this perspective, several defining characteristics of music are identified: i) music is an art form and an aesthetic manifestation, intentionally crafted to convey emotional content; ii) music acts as a mean of communication, serving as one of the forms of language to transmit and receive certain messages between individuals, or between the emotions and senses of the individual performing a song; iii) music presupposes the presence of sound. Although silence may be employed as a structural or expressive element within musical discourse, it cannot constitute music in isolation. Grounded in these premises, proponents of the functionalist approach understand music as a semiotic phenomenon.

From a sociological standpoint (ROY; DOWD, 2010), music can be viewed in two ways. From a textualist perspective, music is an object that originates from a specific moment of creation, retains relatively stable characteristics across time and space, and holds potential for varied uses and effects. From a contextualist perspective, music is an ongoing activity or process — dynamic, mutable, and inherently open to interpretation and reconfiguration in different social settings.

Based on principles of cognitive psychology, music is understood as a complex mental activity that encompasses auditory perception, memory, attention, and emotion. Studies by (SLOBODA, 1999) and (HARGREAVES; NORTH, 1999) demonstrate that music influences listeners in multiple ways. It can regulate emotional states, facilitate social interactions, reinforce cultural identities, and provide aesthetic pleasure. Additionally, musical cognition reveals universal

patterns, such as sensitivity to tonal scales and rhythm, although these patterns are influenced by cultural and individual differences (PATEL, 2010).

Simultaneously, neuroscientific approaches suggest that music engages a wide network of brain regions that are linked to reward, emotion, movement, and language (ZATORRE; SALIMPOOR, 2013). Research by (KOELSCH, 2014) demonstrated that music can evoke powerful emotional responses without the need for words, through neural mechanisms involving the limbic system, auditory cortex, and dopaminergic pathways. These findings support the idea that music has evolved as an adaptive tool with social and emotional functions (FITCH, 2006; CROSS, 2001).

In addition to these perspectives, the philosophy of music addresses fundamental questions concerning the nature and ontological status of musical works (IAZZETTA, 2001). Within this domain, music is commonly classified as a performing art or even a sublime art, given its ephemeral nature and expressive power. According to (KIVY, 2002), music is an art form that expresses emotions in a paradigmatic manner, although it does not necessarily convey specific or determinate emotional states. Other theorists, such as John Blacking (BLACKING, 1973), emphasize that music is, above all, a human activity, whose meanings are deeply embedded in social practices and corporeal experience, underscoring the cultural and embodied dimensions of musical expression.

Given its polysemic nature, Jean Molino (MOLINO, 1975) synthesizes this discussion by conceptualizing music as a set of interdependent and inseparable factors. The complexity of the relationships established among these factors precludes the notion of a single definition serving as a universal model for all music types. Instead, Molino proposes the idea of musical facts, understood as contextualized phenomena that are intrinsically linked to the specific moment and the sociocultural environment in which they are created and performed (IAZZETTA, 2001).

2.1.4 Conceptual and Epistemological Considerations

This section recognizes that music and audio are treated similarly by computer systems, while also emphasizing the conceptual and functional differences between them. Music is an artistic and cultural expression, and its classification as audio, noise, or musical composition depends on various aesthetic, symbolic, subjective, and contextual factors recognized by both the composer and the listener. Because of its subjective and symbolic nature, music requires a different epistemological approach compared to audio.

Conversely, audio is conceived as any sound stimulus that is not necessarily musical, encompassing noises, speech, sound effects, and other auditory signals, falling under the umbrella of multimedia, which focuses solely on the acoustic characteristics of music, while neglecting

its aesthetic, semiotic, and expressive qualities.

In light of this scenario, this work adopts the following terminology: “music” refers specifically to sound stimuli that possess aesthetic intent and a musical structure, following the definition of “musical fact” proposed by Molino (MOLINO, 1975). The term “audio” designates non-musical auditory stimuli, while “sounds” encompasses both categories (TURCHET; FAZEKAS, et al., 2020).

Similarly, mulsemedia focuses on the technical and conceptual interconnection between traditional multimedia content and additional sensory channels, without encompassing musical aspects in their artistic, symbolic, and structural dimensions. For this reason, Io3MT domain broadens this scope by explicitly highlighting the interaction between sensory, multimedia, and musical elements, recognizing the unique impact that each piece of information exerts — and suffers — in this ecosystem, overcoming the limitations imposed by approaches strictly focused on multimedia content and/or its relationship with sensory elements.

2.2 Networked Music Performance (NMP)

Musical practice has traditionally been grounded in in-person interactions, with all participants sharing the same physical space during the process of sound creation (LOVERIDGE, 2020; MONTAGU, 2017; VIEIRA; GONÇALVES; SCHIAVONI, 2020). However, this paradigm began to change in the mid-1970s, when a group of artist-scientists known as the League of Automatic Music Composers started exchanging musical information over network. They mapped frequencies and intervals between computers to control rhythmic patterns (WEINBERG, 2002).

Following this initial development, the field of computer-mediated musical collaboration began to be explored, aiming to investigate innovative techniques capable of generating novel acoustic phenomena (ROTTONDI et al., 2016). In this context, the concept of Networked Music Performance (NMP) emerged, referring to musical interaction between geographically distributed musicians through network connections that support real-time collaboration (LAZZARO; WAWRZYNEK, 2001).

The primary objective of NMP is to recreate realistic conditions for musical interaction over networks, thereby enabling a wide range of applications. These include tele-auditions (GU et al., 2005), remote music education (LOVERIDGE, 2020), rehearsals (IORWERTH; MOORE; KNOX, 2015; IORWERTH; KNOX, 2019), jam sessions, and distributed concerts. NMP supports educators, students, professional musicians, and enthusiasts, by providing a flexible and accessible environment that promotes cost-effectiveness, efficiency, productivity, and creativity in diverse musical activities (TAMPLIN et al., 2020).

To enable these services, NMP relies on an architecture composed of two primary components: the server, which functions as a centralized management unit, and the client, which provides musicians with access to the system. It is important to emphasize that this classification refers to the behavioral roles of the components within the environment, rather than the communication model adopted between them (GU et al., 2005).

Several critical aspects must be considered when implementing such environments. These are outlined in the following sections (TURCHET; FISCHIONE, et al., 2018).

2.2.1 Low Latency

NMP applications require low latency to simulate usage conditions similar to natural music environments, where the sound produced is heard almost instantaneously due to the propagation of acoustic waves in air. Several studies (CAROT; WERNER, 2009; WINCKEL, 2014) estimate that the maximum tolerable delay in networked musical performance lies between 20 and 30 milliseconds. This latency corresponds to a physical separation of approximately 8 to 10 meters, which is commonly considered the upper limit within which musicians can perform together while maintaining a consistent tempo, without the need for explicit synchronization cues.

Nevertheless, this threshold may vary depending on several factors, including the performers' skills, the musical style being played, the intrinsic listening delay introduced by the instruments themselves, the presence of other time-dependent feedback mechanisms involved in the interaction, and even deliberate latency insertion for achieving specific aesthetic or expressive effects.

Latency may be influenced by multiple causes, including the stages of signal transmission across the network, which involve processing, transmitting and receiving packets on both sides of each link. Additionally, there is propagation delay in the physical medium, delays generated by intermediate nodes, packet queues and the use of a playback buffer (ROTTONDI et al., 2016).

Another factor contributing to latency is the processing and reproduction of audio performed by sound cards. This process comprises the packaging of audio data, its fragmentation, the transfer to kernel space, the subsequent copying back to user space, and the final unpacking stage (CARÔT; WERNER, 2007).

2.2.2 Synchronization

Another important aspect of Networked Music Performance systems is the synchronization of audio streams, particularly in scenarios where devices do not share a common clock frequency. This lack of synchronization can lead to issues such as buffer underrun, which occurs when the receiver's buffer is filled at a lower bit rate than the application's playback rate. As a result, the audio stream is periodically interrupted to allow the buffer to refill. Conversely, buffer overrun may occur when the incoming bit rate exceeds the reading rate, causing the buffer to fill too quickly and resulting in data loss due to overflow.

2.2.3 Transparent Integration and Ease of Participation

Given the capacity of such environments to include users without technological expertise, it is essential that they incorporate mechanisms for transparent integration. This can be achieved through the adoption of discovery services and zero-configuration methods[1] (REPP, 2005; CÁCERAS; CHAFE, 2010).

Easy participation can be achieved through the local distribution of the software required for the performance, combined with support for standardized network protocols. This approach simplifies deployment and ensures broader accessibility (ROTTONDI et al., 2016).

2.2.4 Scalability

NMP also envisions a highly distributed and scalable communication infrastructure. To achieve these goals, it is possible to leverage cloud computing or other forms of network-based resource sharing, which enable dynamic allocation of processing power, storage, and connectivity across geographically dispersed participants.

2.2.5 Final Considerations on Networked Music Performance

Following the discussion on the origins, key requirements, and technologies associated with NMP, it becomes clear that this field encompasses a diverse set of relevant capabilities. Among these possibilities are the support for a potentially large number of participants, the transmission of both audio signals and control data to enable the synchronized coordination of distributed elements, the virtualization of spaces and musical instruments to accommodate performers' preferences and to transpose physical devices into the digital domain, and the development of cohesive environments through the integration of heterogeneous microsystems.

[1] An approach in which a system or application is designed to operate automatically, without requiring manual configuration.

These functionalities underscore the transformative potential of NMP for redefining musical collaboration in technologically mediated contexts (ROTTONDI et al., 2016).

2.3 Wireless Multimedia Sensor Networks (WMSNs)

With the advancement of electromechanical microsystems, the miniaturization of low-power circuits, small batteries, and the widespread use of Complementary Metal-Oxide Semiconductor (CMOS) technology have led to the rise of wireless multimedia sensor networks (WMSNs).

The architecture of a WMSN differs from that of traditional sensor networks, primarily due to its requirement to deliver multimedia content while ensuring a predefined level of Quality fo Service (QoS). WMSNs present some particularities, including resource constraints, as nodes are typically limited in terms of battery resource, memory, and processing power; high bandwidth demands, given that multimedia content, especially video, requires elevated data rates for real-time transmission; tight integration among multimedia data generation, processing, and delivery components; adaptive content resolution, allowing for adjustments based on the capabilities of the receiving device; and application-specific requirements, which necessitate diverse mechanisms for handling streaming data, flexible architectures to support heterogeneous services, and interoperability with both the Internet and other wireless communication technologies (AKYILDIZ; MELODIA; CHOWDHURY, 2007, 2008; ALMALKAWI et al., 2010).

The network's architecture is designed to be scalable and efficient, with resources distributed to support various services. It can be categorized into three reference models, each reflecting a distinct approach to data processing and transmission within the network (ALMALKAWI et al., 2010).

The first model is the single-layer flattened architecture, which is composed of homogeneous sensor nodes, that is, all nodes possess the same capabilities and perform identical functions, including multimedia data acquisition, processing, and multi-hop transmission to the sink node. This architecture is relatively simple to manage, and its fully distributed structure contributes to enhanced scalability and prolonged network lifespan.

The second reference model corresponds to the single-layer clustered architecture, which is defined by the deployment of heterogeneous sensor nodes structured into logical groups. Within each group, individual nodes transmit their data to a designated cluster head, typically equipped with superior processing capabilities, memory, and energy reserves. The cluster head is responsible for performing computationally demanding operations, such as data aggregation and preprocessing, prior to forwarding the processed information to the sink node. This architectural configuration enhances energy efficiency, facilitates localized decision-making, and

improves overall network scalability by minimizing redundant transmissions and distributing computational load across the network.

The third structure type is the multilayer architecture, which also relies on heterogeneous sensor nodes, but organizes them into hierarchical levels based on functionality and processing capability. The first layer is responsible for executing basic tasks, such as motion detection. The second layer performs more complex operations, including object recognition, while the third layer comprises nodes with the highest computational capacity, enabling advanced functionalities such as device tracking. This hierarchical distribution of responsibilities enhances system efficiency, promotes task specialization, and supports the scalable integration of services with varying degrees of complexity.

The devices that compose this network are structured around a set of components that provide capabilities for sensing, processing, communication, and, when applicable, actuation. These components include a detection unit, responsible for capturing environmental data through sensors and converting analog signals into digital form via analog-to-digital converters; a processing unit, which executes the embedded software responsible for coordinating local tasks and managing data; and a communication subsystem, which ensures connectivity with the network by handling the appropriate protocol stack. In addition, a coordination subsystem is often incorporated to synchronize the activities of different nodes within the network. A storage unit provides memory resources for temporary or persistent data retention. Optionally, devices may also include a mobility or actuation unit, which enables physical movement or the manipulation of external objects, depending on the application requirements (AKYILDIZ; MELODIA; CHOWDHURY, 2008).

2.4 Extended Reality (XR)

The development of immersive technologies has brought about changes in the forms of human interaction within digital environments. Central to this paradigm shift is the concept of Extended Reality (XR), an inclusive term that encompasses a spectrum of immersive experiences that blend the physical and digital realms. This includes Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) (ROSSI, 2022; RAUSCHNABEL et al., 2022).

Augmented Reality (AR) refers to the integration of digital elements into the physical environment, allowing computational information, such as images, text, and 3D objects, to be superimposed onto the user's perception of reality. To achieve this integration effectively, AR systems rely on spatial tracking and context-aware recognition techniques (AZUMA, 1997; FEINER; MACINTYRE; SELIGMANN, 1993; MILGRAM et al., 1995; RAUSCHNABEL et al., 2022). The term gained prominence in the 1990s, initially through applications in domains

such as the aerospace industry and medicine. Over time, AR technologies have become increasingly prevalent across a broader range of sectors, including retail, education, and industrial maintenance (FURHT, 2011).

The implementation of this technology relies on a range of specialized hardware components, including depth sensors, eye-tracking systems, and retinal displays, among others. These devices not only support the accurate alignment of virtual and physical elements but also enable novel modalities of human-computer interaction, such as hand/finger tracking, voice commands, eye gaze control, and brain-computer interfaces (BCIs). These interaction paradigms, coupled with the operational characteristics of AR systems, underscore the technology's potential to amplify human perception by seamlessly integrating digital content into the user's sensory experience (RAUSCHNABEL et al., 2022).

Virtual Reality (VR), in turn, is the most established area of XR. It involves creating a fully synthetic, immersive, and three-dimensional environment that completely replaces the user's sensory perceptions. This experience is facilitated by devices such as HMDs and, in some cases, haptic equipment and motion platforms, which enable users to actively navigate through the scene. Experiences in VR can vary along a continuum of telepresence, ranging from low (atomistic) to high (holistic) levels of immersion (ROSSI, 2022; RAUSCHNABEL et al., 2022).

Although originally developed for the gaming industry, VR has progressively broadened its application scope to encompass domains such as professional training, prototyping, marketing, and tourism (SHAHAB; GHAZALI; MOHTAR, 2021). Recent research has further demonstrated VR's potential in a variety of commercial and industrial contexts, including supermarkets environments (KRASONIKOLAKIS et al., 2018), the fashion industry (YAOYUNEYONG et al., 2018), manufacturing (BERG; VANCE, 2016), and healthcare (FERTLEMAN et al., 2018). Moreover, VR has gained increasing attention as a methodological tool for scientific research (HOLLÄNDER et al., 2019; STADLER et al., 2019).

The fundamental distinction between Augmented Reality and Virtual Reality lies in their respective relationships with the physical environment. While VR immerses the user in a fully synthetic digital space, AR enhances the real world by overlaying digital content onto the user's physical surroundings (FEINER; MACINTYRE; SELIGMANN, 1993; ALIPRANTIS; CARIDAKIS, 2019). Another relevant difference concerns device accessibility. AR applications can be deployed across a broader range of equipment, including smartphones, tablets, and optical see-through displays, whereas VR experiences typically require HMDs to ensure full immersion. Moreover, AR exhibits a greater potential for universal and continuous integration into everyday contexts, whereas VR remains largely restricted to temporary or situational use, given its immersive nature and the isolation it imposes from the physical environment (RAUSCHNABEL et al., 2022).

Another relevant distinction lies in the way each technology interprets and interacts with the physical environment. In Augmented Reality, the system must employ advanced tracking and spatial mapping techniques to accurately anchor virtual objects in the real world, ensuring their coherent integration with the user's surroundings. In contrast, in Virtual Reality environments the interaction with the physical space is limited to collision avoidance mechanisms, which are primarily designed to detect and signal the presence of real-world obstacles that may pose safety risks to the user (RAUSCHNABEL et al., 2022).

These differences also contribute to distinct experiential barriers. In AR, although the user remains aware of the physical environment, distractions or misinterpretations of virtual overlays can lead to misjudgment of real-world hazards, thereby compromising the quality and safety of the experience. In VR, users may experience disorientation, motion sickness, or anxiety related to physical collisions, due to the disconnect between visual input and bodily perception (RAUSCHNABEL et al., 2022).

The content presented in both areas can also vary significantly. An application that is well-received by AR users may not be suitable for VR users, and the reverse can also be true. Therefore, developing applications tailored for a specific domain necessitates a thorough understanding of the relevant use cases and the availability of devices for that particular service (RAUSCHNABEL et al., 2022; ZELLERBACH; ROBERTS, 2022).

The final paradigm within the XR spectrum is Mixed Reality (MR), which synthesizes characteristics of both Augmented Reality and Virtual Reality, enabling seamless integration between physical and digital elements within a shared environment. As with VR applications, MR experiences are typically mediated by dedicated hardware, namely holographic headsets, such as the Apple Vision Pro or Microsoft HoloLens. However, unlike traditional HMDs, these devices preserve the user's view of the physical surroundings, overlaying digital content that appears anchored within the real world. Similar to AR, users in MR can interact with virtual objects; however, in Mixed Reality, this interaction occurs through physical movements in real space (ROSSI, 2022).

In summary, Mixed Reality is distinguished by its capacity to construct a hybrid environment in which physical and virtual elements coexist and interact in real time. This bidirectional enhancement fosters a reciprocal relationship wherein each domain — real and virtual — augments the perceptual and functional qualities of the other. Consequently, MR establishes a dynamic interplay between the user, the physical environment, and digitally generated content, promoting a more integrated and immersive experience (BEKELE; CHAMPION, 2019; ZELLERBACH; ROBERTS, 2022).

This operational paradigm represents a more advanced stage in the continuum of integra-

tion between the physical and digital realms, as it enables virtual objects to interact responsively with both the real environment and the user. Such fusion demands the implementation of advanced sensing technologies, real-time spatial mapping, and artificial intelligence (AI) algorithms, which collectively support the dynamic coexistence and bidirectional interaction between real and virtual entities within a shared perceptual space.

In general, Extended Reality and its related technologies are defined by three fundamental characteristics: presence, immersion, and interaction (ROSSI, 2022). Presence refers to the subjective sensation of "being there", that is, the user's psychological experience of inhabiting a virtual environment distinct from their actual physical surroundings. For presence to emerge, a necessary condition is immersion, which encompasses the technical affordances, such as high-fidelity graphics, spatialized audio, and real-time responsiveness, capable of producing a coherent and perceptually convincing simulation.

Interactivity denotes the user's capacity to engage with, manipulate, and modify the virtual environment through embodied actions, such as gestures, head movements, or controller inputs. The synergy among these three dimensions is critical for fostering compelling and meaningful experiences within XR systems.

The concepts discussed throughout this section are synthesized in the reality–virtuality continuum, illustrated in Figure 5. At the leftmost end of the spectrum lie environments that are predominantly physical, yet enhanced with superimposed virtual elements (e.g. Pokémon Go or navigation systems with real-time map overlays). These configurations often support natural user interfaces (NUIs), enabled by devices like Microsoft Kinect or Leap Motion, which detect gestures and bodily movements. Such interfaces facilitate interaction modalities that closely resemble those found in the physical world, reinforcing the sense of realism and intuitiveness within augmented environments.

At the opposite side of the spectrum are applications that are predominantly digital, yet incorporate selected elements from the physical world, such as VR systems that integrate real-time camera feeds or employ motion capture sensors to animate virtual avatars based on users' bodily movements. In these contexts, interaction is typically mediated by conventional input devices, such as joysticks, controllers, or button-based interfaces, which, although less natural, offer high precision and responsiveness within fully immersive virtual environments (JOHNSON, 2019).

At the top of the figure, the term "Mixed Reality" is presented as an overarching category that encompasses both Augmented Reality and Virtual Reality. This representation reinforces the understanding that MR does not refer to a single, fixed technology, but rather to a conceptual terminology that spans the intermediate region of the continuum. Within this

range, physical and virtual elements coexist and interact dynamically, highlighting the fluid and hybrid nature of experiences that transcend the boundaries of purely real or purely virtual environments.

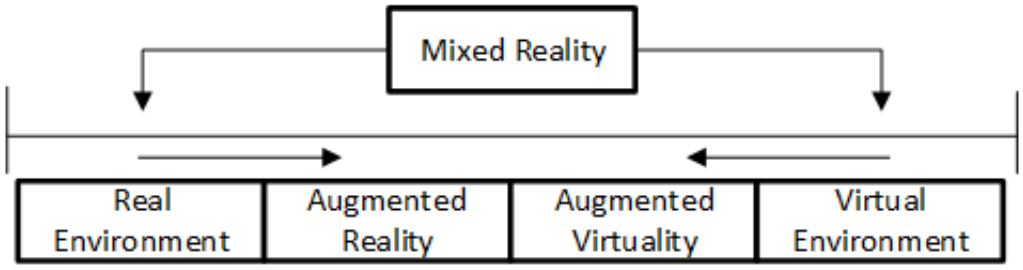


Figure 5: Virtual-reality continuum representing the distinctions and correlations between real and virtual environments (MILGRAM et al., 1995).

2.4.1 Musical Metaverse (MM)

Recent advancements in technology and the decreasing costs of the equipment necessary for developing and using XR applications culminated in the creation of the Metaverse concept. The Metaverse refers to a collection of collaborative environments facilitated by technology that dissolves the boundaries between the physical and virtual worlds. This enables users, even when geographically separated, to interact in shared and interoperable environments using immersive media (RAJ, 2021; O'DAIR; BEAVEN, 2017; LEVSTEK et al., 2021; TURCHET, 2023).

The Metaverse offers new opportunities across various fields, including education (MYSTAKIDIS, 2022) and healthcare (ONDERDIJK; ACAR; VAN DYCK, 2021). Within this context, the Musical Metaverse (MM) (TURCHET, 2023) emerges as a specialized subset dedicated to the musical domain. This research area can be characterized as a multidisciplinary ecosystem arising from the convergence between Musical XR (BOEM; TOMASETTI, et al., 2024) and IoMusT (TURCHET; FISCHIONE, et al., 2018). It enables multisensory and interactive musical experiences, facilitating real-time communication not only among musicians and audiences but also between these individuals and responsive virtual environments and smart musical objects.

As in XR environments, the Musical Metaverse is structured around interactivity, presence, and immersion, along with low-latency streaming and audio quality. Regarding multiuser interactions, it can occur either in the same physical location or remotely, facilitated by the use of HMDs and Internet connectivity. Moreover, connected users can play electric, acoustic, or entirely virtual instruments from anywhere on the network.

A functional MM archetype consists of three layers: the physical layer, the link layer,

and the virtual layer (TURCHET, 2023). The physical layer is responsible for collecting data generated in the real environment by musical things and transmitting it to the virtual layer. This layer can also receive feedback from its virtual counterpart.

The stakeholders involved in this stratum are users, virtual music service providers, and physical music service providers. The user group comprises various profiles within the musical ecosystem, including composers, music students and educators, audio engineers, and performers, as well as audience members. These individuals engage with virtual environments through HMDs or XR-based musical instruments, enabling them to interact with digital content in real time. Typical user activities include musical performances, instructional sessions, and music production tasks, all of which may occur in collaborative or individual settings.

Virtual music service providers are responsible for creating content for virtual worlds, including music, equipment, applications, objects, and scenarios. This category also includes labels, publishers, and copyright institutions.

Physical music service providers are responsible for the development, maintenance, and management of the physical infrastructure that underpins MM applications. Their roles encompass the operation of communication networks, computational resources, and logistics systems that support both the functioning of virtual platforms and the delivery of physical goods transacted within these environments. This category includes electronic device manufacturers, network service providers, concert venues, distributors, and other actors involved in the music industry's physical supply chain.

The data link layer serves as an interface between the physical and virtual layers. It enables bidirectional communication, allowing the physical layer to transmit information to the virtual layer, while also permitting the virtual environment to send feedback or commands back to the physical layer. This tier is composed of two interconnected sublayers: infrastructure and the musical metaverse engine. The first segment handles networking aspects necessary to support synchronous musical interactions, including low latency, quality of experience, and reliability, as well as the ability to manage large volumes of data. It is also responsible for data integration, storage, and computation.

The second sublayer focuses on the system's additional functionalities, which include musical objects like virtual reality musical instruments (VRMIs) (SERAFIN; ERKUT; KOJS; NILSSON, et al., 2016) and digital twins. It also offers services that ensure context sensitivity and management, such as a blockchain network for virtual negotiation.

Finally, the virtual layer is responsible for delivering an immersive or augmented musical experience. This module is composed of avatars, which function as the virtual representations of users within the environment; virtual musical environments, which constitute the digital spaces

in which musical activities occur; and virtual musical goods and services, which encompass both the digital content produced by service providers and the activities executed within the platform, such as live concerts, collaborative composition, or music production.

Given these functionalities, MM has the potential to redefine the form of musical activities in immersive and technology-mediated environments. This paradigm shift enables the emergence of novel systems to support a wide range of activities, including musical composition, education, performance, entertainment, and sound engineering. In parallel, it fosters the development of specialized software and hardware tools tailored to immersive musical experiences, along with the formulation of design methodologies and analytical frameworks capable of guiding the creation, evaluation, and refinement of such system (TURCHET, 2023; BOEM; TOMASETTI, et al., 2024).

2.5 Interactive Art

Musical performances have historically been guided by the European tradition (also known as the Western standard). This model involved the composer, responsible for conceiving and notating the musical work; the performer, tasked with interpreting and executing the piece; and the audience, whose role was largely passive and observational, limited to the reception and aesthetic appreciation of the performance (ARAÚJO et al., 2019; TURCHET; FISCHIONE, et al., 2018).

A departure from this paradigm emerged in the mid-1960s, fueled by artistic movements led by Allan Kaprow, John Cage, and the Fluxus and Guntai collectives. These initiatives introduced interactive dynamics into the artistic process, demanding active engagement from the audience members. This represented a break from the traditional view of the performance as a hermetic system, giving rise to a new perspective in which the spectator assumed the role of co-author. As a result, the creative process became open to multiple interpretations and unforeseen outcomes. These *avant-garde* practices would later be encompassed by the concept of Interactive Art (OLIVEIRA, 2015; CARDOSO, 2019; EDMONDS, 2010).

Such interactions can manifest in two distinct modes: immediate and reflexive. The former highlights the emergent qualities of participation, wherein audience input can directly and substantially influence the unfolding of the work. The latter emphasizes individual perception, framing participation through the lens of personal experience and interpretation. Although conceptually distinct, these modalities are intrinsically interconnected, as aesthetic experience often involves an interactive interplay between collective engagement and subjective reflection (CERRATTO PARGMAN; ROSSITTO; BARKHUUS, 2014).

Participation in an artwork can also be understood in terms of degrees of openness, which reflect varying levels of audience involvement. First-degree openness refers to the plurality of meanings that a work may evoke, allowing for diverse interpretations without altering its structure. In contrast, second-degree openness involves effective interactivity, in which physical interventions, either within installations or in direct interaction with the artists, provoke structural and thematic transformations, leading to the emergence of entirely new material (CARDOSO, 2019).

The technological advances of the past two decades, particularly in multimedia systems, the web, and the Internet, have led to the emergence of a wide array of tools and devices that support the creative process. Notably, these developments have expanded the use of technical interfaces not only as mediating tools but as constitutive elements of the artwork itself. This evolution has also intensified the exploration of interactivity, given that digital materials exhibit a high degree of flexibility, allowing them to adapt dynamically to the interactions established between artists, audiences, and systems.

As a result, new forms of artistic expression have emerged, including literary hypertexts, immersive photography, and interactive video, which collectively contributed to the development of a third degree of openness in aesthetic interaction. In this configuration, machine intervention becomes a decisive agent in the creative process (CARDOSO, 2019).

It is important to emphasize that digital devices should not be regarded merely as tools for information representation, but rather as translators of artistic intent, capable of conveying conceptual and expressive dimensions embedded in the work. Similarly, these devices should not be constrained to predefined actions or rigid functionalities; instead, they ought to remain open to modification, adaptation, and experimentation (PARIKKA, 2010; CARDOSO, 2019).

Communication in this scenarios can occur in either synchronous or asynchronous modes, enabling interactions not only between humans and machines, but also exclusively among machines. This field is not defined by specific technological implementations, but is instead grounded in empirical methodologies that prioritize observation, experimentation, and iterative refinement as means to understand and shape interactive processes (TURCHET; ROTTONDI, 2022).

In the context of interactive performance systems, there is no established architectural standard, as the design and implementation of such systems are shaped by a variety of factors, including the aesthetic intentions and technical preferences of individual developers and artists (EDMONDS; TURNER; CANDY, 2004). Despite this variability, certain recurring principles can be identified, such as accessibility, ease of integration and usability, interoperability across heterogeneous platforms and devices, and the capacity to manage latency and

data transmission delays, besides the creation of micro-systems (NARDIM, 2009; FORNY, 2006).

This chapter presented a comprehensive theoretical foundation for the areas that provide the conceptual pillars supporting the structure of the Io3MT domain and its reference model. It began by discussing the theoretical definitions of multimedia, mulsemedia, and music — concepts that, although interrelated and often abstracted in similar ways within computational systems, exhibit crucial distinctions that significantly influence their applications within the Io3MT domain. Subsequently, it examined related research areas and their key conceptual and technological entities, which contribute to defining and delineating the scope of Io3MT, including Networked Music Performance (NMP), Mobile Music, Wireless Multimedia Sensor Networks (WMSNs), Extended Reality (XR), and Interactive Art. The following chapter provides an overview of related work, encompassing artistic applications that explore musical and multimedia concepts over networked environments, as well as a wide range of mulsemedia systems, comparing these proposals with those presented throughout this thesis.

3 Related Work

This chapter presents a comprehensive review of the literature relevant to the scope and objectives of this research. To identify patterns and common principles underlying the practices that inform the design of the environment proposed in this thesis, an extensive investigation was carried out across nine distinct thematic areas. The first area focuses on analyzing the ecosystems encompassed by the Internet of Musical Things (IoMusT).

This investigation emphasizes the artistic aspirations articulated within these systems, the strategies proposed for their realization, the operational behavior of the devices involved, and their spatial distribution across the environment. These conceptual foundations, along with a comparative evaluation of the principal models, are thoroughly discussed in Section 3.1.

In a similar vein, Section 3.2 explores the evolution of Digital Musical Instruments (DMIs) and Smart Musical Instruments (SMIs), which have progressively incorporated high-resolution sensing, embedded processing, and multimodal feedback into their designs. This examination aims to identify how these instruments embody principles of interactivity, networked performance, and multisensory integration — key aspects aligned with the requirements of Io3MT environments.

Section 3.3 investigates frameworks that formalize the use of technology in musical practice. This examination highlights design principles, methodological approaches, and conceptual models that have guided the development of expressive, accessible, and interconnected musical systems.

Section 3.4 addresses frameworks and systems developed for creating musical experiences in XR environments. The analysis focuses on how immersive technologies have been employed to enhance presence, embodiment, and interactivity in music performance and composition, as well as on the design principles that underpin such applications.

Section 3.5 inquires into artistic applications that integrate haptic feedback as a core expressive element. This examination considers how tactile interaction contributes to sensory immersion, emotional engagement, and accessibility within artistic contexts.

Section 3.6 explores air drumming systems, encompassing both commercial and academic

solutions. These systems are analyzed with respect to their sensing strategies, interaction models, and feedback modalities, which collectively illustrate how free-space gestural performance can be transformed into expressive percussive interaction.

Section 3.7 delves into the environments and applications dedicated to the manipulation of multimedia content for the creation of artistic and musical works. Particular attention is given to the role of immersive media in enhancing these experiences.

Section 3.8 examines the use of immersive media in artistic creation, drawing primarily on VR and 3D environments while also incorporating volumetric video, wearable devices, and other sensing technologies to extend artistic capabilities and explore multiple channels of interaction.

Lastly, the requirements for implementing scenarios based on the principles of mulsemedia are examined. Section 3.9 offers an overview of these conditions, highlighting relevant studies and systems that contribute to their fulfillment.

3.1 Internet of Musical Things Environments

This section presents a review of the main studies related to the IoMusT, with the aim of describing the underlying architectures, technologies, and methodological approaches employed across the scenarios that define this subfield.

The proposal presented by (TURCHET; VIOLA, et al., 2018) seeks to establish an efficient and asynchronous semantic architecture through the use of message-oriented middleware based on a publish/subscribe communication paradigm. This approach facilitates loosely coupled and time-sensitive interactions, thereby promoting the continuous and reliable exchange of information within the system.

The network implementation adopts a client–server architecture and employs directed, labeled graphs for the semantic encoding of information. To retrieve relevant data, the SPARQL language is utilized, enabling efficient and flexible access to structured content. This strategy supports communication among devices sharing common attributes, thereby reducing both data transmission overhead and computational processing demands, ultimately enhancing the overall system performance.

The proposed environment comprises five prototype musical artifacts designed to emphasize wireless connectivity via IEEE 802.11ac Wi-Fi. Each device integrates onboard processing capabilities through the Bela prototyping platform[1], and utilizes the Pure Data programming language as the core audio engine. Data exchange within the system is facilitated by SPARQL queries and Python scripts, while musical information is transmitted using the Open Sound

[1] https://bela.io/

Control (OSC) protocol.

Those prototypes are organized into three functional categories: producers, consumers, and aggregators. The producer class is responsible for updating the system's knowledge base by inserting, removing, or modifying information within the graph structure. At least one producer is required for system operation, although there is no upper limit on their number. Examples of producers include smart musical instruments, wearable technologies, or mobile devices executing music-related applications capable of publishing audio resources.

Consumers, in contrast, serve exclusively as recipients of information from the database. These entities include digital musical instruments that adapt their sound generation parameters in response to contextual information, as well as auxiliary systems for musical practice and performance, such as lighting setups, display screens, fog machines, and HMDs.

The third category, aggregators, encompasses entities that function simultaneously as producers and consumers. Although not essential for the system's operation, their inclusion enables more dynamic and bidirectional information flows. Aggregators may take the form of laptops, wearable devices, or any of the aforementioned performance support technologies capable of both retrieving and contributing data to the networked environment.

The model proposed by Turchet and Barthet (TURCHET; BARTHET, 2018) aims to facilitate interaction not only between artists and their musical instruments but also among musicians and audience members, enabling the generation of musical accompaniments through a process of collective construction. To this end, the designed environment is structured into two primary stages.

The first stage is dedicated to enhancing the musician's expressive capabilities by fostering a more dynamic and responsive interaction between the performer and a smart musical instrument — the Smart Mandolin (TURCHET, 2018a). That instrument integrates acoustic and digital components, including the Bela prototyping board, various embedded sensors, and wireless communication interfaces. The architecture leverages a Wi-Fi network and employs OSC messages encapsulated within User Datagram Protocol (UDP) packets to facilitate the continuous transmission of musical data.

The audio engine, implemented using Pure Data, supports real-time audio processing and enables a wide range of interactive sound effects and behaviors that respond dynamically to the instrument's physical gestures. Additionally, a mobile application developed using the TouchOSC platform[2] allows performers to trigger pre-recorded tracks and transmit control messages to the instrument. The system also incorporates a Python-based software module responsible for retrieving audio files and transmitting them to the Bela board, thereby ensuring

[2] https://hexler.net/touchosc

efficient and seamless communication within the system's architecture.

The second stage of the model focuses on facilitating interaction between the performer and the audience by employing collaborative strategies using smartphones. The objective is to actively engage audience members in the musical experience, incorporating their participation as a form of feedback that can influence and shape the unfolding of the performance.

The third proposal within the context of IoMusT scenarios functions simultaneously as a proof of concept for a communication protocol known as Sunflower (VIEIRA; SCHIAVONI; SAADE, 2022), which defines a set of standardized messages to support interoperability among devices engaged in artistic and musical practices.

The implementation adopts the pipes-and-filters architectural pattern, wherein data processing is modularized into discrete components, referred to as filters, while the communication between these units is mediated by unidirectional conduits known as pipes. A salient characteristic of this architecture is that data exchange occurs exclusively through the input and output interfaces of each module, thereby obviating the need for devices to possess prior knowledge of adjacent entities. As a result, the system supports a high degree of heterogeneity among components, fostering flexible coexistence and extensibility.

The overall structure is stratified into distinct layers, organized according to the types of musical things, data formats, and protocols associated with each functional domain. In this sense, the digital audio layer is responsible for the generation of auditory content using Pure Data. The graphics layer manages visual outputs, primarily developed using the Processing language. The control layer enables remote manipulation of communication parameters, such as volume, frequency, and beats per minute (BPM), and governs the dynamic configuration of participating devices. Finally, the management layer provides system administrators with tools to monitor, configure, or remove connected devices. This layer is implemented through Python-based scripts.

With respect to network communication, a hybrid transmission model is employed, combining Wi-Fi connectivity compliant with the IEEE 802.11n standard and Ethernet for wired data exchange. UDP is used for the general transmission of non-musical data across the network, while the OSC protocol is specifically designated for the exchange of musical and performance-related information.

The architecture proposed by (CENTENARO; CASARI; TURCHET, 2020) is designed to provide reliable, low-latency communication for mobile devices within IoMusT environments. It establishes a set of architectural guidelines, services, and requirements aimed at supporting efficient data exchange in scenarios characterized by high mobility and stringent timing constraints. To achieve this, the model integrates a Next-Generation Radio Access Network (NG-

RAN) with a 5G core infrastructure, thereby enabling the effective transmission of digital audio traffic between musical devices while concurrently managing service provisioning.

The musical things in this model consist of specialized hardware units for audio input, output, and processing. These units are equipped with the Elk Audio operating system and a 5G communication module, allowing for high-performance, low-latency interactions in real time.

Cloud computing functions as the hosting infrastructure for applications and services. These services must either exhibit tolerance to latency or be deployed at the network edge, particularly when responsible for executing time-sensitive operations that critically influence system responsiveness and user experience.

Although the study does not explicitly address details regarding audio file formats, musical data representations, or communication protocols optimized for mobile transmission, the proposed model demonstrates several key advantages. These include native support for heterogeneous data traffic and the capacity to accommodate devices with varying QoS requirements, thus offering a flexible and scalable foundation for real-time musical collaboration in 5G-enabled environments.

The study presented by (MERENDINO; RODÀ; MASU, 2024) explores the potential of IoMusT systems to support artists with specific health conditions in performance settings. The research is motivated by the case of an opera singer who experienced a carotid aneurysm, highlighting the need for technological solutions capable of monitoring physiological parameters, particularly heart rate, in real time during musical performances. The objective is to enable immediate self-regulation interventions, such as deep breathing and meditative practices, thereby enhancing the artist's well-being without compromising artistic expression.

The proposed system comprises a wearable IoT device built on the ESP32 platform, incorporating a photoplethysmographic sensor for heart rate monitoring, vibration motors for constant haptic feedback, and light-emitting diodes (LEDs) for visual signaling. The haptic feedback is designed to be non-intrusive, avoiding any sensory interference with musical performance, and is activated whenever the heart rate exceeds preconfigured safety thresholds. The LEDs indicate system status, including Bluetooth connectivity and power state.

Additionally, the system features a software module capable of real-time processing of the sounds produced by the performer during self-care moments, aesthetically integrating them into the musical context. This functionality allows the singer to alternate between highly demanding operatic vocal techniques and less strenuous vocal expressions, thereby enabling periods of rest without interrupting the performance.

From a computational perspective, the system adopts an edge computing architecture.

Physiological data are transmitted via Bluetooth Low Energy to client-side software running on a laptop. Real-time audio processing is conducted using a Pure Data patch, which incorporates effects such as pitch shifting, sideband modulation, reverb, and comb filtering to enable expressive manipulation of the performer's voice. Communication between the wearable device and the audio engine is facilitated through Musical Instrument Digital Interface Continuous Controller (MIDI CC) messages, allowing effect parameters to be adjusted via potentiometers embedded in the wearable collar. The firmware for the ESP32 microcontroller was developed in C/C++ using the Arduino Integrated Development Environment (IDE), while Python scripts were employed for the acquisition and analysis of physiological data.

The study conducted by (TURCHET; CASARI, 2024) presents an empirical investigation into the feasibility of employing low Earth orbit (LEO) satellite connection for networked music applications in rural regions. The work involves the experimental evaluation of two distinct communication scenarios to assess performance under varying infrastructure conditions.

In the first case, both communication nodes, each consisting of dedicated Elk LIVE Box devices, were geographically separated and connected exclusively through satellite links. In the second scenario, a hybrid configuration was adopted, in which one node remained connected to the satellite network, whereas the other relied on a conventional terrestrial wired connection.

In both experiments, continuous audio transmissions were simulated through locally generated sound sources, specifically drum and electric bass signals. These tracks were produced in Pure Data. This setup enabled the assessment of key performance indicators, including end-to-end latency, packet error rate, and the frequency of consecutive packet losses. All transmissions were conducted using UDP.

BCHJam (ROMANI et al., 2024) is an integrated ecosystem designed to support real-time collaborative musical performances by combining brain-computer music interfaces (BCMIs), digital musical instruments, and mixed reality headsets. Its primary objective is to investigate novel modes of interaction between performers and audiences by leveraging auditory and visual stimuli modulated through brainwave activity.

The BCHJam architecture integrates both active neural signals, such as event-related potentials, and passive signals, such as alpha and beta frequency bands, which are commonly associated with emotional states including arousal and relaxation. Those signals are captured in real time using the Unicorn Hybrid Black, a commercial electroencephalogram (EEG) headset. Following neural feature extraction, the physiological data is transmitted using the OSC protocol to Reaper[3], a digital audio workstation (DAW) that functions as the system's audio engine. Within Reaper, sound effects are dynamically modulated according to the user's

[3] https://www.reaper.fm/

neurophysiological state.

Concurrently, EEG data are transmitted via Bluetooth to mixed reality applications executed on the Meta Quest 3 headset, which enhance the audience's sensory experience by overlaying immersive visual content. The virtual environment is implemented in Unity with C#, incorporating a GUI that allows customization of the mappings between neural signals and audiovisual parameters. The interface also provides real-time feedback on EEG signal quality and overall system performance. Communication among modules within the ecosystem is established through Wi-Fi connectivity.

Table 1 provides a summary of the discussion, presenting the primary programming languages, communication protocols, and core functionalities associated with each of the IoMusT environments described.

Attribute	**Turchet; Viola et al. Model (2018)**	**Turchet & Barthet Model (2018)**	**Sunflower Environment (2022)**	**Centenaro; Casari & Turchet Model (2020)**	**Below 58 BPM (2024)**	**Turchet & Casari Model (2024)**	**BCHJam Ecosystem (2024)**
Network	Wi-Fi	Wi-Fi	Hybrid (Wi-Fi and Ethernet)	5G	Wi-Fi	Hybrid (LEO and Ethernet)	Wi-Fi
Protocols	HTTP, UDP, OSC	UDP, OSC	UDP, OSC, MIDI	Not specified	BLE, MIDI CC	UDP, IP	OSC, Bluetooth
Audio Engine	Pure Data	Pure Data	Pure Data	Elk Audio OS	Pure Data	Pure Data; Elk Audio OS	Reaper
Audio Format	WAV	WAV, MP3	PCM	MP3	PCM	PCM	Not specified
Multimedia Elements	N/A	N/A	Video and Animation	N/A	N/A	N/A	3D Elements
Multisensory Feedback	N/A	N/A	N/A	N/A	Haptic and Visual Feedback	N/A	Visual Feedback
Programming Languages	SPARQL, Python, Pure Data	Python, Pure Data	Python, Pure Data, Processing	Not specified	Pure Data, Python, C++	Pure Data	C#, Python

Table 1: Technical attributes of the Internet of Musical Things (IoMusT) environments analyzed.

3.2 Digital and Smart Musical Instruments: Toward Networked and Multisensory Music Systems

The Smart Cajón (TURCHET; MCPHERSON; BARTHET, et al., 2018; TURCHET; MCPHERSON; BARTHET, 2018) is an SMI-based percussion instrument that integrates multimodal sensing, haptic feedback, and networked connectivity to enhance expressive performance. Its

design incorporates two piezoelectric sensors embedded in the front plate for strike detection, a force-sensitive resistor (FSR) located on the upper surface for expressive modulation, four bilaterally arranged foam-based vibrotactile motors for haptic feedback, and a low-latency microcontroller board (Bela BeagleBone Black), complemented by an integrated loudspeaker. Audio synthesis and processing are managed by Pure Data, which performs real-time audio effects (e.g., delay, reverb), gesture-to-sound mapping, including both triggered and continuous synthesis via the FSR, and pulse-width modulation (PWM) control of vibrotactile profiles. The transmission of control data is carried out using OSC protocol, encapsulated in UDP packets and optimized to reduce latency and jitter. Networking relies primarily on Wi-Fi IEEE 802.11ac, with additional support for 4G connectivity.

In contrast to the previously described prototype, the Sensus Smart Guitar (TURCHET; MCPHERSON; FISCHIONE, et al., 2016; TURCHET; BENINCASO; FISCHIONE, 2017) represents an industrial-origin SMI, distinguished by the integration of multimodal sensors within the instrument's body and actuators coupled to the soundboard that enable acoustic-electronic signal amplification, effectively transforming the soundboard itself into a transducer. The equipment incorporates an onboard processing unit capable of performing audio synthesis, signal processing, and real-time effects without reliance on an external laptop. This configuration is supported by a real-time, audio-oriented software stack, which provides modular effects, OSC/MIDI interfaces, and APIs for integration with mobile applications and cloud-based services. Furthermore, the system features a bidirectional connectivity model, allowing for point-to-point communication between instruments and audience devices, as well as remote interaction topologies encompassing VR/AR environments and interoperability with a range of DAWs. Connectivity options include Bluetooth Low Energy, Wi-Fi, and 4G, supporting both local and Internet-based communication.

Regarding DMI models, the Multimedia Digital Instrument (MuDI) (PATRÍCIO, 2012) is conceived to facilitate real-time musical composition and performance directly synchronized with cinematic content. Its technological configuration combines a multitouch device (iPod) with a laptop functioning as the audio engine, interconnected via Wi-Fi and communicating through OSC messages transmitted over UDP. The system affords both continuous and discrete gesture input through the handheld interface, while the laptop executes a Pure Data application responsible for audio synthesis and processing. In addition, Pure Data provides a graphical interface that enables real-time monitoring of the system, facilitates performance recording, and generates a synchronized video score at the conclusion of the session.

Illusio (BARBOSA et al., 2013) introduces a system that integrates a multi-touch surface with a physical pedal, aiming to expand expressive possibilities, support flexible interaction, and foster audience engagement through a visual grammar based on drawing and hierarchical live

looping (HLL). The performer interacts directly with the surface by sketching graphic elements, a functionality implemented using the Processing programming language. These visual traces are then mapped to loops captured in real time and subsequently manipulated via the pedal interface. The underlying concept of HLL enables operations applied to a given musical node to be propagated across its descendant nodes, thereby facilitating cohesive structural transformations within layered sound trees. The audio looper was developed in openFrameworks (C++) and is designed to operate locally, without reliance on network connectivity.

The TANC (O'NEILL; ORTIZ, 2024) interface seeks to reframe the discussion on design methodologies by advancing a DMI developed and reshaped through direct interaction with the community. Its design process emphasizes ensemble improvisation, ecological deployment, and the principle of continuous system iteration. From a technological perspective, the instrument integrates a joystick, three FSR pads, and an Arduino Nano, with the input data mapped onto six distinct granular synthesis processes implemented in Max/MSP. Communication between the hardware components and the synthesis environment is facilitated via Bluetooth, ensuring wireless data exchange and portability.

3.3 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, 2017). 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 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 (COOK, 2009). 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 contemporary energy storage devices 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, Ge Wang (WANG, 2014) 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.

Similarly, the Soundworks framework (MATUSZEWSKI, 2020) identifies patterns that aid in the development of distributed musical applications employing multimedia concepts and web technologies. Key elements of this framework comprise a star-topology architecture, designed to ensure modularity and extensibility, a communication protocol capable of enabling any node to generate new states from a declared one, and maintain synchronization with the server.

3.4 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; ERKUT; KOJS; NILSSON, et al., 2016), which focuses on presenting and elucidating design principles, evaluation methods, case studies, and considering future challenges concerning 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 (VALBOM; MARCOS, 2005) 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; HAMILTON; ÇAMCI, 2021), 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 field as well as a reference framework. In addition, (CICILIANI, 2020) 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 (BOEM; TURCHET, 2023). This project consists of two prototypes of playgrounds using Web Audio technologies to create and test immersive sound experiences conveniently executed in web browsers integrated into commercially available standalone HMD.

(DZIWIS; COLER; PORSCHMANN, 2023), in turn, 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.

3.5 Haptic Elements Applied to the Arts

Initial evidence suggests that haptic devices may influence and augment art experiences. Building on this premise, a growing number of studies have investigated the role of haptic feedback in artistic creation and performance (VENKATESAN; WANG, 2023). One illustrative example is SensArt (FAUSTINO et al., 2017), a multimodal device that integrates music, vibrations, and temperature variations to convey the emotional qualities of visual artworks, with the goal of strengthening the audience's emotional connection to the art.

In a similar fashion, (MCDOWELL; FURLONG, 2018) propose the concept of haptic listening by investigating how vibrational feedback captured directly from a classical guitar can be reproduced to listeners through body actuators. This research is based on the theory of embodied cognition, suggesting that simulating instrumental vibrations intensifies musical mediation and promotes a more immersive experience.

Some research also suggests that haptic devices can increase the enjoyment of the music-listening experience. For example, (GIROUX et al., 2019) reported that participants who listened to music while seated in chairs delivering vibrotactile feedback synchronized with the audio exhibited heightened psychological arousal and an increased appreciation of the experience. Additionally, (HODGES, 2018) demonstrated that the use of a haptic device while

playing the video game Dance Dance Revolution not only enhanced participants' enjoyment but also improved their performance in the game.

Other studies, such as the one conducted by (TURCHET; ROSAIA, et al., 2025), suggest that combining haptic and artistic elements can enhance healthcare applications. In this particular case, the authors explore whether exposure to music synchronized with vibrotactile stimulation can improve the audiometric performance (both tonal and speech recognition) of participants with cochlear implants. The research also examines the extent to which these multisensory stimuli influence emotional responses during music listening, affect subjective experience, and produce immediate effects on standardized audiometric assessments.

Likewise, (NANAYAKKARA et al., 2013) designed and evaluated a system consisting of a haptic chair and a computer display to enrich the musical experience of people with hearing impairments. The chair transmits vibrations derived from the sound signal directly to the body. At the same time, the display provides visual effects synchronized with musical characteristics, highlighting the potential of haptic technologies for musical accessibility.

3.6 Air Drumming Applications

Air drumming refers to the simulation of percussive gestures in mid-air, without the need for a physical drum kit. Various computational approaches have been proposed to translate these gestures into real-time drum sounds, enabling expressive musical interaction. Such applications are gaining prominence both in industry and academia (SURASINGHE; HERATH; THANIKASALAM, 2023).

One of the most notable examples of the first category is Aerodrums[4], which utilizes drumsticks equipped with reflective markers that are captured by a high-speed camera. The user's movements are processed by proprietary software that synthesizes sounds accordingly and provides an interactive graphic representation of the drum kit, visually indicating the specific impact points needed to trigger each component. Additionally, reflective markers are attached to the user's feet, enabling the simulation of bass drum and hi-hat pedals. The system was later expanded to incorporate a 3D/VR environment while maintaining the same set of input devices.

Another commercial solution is Aeroband PocketDrum 2 Pro[5], a portable air drumming system that offers three degrees of freedom for tracking drumstick movements. The device also supports up to 128 levels of velocity sensitivity. In addition, it includes eight distinct drum kit presets and communicates via Bluetooth with a mobile application responsible for

[4] https://aerodrums.com/
[5] https://www.aeroband.net/products/pocketdrum2-plus

real-time sound synthesis. Furthermore, it supports MIDI transmission, allowing integration with different DAWs. Haptic feedback embedded in the drumsticks provides immediate tactile responses to user actions, reinforcing the sense of interaction. The system also includes two physical pedals, enabling control of bass drum and hi-hat elements.

An additional example is Paradiddle[6], an application developed specifically for VR environments, which uses headset controllers as input devices. Thus, the bass drum and hi-hat are activated through the controller triggers. Given the support for MIDI integration, electronic drum pedals can be connected to the system to perform these same functions. The platform further supports customization of the drum kit and its acoustic properties, offers both practice and performance modes, and provides functionalities for importing musical scores as well as recording and sharing sessions.

In academic research, several air drumming systems have been proposed, many of which leverage computer vision techniques. One example is the Anywhere Anytime Drumming (A2D) system, which employs computer vision and deep learning algorithms to track drumstick movements without requiring additional hardware or power sources (YADID et al., 2023). Other prototypes based on similar approaches include Augmented Virtual Drums (ZAVERI et al., 2022), and Air Drums (TOLENTINO; UY; NAVAL, 2019). On the other hand, (YASEEN; CHAKRABORTY; TIMONEY, 2022) propose a system that combines computer vision and IoMusT concepts to recognize percussive gestures from hand, arm, and body movements, enabling interactive drumming in networked environments.

Among systems that integrate percussion and haptic feedback, DigiDrum (WILLEMSEN; HORVATH; NASCIMBEN, 2020) employs a physical membrane within a VR setting, serving simultaneously as an interaction interface and a source of tactile feedback, to investigate how variations in stiffness affect musical expressiveness.

3.7 Multimedia Services Applied to Artistic Creation

The increasing convergence of art, technology, and interactivity has catalyzed the development of systems and installations in which multimedia services function as foundational elements of aesthetic experience. At the intersection of art and technology, a variety of artistic initiatives has been developed that integrate sound, visual components, movement, and sensory feedback into hybrid, interactive, and immersive environments, thereby expanding the scope of conventional performance and installation practices. These initiatives draw upon the technical affordances of digital media, including VR, wireless communication, gestural interfaces, and audiovisual rendering engines, while also introducing new forms of audience engagement

[6] https://paradiddleapp.com/

that foster experiences characterized by multisensory integration, participation, and subjective construction. This section presents a selection of works that exemplify these interdisciplinary approaches and critically engage with their underlying concepts.

"*O Chaos das 5*[7]" performance (ARAÚJO et al., 2019) was conceived with the objective of reviving *avant-garde* performance practices of the 20th century, particularly the Happenings of Allan Kaprow and the experimental works of John Cage, by fostering collaborative and immersive experiences between artists and audience members. To achieve this, the performance integrates sonic, visual, and gestural elements within an environment that eliminates physical boundaries between performers and spectators.

Audience members are invited to engage directly with the performance through digital musical instruments accessible via web interfaces on their smartphones. The interactive system leverages a combination of HTML5 technologies, the Web Audio API, and a client–server architecture based on Java and Jakarta Server Pages (JSP). Connectivity is facilitated through Wi-Fi access points and a local DNS infrastructure, enabling transparent communication between users and the system.

During the performance, three distinct sets of instruments were provided to the participants, each associated with a specific narrative segment of the show. These instruments incorporated diverse modes of interaction, including accelerometer-based sound synthesis, playback of prerecorded audio tracks, and button-triggered sound events. Each modality was designed to elicit particular forms of audience engagement with the auditory dimension of the work.

The underlying technical infrastructure supported server-side logic for centralized control, enabling the synchronization of interactive events and providing a scalable and device-agnostic platform accessible to users with diverse technological skills. This configuration ensured robust system performance while enabling meaningful, real-time participation in the aesthetic experience.

The study conducted by (BIN et al., 2023) examines the role of digital media in the field of environmental art design[8], proposing both a conceptual and technical reformulation of the creative process in response to the technological transformations of the digital age.

The research adopts a combined analytical and propositional methodology, integrating theoretical perspectives from communication studies with practical strategies for the application of

[7]The title of the performance operates as a linguistic pun in Portuguese, drawing on the phonetic similarity between *caos* (chaos) and *chá* (tea). This wordplay introduces a layered ambiguity that juxtaposes the notion of disorder with the ritual of 5 o'clock tea, evoking symbolic references to Alice in Wonderland and its surreal, destabilizing treatment of time, etiquette, and logic.

[8]An interdisciplinary practice that merges elements of art, design, and ecology to create interventions that engage with the natural or built environment. It may take the form of installations, sculptures, interactive landscapes, or multisensory experiences.

digital media in spatial design. This discussion is organized around three central axes: spatial activation and morphological transfer; the deployment of VR technologies; and the integration of generative AI, particularly through the use of image generation models.

The study highlights a range of techniques employed within this framework, including computer simulations, 3D renderings, holographic imaging systems, real-time projection, interactive sensors, and immersive modeling.

The findings indicate that the strategic adoption of these technologies provides benefits in terms of efficiency, artistic expressiveness, and perceptual engagement. VR, for instance, facilitates multisensory and 3D exploration of designed environments, while generative AI enhances creative possibilities and accelerates the conceptual development process. Moreover, digital design experiences are identified as catalysts for novel forms of socialization and audience engagement, fostering the emergence of a "hybrid space" that operates at the intersection of the physical and the virtual.

The work presented by (ZHUO; SIRIVESMAS; PUNYALIKIT, 2023) investigates the use of LED panels as an expressive audiovisual medium that fosters emotional engagement between artistic creation and viewers. As a proof of concept, the authors developed the installation "*Life with Water*", which draws inspiration from the forms and acoustic textures of ocean waves. The artwork integrates LED panels and computer-controlled sound devices to create a synchronized audiovisual experience, whose temporal variations are designed to emulate the rhythmic flow of ocean tides. Utilizing informational models and curated databases of visual and auditory content, artists are afforded the ability to configure the installation's sensory responses in alignment with specific emotional parameters.

The findings indicate that the incorporation of digital media into spatial artworks facilitates a profound transformation in both the aesthetic and functional dimensions of the work. In particular, the interplay of visual and sonic elements is shown to be especially effective in eliciting emotional responses and encouraging active audience engagement within public and urban environments.

3.8 Use of Immersive Media in Artistic Experiences

A growing body of research has explored strategies to enhance expressiveness and immersion in musical performance within XR environments. An illustrative example is the AirPiano system (HWANG; SON; KIM, 2017), which implements a VR-based piano interface incorporating haptic feedback through ultrasonic transducers. The system was designed to support multi-finger interactions on touch-sensitive virtual keys using Leap Motion hand tracking. Two

haptic rendering strategies were implemented: a constant feedback mode, which delivers sustained tactile stimulation during key presses, and an adaptive feedback mode, which simulates the variable resistance characteristic of physical piano keys based on the depth and speed of touch.

The Cyberdreams application (WEINEL, 2020) explores the simulation of altered states of consciousness, such as dreams and psychedelic experiences, within immersive virtual environments, positioning these phenomena as aesthetic resources in the domain of digital art. Drawing from the aesthetics of abstraction and synesthesia, the project investigates new forms of sensory expression in computational media. The methodology combines audiovisual composition techniques, virtual reality, and cinematic practices. Developed in Unity, the system integrates digital video, soundscapes, electronic music, and abstract 3D models into an interactive navigational experience. Conceived as a cybernetic "dream world", the application enables user interaction through bodily movement and motion controllers. Audiovisual composition is guided by principles of live improvisation and experimental performance, emphasizing subjectivity, real-time expression, and synesthetic perception in the construction of immersive artistic works.

In a more recent investigation, (YOUNG; O'DWYER, et al., 2023) propose an immersive musical experience that integrates volumetric video, VR, and haptic feedback to examine how vibrotactile stimulation influences audience perception and engagement during digital music performances. The primary motivation lies in exploring alternative modalities of musical mediation that incorporate multiple sensory channels, thereby challenging traditional audiovisual paradigms by positioning tactile stimuli as integral to the aesthetic and somatic appreciation of musical experiences.

The experimental system consists of a six-degrees-of-freedom (6DoF) immersive music video and a wearable device equipped with vibrotactile actuators affixed to the participant's hand. The experience was implemented in Unity, featuring volumetric capture of musicians and playback via the Valve Index headset. Participants were assigned to two groups, with one experiencing the performance with haptic feedback, while the other experienced it without this modality.

The results indicate that the inclusion of vibrotactile feedback significantly enhanced users' perceptions of attractiveness, stimulation, novelty, and overall efficiency. Qualitative data suggest that participants valued the spatial synchrony between auditory and tactile cues and reported heightened emotional and sensory engagement with the music. Furthermore, the use of volumetric video in VR was regarded as promising, albeit still constrained by limitations in model resolution and the absence of social co-presence.

Collectively, these studies underscore the expanding role of immersive technologies in musical aesthetics by revealing the expressive potential of multisensory interaction in digital performance contexts. Notably, the incorporation of haptic feedback emerges not merely as a functional enhancement, but as a cognitive and affective dimension that contributes to novel modes of musical perception.

3.9 Mulsemedia Applications

In the context of mulsemedia environments, several factors must be taken into account to ensure satisfactory performance and user experience, such as (JOSUÉ; MORENO; MUCHALUAT SAADE, 2019; JULURI; TAMARAPALLI; MEDHI, 2016):

- **Content Immersion:** refers to the extent to which an experience is capable of engaging the user's senses and sustaining their attention. It encompasses the subjective perception of being fully immersed within a virtual or narrative environment, thereby fostering a sense of presence and integration into the simulated or constructed reality;
- **Low Latency in Media Presentation:** concerns the implementation of techniques aimed at reducing delays and interruptions during the rendering of multimedia and multisensory content, thereby preserving temporal coherence and enhancing the fluidity of user experience;
- **Synchronization of Sensory Effects with Media Objects:** involves the deliberate and precise alignment of sensory stimuli with corresponding media elements, ensuring perceptual consistency and reinforcing the intended affective or narrative impact;
- **Device Capability Description:** pertains to the specification of device functionalities in order to guarantee system compatibility, support adaptive and personalized interactions, optimize computational and network resources, and foster creative exploration in multimedia content design;
- **Interactivity:** encompasses mechanisms that enable the user to actively engage with the system, establishing dynamic relationships not only with audiovisual content but also with associated sensory effects, thereby promoting agency and participatory immersion.

In response to these requirements, numerous studies have proposed strategies to address each of the aforementioned factors, such as those proposed by (GHINEA et al., 2014; KIM et al., 2013), that advocates for the integration of sensory effects into multimedia applications as a means of enhancing immersion and, consequently, improving the overall QoE. In addition,

(MEIXNER; EINSIEDLER, 2016) introduces a prefetching mechanism aimed at minimizing delays and interruptions during the presentation of heterogeneous media content. Alternatively, (DAVISON, 2002) and (SU; YANG; ZHANG, 2000) propose automated methods to address this issue, thereby relieving the multimedia application author from the task of manually specifying which content should be prefetched (JOSUÉ; MORENO; MUCHALUAT SAADE, 2019).

To ensure synchronization between sensory effects and media objects, (YUAN; BI, et al., 2015) introduce the concept of synchronization regions, which define temporal intervals during which sensory stimuli should be activated by actuating devices. When the effect is triggered at any point within this designated interval, it is perceived by users as being synchronized with the corresponding visual content.

In a related study, (YOON, 2013) propose the use of effect metadata encapsulated within an MPEG-2 Transport Stream to achieve precise synchronization of media content. (WALTL; TIMMERER; RAINER, et al., 2011) further emphasize that the timing of effect metadata transmission must account for multiple variables, including the preparation time of the sensory devices, network transmission latency, and the delay associated with rendering the effect in the application environment. Complementarily, (SU; YANG; ZHANG, 2000) advocate for the implementation of synchronization mechanisms during the content transmission phase, with the objective of minimizing temporal discrepancies between audiovisual content and the associated sensory effects.

Regarding device description, (CHOI; LEE; YOON, 2011) suggest the use of Part 2 of the MPEG-V standard to retrieve detailed information concerning device functionalities, thereby facilitating the appropriate adaptation and provisioning of streaming services.

Accordingly, response time emerges as a critical factor in the evaluation of QoE. Building on this perspective, (SALEME, Estêvão; SANTOS; GHINEA, 2020) present a solution designed to optimize response time in event-based multimedia applications, contributing to improved system reactivity and more seamless user interaction.

Finally, the work presented by (JOSUÉ, 2021) addresses multiple dimensions of multimedia delivery, encompassing aspects such as synchronization between TV broadcast and broadband streams, as well as the preparation of broadband content through a prefetching mechanism that is partially governed by both the presentation engine and the application author. This approach also offers a generic and flexible solution applicable to both conventional media objects and sensory effects. Its execution is conditioned upon the buffering state of the media player or the loading of the respective media content into that player. This operational model represents a departure from traditional prefetch events, as it introduces a more context-sensitive and adaptable mechanism. Consequently, the advantages of this method extend beyond its suit-

ability for diverse sensory effects, each with their own specific characteristics and requirements, and include its independence from the storage constraints of the playback environment.

3.10 Final Remarks

This chapter offered a comprehensive analysis of the key works that supported the present research, with the discussion organized around nine main areas, namely IoMusT environments, digital and smart musical instruments, frameworks for specifying musical scenarios and XR musical environments, haptic elements applied in arts, air drumming applicaitons, multimedia applied to artistic creation, use of immersive media in artistic experiences and mulsemedia applications. The adopted approach sought to map the architectures, technologies, and implementation strategies employed in each domain, while also identifying points of convergence and uncovering both conceptual and technical gaps that this thesis aims to address.

In the field of IoMusT, the works analyzed highlighted the use of smart devices, low-latency protocols, and real-time audio engines, which enable networked performance experiences. The variety of proposals and technologies illustrated the need for a formal model to specify how this environment and its actors should interact. Moreover, the incorporation of immersive and multisensory elements has received limited attention.

The analysis of digital and smart musical instruments revealed the convergence of acoustic, electronic, and computational paradigms. These devices integrate sensing, embedded processing, and multimodal feedback to enhance expressivity and responsiveness. However, most remain self-contained and lack mechanisms for large-scale interoperability or multisensory data exchange.

The examined frameworks for musical practice emphasize immediacy, accessibility, and expressive interaction as central design principles. While these contributions remain foundational, they focus largely on local interface design rather than on distributed, networked architectures. Few address how sensing and feedback layers interoperate in multisensory or multiuser contexts.

The reviewed frameworks for music creation in XR environments highlight presence, embodiment, and audiovisual integration as crucial to immersive musical experiences. Despite their advances, most lack robust strategies for multisensory synchronization and interoperability. Broader integration of haptic and physiological feedback within low-latency, adaptive architectures remains limited.

Studies in haptic elements applied to the arts confirm the expressive and emotional potential of haptic feedback in this context. By engaging touch, these systems enhance embodiment and accessibility, yet often operate as isolated setups. Few incorporate adaptive or networked tactile

communication.

Air drumming systems demonstrate advances in gestural tracking, synthesis, and immersion. They effectively map free-space gestures into expressive percussive actions but remain largely monolithic and non-collaborative. Feedback is often limited to sound or simple vibration, lacking multisensory coherence.

Multimedia initiatives demonstrated the expressive potential of combining sound, image, movement, and interaction. These systems range from web-based participatory performances to sensory installations mediated by virtual reality and artificial intelligence. However, many of these environments do not fully integrate multisensory feedback or advanced sensory synchronization and adaptation mechanisms.

Mulsemedia environments offer a complementary perspective by emphasizing key requirements such as immersion, sensory synchronization, interactivity, and detailed device specification, thereby contributing to the development of high-quality experiences in systems that transcend conventional media channels. Nonetheless, networking aspects, as well as musical and artistic elements, remain underrepresented in this domain.

This discussion underscores the scientific contributions of the present thesis, particularly its integration of connectivity, multimodality, musical interaction, sensory coordination, and immersive technologies within a unified architectural model. Such a comprehensive convergence of these dimensions has not been systematically addressed in prior research. The following chapter elaborates on the mechanisms and design principles underpinning this integration.

4 Io3MT Reference Model

An IoT environment can be defined as a structured set of tools, standards, and guidelines that facilitates the development, deployment, and management of solutions involving the interconnection of physical and digital devices through the Internet. Such environments enable the collection, exchange, and analysis of data to support decision-making and automation processes. They serve to accelerate development and reduce the inherent complexity of these ecosystem, while also providing mechanisms to capture and quantify user experience. In doing so, they establish a shared vocabulary, methodological guidelines, and models of interrelation and interaction among participating entities, as well as an archetypal reference for a domain-specific architecture (COALLIER, 2022).

Although a wide range of environments and proposals for IoT already exists, many of them are designed to address highly specific scenarios and therefore exhibit limited generalizability to other contexts. This fragmentation leads to challenges in standardization and restricts the reuse of previously established solutions in the development of new domains. Within this context, this chapter introduces requirements specifically devised for Io3MT, aimed at addressing the heterogeneity of applications and use cases that characterize this domain. The proposal encompasses the definition of data types and tools compatible with its requirements, thereby fostering the adoption of consolidated solutions, the reuse of components, and the anticipation of evaluative processes.

Given that this domain is still at an incipient stage and subject to ongoing exploratory inquiry, reaching a consensus on its desirable characteristics remains a complex task. To overcome this challenge, the author has relied on definitions drawn from a substantial body of academic (FLORIS; ATZORI, 2016b, 2015; TURCHET, 2023; BASSI et al., 2013; ZHENG et al., 2022; GUTH et al., 2016) and industrial literature (AMAZON, 2023; AZURE, 2023; MEHTA et al., 2017; DAVIDSON, 2017; ECLIPSE, 2023), examined in greater depth in Chapters 2 and 3, together with a set of technical standards, including:

- **ISO/IEC 30141:** establishes the foundations for defining a reference architecture for IoT systems, describing functional elements, interfaces, and interrelationships (ISO/IEC 30141:2024. . . , 2024);

- **ISO/IEC 21823-1:** provides rules for IoT interoperability, fostering a common understanding among stakeholders (ISO/IEC 21823-1:2019:..., 2019);
- **ISO/IEC 20924:2018:** standardizes IoT terminology (ISO/IEC 20924:2018:..., 2018);
- **ISO/IEC TR 22417:2017:** offers an overview of enabling IoT technologies, serving as a technical basis for implementation (ISO/IEC TR 22417:2017:..., 2017);
- **ISO/IEC 30118-1:2018:** outlines approaches to ensure interoperability among heterogeneous devices (ISO/IEC 30118-1:2018:..., 2018);
- **ISO/IEC 23005-5:** emphasizes the representation of sensory devices and environments within the MPEG-V standard, enabling the mapping between physical and multimedia stimuli (ISO/IEC 23005-5:2019:..., 2019);
- **ITU-T Y.4100/Y.2066:** specifies functional and non-functional groups within IoT systems (ITU-T Y..., 2014);
- **ITU-T Y.4000/Y.2060:** presents an overview and a conceptual model of IoT, defining its objectives, essential functions, and the architectural characteristics of its underlying systems (ITU-T Y.4000/Y.2060:..., 2012);
- **ITU-T Y.4400/Y.2063:** details the functional architecture of IoT middleware, addressing both generic and specific services that enable interoperability across platforms (ITU-T Y..., 2012);
- **ITU-T Y.4111/Y.2076:** defines requirements for IoT data management, including collection, storage, processing, quality, and security (ITU-T Y..., 2016);
- **ITU-T Y.2068:** establishes communication requirements for IoT, covering latency, reliability, bandwidth, and adaptation between heterogeneous devices (ITU-T Y..., 2015);
- **ITU-T F.748.0:** presents a functional models for IoT-based multimedia services, encompassing object description, content management, and interoperability in context-aware applications (ITU-T..., 2014).

Fundamentally, Io3MT should be conceived as a systemic concept that integrates multiple foundational principles, including distributed computing and cyber-physical systems, while simultaneously embodying the convergence of operational technologies (OT) and information technologies (IT). Moreover, its primary stakeholders extend beyond the conventional IT sector to encompass domains such as the Creative Industries and Cultural Digital Experiences. These fields introduce a wide spectrum of requirements, which significantly heightens the complexity of establishing common standards (COALLIER, 2022).

4.1 Io3MT Environment Requirements

Io3MT constitutes an emerging research domain that encompasses heterogeneous objects capable of connecting to the Internet and participating in continuous streams of multisensory, multimedia, and musical data (VIEIRA; SAADE; CÉSAR, 2023, 2024, 2025). This interconnection seeks to foster the integration of such sensory dimensions, incorporating synesthetic and syncretic aspects into the broader scope of the IoS. In addition, Io3MT aims to enable and automate a wide spectrum of services, ranging from artistic practices and entertainment experiences to applications in educational and therapeutic contexts. From a functional perspective, Io3MT can be characterized as a persistent, multi-user, decentralized, collaborative, and interoperable network. It is designed to merge physical and digital elements, thereby facilitating device integration and real-time communication, while also allowing one category of data to influence or act upon another. For instance, a sequence of musical chords may alter the color patterns of a video, or the introduction of an image may trigger olfactory stimuli that evoke the environment being represented.

The effectiveness of this model depends on meeting specific network performance requirements. These include low latency; reduced jitter, defined as the acceptable variability in packet delay; high reliability, understood as the capacity to transmit data with minimal packet loss; adequate bandwidth for handling multimedia streams; and multipath networking to mitigate the risks associated with reliance on a single transmission route, which may otherwise result in congestion or premature failures. Furthermore, the environment should demonstrate fault tolerance, lightweight implementation, and efficient service coordination. An Io3MT environment must also be capable of managing and synchronizing heterogeneous data types, including music, audio, video, images, and sensory stimuli, while providing mechanisms that enable participation beyond professional musicians, thereby broadening accessibility and supporting diverse creative practices.

As in the broader IoT domain, Io3MT can encompass multiple micro-systems that may be implemented independently or operate collaboratively with other systems. In this configuration, the computational load is distributed across all participants; however, if a micro-system fails or is deliberately removed, the service remains active and resilient, albeit with reduced resources. This approach renders the overall architecture distributed, modular, sustainable, and scalable, while also facilitating more efficient development.

In terms of network extension, the system should avoid imposing spatial constraints on user interaction other than those attributable to the propagation delays inherent in physical media. Transmission capabilities should not be limited to relatively short ranges. Rather, the system must support both wireless local area networks (WLANs) and wide area networks (WANs),

while maintaining bandwidth, latency, and error rates within acceptable thresholds to ensure a satisfactory quality of experience.

Although Io3MT shares several characteristics with related domains, it also presents specific requirements that must be considered in its development, particularly in terms of skills and resources. The creation of a multisensory, multimedia, and musical system requires competence in multimodal interaction, the integration of sensors with auditory communication, and familiarity with real-time audio programming languages. In some cases, it may also involve the design of instruments and services for embedded systems.

All of these factors directly affect the programmability of the tasks involved, as they require handling heterogeneous data types and working with embedded systems that are subject to significant constraints in memory and battery capacity, as well as prolonged compilation times due to limited computational power. Another aspect to be considered is that most current embedded systems are either Unix-based or rely on proprietary *ad hoc* operating systems, which often leads to incompatibilities or restrictions in adopting applications aimed at enhancing music, audio and video quality. In some cases, these systems are further hindered by outdated interfaces and insufficient documentation, which complicates their integration into new services. Collectively, these challenges have a direct impact on development time and effort.

With respect to computational power, multimedia applications impose substantial demands on processing resources, as the tools traditionally employed in this domain are designed to be robust enough to support audio and video processing with low latency and consistent quality. Such tools generally rely on optimized code to minimize interruptions, errors, and the risk of overheating in processing units. Therefore, appropriate hardware is essential to manage these requirements effectively, ensuring that performance does not compromise expected outcomes or hinder the work of artists and researchers.

Mapping strategies are critical in musical services, where the accurate translation of performers' gestures into interface commands for sound control or generation is required. Designers, composers, and performers need to account for this level of control when planning and using devices. In multimedia applications, simultaneous message delivery to connected devices may also occur, producing multimodal content, although the precision required is generally lower than in musical contexts.

This domain also encompasses artistic and pedagogical concerns, taking into account aspects such as composition, rehearsal, performance, video creation and editing, animations, and images, as well as how these elements can operate in combination to enable the participation and full integration of non-experts in the artistic production.

This conceptualization of the environment enables individuals to experience the services

provided by Io3MT, directly influencing the final outcome of the product with which they interact while receiving exclusive stimuli and information. For professionals, it facilitates the analysis of user behavior, thereby supporting the refinement of future work and informing decision-making processes. For industry, it creates opportunities for data analysis, the development of new applications and services, and improvements in efficiency and cost-effectiveness.

A compendium of the requirements for configuring an Io3MT environment is provided in Table 2, which also identifies the reference domains underpinning each premise. It should be emphasized, however, that system-specific variations are inherent, as implementations depend on the artistic objectives and the technological resources available to individual developers.

Io3MT Requirement	**Reference Area**
Loosely Coupled	IoMusT Environments
Scalability	Networked Music Performance, Wireless Multimedia Sensor Networks
Layered Architecture	IoMusT Environments, Wireless Multimedia Sensor Networks
Ease of Development and Evolution	Interactive Art and IoMusT Environments
Fault Tolerance	Networked Music Performance and IoMusT Environments
Lightweight Implementation	IoMusT Environments
Service Coordination	IoMusT Environments
Low Latency	Networked Music Performance
Synchronization	Networked Music Performance
Transparent Integration and Ease of Participation	Networked Music Performance, IoMusT Environments, and Interactive Art

Continued on next page

Io3MT Requirement	**Reference Area**
Employment of Multiple Human Senses	Mulsemedia
QoE Enhancement through Multimodal Combination	Mulsemedia, XR and Wireless Multimedia Sensor Networks
Integration with Legacy Devices	Wireless Multimedia Sensor Networks and IoMusT Environments
Device Modification and Updating	IoMusT Environments and Interactive Art
Synchronous or Asynchronous Communication	IoMusT Environments and Interactive Art
Content Immersion, Reduced Media Presentation Delay, and Synchronization between Sensory Effects and Media Objects	Mulsemedia and XR Environments
Interactivity	Interactive Art, IoMusT Environments, Mulsemedia and XR Environments

Table 2: Main Requirements of an Io3MT Environment.

4.2 Functional Requirements

Functional requirements specify the expected behaviors of the system, namely the operations it must be able to perform. Within the Io3MT context, the infrastructure should provide remote access to physical devices through network resources, ensure interoperability across heterogeneous networks and operating systems, and maintain compatibility among diverse data

formats. Furthermore, it must support collaborative information processing to address complex tasks, such as detecting and tracking entities in the physical environment, while also ensuring continuous operation over extended periods with minimal maintenance demands and efficient energy utilization (ITU-T. . . , 2014).

The network components should provide multiple levels of functionality, thereby promoting flexibility in connectivity strategies. This approach is intended to prevent devices specialized in single tasks from constraining the configuration or expansion of the environment, which would otherwise result in a static and limited topology (ISO/IEC 30141:2024. . . , 2024).

Although several modules may operate autonomously, it is essential that the network incorporate comprehensive management mechanisms, including the control of devices, systems, networks, security, and interfaces. Such management involves the following requirements: i) network communication through reliable, secure, and continuously operational protocols; ii) the implementation of a manageable infrastructure that ensures consistent interconnection among all devices; and iii) support for real-time operations, with data collection and processing occurring immediately after event detection (ISO/IEC 30141:2024. . . , 2024). Additionally, auxiliary functionalities must be guaranteed, such as access control, resource management for agents, and the mapping of information across devices and network entities (LEE, E. et al., 2021; ITU-T Y. . . , 2012).

Interoperability is structured around three main facets (LEE, E. et al., 2021):

- **Syntactic interoperability:** refers to the ability to exchange data through standardized formats and formal rules. Programming languages and data formats such as Web Ontology Language (OWL), RDFS (Resource Description Framework Schema), Extensible Markup Language (XML), and JSON could be employed to ensure this facet;
- **Semantic interoperability:** concerns the shared understanding of the meaning of data across different contexts within the same domain. This facet is helpful to prevent ambiguities in the interpretation of exchanged information;
- **Behavioral interoperability:** refers to the correct execution of expected operations resulting from information exchange. Achieving this requires a clear specification of input/output interfaces, as well as the preconditions, postconditions, and sequences of operations for each entity involved.

4.3 Non-Functional Requirements

Non-functional requirements establish quality attributes and operational constraints that determine how system functionalities should be implemented and delivered. These requirements do not specify what the system performs, but rather how it is expected to operate (ITU-T Y.2066:. . . , 2014; ITU-T Y. . . , 2015).

Within the scope of Io3MT, non-functional requirements must be addressed to guarantee the technical sustainability of the environment, the stability of its operations, and the consistency of the user experience across heterogeneous application scenarios. Considering the distributed, heterogeneous, and time-sensitive characteristics of Io3MT-based applications, heterogeneity represents a core structural aspect. The system must operate within ecosystems comprising devices and platforms with different computational capabilities, architectures, and communication protocols. This diversity requires support for multiple data formats, network interfaces, and abstraction layers, thereby enabling technical interoperability without imposing uniformity on the underlying infrastructure.

Scalability, in turn, concerns the progressive expandability of the system, both in terms of the number of connected devices and sensors and the volume of users, data streams, and services. This requirement ensures that system performance does not degrade significantly as complexity increases.

Reliability and resilience are required to ensure that the system performs its functions correctly even under adverse conditions. This involves the consistent delivery of data and the ability to recover from partial failures, network instabilities, or component malfunctions. These aspects are directly related to high availability, which refers to the capacity of services and functionalities to remain continuously and predictably accessible with minimal interruptions, particularly in applications that demand real-time responsiveness or uninterrupted operation, such as interactive environments or live performances.

Adaptability refers to the ability of the infrastructure to remain flexible in accommodating technological updates, configuration changes, and the integration of new modules or devices without compromising the integrity of the existing system. This adaptive capacity supports the long-term viability of the system in contexts of rapid technological evolution. Manageability, in turn, encompasses the monitoring, control, and maintenance of system components, providing mechanisms for the insertion, removal, and replacement of devices and users, as well as for anomaly detection, resource administration, and fault recovery.

In contrast to the theoretical postulations presented in Section 4.1, which delineate the conceptual, phenomenological, and epistemological foundations of the Io3MT reference model,

such as the centrality of sensorimotor experience, technological mediation as a perceptual extension, and the articulation between corporeality and digital materiality, the non-functional requirements are of a technical nature. While the postulations establish why certain aspects are relevant from a theoretical perspective of immersive and multisensory reality, the non-functional requirements specify how the system must operate to enable these experiences in an effective, reliable, and sustainable manner. Thus, although situated at complementary levels of abstraction, both categories are indispensable: the postulations ground the critical and conceptual vision of the system, whereas the non-functional requirements ensure its technical and operational feasibility in real-world scenarios.

4.4 Musical and Multimedia Protocol Stack

For the symbolic representation of multimedia information, the Musical Instrument Digital Interface (MIDI) can be employed (ROTHSTEIN, 1992). MIDI is a versatile protocol that supports the layering of sounds, enabling a musician to play two or more instruments simultaneously, interconnect different equipment, and establish connections between instruments and computers. It also allows for the editing of any musical event through software and provides message addressing capabilities via communication channels. However, the MIDI model, based on the representation of notes, channels, and continuous controllers, may not be well suited for representing, organizing, or naming parameters in multimedia environments. This limitation stems from its lower flexibility, as the protocol was originally designed specifically for communication between musical instruments (TURCHET; FISCHIONE, et al., 2018).

An alternative that addresses those limitations is OSC, a protocol that employs symbolic parameters rather than purely numerical information in real-time audio control messages. These messages are concerned with the descriptive aspects of sound, such as octave, intensity, and duration, rather than the audio signal itself (WRIGHT, 2005).

This nomenclature is analogous to the Uniform Resource Locator (URL) and provides high-resolution data, the ability to specify multiple recipients for a single message, and the simultaneous delivery of data packets. It can be employed for the integration and synchronization of homogeneous systems, the design of additional protocols, message conversion, gesture mapping, and spatial audio control. Messages are transmitted, via UDP or TCP, from a device configured as a sender to one or more devices configured as receivers. The receiver then forwards the message to the component or function specified by the OSC address. This design enables lightweight and flexible communication among any number of devices with limited latency, facilitating seamless incorporation into distributed networked music services (JOHNSON, 2019). A format of this type also extends its applicability beyond audio technologies, being employed

in the control of multimedia devices and robotic systems (SCHMEDER; FREED; WESSEL, 2010; MADGWICK et al., 2015; WRIGHT et al., 2001).

Through the integration of computers, controllers, and synthesizers, costs are reduced while metrics of reliability, convenience, and user control are enhanced. For this reason, this protocol is often regarded as a replacement for MIDI. However, this assumption is inaccurate. OSC defines only a communication protocol and does not specify a digital interface or electrical connectors (wired or wireless) for device interconnection. Furthermore, OSC lacks a standardized namespace for device interfaces. For this reason, connected devices are neither aware of each other nor of their respective capabilities. In addition, there is no file format for OSC equivalent to the standard MIDI file, which could otherwise enable data exchange across different applications (TURCHET; FISCHIONE, et al., 2018). Therefore, it is more appropriate to regard OSC and MIDI as complementary technologies that can effectively coexist within the same environments (WRIGHT, 2005).

4.5 Data Requirements

In Io3MT scenarios, data are inherently unpredictable due to their heterogeneity, encompassing variable sizes and both structured and unstructured forms. To address this complexity, the adoption of common formats is recommended, as it facilitates the integration and aggregation of information collected from multiple applications (ITU-T. . . , 2014).

The environment should therefore support audio data based in Pulse Code Modulation (PCM) method, such as WAV, which is widely used for real-time data processing, as well as lighter formats such as MPEG-1/2 Audio Layer 3 (MP3), Free Lossless Audio Codec (FLAC), or OGG. A more recent format that has gained attention is the Dynamic Music Object (DYMO), which, in addition to audio data, incorporates a service package containing analytical information about a given sample. DYMO can also operate in conjunction with Semantic Web technologies, such as OWL and SPARQL, and can be employed in adaptive sound experiences that are context-aware and controllable by multiple entities within an IoT device, such as an accelerometer or geolocation module. Its versatility and suitability for network-based operations make it particularly valuable in Io3MT environments (TURCHET; FISCHIONE, et al., 2018; THALMANN et al., 2016).

Regarding visual information, it has emerged as a key element in artistic perception and expression. It accentuates performers' bodily nuances, facilitates the exchange of information and feedback, and can even be integrated with sound parameters to automatically adjust mixing, volume, and related aspects. Visual data may be represented in multiple formats, including the MPEG-V standard, which is being developed for implementation in both real-world actuators

and sensors as well as in virtual devices, thereby enabling convergence between these domains. Until its full adoption, the H.26x and VPx standards remain the most widely used solutions for video compression and encoding in such scenarios (FLORIS; ATZORI, 2015).

Sensory data can be produced either through sensor-based capture or generated synthetically using authoring tools. In both cases, explicit specification is required to enable subsequent execution in a compatible player. To address this need, recent standardization efforts have been undertaken by regulatory bodies, such as MPEG-V Part 3 – Sensory Information, which aims to establish a common framework for this type of data (INFORMATION. . . , 2019).

From this discussion, textual formats emerge whose main characteristic is their reliance on scripting or markup-based languages. Examples include the Sensory Effect Description Language (SEDL), an XML schema–based language that enables the description of sensory effects, and the Sensory Effect Vocabulary (SEV), which can be applied to any type of multimedia content (e.g., films, music, websites, games) to control sensory devices such as fans, vibrating chairs, or lamps through an appropriate mediation device, thereby enhancing the user experience. It is important to note that in all of these methods the transmitted information does not represent the sensory effects themselves but rather metadata in the form of commands, which are delivered via APIs to activate the different sensory effect actuators.

4.6 Artistic Requirements

For artistic creation, the environment should be pleasant to use, provide an engaging and immersive experience for the audience, and enable seamless integration and participation of users. It should also include features that assist individuals without prior artistic or computational knowledge, thereby fostering creativity across a wide range of users.

To meet these objectives, several requirements must be considered. The environment should facilitate presentations by ensuring accessibility, capture information from the audience and integrate it into a coherent flow while ensuring that the artists' actions produce the expected results (artistic safety), guide the audience toward uninhibited participation (initiation), maintain a captivating character (attractiveness), and provide a clear relationship between participants' gestures and the resulting outputs (transparency).

Such characteristics render the system open and editable, allowing for the combination of multiple elements and providing users with opportunities to gradually learn and master specific skills. In this context, there are no functional errors in a strict sense, but rather outcomes that may be considered aesthetically or musically unsatisfactory.

4.7 Desirable Features of Devices

The vision of Io3MT introduces a new class of Internet-connected devices, referred to as multisensory, multimedia, and musical things (3MT). These devices are characterized by their ability to produce, respond to, track, or observe phenomena associated with at least one of these three types of information. They can be understood as providers of services and information, encompassing hardware, software, sensors, actuators, and cloud-based services, as well as electroacoustic, electronic, or virtual musical instruments. This category also includes electronic gadgets with processing capabilities and integrated peripherals, such as smartphones, and computational platforms equipped with switching resources, such as Raspberry Pi or Arduino, which can control rendering devices. In addition, smart equipment such as lamps, scent generators, and fog machines are also part of this ecosystem. Their structures may be either physical or digital, with virtual entities able to exist independently of physical counterparts, while communication occurs over the network (TURCHET; FISCHIONE, et al., 2018; LEE, E. et al., 2021; ITU-T Y.4000/Y.2060:. . . , 2012).

As this class of devices represents an extension of the IoT paradigm, it shares a range of characteristics with equipment commonly found in that domain. They incorporate embedded electronics, wireless communication, and sensing and/or actuation capabilities, while cooperating with neighboring equipment, including legacy ones, to accomplish specific tasks in accordance with defined project constraints and requirements. In addition, they must be uniquely identifiable and addressable, scalable, persistent, reliable, and loosely coupled.

From a technical perspective, these devices must be capable of routing frames, include their own power supply, and exhibit context-awareness in order to minimize redundant data acquisition. Since they are also designed with artistic concerns in mind, aesthetic, expressive, and ergonomic factors become equally relevant (VIEIRA; GONÇALVES; SCHIAVONI, 2020). Consequently, they should demonstrate usability, being easy to learn, flexible, efficient, and effective in performing their intended tasks, along with accessibility, ensuring inclusiveness for users with different motor and social abilities, and communicability, clearly conveying the design intent and logic of their creators while responding appropriately to user stimuli.

Aligned with these aspects, it is essential that 3MT provide consistent mapping for multisensory feedback and information, thereby ensuring perceptual coherence. They may be designed either to control a single synthesis process or effect at a time, or to enable the simultaneous manipulation of multiple processes, particularly in interactive graphical interfaces with visual feedback. In the latter case, an important feature is the use of reactive three-dimensional visual representations, whose graphical parameters are bidirectionally linked to the corresponding musical, sensory, and multimedia parameters. Techniques employed for

manipulating these elements in virtual environments include spatial transformations such as rotation, scaling, and translation, as well as structural modification and alterations of material properties (BERTHAUT, 2020; BERTHAUT; DESAINTE-CATHERINE; HACHET, 2011; DRASCIC; MILGRAM, 1996; ZELLERBACH; ROBERTS, 2022).

These devices can be classified into three categories: agents, managers, and hybrids. Agents act as data producers, typically represented by sensing devices. Managers, in turn, function as data collectors, whereas hybrid elements perform both roles. Connections between devices may occur in any direction, but in most cases, the agent initiates the connection, as it is aware of when data become available (IEEE 11073:. . . , 2004; SANTOS; ALMEIDA; PERKUSICH, 2013).

4.8 Conceptual Model

A conceptual model is a description of elements and their types of relationships, accompanied by a set of constraints that define how these elements should be used. It enables developers to determine the class to which a system organization belongs, based on the characteristics of its components and connectors, architectural topology, semantic constraints, and mechanisms of interaction among components. A conceptual model also provides a foundation for coherently modeling a given class of systems and serves as a guideline for addressing the challenges associated with designing complex architectures. In the case of IoT-based systems, a conceptual model can facilitate the structuring of functional characteristics, as it defines how components should ideally be combined (SANTOS; SILVA, et al., 2020).

Figure 6 presents the conceptual model with the key entities of Io3MT and their relationships, providing a generic and simplified structure with common definitions to describe the concepts and interactions among entities within a system. The model highlights the IoT-User element, which may represent either a human user or a Digital User, such as robots or automation services acting on behalf of humans. Both interact with the system through applications that communicate over the network. Some applications also interact directly with each other, always mediated by the network.

The Io3MT Entity refers to a real-world object, which may include any musical instrument, multisensory generation/rendering device, actuator, sensor, or similar component. The Virtual Entity represents the digital counterpart of an Io3MT Device. These devices interact over the network and are capable of communicating directly with one another without requiring a server to mediate the communication.

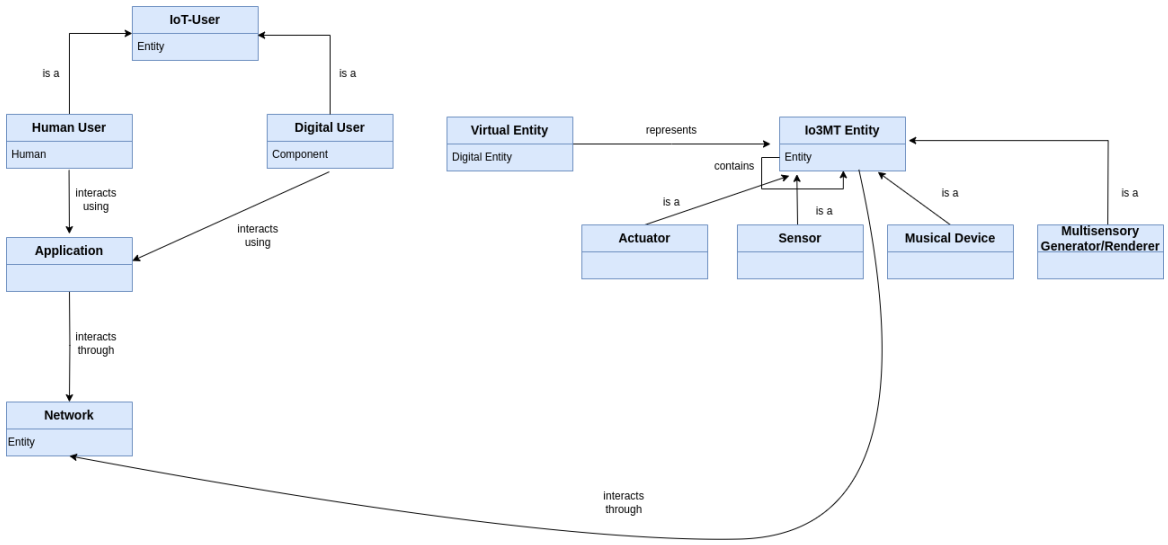


Figure 6: Conceptual model of an Io3MT ecosystem.

4.9 Architectural Model of Io3MT

As is customary in established in several computer science tools, a layered structure can be adopted for this architectural model. The principal advantage of this approach is that the services of each layer are implemented on the basis of the requirements defined by the underlying layers. Conventionally, lower levels provide simpler and more efficient services, while increasingly complex functionalities are located at higher levels. In addition, a layered architecture decomposes solutions in such a way that each substrate can evolve independently (GUTH et al., 2016).

Despite this division, services are generally organized into three categories: device connectivity; data processing, analysis, and management; and presentation and connectivity. Based on this perspective, a three-layer model was adopted, encompassing these pillars in a concise and direct manner while also avoiding a one-size-fits-all approach. Nevertheless, the more detailed each reference architecture becomes, the more heterogeneous they are likely to be as a whole. The layers are presented and explained as follows.

Device Layer: considered the virtualization layer, this level is responsible for collecting and transforming analog data, such as gestures, performative actions, or sound information from unplugged instruments, captured by physical devices into digital formats, thereby initiating the large volumes of information that circulate throughout the environment (FLORIS; ATZORI, 2016a). The techniques associated with this layer focus on the design and implementation of 3MTs optimized for low energy consumption and high performance. Additional concerns include context recognition, large-scale data processing, and device discovery mechanisms (ZHENG et al., 2022).

Network Layer: its purpose is to interconnect devices and applications with the network,

functioning as a bridge between the physical and virtual domains. Accordingly, the data acquired from the device layer must be transmitted securely and reliably. This layer also provides the underlying infrastructure that supports communication within the system, ensuring the fundamental requirements of an IoT environment. These include event-driven, periodic, and automatic communication modes; multiple transmission methods such as unicast, multicast, or broadcast; device-initiated communication; error control; autonomous network operation; and compatibility with heterogeneous communication technologies associated with devices (ITU-T Y. . . , 2015).

Although the Internet does not inherently provide QoS guarantees, operating instead under a best-effort delivery model, these parameters remain essential in Io3MT network performance analysis, as they offer mechanisms that support the transmission and processing of time-critical messages. It is therefore essential to evaluate data transmission performance through metrics such as latency, jitter, packet loss, etc.

Artistic performances impose specific communication constraints that must be taken into account at this layer. In typical real-time use cases, the connectivity infrastructure must ensure low-latency communication, high reliability, high quality, and precise synchronization across connected devices (TURCHET; FISCHIONE, et al., 2018).

These heterogeneous services can combine backbone networks, mobile or satellite communication networks, LANs, WLANs, wireless transmission, 5G, and distributed computing technologies as a whole, including cloud services.

Lastly, this layer manages the large volume of data produced and consumed by Io3MT applications. It is responsible for addressing communication objects and defining the topology of the local network, as well as determining the optimal routing of data. This stratum encompasses resource and traffic management, congestion control, and network integration (ZHENG et al., 2022; LEE, E. et al., 2021; ITU-T Y.4000/Y.2060:. . . , 2012).

Application Layer: this layer provides the core functionality of the Io3MT environment. It orchestrates information services and performs large-scale data processing and intelligent analysis in order to deliver multisensory and multimedia musical environments based on client or user requests. The layer can be further divided into two sub-levels, the first being the engine unit, which encompasses the set of functions responsible for maintaining the overall integrity of Io3MT systems and for provisioning, managing, monitoring, and optimizing their real-time operational performance. Additional functions include data translation, discovery, synchronization, identity management, concurrency control, and state transitions. It is also capable of manipulating the state of physical objects through actuation. AI techniques can be employed to enhance these services by enabling context awareness, real-time recognition of

musical gestures, and the analysis of multimodal and multisensory musical content (ZHENG et al., 2022; TURCHET, 2023).

The second subdivision is the service unit, which processes the services requested by the end user and transforms data from the physical world into cyber-expressions. To support users in performing tasks, it provides interfaces and platforms built upon an extensible service structure.

Figure 7 illustrates and summarizes the discussion developed throughout this chapter, encompassing the layered division of the architecture.

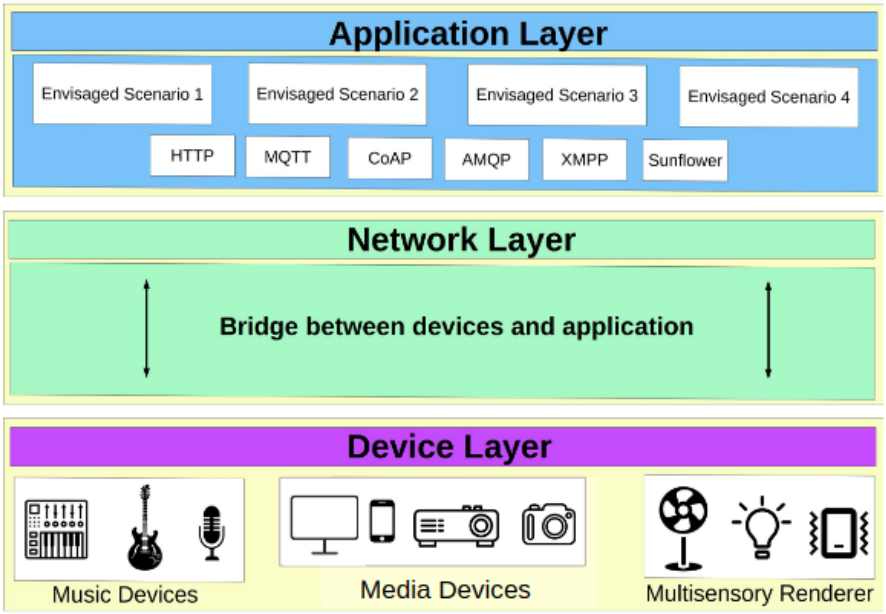


Figure 7: Functional view of the Io3MT architecture.

4.10 Envisaged Scenarios

This section explores a range of use scenarios that emerge from the versatility and applicability of the proposed model. The analysis of these situations demonstrates how the technologies and tools discussed here can be applied across diverse contexts, from musical performances to interactive museums and galleries, thereby illustrating the tangible impact that the proposed concepts can achieve.

The effectiveness of these scenarios depends on several factors, including the types of devices employed (technical factors), the end users who consume the generated resources (social factors), the costs of deployment (business factors), and the physical settings in which the environments are implemented (environmental factors).

4.10.1 Scenario 1: Live Music Performance

In an Io3MT-based performance, musicians can exchange data over the network using a wide variety of devices, ranging from electroacoustic musical instruments with Internet connectivity to scent dispensers, smart lighting, and wearable technologies. At the same time, audience members can contribute to the performance through applications running on their smartphones, generating sounds or controlling other parameters within the ecosystem.

These objects can receive sensory stimuli and respond according to their unique properties, while also having the potential to apply distinctive effects based on the information received. Communication among them should be interwoven and non-hierarchical. For instance, a specific chord progression (musical information) may trigger a light cannon (sensory information), altering brightness, hue, and saturation, or it may terminate the playback of a video (multimedia information).

Examining multimedia information in detail, it can be employed to convey messages to the audience, display videos, animations, and sensory effects that enhance the intended narrative. It may also serve as a support tool for musicians, providing access to lyrics, setlists, scores, or any other information relevant to the performance context.

The control of musical effects, lighting, and large video displays can be managed either by audience members or by specialized technicians. Similarly, traditional sound mixing methods can be replaced by digital systems that are accessible and controllable through the network.

Musicians and dancers should be able to collect data related to the audience's emotional responses, adapting the sequence of songs, choreographies, visual effects, and color schemes used in the performance to better align with those reactions.

Remote audiences should also be enabled to participate in the performance, receiving real-time audio and video streams while being able to control selected parameters over the network.

4.10.2 Scenario 2: An Improvisation Session Combining Multisensory, Multimedia, and Musical Elements

In this use context, electroacoustic instruments are enhanced with multimedia and multisensory elements, thereby improving communication among participants and extending both the quality and the dissemination of the performance.

The graphical aspect is enriched by videos captured through smartphone and laptop cameras, which are displayed to the general audience on Smart TVs or display screens. This introduces an additional dimension of information into the environment, enhancing the audio-visual experience for all participants. In synergy, the manipulation of devices, such as triggering

note sequences on synthesizers or modifying drum rhythms, is also orchestrated through video equipment, which assumes a multifaceted role as both controller and manager.

Remote users can interact with the system by controlling the volume and effects of the instruments, including gain, modulation, and related parameters. They should also be able to modify the color patterns of images, adjust the resolution and format of videos and visual artworks, and trigger sensory effects.

The management of this interaction can be carried out by a sound engineer or technician, who is responsible for handling the connections, enabling or restricting communication between specific instruments and/or users, and selecting which audio tracks will be delivered to the general audience.

4.10.3 Scenario 3: Smart Studio Recording

Another scenario emerges from the integration of Io3MT principles into a smart studio environment, which can accommodate a wide range of musical contexts, from solo performers and duos to bands and orchestras, encompassing diverse instruments. To enable this, the recording interface must be adaptable, providing the optimal number of channels and adjusting dynamically to each musical situation. In addition, it should be capable of storing individual configuration and effect preferences for each artist, thereby personalizing the recording experience. Remote artists and technicians may also collaborate in the processes of recording, mixing, and mastering.

Through multimedia resources, network-related technical information is shared, providing insights into connected objects. In this context, the musical environment is enriched by the display of lyrics, scores, and other relevant information, thereby amplifying both understanding and creativity during the recording process. Pre-recorded tracks, accessible via the Internet, can also be manipulated remotely, blending seamlessly with live performance.

The multisensory sphere emerges as a key-feature, providing immersive feedback to participants. This may take the form of precise visual cues that guide performers to initiate execution at the appropriate moment, or the ability to adjust musical parameters through combined tactile and visual interaction, thereby creating a synesthetic ecosystem.

4.10.4 Scenario 4: Applications in Cinema, Home Entertainment, Education, Healthcare, Immersive Artistic Spaces, and Beyond

Io3MT is primarily designed to support musical and artistic performances. Nevertheless, its scope extends beyond these domains. The reference model demonstrates extensibility to a variety of other sectors, such as cinema, home entertainment, education and healthcare ap-

plications, thereby reinforcing its potential as a versatile paradigm applicable across multiple contexts.

Contemporary cinema already demonstrates how sensory augmentation can transform the viewing experience, with auditoriums equipped with seats capable of vibrating or delivering other forms of haptic feedback in synchrony with each scene. The integration of Io3MT would extend these capabilities by enabling the reproduction of thermal sensations experienced by the main character, such as cold or heat. Furthermore, Io3MT could allow each viewer to receive individualized and context-sensitive stimuli throughout the film, while also supporting collective interaction in which the audience, through a voting system or similar mechanism, could modify parameters such as color patterns, aspect ratio properties, or even the storyline itself.

Home entertainment encompasses streaming services for TV shows, films, books, and music, as well as video games. In this context, unique stimuli and responses can be delivered to each user, such as adjustments to ambient lighting, modifications to musical tempo that alter color patterns in a video, or the triggering of sensory effects. These interactions enhance immersion and user engagement with the content, while also ensuring that each access to the material offers novel and distinctive features.

In the educational domain, Io3MT offers significant potential for enhancing learning experiences, particularly in the field of music education. By integrating multisensory and multimedia feedback into instructional environments, students can interact with virtual or augmented instruments that provide haptic, auditory, and visual responses in real time. Such interactions allow for more engaging and accessible pedagogical strategies, enabling learners to better understand technical concepts while simultaneously fostering creativity and collaboration. Additionally, adaptive feedback mechanisms can be incorporated to accommodate diverse learning styles and abilities, making musical training more inclusive and effective.

Regarding healthcare applications, Io3MT can be applied to therapy and rehabilitation contexts. Patients may engage in exercises that combine musical, visual, and haptic feedback to improve motor skills, cognitive functions, and emotional well-being. For example, multisensory rehabilitation protocols can leverage rhythmic stimuli to aid movement coordination or use immersive soundscapes to support stress reduction and pain management. The adaptability of Io3MT-based systems allows healthcare professionals to tailor therapeutic interventions to individual patient needs, thereby improving efficacy and fostering long-term adherence to treatment programs.

Museums and galleries can leverage this technology to transform the way visitors engage with art. Actuators may trigger distinct musical tracks or sensory effects for each exhibited

piece, while multimedia content can provide contextual information such as interviews with the creators or detailed explanations of the techniques employed. These elements may also become integral components of the artwork itself, shaping the way it is experienced. Besides that, this approach enables interactive works that can be modified by the audience, thereby enriching the overall experience.

Extended Reality environments also provide fertile ground for the application of Io3MT, especially in the context of musical practice. Through immersive and interactive systems, musicians can rehearse in virtual or augmented settings enriched with multisensory feedback, where gestures are coupled with visual and haptic cues that reinforce timing, expressivity, and coordination. Such environments not only replicate the dynamics of traditional practice but also expand them by enabling scenarios that would be difficult or impossible to achieve in the physical world, such as simulating orchestral performances or interactive improvisation sessions with virtual agents. By bridging musical, multisensory, and spatial dimensions, Io3MT also enhances the pedagogical and creative possibilities of XR-based musical training.

4.11 Final Remarks on Io3MT Theoretical Foundations

This chapter presented the theoretical and technical foundations of Io3MT, articulating this domain as a systemic, layered, and horizontally oriented paradigm that integrates cyber-physical principles with multisensory, multimedia and musical interaction. Drawing upon international standards (e.g., ISO/IEC 30141, ITU-T Y.2060/Y.2066, MPEG-V) and pre-existing academic and industrial proposals, the chapter synthesized the functional and non-functional requirements related to connectivity, interoperability, synchronization, scalability, and manageability across heterogeneous devices and networks. It also formalized requirements concerning data, devices, and protocols, in addition to formulating a conceptual and architectural model that clarifies the entities, relationships, and responsibilities within Io3MT ecosystems. Finally, representative application scenarios were outlined, encompassing musical performances, education, healthcare, studio production, and immersive artistic spaces, demonstrating both the feasibility and the broad impact of this domain.

Collectively, these foundations provide a common vocabulary, a reference architecture that separates responsibilities while ensuring end-to-end communication, and concrete guidelines for implementing interoperable multisensory and musical experiences.

The next chapter operationalizes these principles through RemixDrum, the first proof-of-concept artifact of Io3MT. RemixDrum materializes the proposed architecture and emphasizes the behavior of 3MT devices by implementing gesture-to-sound mappings and multimedia control. It functions as an empirical validation prototype, allowing the observation, in real-world

contexts, of how Io3MT principles translate into measurable results and tangible experiences.

5 RemixDrum: A Smart Musical Instrument for Music and Visual Art Remix

In light of the various specificities associated with the components constituting an Io3MT environment, this chapter introduces a prototype of a smart musical instrument that aligns with the principles of 3MT. This equipment, referred to as RemixDrum (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023), enhances a traditional drumstick through the integration of sensors, microcontroller boards, and wireless connectivity. Its primary goal is to amalgamate the acoustic sound produced by the drum with digitally synthesized sounds generated in Pure Data, which are influenced and modulated by the movements of the device. This interaction aims to align both musical and perceptual experiences based in the concepts of Remix Culture.

Moreover, the prototype exhibits the capability to interface with visual art created in Processing, thereby fostering creative thinking and encouraging the generation of works that explores multimedia hybridization, as advocated by the Io3MT domain. Subsequent sections will provide comprehensive details on the device's design and a thorough analysis of the broader macroenvironment in which it operates.

5.1 Remix Culture

The practice of remixing began to gain attention during the 1970s, marked by the emergence of the Disc Jockey (DJ), who utilized techniques such as sampling[1] and mashup[2] to produce novel musical creations. This approach became common in genres such as Hip Hop and Disco Music (NAVAS, 2012; TRAGTENBERG; ALBUQUERQUE; CALEGARIO, 2021). Nevertheless, expressions of these creative and critical impulses have been observable throughout history, wherein artistic forms and styles from one society have been assimilated into another. For instance, Ancient Rome adopted various concepts from Greek culture; the Renaissance drew inspiration from classical antiquity; 19th-century European architecture integrated elements

[1] Sampling refers to the extraction of sound excerpts from pre-existing recordings, which can subsequently be reorganized or recontextualized in new compositions.

[2] A mashup denotes a musical composition constructed through the combination of two or more pre-existing songs, often resulting in hybridized forms that juxtapose or blend distinct stylistic elements.

from multiple historical periods; and contemporary graphic and fashion designs amalgamate diverse cultural influences, ranging from Japanese *manga* aesthetics to experimental techniques of collage and photomontage (MANOVICH, 2005).

Based on this contextualization, it is possible to observe that the notion of remix extends beyond the domain of musical practice, permeating a wide range of artistic manifestations. In this sense, it subverts the Cartesian tradition of compartmentalization, which historically classified the world into distinct and mutually exclusive categories (LEÃO, 2012). From this perspective, the concept has developed into what is commonly designated as Remix Culture, a construct originating in communication theory that encompasses cultural practices, social configurations, and modes of living constituted through processes of appropriation, transformation, and reconfiguration of pre-existing works. Such processes give rise to new forms, concepts, ideas, and services, fostering innovative modes of cultural production and circulation (PARSONS, 2010; MANOVICH, 2007).

The advent and swift expansion of Web 2.0 has led to the emergence of new tools designed to explore the concept of remix in the digital environment. These tools possess a degree of flexibility and modularity, enabling collaborative remixing activities (LÉVY, 2010).

Drawing on the multiplicity of combinations inherent to Remix Culture, new opportunities for artistic exploration emerge, particularly through the integration of diverse sonic elements with the control of visual productions. Such practices directly address several of the requirements established by the Io3MT reference model. As a proof of concept, an instance of a SMI was developed, which, in this case, is conceptualized as a 3MT object. This prototype integrates sensors, actuators, and interconnection capabilities, providing a practical validation of hardware functionalities within the proposed paradigm. In parallel, a dedicated artistic environment was designed to accommodate the instrument, functioning as an experimental testbed to assess factors such as network performance, quality of user experience, and the potential of these concepts to advance the state of the art in both SMI research and multimedia artistic creation.

5.2 The RemixDrum Design

The practical development of RemixDrum was guided by a set of design principles, namely: i) implementation of a modular and adaptable architecture to facilitate device interconnection; ii) incorporation of two sensor-based interfaces — one based on tactile pressure for functional control (e.g., sample triggering and preset selection) and another based on spatial information for the manipulation of aesthetic parameters (e.g., movements along the X, Y, and Z axes); iii) adoption of lightweight and compact technologies to ensure portability and ease of instal-

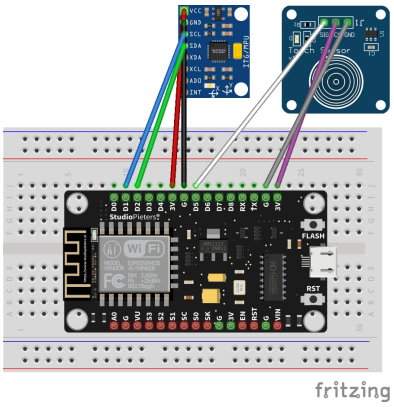
lation; iv) scalability through the use of a wireless transmission system; v) interoperability supported by standard communication protocols; vi) accessibility in terms of programming and software updates; and vii) cost efficiency through the use of low-cost and/or open-source technologies (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

To meet these requirements, ST235 capacitive touch sensors and MPU-6050 accelerometers were employed. The ST235 operates by modulating circuit capacitance in response to variations in accumulated charge at the reference point, enabling commands such as the activation or deactivation of audio tracks. The MPU-6050, in turn, integrates a control chip capable of measuring acceleration, rotation, and vibration. Rotational movements along its axes are mapped to the control of audio parameters such as flanger, reverb, and volume, as well as to the modulation of visual elements, including color schemes and playback speed. Data processing and wireless communication were handled by ESP8266 NodeMCU v2 boards, selected for their 32-bit processor capacity, which is sufficient to capture input signals from the touch sensor and accelerometer while transmitting them over the network. These boards also ensure compliance with the TCP/IP stack and multiple Wi-Fi IEEE 802.11 standards (b/g/n).

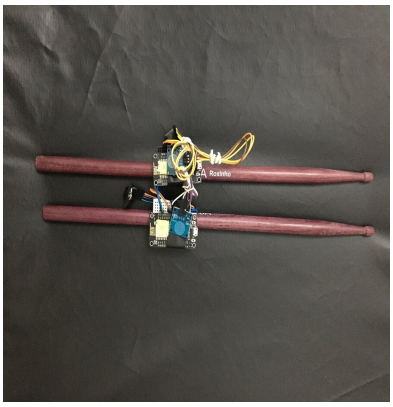
Musical information is transmitted using OSC messages, while the UDP transport protocol ensures efficient network communication. Pure Data was employed for audio synthesis and Processing for the creation of visual art. All technologies adopted in the project are based on open-source platforms, thereby ensuring flexibility for adaptation, extension, and modification according to the requirements of specific environments. Furthermore, the construction of the prototype was carried out using Do-It-Yourself (DIY) methodologies, which contributed to self-sufficiency, reduced costs, and the unrestricted dissemination of the resources underpinning the creative process.

Figure 8 illustrates the electrical circuit, the final prototype, and its implementation in a real-world performance scenario, while a demonstration of its operation is available in a YouTube video[3]. The project's source code is publicly accessible in its official GitHub repository[4].

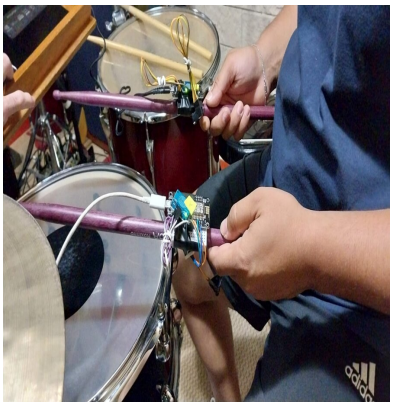
[3] https://www.youtube.com/watch?v=RQgIk2Z2Vxg
[4] https://github.com/romulovieira-me/RemixDrum



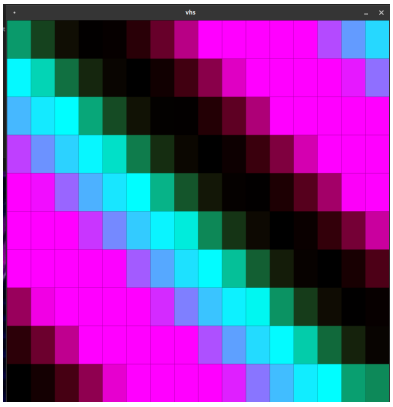
(a) Electrical circuit of RemixDrum.



(b) Physical structure of RemixDrum.



(c) RemixDrum in practical application.



(d) Multimedia artwork controlled by RemixDrum.

Figure 8: Structural composition and practical application of RemixDrum (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

5.3 Practical Evaluation of RemixDrum

The evaluation of the prototype was conducted in two distinct stages. The first phase examined network performance, while the second addressed QoE. The subsequent sections present a detailed account of these analyses, conducted through testing sessions with a professional drummer possessing more than 15 years of performance experience, who engaged with RemixDrum in a studio environment.

5.3.1 Network Performance Analysis

The evaluation of network behavior, carried out to ascertain how effectively the system fulfills its operational requirements, was grounded in prior studies that articulate performance thresholds for networked musical environments. Accordingly, the analysis concentrated on key indicators of efficiency and reliability, namely latency, jitter, and throughput (VIEIRA; SCHIAVONI;

SAADE, 2022; TURCHET; CASARI, 2024).

Latency is calculated as the mean difference between the actual relative time of the sound and the expected time, as expressed mathematically in Equation 5.1 (SCHIAVONI; QUEIROZ; WANDERLEY, 2013). In the context of networked artistic performances, latency should not exceed 40,000 µs (40 ms) (VIEIRA; SCHIAVONI; SAADE, 2022).

$$latency(\Delta t) = \frac{1}{n}\sum_{i=1}^{n}(t(i) - expected_t(i)) \tag{5.1}$$

Jitter, in turn, can be quantified as the standard deviation of latency, as expressed in Equation 5.2 (SCHIAVONI; QUEIROZ; WANDERLEY, 2013). For percussive instruments that integrate both tactile components (e.g., touch sensors and accelerometers) and non-tactile elements (e.g., Pure Data and Processing), this metric should remain below 55,000 µs (55 ms) in order to prevent perceptible disruptions to the user experience (ZAVERI et al., 2022).

$$jitter = \frac{1}{n}\sum_{i=1}^{n}|t(i) - \Delta t| \tag{5.2}$$

Throughput, defined as the expected number of messages successfully received within a given time interval, constitutes another critical performance metric. In the present study, the analysis focuses on the packet transmission rate per second for each drumstick, rather than adopting the conventional measurement in bits per second. Throughput is computed by dividing the total number of packets transmitted by the duration of the test, as expressed in Equation 5.3. It is important to emphasize that the time measurement spans from the first to the last sample, thereby accounting for any gaps or interruptions observed between transmissions.

$$throughput = \frac{\sum packets}{test\_duration} \tag{5.3}$$

To evaluate the technical dimensions of the aforementioned metrics, ten testing sessions were conducted using a wireless network based on the Wi-Fi IEEE 802.11ac, with a theoretical transmission limit of 300 Mbps. The network was provided by a TP-Link Archer C5 router with security requirements disabled. Figure 9 provides a representation of this setup, in which OSC messages, encapsulated in UDP packets, were transmitted to Pure Data for the manipulation of sound parameters, thereby integrating digital processing with the acoustic output of the drum performance. When routed to Processing application, the same information was used to generate modifications in a visual artwork, specifically altering its color schemes and motion dynamics. Network information generated during these sessions was captured and subsequently

analyzed using Wireshark[5].

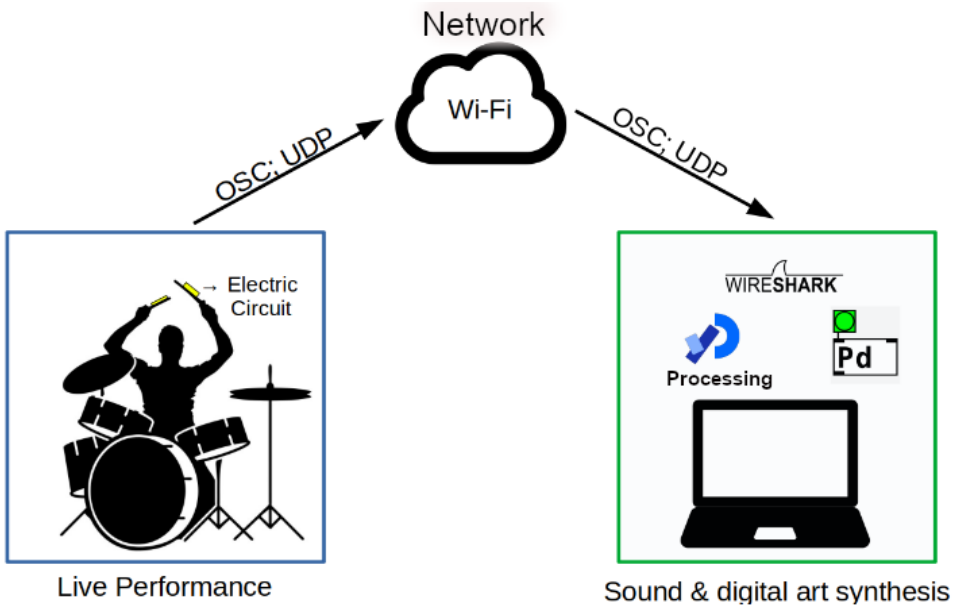


Figure 9: Composition of the RemixDrum test environment (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

The expressiveness of the musician, together with the diverse musical nuances manifested in each session, influenced the outcomes of the performances conducted during the tests. Consequently, variations in the number of packets transmitted across sessions were to be expected. The total number of packets recorded in each session is summarized in Table 3.

Test	**Drumstick A**	**Drumstick B**
Test 1	2630	2325
Test 2	4132	4086
Test 3	2031	2065
Test 4	1690	1447
Test 5	1708	2254
Test 6	1802	1542
Test 7	1702	1403
Test 8	1729	1393
Test 9	1835	1570
Test 10	1590	1342

Table 3: Number of packets transmitted by the drumsticks in each test (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

The results obtained for each of the three metrics across both drumsticks over the ten

[5] https://www.wireshark.org/

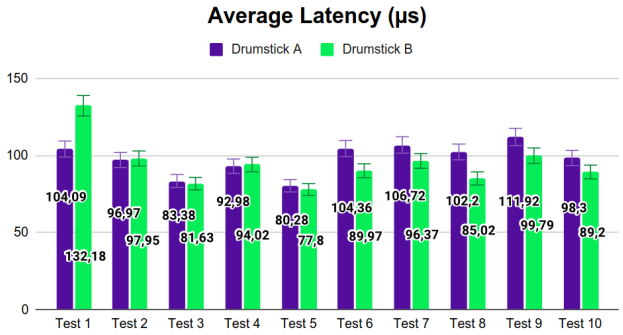
testing sessions are presented in Figure 10, with a 95% confidence interval. The analysis indicates that the latency values remained well below the threshold considered ideal for this type of performance. Several factors contributed to this outcome, including the adoption of a dedicated network infrastructure for the system and the exclusive connectivity of the drumsticks to the access point. Packet size and content also proved decisive in sustaining low latency. With an average size of 70 kBytes and transmission restricted to values supported by the sensors and actuators, no bottlenecks or processing queues emerged that could compromise data transmission.

As a result, the average jitter was approximately 46.45 µs for Drumstick A and 49.07 µs for Drumstick B, both of which fall within acceptable limits for percussive instruments. Similarly, the average throughput reached 11 packets per second for Drumstick A and 10 packets per second for Drumstick B. These values mitigate common challenges associated with excessive packet transmission, such as network congestion, performance degradation, packet loss, improper service prioritization, and resource exhaustion. It should also be noted that the Wi-Fi network employed in the tests operated within the public frequency spectrum, rendering it susceptible to potential interference from coexisting networks. Nonetheless, no further investigations or corrective measures were undertaken in this regard, since the current infrastructure satisfactorily met the operational requirements of the application. This robustness can be attributed, in large part, to the characteristics of the transmitted data.

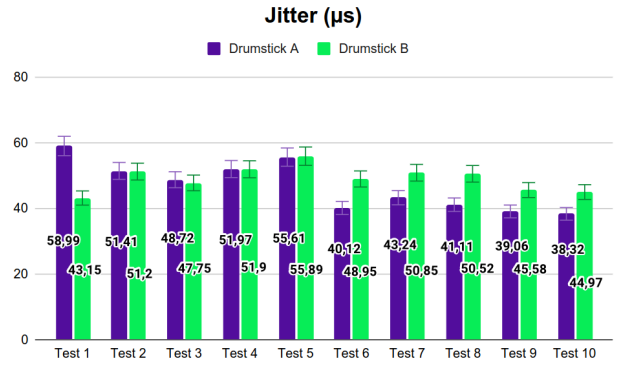
5.3.2 Quality of Experience (QoE) Analysis

The second stage of analysis addressed the QoE associated with the prototype. This metric reflects the overall acceptability of an application or service, with particular emphasis on hedonic dimensions such as aesthetics and self-fulfillment that arise during the use of a given system. For the purposes of this study, a semi-structured interview was employed (WILSON, 2013). This qualitative research method combines predefined guiding questions with a flexible structure, allowing for an in-depth exploration of aspects related to user–system interaction. As such, it served as an effective technique for eliciting participant's perceptions, opinions, and interpretations concerning specific elements of the prototype's design and practical use.

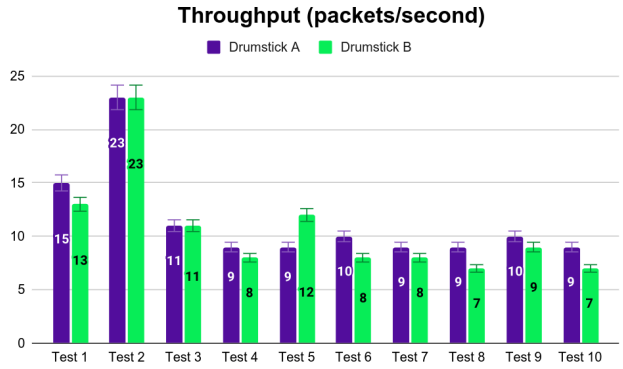
The decision between qualitative and quantitative methodologies is contingent upon the specific objectives of the research. Qualitative approaches are particularly appropriate in exploratory contexts, such as new product development or artistic applications. They are especially valuable when the aim is to capture subjective experiences, interpret the meanings attributed by users, and investigate the aesthetic and creative dimensions of use, as exemplified in the case of RemixDrum (DENZIN; LINCOLN, 2011). On the other hand, quantitative methodologies are particularly suited to contexts in which the primary goal is the objective



(a) Latency for Drumstick A and Drumstick B.



(b) Jitter for Drumstick A and Drumstick B.



(c) Throughput for Drumstick A and Drumstick B.

Figure 10: Network performance for RemixDrum (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

measurement of variables and the comparative evaluation of the performance of established products or services. These approaches provide a higher degree of generalizability of findings and support the formulation and empirical testing of hypotheses through the application of statistical procedures (CRESWELL; CRESWELL, 2017).

The system was evaluated by a single expert. This choice corresponds to the profile of a key stakeholder in the design and use of the proposed equipment, ensuring that the assessment was conducted by a highly qualified representative of the intended user group. The adoption of a single-expert evaluation protocol is a well-established practice in the fields of computer music and IoMusT. Frequently, such systems remain at an early stage of development or are designed primarily for individual interaction. Under these conditions, the participation of an expert is essential to thoroughly investigate the expressive affordances of the instrument and to provide informed feedback regarding its usability and performance potential (REFSUM JENSENIUS; LYONS, 2017; TURCHET, 2018a; MERENDINO; RODÀ; MASU, 2024).

This evaluation method enables a deeper level of analysis, as the expert is capable of assessing both technical and artistic nuances that might not be perceived by inexperienced or generalist users. It also enhances the reliability of identifying critical issues related to musical performance and optimizes resources, since longitudinal studies involving multiple participants generally demand considerably more time, infrastructure, and instrumentation. Moreover, involving a professional musician enables the validation of the system's suitability in demanding usage scenarios, which demonstrates that its design addresses both artistic requirements and technical quality criteria.

Naturally, this approach entails methodological limitations, most notably the inability to statistically generalize the findings and the absence of different user profiles. Nevertheless, within the specific scope of this study, such limitations do not compromise the validity of the analysis, since the main goal of the evaluation was to assess the practical feasibility and artistic potential of the system when used by experts, rather than to establish comparative performance benchmarks for the entire population.

Regarding the constructs evaluated in this research, they are presented and explained below (MERENDINO; RODÀ; MASU, 2024):

- **Input:** refers to the actions required for the performer to interact with the device, encompassing the degree of adaptation necessary for effective use as well as the perceived level of difficulty during performance;
- **Output:** corresponds to the sonic and artistic results generated by the system, with particular emphasis on its expressiveness, versatility, and alignment with the performer's creative intentions, in addition to the satisfaction and emotional quality derived from the

experience;

- **Control:** concerns the manner in which performance is executed, highlighting the mapping between input actions and the resulting outputs, as well as the technical and musical skills required to fully explore the expressive potential of the instrument;
- **Body:** relates to the physical configuration of the device, including ergonomic aspects, aesthetic, portability, and any design constraints, such as the arrangement of circuits;
- **Adherence:** examines the factors that influence a user's inclination to adopt the system, including motivation, required effort, repertoire compatibility, likelihood of recommending it to other musicians, and perceived barriers to use;
- **General:** encompasses broader dimensions of UX, such as overall usability, engagement, frustration, and the holistic evaluation of the device's functionality and design.

At this stage, the user was not assigned specific tasks but was instead encouraged to articulate thoughts and reflections regarding the experience with the instrument. This process employed a modified version of the think-aloud method, implemented through a continuous verbal protocol (KRAHMER; UMMELEN, 2004). The objective of this technique was to capture the participant's internal reasoning during interaction, thereby eliciting reflections on dimensions such as expressiveness, system behavior, exploratory use, and musical creation.

The interview content and the corresponding responses are summarized in Table 4. For the sake of conciseness and clarity, only abridged versions are presented, while the complete material is available in Appendix A. The interview protocol and evaluation criteria were adapted from methodologies reported in previous studies with comparable objectives in the design and assessment of novel SMIs (BARGAS-AVILA; HORNBAEK, 2011; BROWN; NASH; MITCHELL, 2017; TURCHET, 2018b; O'MODHRAIN, 2011).

Construct	Questions	Answers
Input	How do you evaluate the ease of triggering the device?	Easy. Based on responsive sensors with direct activation
	Did you encounter any difficulty in execution or adaptation?	No, but subtle nuances were less precise

Continued on next page

Construct	Questions	Answers
Output	How expressive and versatile were the sounds produced?	Very expressive. Convincing sounds; expands sonic palette
	Did the device meet your creative needs?	Yes; Help to expand repertoire beyond the acoustic drum set
	Did the results stimulate your creativity?	Yes; motivated new artistic combinations
	How do you evaluate the overall quality of the artistic experience?	Positive and satisfactory
Control	How natural or intuitive was the initial interaction?	Intuitive after adaptation; latency in subtle technical variations
	What musical or technical skills were required?	Basic drumming skills; minimal technological familiarity is helpful
Body	How did you perceive ergonomics and physical comfort?	Reasonable; cables and circuit limit movements
	What did you think of the system's appearance and portability?	Good design; physical constraints reduce movement freedom
Adherence	What would most motivate you to use the system in a real performance?	Sonic expansion and artistic innovation
	How much effort was required to use it?	Low physical and mental effort
	Would you recommend the system to other musicians? Why?	Yes; it introduces innovation and expands creativity
	What barriers or limitations did you identify?	Cables and limited ergonomics

Continued on next page

Construct	Questions	Answers
General	Did the system perform as expected?	Yes; functional and innovative
	Did the device capture your attention and engagement?	Yes; sonic variety attracted attention
	Was there any frustrating aspect in the experience?	Presence of cables and position of the electrical circuit
	What suggestions for improvement would you provide?	Wireless version, improved ergonomics, customizable mappings and soundtracks

Table 4: Summary of the semi-structured interview conducted during the RemixDrum evaluation (complete responses in Appendix A).

Based on the participant's answers, the input quality emerged as a noteworthy strength. Motion detection was described as responsive and sufficiently precise for rhythmic performance, particularly in linear and straightforward patterns. Nonetheless, the performer observed that delicate dynamic variations were not always captured with equal immediacy, as the sound synthesis application required additional time to register and reproduce them. This observation underscores the need for further technical refinement to enhance sensitivity and responsiveness.

With regard to the output, the musician emphasized the expressive character of the generated sounds and the capacity of RemixDrum to extend the sonic palette beyond that of the acoustic instrument. The integration of traditional percussive timbres with digitally synthesized elements was perceived as conducive to creativity and experimentation, particularly within hybrid artistic contexts.

In terms of control, the relationship between gesture and response was generally considered intuitive and transparent. The learning curve was regarded as low. However, fully exploiting the expressive potential of the instrument required additional competencies, particularly a degree of technological familiarity, such as knowledge of sensor mappings. This suggests that while RemixDrum remains accessible to experienced percussionists, it simultaneously introduces layers of creative complexity that broaden its range of potential applications.

Concerning the physical structure of the device (body construct), the performer praised its resemblance to a conventional drumstick, recognizing this feature as beneficial for adoption and acceptance by professional musicians. Nevertheless, ergonomic issues were highlighted: the placement of the circuit and the presence of external cables restricted freedom of movement during performance and, in longer sessions, caused discomfort. These limitations reinforce the importance of developing a wireless version or adopting more efficient encapsulation strategies to ensure both protection and freedom of

movement.

Regarding adherence, the musician's motivation to employ RemixDrum in live performances was strongly associated with its potential for artistic innovation. The effort required to adapt to the instrument was deemed minimal, supporting its feasibility for use in professional performance settings.

With respect to general user experience, RemixDrum was described as functional, innovative, and capable of generating a high degree of engagement. The device captured the musician's attention, offering an aesthetically appealing and creatively distinctive experience. Its versatility was also emphasized, particularly the ability to process audio in real time, to navigate across samples, and to play multiple tracks at the same time. These features reduce the need for additional equipment and minimize setup time, while the ease of programming and updating supports efficient transitions between presets.

The main limitations identified — cabling, ergonomics, and unfinished body aesthetics — were regarded as surmountable and did not undermine the overall positive evaluation. It was also noted that the prototype is not a fully self-contained system, as it incorporates sensors, processing units, and wireless connectivity in the drumsticks, while audio rendering and visual generation are handled by a separate computer within the same network. However, this configuration does not pose inherent challenges, since the distribution of tasks across devices, commonly referred to as crowdcomputing, is a well-established approach in networked music performance, interactive art, and new musical interface research.

Finally, the expert suggested enhancements such as greater flexibility in mapping gestures to audiovisual responses and the inclusion of customizable sound content, which would further expand the instrument's creative potential.

5.4 Analysis of Desirable Characteristics for the Io3MT Environment

This section analyzes the extent to which RemixDrum fulfills Io3MT requirements. From the perspective of its general characteristics, the system can be described as loosely coupled (R1). This property manifests in the hardware layer, where each drumstick operates independently without depending on resources or information from its counterpart, and in the software layer, where the Pure Data application (responsible for audio) and the Processing application (responsible for visuals) function autonomously, regardless of the presence of the drumsticks. Although the actions of the drumsticks modify the auditory and visual outputs, the applications themselves retain structural independence and operational autonomy. This arrangement also reveals the presence of micro-systems, wherein each drumstick–application pair establishes message exchanges and generates localized actions, which subsequently combine to form the broader auditory and visual landscape of the environment.

Regard to scalability, RemixDrum exhibits a modular and configurable architecture, which allows

the integration of new components and functionalities without requiring major structural reconfigurations. Its open-source nature enables continuous modification and updating of devices, fostering community-driven evolution and system replicability. The same applies to logical applications, particularly Pure Data, which can be reprogrammed and updated on the fly, that is, while still running. This mode of operation not only facilitates the gradual expansion of the system but also enables the incorporation of new artifacts without compromising stability. Such characteristics contribute to the definition of RemixDrum as a distributed, modular, sustainable, and scalable architecture.

Service coordination is expressed through the mapping between gestures and multimedia effects. The accelerometer and touch button information of the right drumstick are mapped to musical parameters in Pure Data, while corresponding information from the left drumstick modulates visual properties in Processing, such as color and speed. This direct correspondence between services ensures perceptual coherence and exemplifies a one-to-one mapping between sensory input and multimodal response.

With respect to integration, the current configuration is established manually (hard-coded), using predefined IP addresses, ports, and connection parameters. While functional, this approach may hinder transparent integration, as it requires users to be aware of available addresses or to have sufficient technical knowledge to adjust network parameters on their own devices. On the other hand, the adoption of multicast reception mitigates this limitation. As long as devices are connected to the same network, they can receive transmitted data without additional configuration, thereby facilitating participation. This configuration also enables users, including non-musicians, to interact with the system by manipulating auditory or graphical properties, thus enhancing accessibility and promoting collaborative creative participation.

Regarding network performance, discussed in greater detail in Section 5.3.1, the experiments demonstrated low latency and low jitter, combined with sufficient bandwidth and lightweight implementation within a local network. These factors ensured system reliability, with minimal packet loss and throughput rates consistent with the requirements for networked musical performance. The system further benefits from operating over a Wi-Fi local network and employing OSC and UDP protocols, which provide low cost and simplicity. In this way, no significant issues were observed in user-based evaluations.

The system's ability to synchronize multiple signal types, both auditory and visual, ensured multimodal consistency. Robustness was further demonstrated through its fault tolerance mechanisms. Specifically, the intrinsic properties of the ESP8266 board, in conjunction with the Pure Data and Processing applications, enabled automatic reconnection in the event of communication failures. As a result, the system exhibited resilience, seamlessly resuming data transmission once connectivity was restored. Moreover, even in cases of drumstick malfunction, the applications remained operational in a reduced mode, thereby safeguarding their core functionalities and preserving system stability.

From a multimodal perspective, RemixDrum engages multiple human senses by coupling gestures inherent to musical performance with real-time modifications in the visual application, alongside the tactile vibration of the drumsticks when hitting the acoustic drum kit. This integration enhanced

QoE, as identified in Section 5.3.2, by fostering creativity, artistic exploration, and a heightened sense of immersion.

5.5 Comparative Analysis with Related Work

A comparative analysis with the musical devices exhibiting similar operational modes presented in Section 3.2, whether classified as SMIs or DMIs, reveals RemixDrum's distinctive characteristics in terms of input, processing, connectivity, multimodal integration, and applicability. Such a comparison serves to validate the innovative aspects introduced by the proposed system, not only within the Io3MT reference model but also in the broader field of digital musical interface design.

When examining input and sensing modalities, a preference emerges for the use of piezoelectric sensors, FSRs, joysticks, and multitouch interfaces. Exceptions include the Sensus Smart Guitar (TURCHET; MCPHERSON; FISCHIONE, et al., 2016; TURCHET; BENINCASO; FISCHIONE, 2017), which integrates multimodal sensors distributed along the instrument to capture detailed information on position and applied force, and RemixDrum, which employs accelerometers and touch sensors embedded in the drumsticks. This approach ensures high responsiveness and leverages the natural movements of the instrument itself to generate digital-domain actions, thereby establishing a direct connection between gesture, sound, and multimedia elements.

When considering the processing unit, a greater diversity of approaches is observed. The Smart Cajón (TURCHET; MCPHERSON; BARTHET, et al., 2018; TURCHET; MCPHERSON; BARTHET, 2018) and Smart Mandolin (TURCHET, 2018a) employ embedded low-latency processors (Bela BeagleBone and dedicated microcontrollers), supporting Pure Data audio synthesis. The Sensus Smart Guitar distinguishes itself by executing both analysis and synthesis entirely within the instrument, thereby eliminating reliance on external devices and attaining a high degree of computational autonomy. By contrast, the MuDI (PATRÍCIO, 2012) relies on a laptop as an audio engine, which expands processing capacity but reduces portability. The Illusio (BARBOSA et al., 2013) integrates Processing and openFrameworks (C++) for combined visual and audio processing, with a particular emphasis on structural and graphical transformations. The TANC (O'NEILL; ORTIZ, 2024) couples Max/MSP with an Arduino Nano, enabling configurable granular synthesis, albeit with limited scalability. RemixDrum adopts a hybrid architecture in which dedicated microcontrollers perform data capture and preprocessing, while synthesis and multimodal mapping are managed through Pure Data and Processing modules. Although it does not achieve full autonomy like the Sensus Smart Guitar (TURCHET; MCPHERSON; FISCHIONE, et al., 2016; TURCHET; BENINCASO; FISCHIONE, 2017), RemixDrum leverages distributed computing principles, distinguishing itself from other systems by utilizing crowd-computing and distributed processing rather than relying on a single machine for the synthesis of artistic elements.

In terms of connectivity, the majority of systems adopt similar solutions. The Smart Cajón (TURCHET; MCPHERSON; BARTHET, et al., 2018; TURCHET; MCPHERSON; BARTHET,

2018), Smart Mandolin (TURCHET, 2018a), Sensus Smart Guitar (TURCHET; MCPHERSON; FISCHIONE, et al., 2016; TURCHET; BENINCASO; FISCHIONE, 2017), and RemixDrum all employ wireless networks based on the Wi-Fi and transmit data using UDP. The Sensus Smart Guitar (TURCHET; MCPHERSON; FISCHIONE, et al., 2016; TURCHET; BENINCASO; FISCHIONE, 2017) further supports 4G and Bluetooth, the latter also employed by the TANC (O'NEILL; ORTIZ, 2024). The Illusio (BARBOSA et al., 2013), by contrast, does not implement networking capabilities or any protocol that enables interconnection with other systems. This general convergence indicates a shared technological direction in the design of networked musical instruments, situating RemixDrum in alignment with the state-of-the-art.

The dimension of multimedia and multisensory integration is also present across most systems, albeit with varied approaches. The MuDI (PATRÍCIO, 2012) derives its sonic creation from visual information extracted from a film, whereas the Illusio (BARBOSA et al., 2013) relies on graphical inputs to control loops. Among SMIs, the Smart Mandolin (TURCHET, 2018a) is equipped with sensors for interaction with projectors and lighting systems, while the Smart Cajón (TURCHET; MCPHERSON; BARTHET, et al., 2018; TURCHET; MCPHERSON; BARTHET, 2018) provides haptic feedback to the performer. The Sensus Smart Guitar (TURCHET; MCPHERSON; FISCHIONE, et al., 2016; TURCHET; BENINCASO; FISCHIONE, 2017) is compatible with XR environments. Despite these multimodal features, they are generally treated in a fragmented manner, with each subsystem functioning independently. RemixDrum, in contrast, employs gestural input as the central mechanism for controlling both auditory parameters in Pure Data and visual parameters in Processing. This establishes a clear and interdependent linkage among tactile, auditory, and visual dimensions.

In terms of applicability, the Smart Cajón (TURCHET; MCPHERSON; BARTHET, et al., 2018; TURCHET; MCPHERSON; BARTHET, 2018) and Smart Mandolin (TURCHET, 2018a) are primarily designed for live performance, emphasizing hybrid acoustic–digital expressiveness. The Sensus Smart Guitar, developed as a commercial product, aims to bridge mainstream musical practices with advanced digital ecosystems. The MuDI (PATRÍCIO, 2012) focuses on composition and real-time performance, whereas the Illusio (BARBOSA et al., 2013) is oriented toward experimental visual–sonic improvisation. The TANC (O'NEILL; ORTIZ, 2024) emphasizes improvisation and pedagogical experimentation. RemixDrum, beyond its acoustic–digital integration, is primarily directed toward real-time performance and multimedia remixing.

Based on the author's knowledge, RemixDrum represents the first SMI/3MT that directly correlates musical actions with multimedia outcomes, while simultaneously incorporating the concept of remix. This positions it as a paradigmatic prototype for exploring new forms of immersive, collaborative, and multisensory musical interaction. A synthesis of this discussion is presented in Table 5.

Attribute	Smart Cajón	Smart Mandolin	Sensus Smart Guitar	MuDI	Illusio	TANC	RemixDrum
Input	Piezoelectric + FSR	Piezoelectric + multimodal sensors	Position and force sensors	Multitouch screen (iPod)	Multitouch screen + pedals	Joystick + sensors	Gestures + touch sensor + accelerometer
Processing	Bela BeagleBone + Pure Data	Embedded microcontrollers + Pure Data	Autonomous synthesis	Pure Data	Processing + openFrameworks (C++)	Max/MSP	Microcontrollers + Pure Data + Processing
Networking / Connectivity	4G; Wi-Fi + UDP	Wi-Fi + UDP	Wi-Fi; 4G; Bluetooth	Wi-Fi	N/A	Bluetooth	Wi-Fi + UDP
Multimedia / Multisensory Integration	Haptic feedback	Interaction with projectors/lights	XR integration	N/A	Visual-sonic loops	N/A	Gestural control linking sound (Pure Data) and visuals (Processing)
Application	Live performance; hybrid acoustic–digital environments	Live performance; hybrid acoustic–digital environments	Live performances; professional music ecosystem	Real-time composition + live performance	Experimental audiovisual improvisation	Improvisation; pedagogy	Real-time performance; remix of multimedia content; Io3MT integration

Table 5: Comparative overview of musical digital instruments presented in this chapter.

5.6 Final Remarks on RemixDrum

This chapter presented RemixDrum, conceived as a proof of concept for a Multisensory, Multimedia, and Musical Thing (3MT), drawing upon the concepts of SMIs, DMIs, and musical things. The device consists of a traditional drumstick equipped with a tactile sensor, accelerometer, embedded processing resources, and wireless connectivity, thereby enabling both audio manipulation and the mediation of multimedia content. This configuration demonstrates the system's ability to function not only as a musical instrument but also as an active node within Io3MT ecosystems, simultaneously acting as a sensing device and a multimodal agent.

The evaluation of RemixDrum was conducted across two complementary dimensions — network performance and QoE — providing essential insights for the technical and artistic validation of the prototype. Regarding the first dimension, results obtained from the analysis of latency, jitter, and throughput indicated that the device satisfactorily meets the technical requirements for networked musical performance. In particular, latency values remained below the 40 ms threshold, ensuring that both auditory and visual responses were perceived as immediate by the performer, without compromising gesture–sound–image synchrony. This finding is especially relevant in percussive contexts, where microtemporal variations can drastically affect rhythmic perception and performance expressivity.

Along similar lines, the jitter analysis revealed stable average values, within acceptable limits for percussion instruments with tactile elements. This suggests that the system maintains temporal consistency between successive events, a critical factor in preserving dynamic nuances and accentuation exploited by experienced musicians. The packet transmission rate, which varied between 10 and 11 packets per second, reinforces this conclusion, showing that the network adequately supported the transmitted data volume without congestion or perceptible message loss.

Taken together, these results position RemixDrum in line with consolidated practices in next-generation of musical performance. Although further testing in environments with more users and

devices is desirable, the initial outcomes demonstrate that the system fulfills network requirements for application in real artistic contexts.

The qualitative analysis, conducted with an expert user, complements the technical evaluation by highlighting how network parameters helps to achieve perception, usability, and artistic expressivity. The musician's account confirmed that low latency and network stability enabled fluid interactions, with immediate auditory responses to gestural actions, which in turn encouraged experimentation and fostered creative immersion.

With respect to input and control, it was observed that initial learning required a short adaptation period, particularly in associating gestures with auditory and visual responses. However, once this familiarization curve was overcome, the performer reported that the gesture–outcome relationship became natural and intuitive, enabling the exploration of expressive variations.

In terms of output, results were considered highly expressive, especially in the auditory dimension, where RemixDrum was definied as a "pedal stomp for drummers". The ability to trigger and modify sounds in real time expanded the performer's sonic palette, motivating the creation of novel combinations of gestures and timbres. In the visual domain, although gesture–image mapping was deemed less relevant, multimodal integration was regarded as positive and as having the potential to enrich hybrid performances.

In the body construct, critiques emerged regarding ergonomics, particularly the positioning of the circuit and the reliance on power cables. These factors caused discomfort during longer sessions, indicating the need for physical design improvements and, ideally, the adoption of wireless power solutions to enhance gestural freedom.

As for engagement and motivational factors, the performer highlighted innovation and compatibility with repertoire as decisive for adopting the system in live performances. The effort required was assessed as low, reinforcing the feasibility of adoption by professional musicians, while artistic attractiveness was identified as the decisive factor for sustained use.

From a functional standpoint, RemixDrum addressed several requirements formalized by the Io3MT requirements, such as embedded sensing (touch and motion), distributed processing, and real-time communication. Its architecture allows reprogramming and continuous adaptation, characterizing it as an open and scalable system. Although limitations remain in areas such as security and more complex network topologies, the solution already demonstrates partial interoperability and potential for evolution toward broader ecosystems. From a non-functional perspective, reliability and adaptability stand out, supported by automatic reconnection and incremental updates, although aspects such as hardware redundancy and autonomous power supply remain areas for refinement.

Furthermore, the comparative analysis with related works revealed that RemixDrum distinguishes itself by proposing a deeper integration of gesture, sound, and multimedia. By doing so, it transforms performative gestures into direct vectors for auditory and visual manipulation, inaugurating new perspectives for immersive, connected, and multimedia remix-oriented artistic practices.

RemixDrum not only validates, on an experimental basis, the concept of 3MT but also extends the scope of application for smart musical instruments. Equally important, the methodologies of design, evaluation, and analysis employed in this study provide valuable foundations for the development of future musical devices.

The following chapter presents an extended version of the RemixDrum, developed based on the feedback collected from the experiments with the specialist. This evolution led to the creation of a new environment, named PhysioDrum, which not only enhances the SMI used as an input device but also transitions the system into a fully virtual environment. To achieve this, a set of design guidelines is specified for the development of the application, together with a user experience evaluation protocol. Furthermore, the proposed environment also serves as a testbed for investigating the impact of multisensory feedback, particularly haptic cues, on user interaction and experience within immersive musical environments.

6 PhysioDrum: Bridging Physical and Digital Realms in an Immersive Io3MT Environment

This chapter presents PhysioDrum, a practical scenario of an Io3MT immersive environment. Initially, a focus group was conducted with four specialists in virtual reality and music, aiming to establish guidelines to design the envisioned environment. Subsequently, it introduces PhysioDrum, a proof of concept that implements the proposed guidelines, with particular emphasis on examining the role and influence of haptic elements in immersive contexts.

6.1 Designing an Immersive Io3MT Environment

The operating model proposed by Io3MT paves the way for the creation of environments capable of detecting, acquiring, processing, and exchanging data that facilitate the connection between the digital and physical worlds, fostering the emergence of new creative applications and services that explore multimodal, multiplatform, and transmedia relationships. A promising area for development within this paradigm is XR, specifically its branch known as the Musical Metaverse (BOEM; TURCHET, 2023).

However, current applications in XR have predominantly prioritized enhancing audience experiences, with a pronounced emphasis on visual components, often at the expense of musical production and multisensory effects, which are frequently relegated to secondary status. Moreover, there is an absence of a comprehensive model capable of systematically categorizing the design dimensions of XR systems with a focus on artistic aspects, alongside a limited understanding of musical interactions within shared and collaborative virtual environments.

Concurrently, only a limited number of studies have identified and defined the key constructs necessary for evaluating immersive musical applications, and there is a lack of systematic methodologies — both quantitative and qualitative — capable of assessing these dimensions in a robust and reproducible manner.

Also, it is known that haptic elements help develop musical skills, in addition to being significant for artistic performance and contributing to the creation of more intuitive and easy-to-use applications (O'MODHRAIN; CHAFE, 2000; MARSHALL; WANDERLEY, 2011). In light of this, there is

a growing body of research investigating the role of tactile feedback in enriching musical experiences (TURCHET; ROSAIA, et al., 2025; TURCHET; WEST; WANDERLEY, 2019). This approach spans a diverse array of applications, including the use of low-frequency vibrotactile stimulation to intensify sensations of aesthetic enjoyment (TURCHET; WEST; WANDERLEY, 2021), as well as the analysis of how subwoofer-generated vibrations influence motor behavior in artistic contexts (VENKATESAN; WANG, 2023). Furthermore, several studies highlight the enhancement of musical engagement through haptic elements when compared to the exclusive use of headphones (BALANDRA et al., 2019), along with improved performance in audiometric tasks facilitated by tactile stimulation (TURCHET; ROSAIA, et al., 2025).

Nevertheless, developing compelling and convincing haptic experiences through vibrotactile feedback still represents a significant challenge in immersive musical performances. Even though recent work (SCHNEIDER; MACLEAN, 2016; STROHMEIER et al., 2020) has focused on manipulating low-level parameters such as frequency and amplitude, offering precise technical control, it remains challenging to translate these abstract parameters to musical contexts (SCHNEIDER; MACLEAN, et al., 2017; SEIFI; MACLEAN, 2017). In addition, such tools rarely support rapid prototyping methods and do not allow for the direct and intuitive mapping of vibrotactile feedback to users' spatiotemporal interactions in VR environments (DEGRAEN et al., 2021).

To address the aforementioned challenges, this chapter first establishes a theoretical foundation that outlines the conceptual guidelines for designing an immersive Io3MT environment, serving as a unifying metaphor for its structure and operation. This conceptual model specifies the interaction design, defines the relationships among the different types of information that constitute the system. Subsequently, the behavior of the haptic elements is defined, detailing their operational role within the virtual space and the enabling technologies required to implement these functionalities. Finally, a proof of concept, named PhysioDrum, is presented to operationalize the proposed guidelines and features,

6.2 Focus Group

A focus group is a qualitative research technique commonly employed in the fields of social sciences, communication studies, and user-centered design. Its primary objective is to examine participants' perceptions, opinions, beliefs, and attitudes toward a specific topic, product, or experience. The method typically involves a structured or semi-structured discussion with a small group of individuals, enabling an in-depth exploration of subjective and interactive dimensions that may not be readily captured through quantitative approaches. One of its principal strengths lies in the collective dynamics it fosters, which can stimulate the emergence of new ideas, support the joint construction of meanings, and reveal points of disagreement or consensus among participants. This approach is particularly valuable for generating insights into user needs and informing design decisions (KRUEGER, 2014).

In this study, the focus group comprised four experts, three of whom were women and one a man.

Specialist 1 (S1) is a postdoctoral researcher with experience as a developer and evaluator of XR systems, and also a drummer. Specialist 2 (S2) is a master's student, proficient in VR, and a guitarist. Specialist 3 (S3) holds a master's degree, is a bass player, and has experience in the development and use of XR for gaming. Specialist 4 (S4) holds a master's degree, is a 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. 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 specialists introducing themselves, followed by the moderator presenting an outline of the workshop flow. Subsequently, specialists proceeded to sign the consent form;
- **Warm-up activity (10 minutes):** during the warm-up, specialists 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, specialists 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 specialists 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 specialists reflected on XR-based artistic experiences, addressing implementation challenges and future research directions.

A thematic analysis (CLARKE; BRAUN, 2014) was performed to extract insights from the focus group data, providing a comprehensive understanding of specialists' perceptions and experiences. Dovetail[1] 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 applied the same eight tags used by (LEE; STRINER;

[1] https://dovetail.com/

CÉSAR, 2022) to categorize the opinions and feelings expressed by the focus group specialists. These tags were adopted because the referenced study addresses a similar focus — the intersection of VR and musical practice — and employs a method that has been scientifically validated. Table 6 provides a summary of each theme and primary functionalities.

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

Table 6: Summary of thematic analysis.

Following qualitative analysis of the focus group, the author repeatedly referred back to the existing literature presented in Sections 3.3 and 3.4 to compare the findings with documented perspectives. This process led to adjustments in the model, revisions of terms, and the incorporation of new concepts based on additional suggestions received. After the workshop, two experts provided supplementary documents, videos, and ideas. Data collection was ceased once saturation was reached, as further analysis of the materials and data processing did not result in any new insights.

6.2.1 Design Guidelines for Immersive Io3MT Environments

Several common points were observed both in the literature analysis and in the focus group discussions. Overall, the positive aspects of XR interactions encompass the ability to explore various applications from multiple perspectives, engaging environments, freedom of navigation, and immersive experiences leading to prolonged enjoyment. On the other hand, areas for improvement include enhancing the precision and accuracy of control to ensure effective performance. Discomfort in use, attributed to weight, heat generation, and wired connections of the HMD, was also identified as an area for improvement, along with limited interactions and features, and mobility restrictions. As gaps to be addressed, the inclusion of progressive stages of interaction and integration of multimodal resources

and physicality to enhance immersion were observed. The outcome of this discussion led to the formulation of five design principles aimed at creating an immersive Io3MT environment. They are presented and elaborated subsequently (VIEIRA; WEI, et al., 2024).

6.2.1.1 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.

Another meaningful aspect to consider is the interaction with the system, which should be realistic and close to real-world practice. For this reason, relying solely on simple button clicks as the only form of command input should be avoided. As traditional sound creation often involves physical interactions such as blowing, plucking, strumming, squeezing, stroking, hitting, or bowing, it is recommended that these aspects be incorporated into the proposed environment as a means of expressing intentionality. The advantages stemming from this approach encompass improvements and expansions in performance possibilities (COOK, 2009). This also enhances communication with the audience, providing visual cues about the correlation between gestures and musical outcomes, offering a channel for a better interpretation of the performer's energy and intent. To achieve these results, it is possible to draw inspiration from existing instruments and expand their capabilities, rather than solely creating entirely new applications (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

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 (SERAFIN; ERKUT; KOJS; NILSSON, et al., 2016).

6.2.1.2 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 (LI; CÉSAR, 2023).

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 (SERAFIN; ERKUT; KOJS; NILSSON, et al., 2016). 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 (SERAFIN; ERKUT; KOJS; NILSSON, et al., 2016).

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.

Numerous additional facets contribute to enhancing the immersiveness of the system, encompassing user interactions facilitated by multisensory, multimedia, or musical elements. Also, user-object interactions may occur through controls, joysticks, VRMIs, sensory devices, and new interfaces for musical expression. Further, aesthetic factors, such as visual attractiveness and colorfulness, influence the immersive experience. The coupling of gesture and sound within the digital space allows these interactions to be recontextualized based on the specific environment in which they take place. For instance, the audio response may vary depending on whether the illustrated scenario is an opera, a studio, or a garden. The symbiosis between sound and visual experience is also a significant consideration. Thus, the capacity for customization and the creation of unique profiles, reflecting user preferences, the ability to choose instruments or creative tools (software or programming languages), and pre-defined user settings, contributes to a personalized and engaging experience in Io3MT environments. By seamlessly integrating superior audio with visual elements, artists not only enhance the immersive quality of the experience but also demonstrate a commitment to delivering a polished and professional service.

6.2.1.3 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.

6.2.1.4 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 (LI; CÉSAR, 2023). 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.

6.2.1.5 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 different scenarios, 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.

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

6.3 Technical Implementation of PhysioDrum

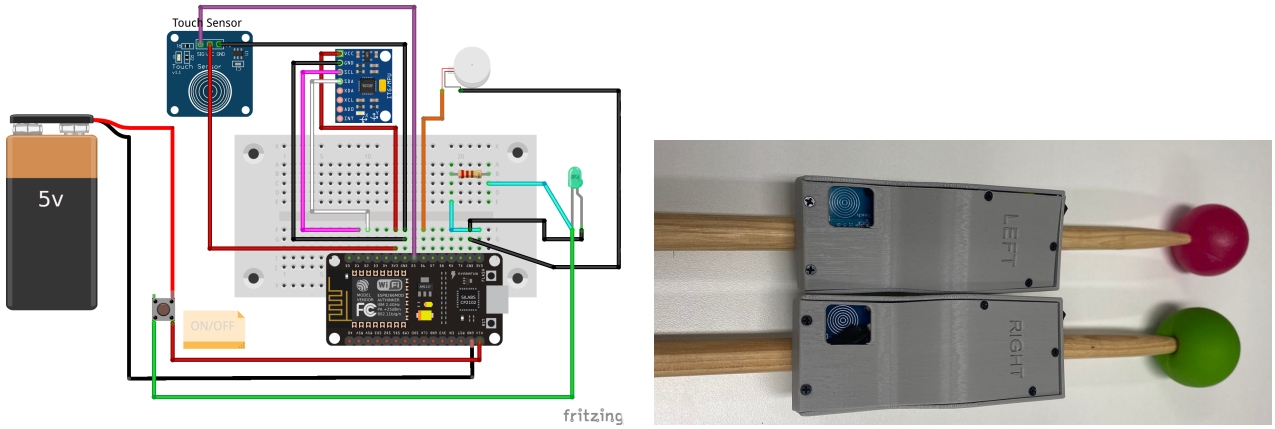
In accordance with the design guidelines for creating an immersive Io3MT environment, PhysioDrum was prototyped[2][3] (VIEIRA; WEI, et al., 2024; VIEIRA; MUCHALUAT SAADE; CÉSAR, 2025). This multisensory and immersive musical platform served as both a central experimental tool and the primary source of empirical data.

For the implementation, Unity (version 2022.3.30) was used in conjunction with the Meta Quest 3 and the Meta Interaction SDK (version 1.3.2). The virtual environment provides a first-person perspective of a drum kit that replicates the components of a traditional acoustic set, including a bass drum, snare drum, floor tom, two rack toms, a ride cymbal, and a hi-hat. Key functional requirements include minimal latency between hitting the drum and sound generation, minimal striking errors, and appropriate sound volume and quality.

Interaction with the virtual drum is mediated by two foot pedals and RemixDrum (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023), presented in Chapter 5. RemixDrum was adapted for this specific usage scenario by integrating colored spheres at the top of the drumsticks. The movements of these markers are tracked and processed through a computer vision algorithm developed in Python 3.8, which analyzes their displacement and converts it into digital commands. This process enables real-time synchronization between physical actions and the corresponding virtual environment interactions.

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

Additionally, a coin-type vibration motor was incorporated into the instrument's electronic circuit to provide haptic feedback, enhancing users' sensory perception of collisions within the virtual drum set. This new version also minimizes user inconvenience by simplifying actuator placement and reducing excessive cables. Furthermore, it is easily re-configurable with straightforward input-output mapping. Figure 11 illustrates the electrical circuit that integrates all components, along with the final version of the high-fidelity RemixDrum prototype.



(a) RemixDrum electrical circuit adapted to include vibration motor.

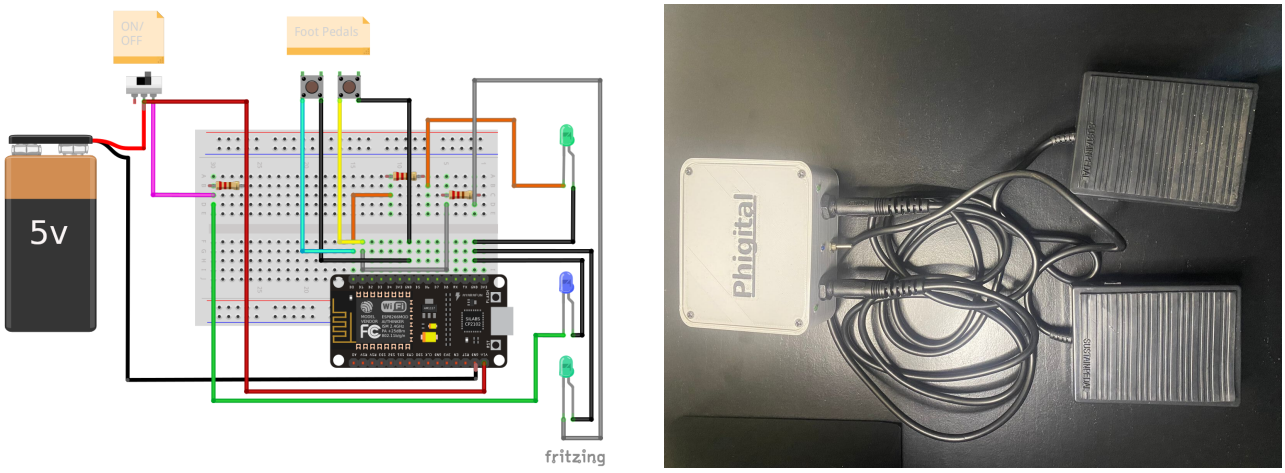
(b) New version of RemixDrum.

Figure 11: Complete RemixDrum system developed for this study (VIEIRA; MUCHALUAT SAADE; CÉSAR, 2025).

The foot pedals used in the system are connected to an electrical circuit that detects when they are pressed and sends this information to the virtual application through OSC messages, encapsulated in UDP packets. These devices simulate the behavior of traditional drum pedals, triggering the sound of the bass drum and hi-hat. Although the physical structure and ergonomics of the pedals differ from those of traditional models, they were designed to replicate the intended functions in a consistent manner, contributing to a responsive interaction within the system. Figure 12 illustrates the electrical circuit employed in this setup, along with the final configuration of the system with the pedals connected to the enclosure.

This operational mode utilizes physical movements to control and activate elements in the digital world, grounded in the concept of phygital (physical + digital) (MELE et al., 2023). As the term suggests, this approach merges physical and digital processes, creating connections and networks that link these two domains to enable new functionalities and forms of interaction. Consequently, the system offers a transparent and intuitive interaction experience, connecting physical actions with their corresponding virtual responses.

The sound generated by the drums is combined with the synthesized audio in Pure Data, while graphical interactions in Processing create a multimodal remix of content. 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 13 illustrates the architecture of PhysioDrum.



(a) Electrical circuit for pedal input detection. (b) Case with connected electronic pedals.

Figure 12: Pedal system developed for PhysioDrum environment (VIEIRA; MUCHALUAT SAADE; CÉSAR, 2025).

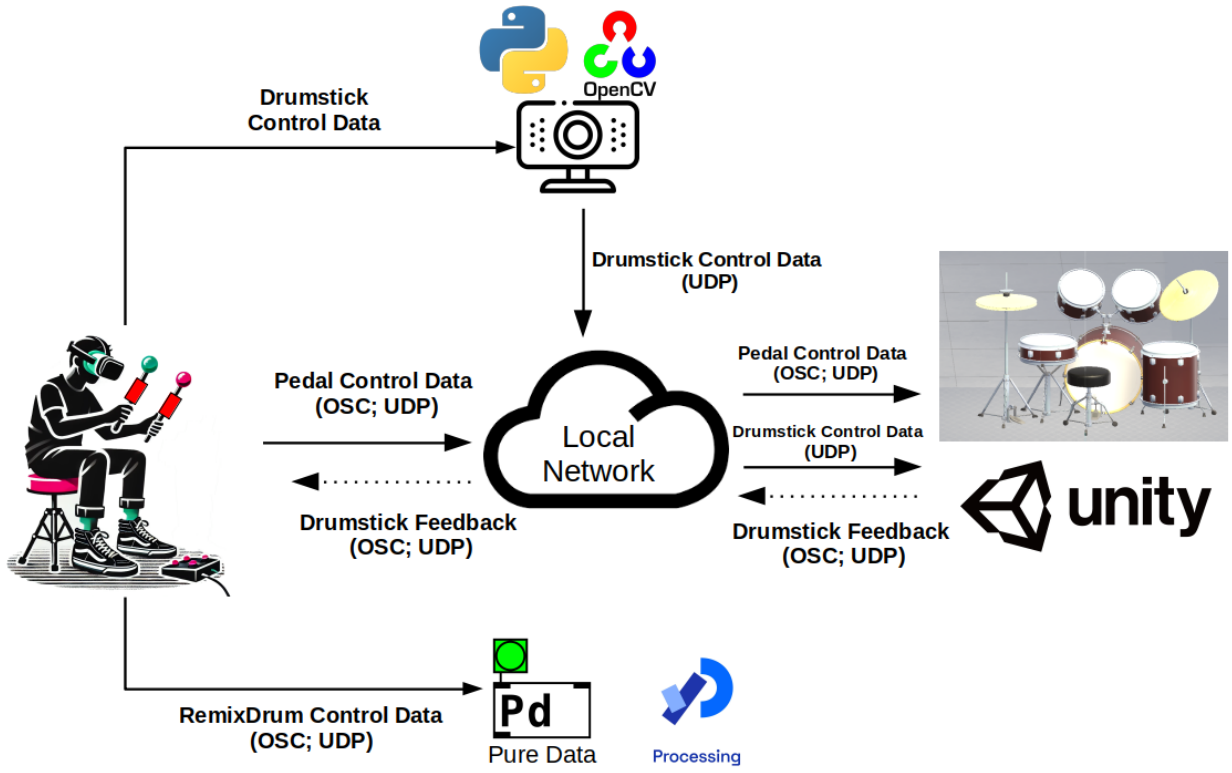


Figure 13: System architecture of PhysioDrum, depicting the integration and communication flows between software modules, hardware components, and supporting technologies that constitute the platform (VIEIRA; MUCHALUAT SAADE; CÉSAR, 2025).

6.4 Discussion & Lessons Learned

Upon completing the development of PhysioDrum in accordance with the proposed design guidelines, several key aspects can be identified based on the application's features. 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 (WILLEMSEN; HORVATH; NASCIMBEN, 2020; FEICK et al., 2021).

Simultaneously, (HAYES, 2011) 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 (OKADA et al., 2014; ZAVERI et al., 2022). 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 (SUEN et al., 2015). 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 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 (VIEIRA; SAADE; CÉSAR, 2023). 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.

6.4.1 Influence of the Io3MT Reference Model on the Design of PhysioDrum

The development of PhysioDrum was deeply informed by the Io3MT reference model proposed in Chapter 4. First, the layered orientation of the Io3MT architecture directly influenced the modular organization adopted in PhysioDrum. The separation between sensing, processing, communication, and rendering stages enabled the independent evolution of each subsystem, drumstick sensing, haptic actuation, audiovisual rendering, and network transmission, while preserving interoperability among heterogeneous components. This organization not only facilitated integration with external devices such as RemixDrum and foot pedals but also supported the adaptation of third-party tools, including

Unity, Pure Data, and Python-based computer vision modules, in accordance with the reference model's emphasis on loosely coupled design.

Second, the reference model's emphasis on phygital integration, a core characteristic of Io3MT environments, was operationalized in PhysioDrum through its explicit bridging of physical gestures and digital responses. The smart drumsticks, augmented with motion-tracking markers and vibrotactile actuators, embodied the bidirectional flow defined in the model: physical actions modulated digital events, while virtual collisions triggered physical stimuli. This mirrored the Io3MT principle that sensory and multimedia channels should interact symmetrically to support expressive, embodied, and meaningfully coordinated interactions.

Moreover, the Io3MT requirement of supporting multisensory rendering informed the implementation of haptic feedback as a first-class modality in PhysioDrum. Rather than functioning as a mere add-on, the tactile information was incorporated according to the model's guidelines for sensory alignment, multimodal consistency, and temporal synchronization. This ensured that tactile cues were not only perceptually coherent but also served as an active component in the musical and interactive experience, reinforcing gestural learning and enhancing the sense of presence.

The reference model also highlights the importance of interactivity, expressive control, and low-latency communication, which were treated as key operational constraints in PhysioDrum. For instance, the minimization of the delay between drumstick movement and sound rendering followed directly from the Io3MT network performance guidelines, which stress the necessity for predictable, time-sensitive responses in musical environments. Similarly, the system's architecture was designed to maintain accurate synchronization between auditory, visual, and haptic events, an alignment essential for achieving the high levels of immersion and embodiment described in the Io3MT framework.

Additionally, the proposed model informed decisions on system extensibility and evolvability. PhysioDrum was deliberately implemented with modular mappings, configurable inputs, and interchangeable sensory devices, reflecting the Io3MT emphasis on flexibility to accommodate future musical scenarios, alternative hardware configurations, and multi-user expansions. In this sense, the system functions not only as a single application but also as a demonstrator of how the Io3MT model can support hybrid architectures capable of real-time interaction across music, multimedia, and multisensory components.

Finally, the reference model influenced the design philosophy that underpinned PhysioDrum, especially regarding the role of physicality, user embodiment, and expressiveness in immersive musical environments. By drawing from the theoretical guidelines and ontology of interactions established in earlier chapters, PhysioDrum operationalized the Io3MT concept of a multisensory, multimodal participatory space, permitting performers to explore artistic expression through coordinated layers of gesture, sound, vision, and touch.

6.5 Conclusion

This chapter presented the theoretical foundations and design principles underpinning the creation of an immersive environment guided by the concepts of the Io3MT. To this end, a focus group was conducted with four experts, all with backgrounds in technology and amateur experience in music, with the goal of identifying the key aspects to be considered in the design of an immersive musical environment oriented toward performance.

Based on these guidelines, the PhysioDrum system was developed as a phygital application that enables physical devices to perform actions within a virtual environment while receiving real-time sensory feedback. In this configuration, PhysioDrum preserves the same affordances as traditional drumming practice, employing the smart drumsticks RemixDrum for hand interaction and two electronic pedals for foot control. Special attention was devoted to the implementation of haptic feedback, allowing the drumsticks to vibrate according to their collisions with the virtual drum components.

This operational model concretely implemented the principles of the Io3MT, creating an immersive environment that integrates physicality, interactivity, and musical expressiveness. By articulating theoretical concepts with design decisions grounded in literature and expert contributions, it was possible to establish an operational model capable of bridging physical and digital domains through haptic interfaces, sensors, and multimodal feedback mechanisms.

The conception of PhysioDrum provides a platform that reduces spatial and financial barriers while broadening creative possibilities in both artistic and educational contexts. Moreover, the integration of tactile and visual devices has the potential to enhance the sense of presence, reinforce gestural coherence, and expand the performer's expressiveness within immersive environments.

Concluding this conceptual and implementation phase, the following chapter presents the practical evaluation of PhysioDrum, in which the system is subjected to empirical user experiments. The next chapter discusses the methodological procedures adopted, the data collection instruments, and the quantitative and qualitative analyses that examine the effects of haptic feedback on immersive experience and musical performance.

7 PhysioDrum: Evaluation Protocol, Experimental Design and Results

This chapter reports on the experimental evaluation of PhysioDrum, conducted to assess user experience within an immersive Io3MT environment. A dedicated evaluation protocol was developed to identify the key constructs underlying the quality of experience in such contexts, combining quantitative and qualitative methods within a mixed methodological framework. The study involved 30 participants who were randomly assigned to two experimental groups. Group A received differentiated haptic feedback according to the specific drum component being played, whereas Group B received uniform feedback across all components.

The experimental procedure comprised three sequential phases. Initially, participants engaged in a free exploration period designed to promote familiarization with the system and its interactive dynamics. Subsequently, they performed four rhythmic exercises of increasing complexity, distributed across two conditions — one incorporating haptic feedback and another conducted without it. The presentation order of the tasks was randomized to minimize potential bias and learning effects.

This design enabled a comprehensive examination of PhysioDrum's performance, allowing for the identification of its strengths and limitations, as well as revealing directions for future refinement and further research on the role of haptic feedback in immersive musical interaction.

7.1 Protocol for User Experience Assessment in Immersive Io3MT Environments

Despite the growing body of research on user experience in computer-music-oriented environments, no reference model has yet been established to define the key evaluation constructs, particularly in the context of immersive applications. Moreover, a systematic framework for guiding both qualitative and quantitative assessments of such systems remains absent.

To address this gap, a literature review was conducted using four prominent scientific research databases in the computing field: IEEE Xplore, ACM Digital Library, Springer, and Science Direct. This review sought to identify the key constructs and methodological approaches employed in assessing UX within immersive music applications. This process comprised three stages: mapping the state-of-the-art, categorizing the selected studies, and critically analyzing the resulting findings (VIEIRA; MUCHALUAT SAADE; ROCHA, et al., 2023).

In the first stage, once again, a keyword-based search was conducted across IEEE Xplore, ACM Digital Library, Springer, and Science Direct. The search strings employed were: "*Virtual Reality Musical Instrument AND user experience*", *Virtual Reality Musical Instrument AND metric OR construct*", "*Virtual Reality Musical Instrument AND evaluat*[1]", "*Virtual Reality AND music performance AND user experience*", and "*Virtual Reality AND metric OR construct*".

The inclusion criteria comprised peer-reviewed studies written in English, with more than three pages, published from 2010 onwards, and whose title, abstract, or conclusions clearly addressed the investigated theme, particularly aspects related to user experience in virtual musical instruments or immersive music environments. Exclusion criteria included studies not indexed in recognized scientific databases, publications not written in English, and preliminary documents such as white papers and preprints that had not undergone formal peer review.

After applying these filters, the sample size was 65 publications. Moreover, the snowball technique was employed, analyzing references cited in the selected articles. This process expanded the final sample to 90 publications, as summarized in Table 7.

Scientific Database	**Total**
ACM Digital Library	4
IEEE Xplore	47
Springer Link	9
Science Direct	5
Other Sources	25

Table 7: Articles related to the search for evaluation methods in immersive musical environments.

In the second stage, the selected works were categorized based on their thematic focus or application domain, resulting in six distinct classes. The first category, Extended Reality, includes studies employing VR, AR, or MR technologies to enable immersive musical experiences. These works examined presence, plausibility, spatial perception, emotional response, and embodied interaction in interactive 3D environments.

The New Interfaces for Musical Expression (NIME) category encompasses research dedicated to designing and evaluating novel digital interfaces for musical performance. These interfaces often incorporate sensors, gesture recognition, touch-sensitive surfaces, and computational strategies to expand musicians' expressive capabilities.

Interactive Music Learning comprises studies focused on music education and game-based or playful learning. These works explored tools ranging from AR tutorials to rhythm games and gamified platforms to teach musical concepts, support instrumental practice, and foster cognitive and affective skills.

[1]Only the prefix of the word was searched, in order to encompass its various forms, such as evaluation, evaluating, evaluate, and related variations.

The Music and Healthcare/Therapy category includes research applying interactive musical technologies for therapeutic or well-being purposes, such as emotional regulation, relaxation, sensorimotor rehabilitation, and affective expression.

Networked Collaborative Music Performance refers to studies addressing remote musical collaboration between geographically distributed participants, emphasizing temporal synchronization, copresence, expressivity, and engagement in teleperformance contexts.

Lastly, Generic Multimodal Interfaces category studies propose or analyze systems that integrate multiple sensory modalities (visual, auditory, tactile), even when not directly applied to conventional musical performance.

In the third and final stage, each thematic cluster was analyzed to identify its main contributions and relevance to the research community, particularly regarding the clarification of evaluated constructs and the use of practical models for qualitative and quantitative assessment.

This analysis identified 139 distinct measurement scales and 236 unique constructs. Most of the scales (81.29%) and constructs (80.51%) appeared only once across the reviewed studies. The most frequently used instruments were the System Usability Scale (SUS) (n = 11), the Presence Questionnaire (PQ) (n = 9), the Simulator Sickness Questionnaire (SSQ) (n = 7), and the NASA Task Load Index (NASA-TLX) (n = 6). Other recurrent methodologies included the Game Experience Questionnaire (GEQ) (JOHNSON; GARDNER; PERRY, 2018), the Self-Assessment Manikin (SAM) (BRADLEY; LANG, 1994), AttrakDiff (VIEIRA; PROVIDÊNCIA; CARVALHO, 2023), the Flow State Scale (JACKSON; MARSH, 1996), and the Creativity Support Index (CARROLL; LATULIPE, 2009). Semi-structured interviews were also widely adopted (n = 22).

The results reveal that only a small proportion of the studies presented a complete rationale for the selection of the scales employed, with merely 32.22% doing so. Moreover, only 29.78% reported all the items included in the experiment. In addition, 41.11% of the scales were adapted in various ways, including translation, abbreviation, or semantic modification. Despite these adjustments, only 36.67% of the studies conducted psychometric analyses, which include measures of internal consistency (e.g., Cronbach's alpha), test-retest reliability, or construct validity.

Among the 90 studies analyzed, 32 reported the use of custom-designed scales (35.56%), with five of them describing the development of entirely new evaluation tools (5.56%). In contrast, 86 studies (95.56%) employed at least one established methodology. Of these, 52 studies (57.78%) relied exclusively on questionnaires previously validated in the literature, while 30 (33.33%) combined standardized tools with custom-designed ones. Only four studies (4.44%) used solely self-developed evaluation methods.

Regarding the constructs assessed, a total of 236 unique constructs were coded across 731 instances of measurement. For example, the construct usability appeared 63 times across the articles, representing 8.62% of all occurrences. On average, each study addressed 8.12 distinct constructs (mode = 6; minimum = 1; maximum = 17). The majority of constructs were mentioned only once (190 out of

236, or 80.51%), with an average of 3.10 mentions per construct (mode = 1; minimum = 1; maximum = 63). The most frequently reported constructs were:

- Usability (63 times or 8.62%);
- Enjoyment (51 times or 6.98%);
- Presence (44 times or 6.02%);
- Engagement (41 times or 5.61%);
- Generic UX (37 times or 5.06%);
- Emotion (31 times or 4.24%);
- Frustration (26 times or 3.55%);
- Aesthetics (21 times or 2.87%);
- Motivation (16 times or 2.19%);
- Enchantment (12 times or 1.64%).

The results of the systematic review indicate that, although standardized instruments such as the SUS and PQ are widely adopted, user experience assessment methods remain highly fragmented. A considerable diversity of evaluation tools was observed, many of which are informal or developed *ad hoc*, highlighting the need for greater methodological convergence and improved replicability in user experience evaluations within immersive and interactive musical contexts.

In light of these findings, the decision was made to assess the four most frequently reported constructs in the literature: usability, enjoyment, presence, and engagement. The scales selected for measuring these aspects were those with the highest incidence in the reviewed studies and supported by evidence of validity, reliability, and sensitivity. The following methods are outlined below.

7.1.1 Simulator Sickness Questionnaire (SSQ)

Interactions in VR, particularly those mediated by HMDs, can induce cybersickness, a condition characterized by symptoms such as nausea, dizziness, vertigo, sweating, among others, which may compromise users' immersion and engagement in such environments. These symptoms are generally attributed to a sensory conflict, in which visual cues signaling motion are incongruent with the vestibular system's perception of the body as stationary (SEVINC; BERKMAN, 2020; BALK; BERTOLA; INMAN, 2013).

Given the potentially adverse effects of this phenomenon on virtual applications, it is essential to assess participants' physiological conditions after interacting with the system. In this context, the Simulator Sickness Questionnaire (SSQ) (KENNEDY et al., 1993) was developed. Originally designed

to evaluate sickness symptoms in fighter jet pilots, the instrument was later adapted for use in computer systems. The SSQ evaluates sixteen symptoms across four categories: nausea (N), oculomotor disturbances (O), disorientation (D), and a total score (TS). Users rate each symptom on a four-point scale, identifying them as absent, mild, moderate, or severe. This provides a quantitative measure of the adverse effects experienced during virtual reality interactions. The questionnaire is administered both before and after the experience, focusing on the physiological responses and perceptual discrepancies induced by Immersive Virtual Environments (IVEs). The goal of administering the SSQ prior to the experiment is to assess the user's "usual state of fitness" before exposure to the stimuli, allowing for a comparison to determine whether the virtual experience negatively impacted them (YONKOV, 2024).

To compute the scale scores, each symptom variable was multiplied by its corresponding weight, and the resulting weighted values were summed within each column to obtain the weighted total. Nausea, oculomotor, and disorientation scores were subsequently derived from these totals using the conversion formulas presented in Equation 7.1[2](KENNEDY et al., 1993). The total score was obtained by summing the totals of all subscales and applying the corresponding formula. This computational procedure yields reliable estimates of the overall severity of simulator sickness, as well as sufficiently robust subscale scores for diagnostic purposes. In general, higher values on each scale reflect stronger perceptions of the associated symptoms, and are therefore considered undesirable outcomes (BIMBERG; WEISSKER; KULIK, 2020).

$$\begin{aligned} N &= [1] \times 9.54 \\ O &= [2] \times 7.58 \\ D &= [3] \times 13.92 \\ TS &= ([1] + [2] + [3]) \times 3.74 \end{aligned} \tag{7.1}$$

Given that the target audience of this study consists of native Brazilian Portuguese speakers, the translated version of the SSQ, presented in Appendix C, was employed (SEVINC; BERKMAN, 2020; GONSALEZ; ALMEIDA, 2015).

7.1.2 Presence Questionnaire (PQ)

The effectiveness of an IVE is intrinsically linked to the user's perception of presence and immersion. Presence denotes the subjective experience of "being there" in a specific environment. This perception is shaped by the interplay of sensory stimulation, application features that foster involvement, and individual predispositions that facilitate engagement. Immersion, in contrast, refers to a psychological state characterized by a sense of engagement, inclusion, and active interaction with an environment that delivers a continuous flow of stimuli and experiences (WITMER; SINGER, 1998).

Owing to the critical importance of these constructs for virtual and immersive applications, the

[2]Parentheses were not part of the original equation and were added here solely to clarify its structure.

Presence Questionnaire (PQ) (WITMER; SINGER, 1998; SILVA et al., 2016) was developed to assess them through 33 items rated on a 7-point Likert scale. The instrument evaluates four primary factors associated with presence: Engagement, comprising items that measure the extent to which attention and mental effort are directed toward coherent or meaningfully related stimuli, activities, or events; Adaptation/Immersion, denoting the sense of being engaged, integrated, and interacting with a variety of stimuli; Sensory Fidelity, concerning the visual, auditory, and tactile perceptions of the VR environment; and Interface Quality, which examines the influence of visual and control interfaces on the immersive experience.

Subsequently, the authors of this method conducted statistical analyses to evaluate the internal consistency of the items. The results indicated that items 26, 27, and 28 reduced the overall reliability of the scale, leading to their removal in order to ensure stable and interpretable subscales, while preserving the instrument's conceptual comprehensiveness. As a result, this study employs version 3.0 of the PQ, which comprises 29 questions and represents the optimal balance between comprehensive coverage of the "presence" construct and robust metric properties in the validation sample (WITMER; JEROME; SINGER, 2005). The Brazilian Portuguese version (SILVA et al., 2016) of the questionnaire was adopted for this research, with both the original and translated versions provided in Appendix D.

Calculating the PQ enables both the individual analysis of each scale and the computation of an overall score for the user experience. In the first method, items are grouped according to the specific construct they measure, namely: Involvement/Control (items 1, 2, 3, 4, 5, 6, 13, 14, 15, 16, 22), Naturalness (items 7, 8, 9, 10, 17, 18, 19), Interface Quality (items 11, 12, 20, 21), Auditory Realism (items 23, 24, 25), Haptic Realism (items 26, 27, 28), and Visual Fidelity (item 29).

After completing the distribution, a simple average is calculated for each subscale. The same method is used to determine the final score, which considers all 29 items in the questionnaire. The values range from 0 to 7. Higher scores reflect better performance on each subscale as well as on the overall result.

7.1.3 System Usability Scale (SUS)

Usability refers to the overall quality of an artifact, system, or application in fulfilling its intended purpose. To evaluate this attribute, the System Usability Scale (SUS) was developed in 1986. The instrument comprises ten items, each assessed using a Likert scale. It is recommended that the SUS be administered immediately after participants have interacted with the application, and before any debriefing or discussion takes place. Respondents should also be encouraged to record their immediate reactions to each statement, avoiding excessive reflection or reconsideration of their responses (BROOKE et al., 1996; BROOKE, 2013).

The SUS encompasses a range of aspects, including the need for support, training requirements, and perceived complexity. Several of its items are aligned with Nielsen's heuristics, addressing elements such as ease of learning (items 3, 4, 7, and 10), efficiency (items 5, 6, and 8), ease of memorization

(item 2), error minimization (item 6), and user satisfaction (items 1, 4, and 9). Its advantages include robustness and versatility, as well as the ability to generate a single, easily interpretable score. In addition, the SUS is straightforward to administer and demonstrates high reliability across different contexts (BROOKE et al., 1996). As with the other questionnaires, the version of the SUS employed in this study was the Brazilian Portuguese translation, as presented in Appendix E (LOURENÇO; CARMONA; MORAES LOPES, 2022).

To calculate the SUS score, for each odd-numbered item, one point is subtracted from the user's rating, whereas for each even-numbered item, the assigned score is subtracted from five. The resulting values for all ten items are then summed and multiplied by 2.5, yielding a final score ranging from 0 to 100. In general, scores equal to or above 68 are considered indicative of acceptable usability.

Using the SUS score as a reference, five complementary interpretations can be derived through comparisons with the SUS benchmark database, which compiles results from a wide variety of applications and systems evaluated using the same methodology (LEWIS; SAURO, 2018; GRIER et al., 2013). The first interpretation relies on percentile ranking, indicating the proportion of systems that achieved equal or lower performance.

The second interpretation applies the academic grading system traditionally used in North American educational institutions to the classification of system usability. Within this method, the evaluation is expressed in letter grades ranging from A, representing the highest level of usability, to F, denoting inadequate performance, while grade C indicates average performance.

The third classification is the adjective rating method, which employs descriptive terms rather than numerical values to express system usability. This approach utilizes six descriptors, such as "Excellent", "Good", "OK", and "Poor", to qualitatively convey the user's perception of the system. In turn, system acceptability is assessed through verbal categories, classifying the system as acceptable, marginally acceptable, or not acceptable.

The fifth and final classification method is the Net Promoter Score (NPS). This metric evaluates the likelihood of users recommending the system to others and classifies them into three distinct groups: promoters, passives, and detractors. Promoters are individuals who express a strong intention to recommend the application within their personal or professional network, reflecting a high level of satisfaction and engagement. Detractors, in contrast, are those inclined to discourage others from using the system, often due to negative experiences or unmet expectations. Passives adopt a neutral position, neither actively recommending nor dissuading potential users.

Table 8 provides a summary of all the classifications discussed, indicating the corresponding SUS score ranges associated with each category.

Grade	SUS	Percentile Range	Adjective	Acceptability	NPS Category
A+	84.1-100	96-100	Best Imaginable	Acceptable	Promoter
A	80.8-84.0	90-95	Excellent	Acceptable	Promoter
A-	78.9-80.7	85-89		Acceptable	Promoter
B+	77.2-78.8	80-84		Acceptable	Passive
B	74.1-77.1	70-79		Acceptable	Passive
B-	72.6-74.0	65-69		Acceptable	Passive
C+	71.1-72.5	60-64	Good	Acceptable	Passive
C	65.0-71.0	41-59		Marginal	Passive
C-	62.7-64.9	35-40		Marginal	Passive
D	51.7-62.6	15-34	OK	Marginal	Detractor
F	25.1-51.6	2-14	Poor	Not Acceptable	Detractor
F	0-25	0-1.9	Worst Imaginable	Not Acceptable	Detractor

Table 8: Mapping of SUS scores to corresponding grades, percentile ranges, adjective ratings, acceptability levels, and Net Promoter Score (NPS) categories (GRIER et al., 2013).

7.1.4 NASA Task Load Index (NASA-TLX)

A user's behavior within a system is influenced not only by the demands imposed by the system itself but also by the individual's perceptions regarding the expected actions, the strategies adopted during use, and the physical and mental effort required for interaction. To assess workload-related aspects in a given environment, the NASA Task Load Index (NASA-TLX) (HART, 2006) is employed. This instrument comprises six subscales, each representing a set of relatively independent variables: Mental Demands (cognitive effort required to perform the task), Physical Demands (extent of physical activity necessary to complete the work), Temporal Demands (degree of time pressure experienced while performing the task), Frustration (emotional factors that hinder performance, such as insecurity, irritation, lack of stimulation, and stress), Effort (combined physical and mental exertion required to achieve a desired performance level), and Performance (self-assessment of the success achieved when using the application or system). The version of the questionnaire employed in this study, translated into Brazilian Portuguese (CIOFI-SILVA et al., 2023), is provided in Appendix F.

Each dimension is rated on a 21-point scale, enabling users to convey the intensity of the workload across various aspects of the task in a stepwise manner. This multidimensional approach offers insights into the elements that contribute to the overall task load, making it easier to identify areas for optimization and improvement (YONKOV, 2024).

The NASA-TLX score can be calculated by either assigning weights to the scales or by calculating the simple averages of the six scales, referred to as the raw score. The latter method has become more

popular in academic studies because it eliminates the need for weight matching, which can be laborious and ineffective. Additionally, this approach reduces inconsistencies that may arise when the same scale is assigned different weights at different moments, thus preserving the logic of the assessment. It also mitigates statistical noise that can occur when combining ordinal scales, a problem associated with the composite method (HART, 2006; SAID et al., 2020; VIRTANEN et al., 2022; BOLTON; BILTEKOFF; HUMPHREY, 2023).

7.1.5 Haptic Questionnaire (HQ)

Haptic technology is increasingly recognized as a key component in optimizing user experience. In virtual reality environments, for instance, haptic feedback has been shown to enhance the sense of presence (AZMANDIAN et al., 2016; BERGER et al., 2018). Nevertheless, there is currently no standardized methodology capable of quantitatively assessing the contribution of this modality to the usability of digital systems, nor one that provides clear guidelines for design improvements grounded in the use of haptic elements (SATHIYAMURTHY et al., 2021).

Building on this premise, (SATHIYAMURTHY et al., 2021) propose a 22-item questionnaire, employing a simple Likert scale, to evaluate the haptic experience (HX) across five constructs: harmony, expressiveness, autotelic, immersion, and realism. The construct of harmony refers to the absence of noise or disruptions during the experience, indicating that the haptic feedback is integrated with the other elements of the environment, rather than appearing disjointed or inconsistent.

Expressiveness refers to the capacity of haptic feedback to exhibit variation and convey subtle nuances, enriching the overall sensory experience. Autotelia assesses the extent to which the feedback is intrinsically enjoyable and satisfying, possessing inherent value independent of the task being performed. Immersion describes the user's degree of involvement, focus, and engagement, as well as the bidirectional influence between the individual and the environment mediated by the feedback. A stronger sense of reciprocity corresponds to a deeper immersive state. Lastly, realism concerns the credibility of the haptic feedback, specifically its ability to appear convincing and congruent with the sensations and responses expected in a real-world context.

Each subscale is computed as the arithmetic mean of the items corresponding to its respective construct, while the overall score is obtained from the arithmetic mean of all 22 items in the questionnaire. This distinction between subscale-level and overall analyses enables a more precise identification of the specific aspects that positively influenced the experience and those that may require improvement.

The questionnaire employed in this study is presented in Appendix G. As this is a relatively recent methodology, a formally validated Brazilian Portuguese version is not yet available. Therefore, the version adopted in this work corresponds to a direct translation carried out by the author of this thesis.

Moreover, the application of appropriate statistical methods in experimental research is essential to ensure the reliability of results and the validity of inferences. These methods reinforce the

methodological rigor of the study, allowing for a more accurate analysis of participants' subjective perceptions collected through quantitative questionnaires. In this regard, the present thesis also employs a set of statistical techniques — including the Mann–Whitney test (MCKNIGHT; NAJAB, 2010), the Wilcoxon signed-rank test (REY; NEUHÄUSER, 2011), effect size measures (ROSENTHAL; COOPER; HEDGES, et al., 1994), Cliff's delta (δ) (RAZUMIEJCZYK; MACBETH, 2011), Spearman's rank correlation coefficient (ρ) (ZAR, 2005), and Cronbach's alpha (α) (TAVAKOL; DENNICK, 2011) — to achieve a more precise interpretation of response variations across all administered questionnaires. Further details on these methods are provided in Appendix B.

7.1.6 Semi-structured Interview

Quantitative questionnaires provide valuable information on user behavior across various dimensions of an application. However, they often lack the capacity to explain the underlying reasons for users' responses. An effective complementary approach is the semi-structured interview. This qualitative research method involves engaging participants in purposeful oral communication, guided by an interview that contains questions aligned with the study's objectives. Rather than following a rigid sequence, the guide is designed to offer structure and direction to the conversational flow, enabling the interviewer to adapt the discussion to the participant's context and responses. Typically, a semi-structured interview protocol comprises open-ended lead questions accompanied by follow-up prompts, which the interviewer can use to explore topics in greater depth as the conversation unfolds (ILOVAN; DOROFTEI, 2017).

One of the main advantages of semi-structured interviews is their ability to balance focus with flexibility, enabling the researcher to maintain a structured line of inquiry while also exploring relevant topics that may emerge during the conversation. This adaptability can enhance the depth and breadth of understanding regarding the evaluated service. Moreover, when combined with other instruments, such as in a mixed-methods approach, this technique can complement quantitative data by providing richer and context-specific perspectives. The set of questions used as the basis for this study is presented in Appendix H.

7.1.7 Experimental Setup

To evaluate PhysioDrum, a laboratory-based study was conducted at Fluminense Federal University, in Brazil, from May 5th to 27th, 2025, involving 30 participants who were randomly assigned in equal numbers to two groups. Group A interacted with a configuration providing uniform haptic feedback, whereas Group B experienced distinct haptic patterns for different elements of the drum kit. The order of these conditions was also randomized to control for order effects and reduce systematic bias.

Table 9 provides a summary of the drum elements and the corresponding haptic feedback patterns assigned to each piece. Given the absence of a standardized mapping between virtual musical instruments and their associated haptic responses, the feedback configurations adopted in this study

were defined empirically by the author.

Item	Function	Haptic Feedback
Hi-Hat	Pair of metal cymbals operated with a pedal, used to mark time with precision.	Single short vibration pulse (0.5s), simulating its sharp and quick sound.
Snare	Drum with metallic snares, producing a sharp and crisp sound.	Double short pulses (0.4s each) with a 0.2s interval.
Low Tom	Low-pitched tom used for rhythmic variations and fills.	Deep vibration burst (1.5s), reflecting its resonant low tone.
High Tom	Higher-pitched version of the Low Tom, used in melodic rhythmic fills.	Medium-duration vibration (1s) with a rising intensity profile, representing tonal ascent.
Ride Cymbal	Large cymbal sustaining groove and musical timing.	Long vibration pulse (2s), capturing its sustained shimmering resonance.
Floor Tom	Large floor tom producing a deep, sustained sound.	Continuous low-frequency vibration (3s), imitating prolonged drum resonance.
Bass Drum	Also known as the kick drum, producing a low-frequency sound via pedal.	N/A

Table 9: Description of drum kit items and corresponding haptic feedback patterns designed for the PhysioDrum platform.

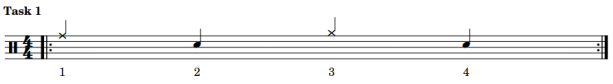
7.1.7.1 Participants

The participants' ages ranged from 18 to 54 years (Mean = 26.67, SD = 10.23), with the sample consisting of 24 men and 6 women. The group reflected diverse educational backgrounds, with concentrations in fields such as Computer Science, Engineering, and Psychology. Two reported advanced expertise in both music and technology (P13 and P22); six had prior VR experience (P02, P13, P20, P22, P23, P28); ten had substantial musical training (P01, P06, P07, P08, P12, P13, P18, P19, P22, P24); the remainder were novices in both domains.

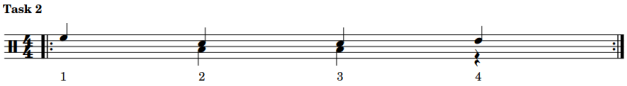
The study is part of the SenseGames Project (CAAE 88638025.0.0000.8160), registered by the Research Ethics Committee of Fluminense Federal University. All participants gave informed consent, and procedures adhered to institutional ethical standards, including confidentiality and the right to withdraw at any time.

7.1.7.2 Procedure

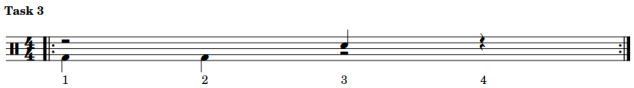
Participants were randomly assigned to two groups. Group A interacted with a configuration that provided uniform haptic feedback, whereas Group B experienced distinct haptic patterns corresponding



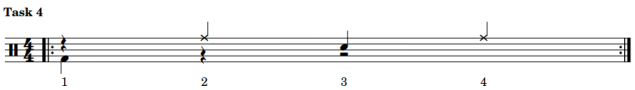
(a) Task 1 - *largo*.



(b) Task 2 - *adagio*.



(c) Task 3 - *andante*.



(d) Task 4 - *moderato*.

Figure 14: Musical patterns performed during the tests with PhysioDrum.

to different elements of the drum kit. Following this assignment, the testing procedures commenced and were organized into three sequential stages.

In the first stage, participants engaged in a three-minute free exploration session with the PhysioDrum system to familiarize themselves with motion dynamics and adapt to the system's audiovisual feedback.

In the second phase, participants were instructed to perform four standard musical tempo markings: *largo* (40 BPM), *adagio* (66 BPM), *andante* (76 BPM), and *moderato* (108 BPM). The rhythmic patterns associated with each tempo were designed to be accessible to beginners, while still introducing a gradual increase in complexity. They are displayed in Figure 14.

Each session had an average duration of approximately 15 minutes. Upon completion of the task, participants were asked to fill out questionnaires and take part in a semi-structured interview.

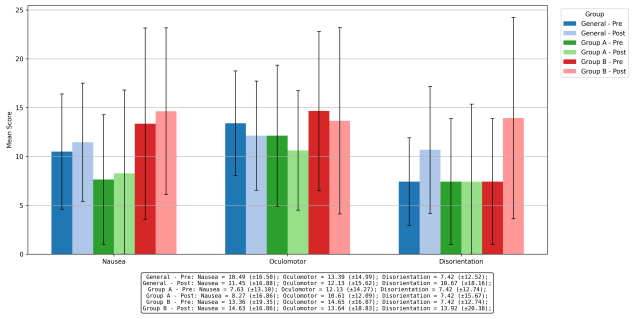
7.2 Data Analyses and Discussion

In this section, quantitative data from the questionnaires are analyzed and discussed, including descriptive measures and inferential statistical tests. The analysis examines patterns, trends, and differences in responses across experimental groups and participants profiles, providing empirical evidence for evaluating the PhysioDrum system.

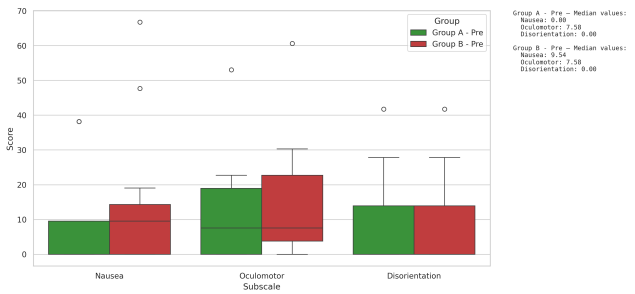
7.2.1 Assessment of Simulator Sickness Symptoms (SSQ)

Figure 15 presents the SSQ results, organized by subscale — Nausea, Oculomotor, and Disorientation — for each experimental group. The illustration contrasts pre and post-experiment measurements through a combination of bar charts and box plots, a dual representation that facilitates the identifi-

cation of distribution patterns, detection of outliers, and assessment of statistical trends. The overall scores remained within thresholds typically regarded as safe, indicating the absence of severe physiological symptoms. Nevertheless, statistically significant differences emerged between groups throughout the experiment, with the most pronounced effects observed in the Nausea and Disorientation subscales.



(a) Bar charts of SSQ subscale scores before and after the experiment for each group.



(b) Box plots illustrating the distribution and variability of SSQ scores.

Figure 15: Comparison of pre- and post-experiment results for the SSQ subscales, showing both aggregated scores by group (a) and the underlying distribution characteristics (b).

In the general analysis of the pre-experimental data, the distributions exhibited marked asymmetry, with substantial discrepancies between mean and median values, as well as high standard deviations indicative of outliers. This pattern is particularly evident in Figure 16, which presents individual responses for each subscale. For the Nausea subscale, the mean was 10.49 (SD = 16.50) compared to a median of 0.00, indicating that while most participants reported mild symptoms, a few extreme cases considerably affected the mean. The Oculomotor subscale showed the highest mean score (13.39), reflecting visual discomfort and eye strain prior to exposure to the virtual environment. By contrast, the Disorientation subscale presented a mean of 7.42 and a median of 0.00, reinforcing that only a small proportion of participants experienced more pronounced symptoms.

The Mann-Whitney U test results indicate no statistically significant differences between Groups A and B on the SSQ subscales assessed in the pre-test. Comparisons for nausea ($U = 87.00$; $p = 0.260$), oculomotor effort ($U = 100.50$; $p = 0.624$), and disorientation ($U = 112.50$; $p = 1.000$) revealed p-values above the adopted significance threshold ($\alpha = 0.05$), indicating that self-reported levels of nausea symptoms were statistically equivalent between the groups prior to the experimental intervention.

This is confirmed by the analysis of the effect size values. The Nausea subscale showed a small effect ($r = 0.1931$), while the Oculomotor subscale had a negligible effect ($r = 0.0909$), and Disorientation showed no difference between the groups ($r = 0$). The Cliff's Delta results corroborate this analysis. Nausea exhibited a small negative effect ($\delta = -0.2267$), indicating a tendency for Group B to present slightly higher scores on this subscale, although not significant. The Oculomotor ($\delta = -0.1067$) and Disorientation ($\delta = 0$) subscales, however, demonstrated a negligible effect, reinforcing the homogeneity between the groups at baseline.

The Spearman correlation analysis, depicted in Figure 17, offers further evidence regarding the par-

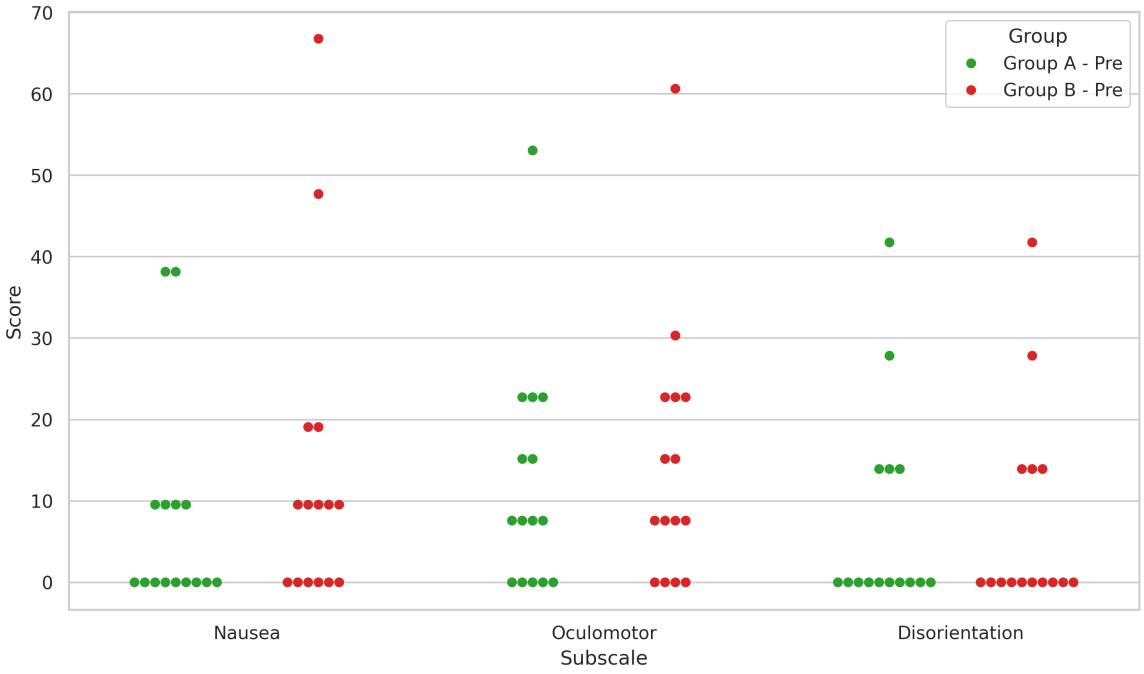
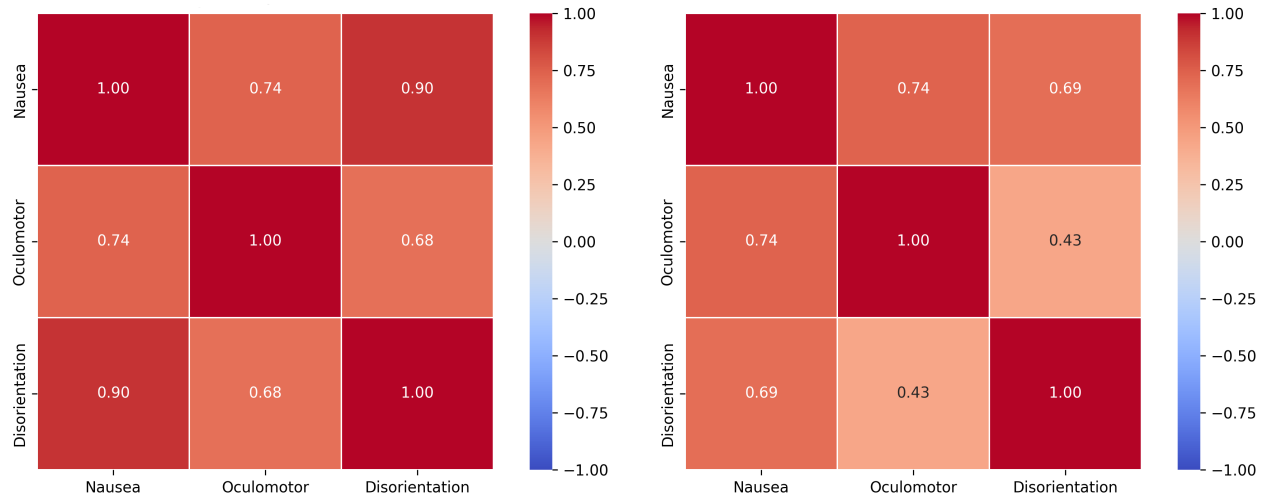


Figure 16: Individual pre-experiment SSQ scores for each subscale.

ticipants' perception of physiological symptoms prior to the experimental intervention. For Group A, a strong interdependence among the three subscales was identified, with all correlation coefficients being high and statistically significant ($p < 0.05$). Notably, the association between Nausea and Disorientation was particularly strong ($\rho = 0.90$, $p < 0.001$), indicating that higher nausea scores are consistently accompanied by higher disorientation scores. The p-value < 0.001 denotes that the likelihood of this correlation arising from random variation is below 0.1%, thereby confirming its high statistical significance. This result suggests that, within this group, cybersickness symptoms tend to emerge in a coupled and mutually reinforcing manner, reflecting a relatively homogeneous symptom profile.

In opposition, the results for Group B revealed a more heterogeneous pattern. The correlation between Nausea and Oculomotor remained high ($\rho = 0.74$), indicating a strong and statistically significant positive relationship between these dimensions. In contrast, the association between Oculomotor and Disorientation was substantially weaker ($\rho = 0.43$) and did not reach statistical significance ($p = 0.111$). This result suggests that, in this case, although a positive trend is present, it may be attributable to random variation and, therefore, cannot be regarded as statistically reliable. Such dissociation between symptoms may reflect heterogeneous physiological sensitivity profiles among participants. These findings indicate that Group B exhibits a less cohesive symptom structure, which may influence individual responses to the sensory and motor demands imposed by the PhysioDrum system. The divergence in the correlation structure between the groups underscores the importance of considering pre-existing individual profiles as potential moderating factors in the response to immersive technologies.

Following the scale-by-scale analysis, the total pre-experiment score was examined, as illustrated in



(a) Spearman correlation matrix for SSQ subscales in Group A (pre-test).

(b) Spearman correlation matrix for SSQ subscales in Group B (pre-test).

Figure 17: Spearman correlation analysis of SSQ subscale scores prior to the experimental intervention for Groups A and B.

Figure 18. The general mean score was 117.10 (SD = 148.34). A Mann–Whitney test was conducted to evaluate potential differences in the total SSQ scores between Group A and Group B prior to the experimental intervention. The analysis revealed no statistically significant difference between the groups ($U = 95.0$; $p = 0.474$), indicating that self-reported cybersickness symptoms were comparable across participants from both groups in the pre-test phase.

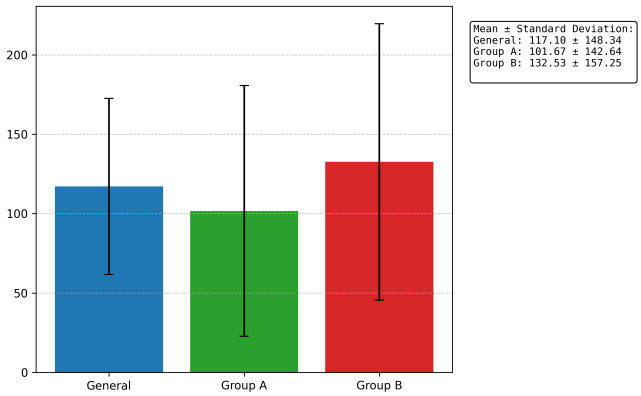
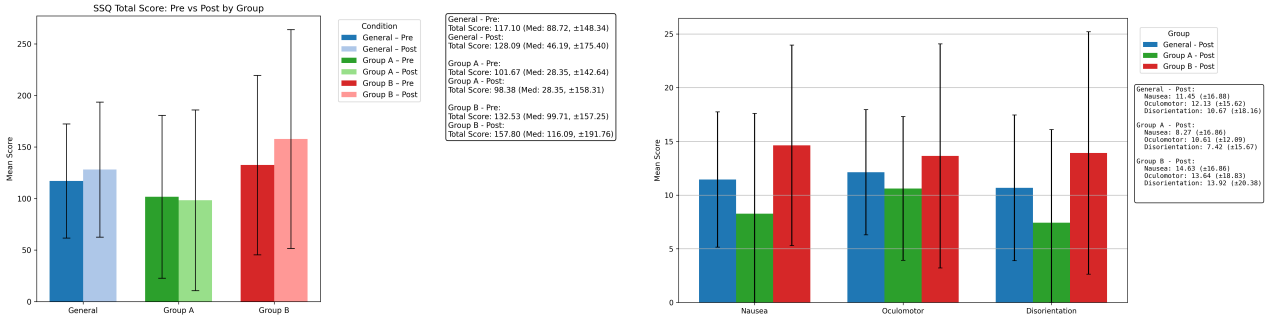


Figure 18: SSQ Total scores in the pre-test phase.

The convergence between the analyses allows for some important interpretations within the context of this study. The absence of significant differences between groups A and B in the pre-test, along with the minor or negligible effects observed, reinforces the equivalence of the groups at baseline, thereby ensuring the validity of the experimental design. At the same time, the symptomatic cohesion observed in group A, in contrast to the dissociation noted in group B, may provide valuable understanding about how different groups respond to subsequent immersive experiences, particularly in relation to perceptual integration or fragmentation. It is also noteworthy that the moderate physiological symptoms reported prior to the immersive experience may have been influenced by external factors,

such as commuting, engagement in physically and/or cognitively demanding activities, or pre-existing fatigue.

After completing the pre-experiment analysis, a similar assessment was conducted using the data obtained after the interaction with the PhysioDrum system. The total score, presented in Figure 19, exhibited a slight increase, rising from 117.10 to 148.34. This variation, however, warrants cautious interpretation. Although the mean value increased, the median score (46.19) indicates that the majority of participants maintained low levels of symptoms, with the observed increase being driven primarily by a limited number of isolated cases. Among the subscales, Disorientation contributed most substantially to the increase, with the mean rising from 7.42 to 10.67. Nonetheless, the median remained at 0.00, suggesting that these symptoms were concentrated in a small subset of participants. The Nausea and Oculomotor dimensions exhibited only minor fluctuations.



(a) Total scores of the Simulator Sickness Questionnaire (SSQ) in the post-test phase.

(b) Subscale scores of the Simulator Sickness Questionnaire (SSQ) in the post-test phase.

Figure 19: Results of the Simulator Sickness Questionnaire (SSQ) for Group A and Group B after the experimental intervention.

The results of the Wilcoxon signed-rank test for the entire sample, comprising all 30 participants, revealed no statistically significant differences between the scores obtained in the pre and post-experiment sessions for any of the SSQ scales analyzed. In the Nausea subscale, the Wilcoxon signed-rank test yielded $W = 25.00$ and $p = 0.1448$, a result that does not reach statistical significance but indicates a slight upward trend in some individuals. Nevertheless, the overall distribution pattern remained consistent across the two evaluation conditions, suggesting that the use of PhysioDrum did not result in an increase in this symptom.

The Oculomotor subscale demonstrated a highly stable performance, with $W = 56.50$ and $p = 0.5481$. This finding suggests that interaction with the system's visual interface did not place excessive demands on participants' visual processing capabilities, nor did it elicit discomfort within this functional domain.

For the Disorientation subscale, the Wilcoxon signed-rank test registered $W = 6.00$ and $p = 0.1605$. Although this result does not reach statistical significance, it may reflect increased individual sensitivity to this symptom, as already illustrated in Figure 15, with certain participants exhibiting elevated deltas, that is, reporting higher symptom scores following interaction with the tool. Nevertheless, the group median remained unchanged, and the overall distribution pattern does not indicate a substantial

risk of disorientation at the group level.

The Total SSQ score, representing the weighted sum of the three subscales, resulted in $W = 89.50$ and $p = 0.3653$, indicating that self-reported symptoms of simulation sickness did not increase significantly subsequent to exposure to the virtual environment. In light of these results, it can be concluded that the mean values remained within the range classified as mild, with no evidence of progression to moderate or severe symptom categories.

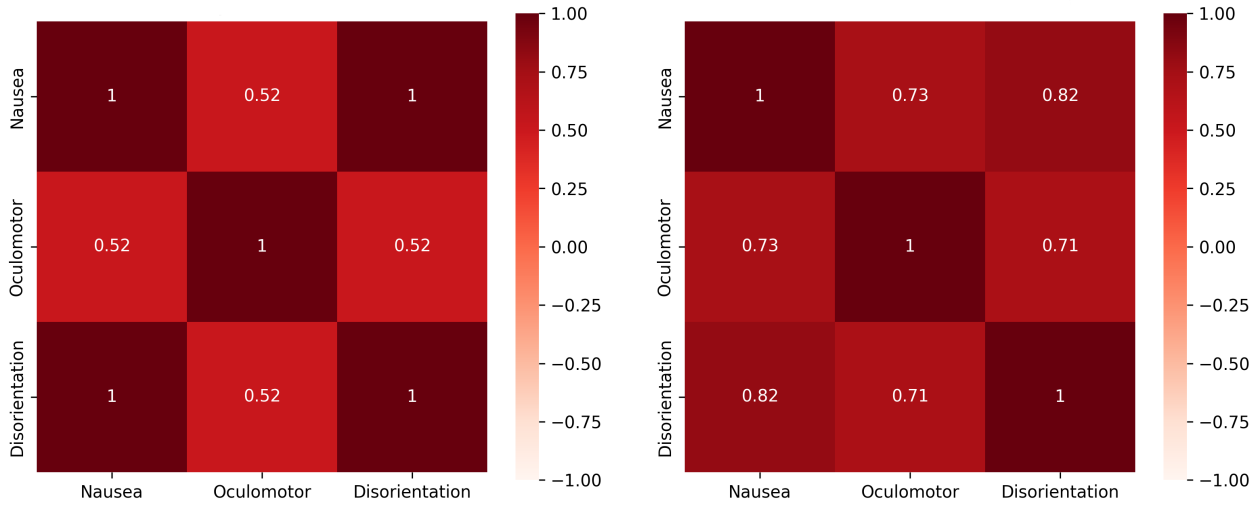
Afterwards, the Mann–Whitney test was applied to examine potential differences in the total SSQ score between the two experimental conditions after the intervention. The test produced $U = 84.00$ with $p = 0.239$, indicating the absence of statistically significant differences. Likewise, no significant differences were observed for the Nausea dimension ($U = 73.50$; $p = 0.081$), the Oculomotor subscale ($U = 104.00$; $p = 0.728$), or the Disorientation measure ($U = 89.00$; $p = 0.268$).

Based on Spearman's correlation analysis for the post-test (Figure 20), an intensification of the associations among symptoms was observed in both experimental conditions, although with different cohesion patterns. In Group A, post-test correlations presented moderate magnitudes among the measured dimensions ($\rho = 0.52$), except for a perfect correlation between Nausea and Disorientation ($\rho = 1.00$). This value, which may be influenced by the sample size, indicates a direct relationship between these two variables in some participants. Compared to the pre-test, the correlations decreased in magnitude. Before the task, all associations were strong ($\rho = 0.68$ to $\rho = 0.90$), with higher values for the relationships between Nausea and Disorientation ($\rho = 0.90$) and between Nausea and Oculomotor ($\rho = 0.74$). This reduction suggests that, after the immersive experience, the symptoms tended to manifest in a less interdependent manner.

In Group B, the correlations between the subscales increased. This configuration implies that participants in this group tended to experience symptoms concurrently and in an interrelated manner. The strong co-occurrence of symptoms may also be associated with increased susceptibility to cybersickness or with a more integrated perception of discomfort during the task.

After the experiment, participants in Group B exhibited a more interdependent configuration of symptoms compared to those in Group A. This variation in perception between the groups, when comparing different sessions, suggests that the type of haptic feedback may influence the perceptual organization of cybersickness. A specific haptic effect could have contributed to system regulation, facilitating user adaptation and, consequently, mitigating physiological symptoms. These observations hold potential relevance for the development of personalized and adaptive sensory systems, particularly in immersive and music-related applications.

Posterior to the general analysis of behavior in different phases, an intragroup analysis was performed to identify variations in individual responses. The Wilcoxon test results for Group A indicated no statistically significant differences between the initial and final experimental sessions. For the Nausea dimension, the statistic was $W = 9.00$ with $p = 0.3795$, demonstrating stability in responses across the two measurement points. The individual-level graphical representation (Figure 21) supports this



(a) Spearman correlation matrix for the SSQ subscale scores in Group A (post-test).

(b) Spearman correlation matrix for the SSQ subscale scores in Group B (post-test).

Figure 20: Spearman correlation analysis of SSQ subscale scores in the post-test phase for Group A and Group B.

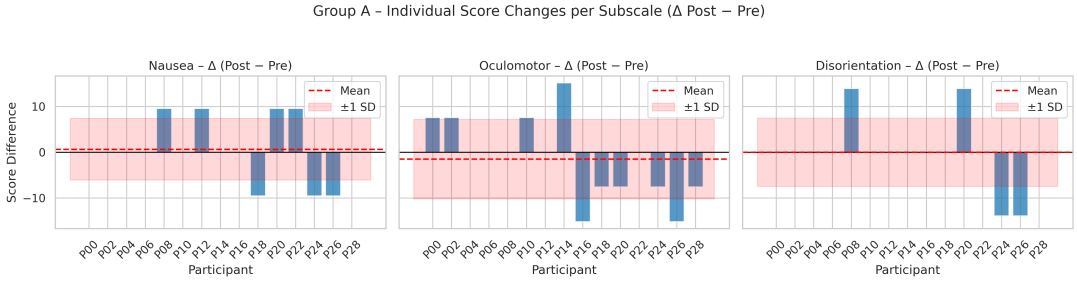
interpretation. Although some participants, such as P08, P12, P20, and P22, exhibited pronounced increases in symptom levels, others — including P18, P24, and P26 — reported substantial reductions. The mean delta values approximate zero, with the majority of scores distributed near to the standard deviation indicating no discernible collective tendency toward either deterioration or improvement. The coexistence of positive and negative deltas reflects individual variability, which offsets the average group behavior, thereby characterizing a globally neutral physiological response to the system.

The Oculomotor dimension also presented no statistically significant differences between the two measurement points ($W = 20.50$, $p = 0.4694$). The mean delta values were marginally negative, indicating a minor tendency toward symptom reduction after interaction with PhysioDrum. Graphical analyses reveal that most individuals either maintained comparable scores or demonstrated modest improvements, with occasional increases, as in the case of participant P14. Nevertheless, the mean variation remained close to zero, and the dispersion of values was relatively uniform, suggesting that exposure to the system did not lead to increased visual strain for the majority of participants.

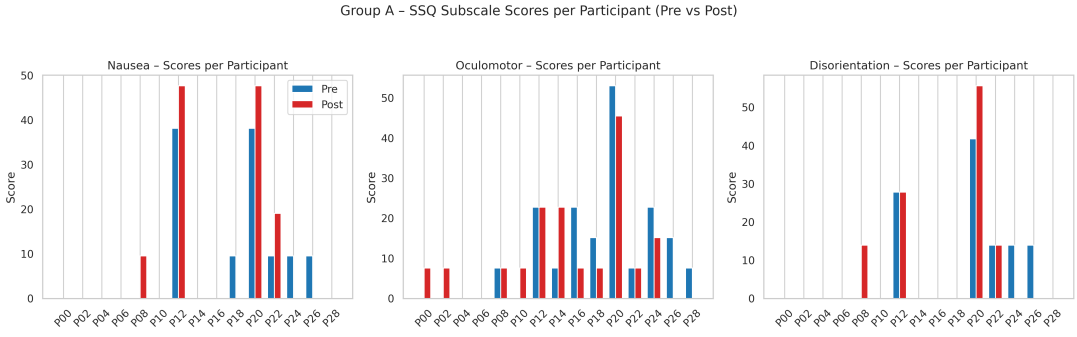
In the Disorientation dimension, the results demonstrate a high degree of stability. The Wilcoxon test produced $W = 4.00$ with $p = 0.7055$, reinforcing the absence of statistically significant changes between the initial and final sessions of the intervention. The graphical representation supports this interpretation: while participants P08 and P20 exhibited increases in symptoms, P24 and P26 reported substantial reductions, with negative deltas exceeding the standard deviation. The distribution of deltas is symmetrical and centered, with no extreme outliers, confirming the robustness of physiological tolerance to the system in this domain.

In the analysis of the total SSQ score for Group A, the Wilcoxon signed-rank test revealed no statistically significant difference between the two measurement points ($W = 45.00$, $p = 0.9720$). The

mean values remained virtually unchanged, with minor individual fluctuations mutually offsetting in the overall average.



(a) Individual changes in SSQ subscale scores for Group A.



(b) SSQ subscale scores per participant in Group A.

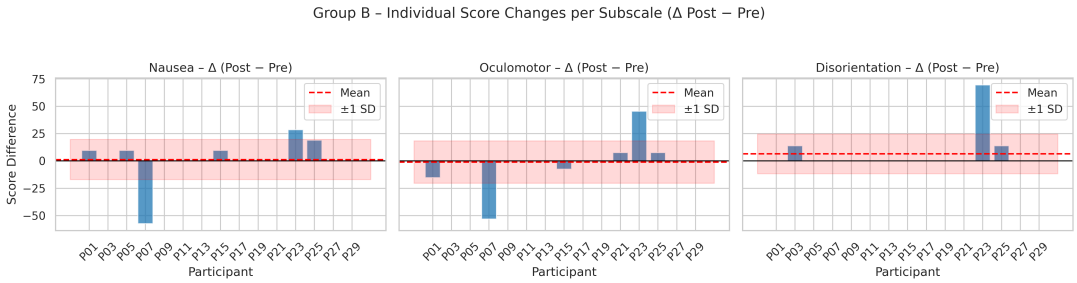
Figure 21: Individual variations and scores of the SSQ for Group A.

For Group B, the evaluation of outcomes after exposure to the PhysioDrum system also indicated no statistically significant differences between the two measurement points, as evidenced by the Wilcoxon test, with $W = 6.00$ ($p = 0.344$) for Nausea, $W = 10.00$ ($p = 0.916$) for Oculomotor, and $W = 0.00$ ($p = 0.102$) for Disorientation.

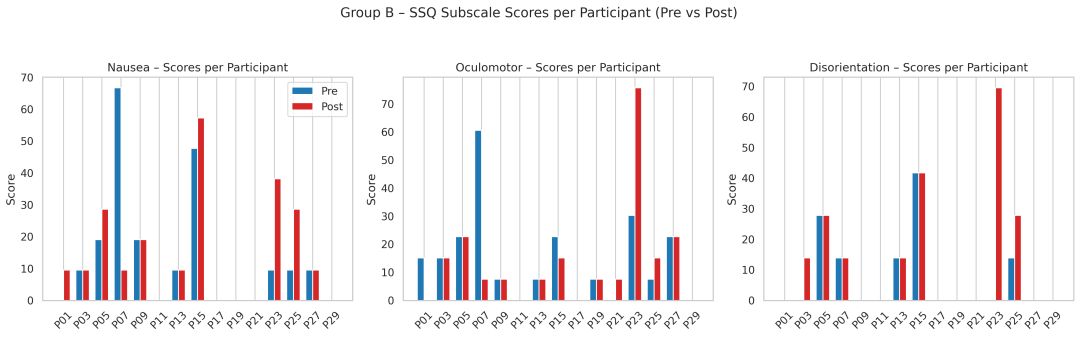
The graphical representations (Figure 22) corroborate the statistical interpretation. In the Nausea dimension, individual variations are observed, such as the decrease in participant P07 and the increases in P23 and P25. The mean delta values remain close to zero, with the standard deviation encompassing most participants, indicating that these variations are within the expected range and do not represent a collective trend toward symptom increase.

In the Oculomotor dimension, the variation was heterogeneous. Participant P07 showed a decrease, whereas P23 registered an increase. Despite these changes, the dispersion of values does not indicate a systematic bias at the group level.

For the Disorientation dimension, most participants maintained stable scores, except for increases in P23 and P25. The absence of statistical significance ($p = 0.102$), combined with the stability or minor changes observed in the remaining participants, supports the interpretation that these effects are participant-specific and not representative of the group.



(a) Individual changes in SSQ subscale scores for Group B.



(b) SSQ subscale scores per participant in Group B.

Figure 22: Individual variations and scores of the SSQ for Group B.

The post-test results showed an increase in the Nausea and Disorientation dimensions, particularly in Group B. This increase was associated with symptom intensification in one subgroup (P05, P15, P25) and with a pronounced worsening in a single case (P23). Nonetheless, no statistical evidence was found to support a collective aggravation of simulation sickness symptoms following interaction with PhysioDrum. These findings underscore the relevance of accounting for individual characteristics and baseline conditions of users in the design and evaluation of immersive systems, as external factors may exert a significant influence on the reported outcomes.

7.2.1.1 Comparison Between Musicians and Non-Musicians

In order to extend the scope of the analysis, a comparative evaluation was conducted between participants who self-identified as having musical expertise (P01, P06, P07, P08, P12, P13, P18, P19, P22, P24) and those without such experience. The procedure adopted was consistent with the methodology applied in the previous comparisons, encompassing the examination of total scores and subscale results, in addition to the assessment of statistical dependence between these measures.

In the aggregated analysis, which does not differentiate participants by experimental group, no statistically significant differences were observed between musicians and non-musicians in any of the subscales or in the total SSQ score. Nevertheless, Cliff's delta values indicate small to moderate effects, particularly for Oculomotor ($\delta = 0.220$), Disorientation ($\delta = 0.190$), and the Total SSQ score ($\delta = 0.255$), with musicians presenting slightly higher scores even during the initial measurement phase (see Table 18 - Appendix C).

Upon stratification of the data by experimental group, distinct patterns were identified. In Group A, comprising six musicians (P06, P08, P12, P18, P22, P24) and nine non-musicians, the musician subgroup consistently exhibited higher mean and median scores across all subscales. Cliff's delta coefficients indicated moderate effects for Nausea ($\delta = 0.407$) and for the Total SSQ score ($\delta = 0.370$), suggesting that, even prior to the experimental manipulation, musicians presented a greater incidence of this specific symptom as well as an overall less favorable condition (see Table 19 - Appendix C). In this context, the absence of statistical significance is mitigated by the effect size estimates, which point to potentially meaningful practical differences warranting further investigation in studies with larger and more balanced samples.

Conversely, in Group B, which comprised four participants with musical training (P01, P07, P13, P19) and eleven without such background, the results were more homogeneous. Cliff's delta values were close to zero across all subscales (e.g., $\delta = -0.045$ for Nausea), and p-values remained high ($p > 0.5$), indicating the absence of a systematic trend in responses between the subgroups (see Table 20 - Appendix C). This homogeneity suggests that, within this cluster, musical experience did not act as a moderating factor in the perception of cybersickness.

Subsequent to the experimental phase, additional analyses were conducted considering the results obtained post-exposure to the PhysioDrum system. In the aggregated analysis, no statistically significant differences were observed between musicians and non-musicians in any of the subscales ($p > 0.5$ in all cases). Cliff's delta values remained small ($\delta < 0.130$), supporting the hypothesis that, although musical training was associated with higher baseline values, it was not sufficient to modulate the intensity of symptoms reported post-immersion (see Table 21 - Appendix C).

It is important to note that, despite the sample imbalance in Group B, the statistical analysis remains valid. This is attributable to the adoption of nonparametric tests, which do not assume normal distribution or homogeneity of variance between groups. Furthermore, the interpretation was complemented by effect size measures, enabling the assessment of the practical impact of the observed differences, even in the absence of statistical significance. Such an approach is recommended in exploratory studies with small samples, as in the present evaluation of PhysioDrum, whose objective is not population-level inference but rather the identification of patterns and trends within the exploratory scope of the study. Accordingly, despite the numerical asymmetry, the comparison yields relevant preliminary insights into potential effects that different participant profiles may have on system perception. Nevertheless, the results must be interpreted with caution, acknowledging the sample size limitation and recommending future replication with more balanced groups to strengthen the statistical robustness of the conclusions.

7.2.1.2 Comparison Between Musicians in Group A and Group B

An additional analysis was conducted to compare the behavior of musicians across the two experimental conditions. Specifically, the scores assigned by the six self-reported musicians in Group A were compared with those of the four counterparts in Group B. The results indicated that none of the SSQ

subscales showed statistically significant differences between the sets, with all p-values exceeding 0.8. Furthermore, the effect sizes were low, with Cliff's delta values ranging from $\delta = -0.125$ to $\delta = 0.083$ (see Table 22 - Appendix C). This minimal variation suggests that the musician cohorts were relatively equivalent in their initial symptom levels, regardless of their allocation in the experimental design.

Upon exposure to the PhysioDrum system, the direct comparison between musicians in Groups A and B revealed small differences in SSQ scores, with a slight advantage for participants in Group B (median = 75.89) compared to those in Group A (median = 86.40). Although the data did not reach statistical significance, the results suggest that the type of haptic feedback may modulate the immersive experience differently for participants with musical training (see Table 23 - Appendix C. These findings highlight the importance of aligning sensory feedback design with users' expectations and perceptual profiles.

7.2.1.3 Comparison Between VR Specialists and Non-Specialists

Consistent with the approach presented in the previous section, comparisons were conducted between participants with prior experience in VR (P02, P13, P20, P22, P23, P28) and those without such background. Considering the total sample of 30 participants, no statistically significant differences were identified in any of the SSQ subscales, with p-values exceeding 0.36 in all Mann–Whitney tests. Cliff's delta for the total SSQ score was $\delta = 0.229$, and for the Disorientation subscale, $\delta = 0.208$, both indicating small to moderate effects (see Table 24 - Appendix C). These results suggest that participants with VR experience, even prior to exposure to the PhysioDrum system, reported slightly higher levels for certain symptoms.

The intragroup analyses provided further evidence for this observation. In Group A, participants with prior VR experience presented higher median scores across all SSQ subscales compared to their non-experienced counterparts. The largest effects were observed for Disorientation ($\delta = 0.273$) and for the total score ($\delta = 0.182$). Although the tests did not reach statistical significance ($p > 0.39$), the effect size estimates a moderate practical difference (see Table 25 - Appendix C).

In Group B, the scores between VR specialists and non-specialists were closer, and Cliff's delta values were low across nearly all subscales (see Table 26 - Appendix C). The only exception was the total score, which still indicated a moderate effect ($\delta = 0.423$). These results suggest more homogeneous initial conditions between the two participant profiles analyzed in this cohort.

In the post-experiment analysis, no statistically significant differences were found between VR specialists and non-specialists (see Table 27 - Appendix C). Nonetheless, specialists presented higher median scores across all subscales and for the total score (e.g., Total SSQ: 133.93 vs. 32.02). The p-values were approximately 0.20, with moderate effect sizes, such as $r \approx 0.3$ and Cliff's delta ≈ 0.35, particularly for Disorientation ($\delta = 0.417$) and for the total score ($\delta = 0.347$). These findings indicate a consistent tendency toward higher symptom perception among VR specialists both before and after system use, suggesting that PhysioDrum did not exacerbate any specific symptom.

7.2.1.4 Comparison Between VR Specialists in Group A and Group B

The performance of participants with prior expertise in VR was also examined across the two experimental conditions. This comparison considered a subset comprising four VR specialists from Group A and two from Group B. In the baseline assessment, none of the SSQ dimensions exhibited statistically significant differences between these subsets, with all p-values exceeding 0.63. Effect sizes were negligible, with Cliff's delta values ranging from $\delta = -0.375$ (SSQ total) to $\delta = 0.125$ (Disorientation), indicating a relatively uniform distribution of initial simulation sickness symptoms among specialists in both experimental settings (see Table 28 - Appendix C). Thus, prior to exposure to the immersive environment, the discomfort profiles self-reported by VR specialists were comparable across the two conditions.

The post-experiment analysis revealed that VR specialists from Group B recorded substantially higher scores across all SSQ subscales. For the total score, the median reported by Group B specialists was 401.30, a value that markedly exceeded the 90.06 observed for their counterparts in Group A. The magnitude of this difference corresponded to a Cliff's delta of $\delta = 0.667$, which is indicative of a large effect size. At the subscale level, the results exhibited consistent patterns. In the Nausea subscale, the median score among Group B specialists was approximately four times greater than that of Group A specialists, with $\delta = 0.583$. Similarly, in the Disorientation subscale, Group B specialists also reported considerably higher scores, resulting in $\delta = 0.667$ (see Table 29 - Appendix C).

These results should be interpreted with caution, particularly for Group B, whose analysis was based on a reduced number of specialists ($n = 2$). The observed increase in cybersickness scores may not necessarily indicate an inherent limitation of the application, but rather a heightened level of interaction with the system by these users, which could have intensified their exposure to vestibular, visual, and haptic stimuli. Rather than serving solely as an assessment of the overall performance of PhysioDrum, this analysis provides insights into individual nuances in user behavior. Such findings offer valuable input for refining the design of haptic feedback and for calibrating the sensory elements of the system to better accommodate diverse user profiles.

7.2.1.5 Synthesis and Implications of SSQ Results

The SSQ analysis indicated that exposure to the immersive environment, whether employing uniform or varied haptic feedback, did not elicit significant physiological symptoms at the collective level. Scores remained predominantly within the range classified as mild, and statistical tests (Wilcoxon test for within-participant comparisons before and after the experiment, and Mann-Whitney test for comparisons between distinct participant groups and/or profiles) revealed no significant differences between conditions. Such evidence demonstrates the robustness of the system in terms of physiological tolerance.

Important nuances emerged from the subgroup analysis. Participants with prior experience in music or VR exhibited greater susceptibility to symptoms in specific conditions, particularly in Group B, which

incorporated varied haptic feedback. Within this group, correlations among SSQ subscales became more pronounced, indicating a more integrated and generalized symptom pattern. In contrast, Group A presented a reduction in correlation strength after the experience, suggesting that standardized feedback may modulate symptoms in a more predictable manner.

The profile-based examination highlighted the relevance of accounting for individual differences in the design of immersive systems. Musicians appeared more sensitive to feedback type, reporting higher discomfort in Group A but lower symptom intensity in Group B, suggesting that alignment between haptic stimuli and musical expectations may influence the perception of physiological responses. VR specialists reported higher scores in both groups, potentially reflecting heightened sensitivity to the coherence and naturalness of multisensory stimuli.

Consideration of users' sensory and experiential profiles is, therefore, crucial in the development of immersive experiences. Approaches encompassing the precise calibration of feedback, in conjunction with personalization mechanisms, continuous physiological monitoring, and adaptive sensory modulation, constitute a promising direction for advancing comfort and optimizing the effectiveness of creative and educational immersive applications. Such approaches aim to ensure that these sensory elements do not operate as potential triggers or amplifiers of physical discomfort.

7.2.2 Assessment of Presence Questionnaire (PQ)

Figure 23 presents the overall results for the groups under analysis, also specifying the mean values and their respective standard deviations. The data indicate that both groups achieved total mean scores consistent with the classification of high perceived presence, according to the conventional interpretation of the PQ scale: Group A: Mean = 5.25 (SD = 0.81) and Group B: Mean = 5.34 (SD = 0.66).

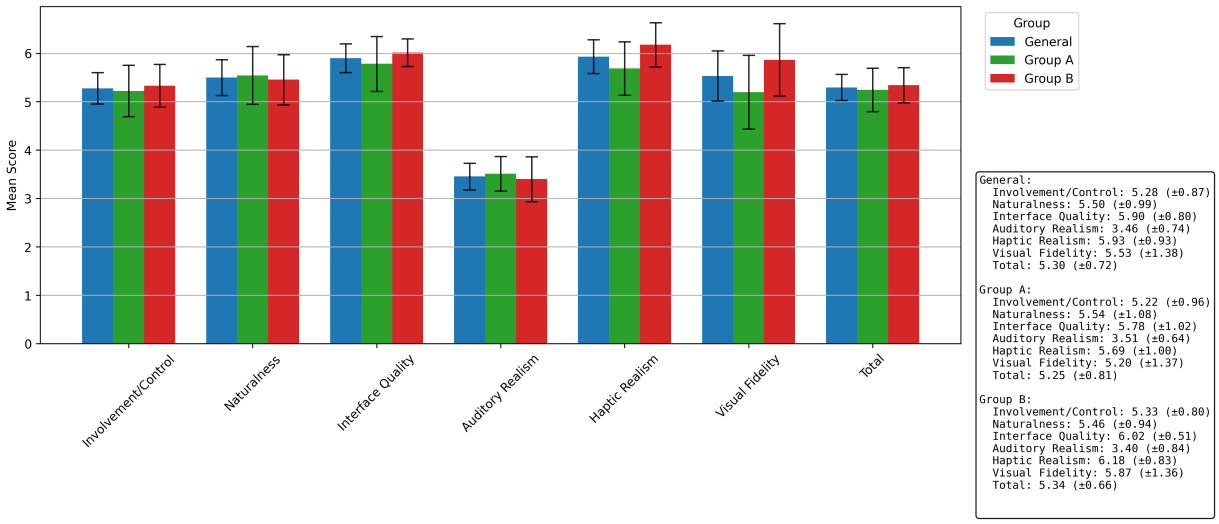


Figure 23: Results of the Presence Questionnaire (PQ).

Although no statistically significant differences were observed, quantitative disparities emerged regarding how each group experienced immersion in the PhysioDrum system. Group B recorded the highest

mean scores in Haptic Realism ($M = 6.18$; $U = 77.50$; $p = 0.146$), Interface Quality ($M = 6.02$; $U = 110.50$; $p = 0.950$), and Visual Fidelity ($M = 5.87$; $U = 77.00$; $p = 0.131$). These results suggest that the variability in haptic feedback may have positively influenced the perception of realism, sensory expressiveness, and coherence between user actions and system responses. The provision of differentiated tactile stimuli may have reinforced critical components for presence, such as the sense of agency, motor engagement, and environmental responsiveness, all essential elements for achieving effective immersive experiences.

In parallel, Group A also achieved high scores in every subscale, particularly in Naturalness ($M = 5.54$; $U = 119.50$; $p = 0.787$) and Involvement/Control ($M = 5.22$; $U = 108.00$; $p = 0.868$). These findings indicate that the consistency and predictability of the haptic feedback likely facilitated users' familiarization with the environment, fostering the development of a sense of mastery over the interaction.

Across both groups, the Auditory Realism dimension consistently received the lowest evaluations (Group A: $M = 3.51$; Group B: $M = 3.40$), indicating a limitation in the integration of the auditory channel into the experience. This outcome may be associated with users' estrangement toward the spatialization techniques implemented by the VR headset, or with perceptual incongruities, such as the static behavior of the hi-hat, which invariably emitted the "open sound" regardless of pedal actuation. Such discrepancies between auditory feedback and motor interaction can disrupt the coherence of multisensory integration, potentially reducing the perceived realism of this specific feature.

The boxplot visualizations in Figure 24 corroborate this interpretation by displaying a centralized distribution and a clear overlap of quartiles between the groups. Dispersion values remain moderate across all subscales, with Auditory Realism presenting both the lowest medians and the highest variability. This pattern suggests a lower degree of consensus among participants regarding the auditory fidelity of the experience, potentially reflecting individual differences in sensitivity to spatial audio cues or expectations about sound–action coherence.

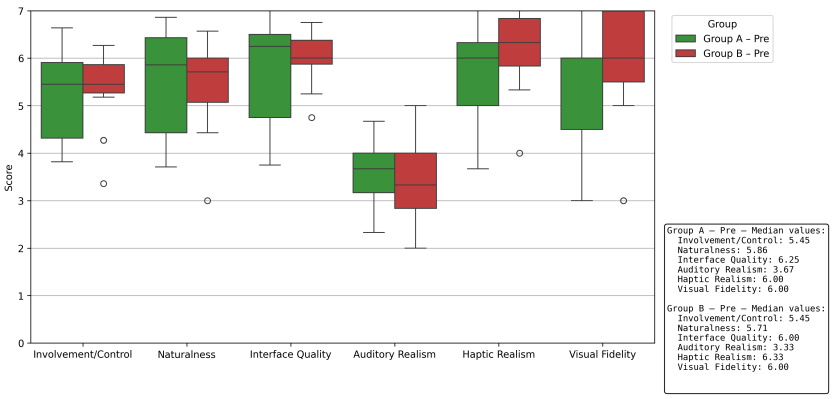


Figure 24: Boxplot representation of the Presence Questionnaire (PQ) subscale scores for Groups A and B.

To deepen the analysis and capture subtle nuances across subscales and between groups, Spearman's rank correlation was employed to investigate monotonic associations among the PQ subscales within each group. This approach aimed to elucidate how different dimensions of perceived presence inter-

relate in the context of the PhysioDrum system. The results, presented in Figure 25, reveal distinct correlation patterns for Group A and Group B, carrying relevant implications for the design and optimization of the user experience.

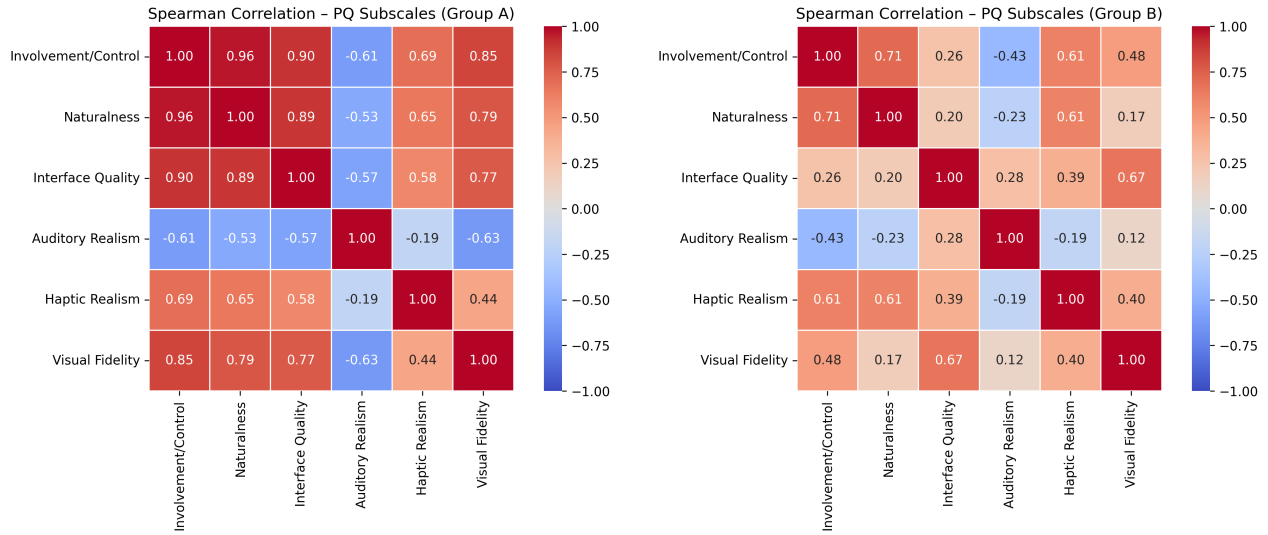
In Group A, strong positive correlations were observed among the Involvement/Control, Naturalness, and Interface Quality subscales, with coefficients ranging from $\rho = 0.89$ to $\rho = 0.96$. This association indicates that the sense of control is closely linked to perceptions of naturalness and interface quality. Additionally, Visual Fidelity showed high correlations with these dimensions ($\rho = 0.85$ with Involvement/Control and $\rho = 0.79$ with Naturalness), reinforcing the hypothesis that visual fidelity plays a critical role in enhancing the sense of presence in immersive environments.

The Haptic Realism subscale exhibited moderate positive correlations with Involvement/Control ($\rho = 0.69$), Naturalness ($\rho = 0.65$), and Interface Quality ($\rho = 0.58$), indicating that tactile stimuli contributed meaningfully, albeit secondarily, to the overall experience. Conversely, Auditory Realism showed negative correlations with nearly all other subscales, most notably with Visual Fidelity ($\rho = -0.63$) and Involvement/Control ($\rho = -0.61$). This pattern suggests a potential perceptual misalignment between auditory stimuli and the other sensory channels in the environment provided to Group A, a phenomenon also observed in the complementary analyses.

In Group B, correlation coefficients were generally lower, indicating reduced interdependence among the subscales. The association between Involvement/Control and Naturalness remained relatively strong ($\rho = 0.71$), whereas all other relationships among the main dimensions were considerably weaker (e.g., Involvement/Control and Interface Quality, $\rho = 0.26$). Interface Quality correlated more strongly only with Visual Fidelity ($\rho = 0.67$), which may reflect a greater emphasis on the visual aspects of the system for this group.

Correlations involving Haptic Realism remained moderate ($\rho \approx 0.61$ with both Involvement/Control and Naturalness). In contrast, Auditory Realism exhibited weak or even negative correlations with the other subscales, as in the case with Involvement/Control ($\rho = -0.43$). Similar to the pattern observed in Group A, the auditory component did not integrate harmoniously with the rest of the experience.

This analysis indicates that, even in the absence of statistical significance, a practical difference emerged in how the groups perceived the tactile and visual realism of the system.



(a) Spearman's rank correlation matrix for the PQ subscales in Group A.

(b) Spearman's rank correlation matrix for the PQ subscales in Group B.

Figure 25: Spearman's rank correlation for the Presence Questionnaire (PQ) subscales.

7.2.2.1 Comparison Between Musicians and Non-Musicians

This section reports the statistical analysis of the PQ scores, segmented according to participants' prior musical experience. For each comparison, descriptive statistics, including measures of central tendency (mean, median, mode) and dispersion (standard deviation), were calculated. In addition, the Mann–Whitney test was applied to assess intergroup differences, and the corresponding effect size was computed to quantify the magnitude of the observed differences (see Table 31 - Appendix D).

For the Involvement/Control factor, non-musicians achieved a higher mean score ($M = 5.47$; $SD = 0.84$) compared to musicians ($M = 4.89$; $SD = 0.82$). The Mann–Whitney test indicated a statistically significant difference ($U = 53.0$; $p = 0.0405$), with a medium effect size ($r = 0.374$). This outcome suggests that participants without formal musical training felt more immersed and in control of the experience. A plausible explanation is that musicians' technical repertoire may have led them to expect control interfaces identical to those encountered in traditional musical performance contexts. Similarly, differing expectations regarding the system's responsiveness to user actions may have contributed to lower scores among this participant profile.

For the Naturalness dimension, a similar trend was observed, with musicians scoring lower ($M = 4.94$; $SD = 1.17$) than non-musicians ($M = 5.78$; $SD = 0.78$). Although the p-value was marginal ($p = 0.0546$), the effect size remained in the medium range ($r = 0.354$), indicating a higher perceived naturalness among non-musicians. This finding supports the hypothesis that prior familiarity with real musical contexts may render musicians more sensitive to discrepancies between the virtual environment and physical performance, thereby influencing their evaluation of the immersive experience.

For the Interface Quality construct, both groups assigned high scores (musicians: $M = 5.75$; $SD =$

0.87; non-musicians: $M = 5.98$; $SD = 0.78$), with no statistically significant difference ($p = 0.578$). The absence of effect indicates that the game-like scenario, the visual cues, and the functional organization of PhysioDrum were perceived as consistent regardless of participants' musical background, underscoring the robustness of the interface design in supporting user engagement across different expertise levels.

The Auditory Realism subscale yielded similarly modest mean scores (≈ 3.5) for both groups, with no statistically significant difference ($p = 0.704$). This outcome highlights a critical limitation: both musicians and non-musicians perceived the auditory realism as insufficient, consistent with the previously identified negative correlations between the auditory channel and other assessed constructs. Enhancements in timbre modeling, spatialization, and dynamic volume response are recommended to increase auditory credibility and, consequently, strengthen immersion in the virtual environment.

Regarding Haptic Realism, both groups assigned high ratings (≥ 5.9), with no statistically significant difference ($p \approx 0.96$, $r = 0.106$). This convergence suggests that the tactile feedback was effective in transcending differences in expertise, thereby validating the central role of the haptic component within PhysioDrum's multisensory architecture. The consistency of these evaluations reinforces the importance of well-calibrated haptic cues in sustaining immersion across diverse user profiles.

Finally, in Visual Fidelity, non-musicians reported slightly higher mean scores ($M = 5.75$) compared to musicians ($M = 5.10$), although the difference did not reach statistical significance ($p = 0.191$, $r = 0.239$). Despite the absence of significance, the small-to-medium effect size suggests that visual aesthetics and modeling criteria may exert a stronger influence on lay users, whereas musicians are likely to concentrate their attention on performance-related dimensions.

The results indicate that non-musicians benefit more immediately from the immersive environment, particularly in terms of involvement and naturalness, whereas musicians tend to adopt a more critical evaluative stance, especially when perceiving auditory inconsistencies. PhysioDrum demonstrates robustness in both interface design and haptic feedback across audiences; however, improvements in auditory realism and visual refinement could help equalize the overall experience, aligning musicians' higher performance standards with the already positive perceptions reported by non-musicians.

The intragroup comparison between musicians and non-musicians in Group A (see Table 32 - Appendix D) revealed no statistically significant differences between subgroups. Nevertheless, certain trends were identified in how each profile experienced presence within the PhysioDrum environment, suggesting subtle variations in perception that may be related to prior musical experience.

For the Involvement/Control construct, non-musicians obtained a slightly higher mean score ($M = 5.39$) compared to musicians ($M = 4.97$), with corresponding medians of 5.82 and 4.86, respectively. However, the Mann–Whitney test did not reveal a statistically significant difference ($U = 19.5$; $p = 0.4085$), and the effect size was small ($r = 0.884$). Although not statistically relevant, this numerical difference suggests a mild tendency toward greater engagement and perceived control among non-musicians within Group A, a trend consistent with the general analysis across groups.

In Naturalness, non-musicians again reported higher scores ($M = 5.81$) than musicians ($M = 5.14$), with similar medians (5.86 vs. 5.00). While the Mann–Whitney test did not yield statistical significance ($U = 18.0$; $p = 0.313$), the effect size ($r = 0.360$) indicates a medium magnitude effect. This pattern suggests that non-musicians may perceive the experience as more natural, potentially due to the absence of technical references or specific expectations associated with traditional musical performance, making them more receptive to the virtual mediation proposed by PhysioDrum.

For Interface Quality, both subgroups assigned high scores (musicians: $M = 5.58$; non-musicians: $M = 5.92$), with low dispersion and no statistically significant difference ($U = 21.5$; $p = 0.5521$; $r = 0.648$). This consistency suggests that the structural and functional quality of the interface was well-evaluated regardless of musical background, corroborating the effective performance of the system's interactive architecture.

For the Auditory Realism measure, the results were similar between subgroups, with musicians reporting a mean score of $M = 3.67$ and non-musicians $M = 3.41$. Both means are modest, and the Mann–Whitney test confirmed the absence of statistical significance ($U = 33.5$; $p = 0.4702$; $r = 0.766$). These findings reiterate that the auditory component was perceived as insufficiently realistic across profiles, with no differential impact associated with prior musical experience within Group A.

Regarding Haptic Realism, musicians assigned a slightly higher mean score ($M = 5.94$) than non-musicians ($M = 5.52$), with identical medians (6.00). The lack of statistical significance ($U = 33.5$; $p = 0.47673$) and equivalence in central tendency measures point to a homogeneous tactile experience across the group, reinforcing the robustness of haptic feedback in sustaining immersion regardless of familiarity with physical instruments.

In Visual Fidelity, non-musicians reported a higher mean score ($M = 5.33$) compared to musicians ($M = 5.00$), with respective medians of 6.00 and 5.00. Although the difference was not statistically significant ($U = 22.0$; $p = 0.5806$; $r = 0.589$), it suggests that non-musicians placed greater emphasis on the visual aspects of the simulation, a pattern consistent with prior comparative findings.

In summary, while none of the subscales exhibited statistically significant differences between musicians and non-musicians within Group A, the results reveal tendencies toward more positive perceptions among non-musicians in involvement, naturalness, and visual fidelity. These tendencies align with the overall analysis, suggesting that prior musical familiarity may critically shape presence evaluation.

For Group B (see Table 33 - Appendix D), in the Involvement/Control measure, non-musicians obtained a higher mean score ($M = 5.54$; $SD = 0.79$) compared to musicians ($M = 4.77$; $SD = 0.58$), with medians of 5.64 and 4.72, respectively. The difference was statistically significant ($U = 10.5$; $p = 0.0303$), with a medium-to-large effect size ($r = 0.480$). This suggests that, within Group B, non-musicians felt substantially more engaged and in control of the experience. Such perception of control may have been more spontaneous among those without technical reference points, indicating that PhysioDrum and its varied haptic feedback are accessible to users with limited prior experience in musical performance.

In Naturalness, non-musicians again outperformed musicians ($M = 5.75$; $SD = 0.63$ vs. $M = 4.64$; $SD = 1.24$), with medians of 5.86 and 4.86, respectively. Although the result did not reach statistical significance ($U = 9.0$; $p = 0.1005$), the effect size was medium ($r = 0.310$), reinforcing the trend observed in other analyses: non-musicians tend to perceive the experience as more natural, whereas musicians may hold higher expectations for gesture fidelity, interaction precision, and sensorimotor responsiveness.

For Visual Fidelity, non-musicians reported higher scores ($M = 6.09$) than musicians ($M = 5.25$), with medians of 6.00 and 5.50, respectively. However, the difference was not statistically significant ($U = 14.5$; $p = 0.3349$), and the corrected effect size was negligible ($r = 0.080$), indicating that the descriptive difference is not meaningful from a statistical perspective.

For the remaining measures (Interface Quality, Auditory Realism, and Haptic Realism), all p-values exceeded 0.50, and effect sizes were small ($r < 0.27$), indicating no relevant differences between subgroups. These results suggest that interface quality, haptic realism, and auditory response were perceived similarly by musicians and non-musicians in Group B.

7.2.2.2 Comparison Between Musicians in Group A and Group B

The comparison between musicians in Group A and Group B (see Table 34 - Appendix D) revealed no statistically significant differences across any of the PQ measures, a result that may be partly attributed to the small sample size, particularly in Group B ($n = 3$). Nevertheless, a few qualitative patterns can be observed

In Involvement/Control, mean scores were highly similar between groups ($M_{\mathrm{A}} = 4.97$; $M_{\mathrm{B}} = 4.77$), with a minimal effect size ($r \approx 0.11$). This suggests that the perceived sense of control over the experience was stable among musicians, regardless of the experimental condition. Similarly, Interface Quality yielded close means ($M_{\mathrm{A}} = 5.58$; $M_{\mathrm{B}} = 6.00$) and a comparably small effect size, reinforcing the notion that interface quality was consistently evaluated across both musician subgroups.

Naturalness, although not statistically significant ($U = 14.5$ $p = 0.6688$), exhibited the largest descriptive contrast between groups, with Group A musicians reporting a higher mean ($M = 5.14$) compared to Group B ($M = 4.64$). This may indicate that musicians in Group A perceived the interaction as slightly more natural.

For Auditory Realism, the effect size was the most notable ($r \approx 0.40$), with Group A showing a higher mean ($M = 3.67$) than Group B ($M = 3.25$). This difference may reflect a more favorable perception of the auditory channel among musicians who experienced the PhysioDrum version with lower additional sensory load, possibly resulting in greater integration and salience of multimodal cues.

Haptic Realism and Visual Fidelity both presented negligible effect sizes ($r < 0.08$), indicating stable evaluations of tactile and visual realism between musician subgroups, even under different sensory conditions. This suggests that for participants with musical training, the visual and haptic components of the system are sufficiently robust to elicit consistent evaluations regardless of experimental

manipulation.

7.2.2.3 Comparison Between VR Specialists and Non-Specialists

In the overall comparison between VR-experienced users and Non-VR specialists (see Table 35 - Appendix D), Interface Quality was the only measure to reach statistical significance ($p = 0.0243$), with higher mean scores among participants without prior VR experience ($M_{\text{noVR}} = 6.03$; $SD = 0.80$) compared to VR-experienced users ($M_{\text{VR}} = 5.38$; $SD = 0.65$). The effect size was large ($r = 0.55$), suggesting that VR novices perceived the PhysioDrum interface as clearer, more accessible, and better integrated. This finding may indicate a favorable learning curve for new users and supports the hypothesis that experienced users tend to be more demanding regarding control precision and interaction fluidity.

For Naturalness, Haptic Realism, Visual Fidelity, and Involvement/Control, mean scores were also consistently higher among Non-VR specialists. However, none reached statistical significance (0.574; 0.101). The corresponding effect sizes were nonetheless considerable (0.45; 0.49), suggesting potentially meaningful differences that could become significant with larger samples. These results indicate that the immersive experience of PhysioDrum may have a stronger impact on VR novices, possibly due to novelty effects or reduced comparative expectations.

In Auditory Realism, both groups reported similar mean scores ($M_{\text{noVR}} = 3.50$; $M_{\text{VR}} = 3.28$), with low variability and no significant difference ($p = 0.528$). The small effect size reinforces the recurrent pattern that the auditory dimension remains a weakness of the system, regardless of VR familiarity.

To sum up, participants without prior VR experience tend to perceive the experience as more natural, immersive, and structurally coherent, particularly in terms of interface quality and visual and haptic realism. PhysioDrum thus demonstrates high accessibility and impact potential for new VR users, while experienced users may adopt a more critical perspective, especially regarding responsiveness and refinement of visual and haptic elements. The lack of differences in Auditory Realism underscores the need for targeted improvements in this component, given its consistent identification as a weakness across all participant profiles.

When comparing VR-experienced users with VR-naïve participants within Group A (see Table 36 - Appendix D), those without prior VR experience reported a notably higher mean score in Involvement/Control ($M_{\text{noVR}} = 5.49$; $SD = 0.86$) than their experienced counterparts ($M_{\text{VR}} = 4.50$; $SD = 0.92$). The medium-to-large effect size ($r = 0.583$) suggests a substantial advantage for VR novices, who felt more immersed and in control when interacting with PhysioDrum.

A similar pattern was observed for Naturalness, where non-VR participants scored higher ($M_{\text{noVR}} = 5.78$) compared to VR-experienced participants ($M_{\text{VR}} = 4.89$). Despite the lack of statistical significance ($p = 0.148$), the medium effect size ($r = 0.528$) indicates that the sense of naturalness was more pronounced among novices.

For Interface Quality, non-VR participants again reported higher scores ($M_{\text{noVR}} = 5.98$) than VR-

experienced participants ($M_{\text{VR}} = 5.25$). The effect size approached the large range ($r = 0.569$), and the p-value ($p = 0.099$) suggests a trend toward significance. This result implies that PhysioDrum's interface design may be particularly engaging for VR novices, potentially due to lower pre-existing expectations regarding interaction paradigms and system responsiveness.

For Haptic Realism, the difference between subgroups was statistically significant ($U = 4.5$; $p = 0.0255$) and yielded the largest effect size observed in this set of comparisons ($r = 0.691$, large). Once again, VR-naïve participants assigned higher scores ($M_{\text{noVR}} = 6.06$, $SD = 0.73$) compared to VR-experienced participants ($M_{\text{VR}} = 4.67$, $SD = 0.98$). This result, highlighted in Table 10, indicates that tactile feedback is perceived as more relevant to immersion among participants with no prior exposure to haptic experiences in VR.

Subscale	Mean (VR-spec.)	Mean (NonVR-spec.)	Median (VR-spec.)	Median (NonVR-spec.)	Mode (VR-spec.)	Mode (NonVR-spec.	SD (VR-spec.)	SD (NonVR-spec.)	U	*p*-value	Effect Size (*r*)
Haptic Realism	4.67	6.06	4.50	6.33	3.67	6.33	0.98	0.73	4.5	0.02552	0.69147

Table 10: Comparison of Haptic Realism subscale scores between participants with (VR-spec) and without (NonVR-spec) prior virtual reality experience in Group A.

In contrast, Auditory Realism showed very similar means between groups ($M_{\text{VR}} = 3.45$; $M_{\text{noVR}} = 3.67$) and a small effect size ($r \approx 0.14$), confirming the previously observed pattern for the auditory channel: a consistent perception of limited realism, independent of VR familiarity.

For Visual Fidelity, participants without prior VR experience reported a higher perception of visual realism compared to VR-experienced users. The analysis resulted in a medium-to-large effect size ($r \approx 0.515$), consistent with the trends observed in other measures, reinforcing a clear tendency for PhysioDrum to deliver a more impactful visual experience to novice users.

For Group B (see Table 37 - Appendix D), the results revealed considerable effect sizes in some subscales, even in the absence of statistical significance. Naturalness stood out with a medium-to-large effect size ($r \approx 0.56$), indicating that participants without VR experience perceived the experience as more natural. A similar trend was observed for Auditory Realism ($r \approx 0.53$), suggesting that the auditory realism had a stronger impact on novice users.

In Interface Quality, the effect size was moderate ($r \approx 0.40$), with slightly higher scores reported by participants without prior VR experience. Conversely, Involvement/Control and Visual Fidelity exhibited small effects ($r \approx 0.16$ and $r \approx 0.13$, respectively), with no indication of a relevant influence of VR familiarity on these measures.

Haptic Realism, in turn, presented a virtually null effect size ($r \approx 0.03$), suggesting that tactile perception was stable between the two subgroups. This consistency highlights the robustness of the haptic response in PhysioDrum, even in a scenario with greater heterogeneity in participants' levels of VR experience.

7.2.2.4 Comparison Between VR Specialists in Group A and Group B

An examination of VR specialists across groups (see Table 38 - Appendix D) revealed that the largest effect was found in Haptic Realism ($r = 0.69$, large), with both the mean and median scores substantially higher in Group B ($M = 6.16$; median = 6.16) compared to Group A ($M = 4.67$; median = 4.50). The mode was likewise higher in Group B (5.33 vs. 3.67), indicating that tactile perception played a more prominent role in enhancing immersion for participants exposed to the variable haptic feedback condition.

Interface Quality showed a moderate-to-large effect size ($r = 0.56$), with Group B scoring higher ($M = 5.62$) than Group A ($M = 5.25$). This suggests a more stable and positive interface evaluation among specialists who experienced multiple haptic stimuli. Involvement/Control also favored Group B ($M = 5.00$) over Group A ($M = 4.50$), with a moderate effect size ($r = 0.52$), implying that the Group B condition enhanced engagement and perceived control for VR specialists.

For Visual Fidelity, Group B again outperformed Group A ($M = 5.00$ vs. $M = 4.25$), with a moderate effect size ($r = 0.52$). While participant ratings in Group B were more polarized, the overall evaluation of visual realism was superior compared to Group A, suggesting a globally more impactful visual experience despite greater variability.

Naturalness scores were close between groups ($M_A = 4.89$; $M_B = 4.72$), with a small-to-moderate effect size ($r = 0.43$). Group B exhibited more consistent evaluations and a subtle tendency toward more positive perceptions, which may indicate that the variable haptic feedback condition provided slightly more intuitive and motor-congruent interactions.

Auditory Realism was the only measure in which Group B performed worse ($M_A = 3.67$ vs. $M_B = 2.50$). The effect size was negligible ($r = 0.09$), but the direction of the data suggests that the sound quality or its integration with the other sensory channels was perceived as less satisfactory in the Group B.

7.2.2.5 Synthesis and Implications of PQ Results

The PQ data analysis provided insights into how different groups and user profiles perceived the immersive experience offered by PhysioDrum. In general, presence and immersion were positively evaluated, with Interface Quality and Haptic Realism consistently obtaining the highest mean scores across groups. These results indicate that interaction interfaces, graphical elements, and tactile feedback had a significant positive impact on the user experience, contributing to enhanced immersion and presence.

Differences were also identified according to participant profiles, segmented by prior musical and VR experience. Non-musicians and VR-naïve participants assigned higher ratings to dimensions such as Naturalness and Auditory Realism, possibly due to the absence of comparative references from physical musical systems or prior immersive experiences. Conversely, musicians — particularly in Group A —

displayed greater criticality regarding the auditory component, indicating the need for improvements in the system's sound fidelity.

Variable haptic feedback, as implemented for Group B, was often associated with higher scores in Haptic Realism, Interface Quality, and Involvement/Control, especially among VR specialists. This suggests that the variable haptic condition was particularly effective for more experienced users, offering greater technical sophistication and sensory responsiveness.

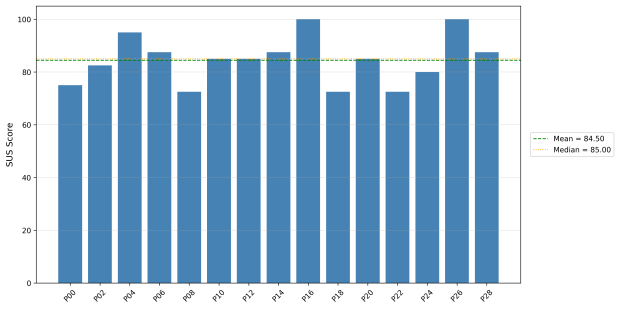
The convergence of results across analyses leads to two key implications for PhysioDrum's development and refinement. First, presence appears to be strongly anchored in control, naturalness, and visual/-tactile fidelity, which represent perceptual pillars of the immersive musical experience. Enhancements to gesture mapping, interface responsiveness, and visual detailing of instruments are likely to reinforce users' sense of presence. Second, the auditory channel requires reengineering. The low correlation between Auditory Realism and other subscales, along with its negative impact on perceptual cohesion, underscores the need to improve sound quality, spatialization, and multimodal integration.

Moreover, findings reinforce the hypothesis that haptic feedback is a central component in building presence within virtual musical environments. Manipulation of tactile characteristics demonstrated a direct impact not only on sensory perception (e.g., tactile realism) but also on cognitive and affective presence dimensions, such as perceived control, interaction fluency, and overall environmental realism.

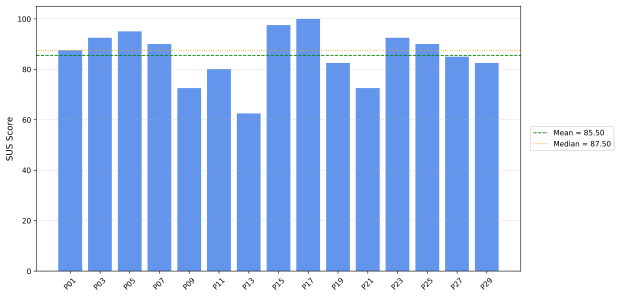
From an immersive experience design perspective, the results point toward two complementary and equally valid strategies: i) consistent haptic feedback, as in Group A, can promote more stable, predictable experiences conducive to immersion; and ii) haptic variability, as in Group B, can foster more dynamic, engaging, and realistic interactions, amplifying the perceptual and emotional impact of the experience.

7.2.3 Assessment of System Usability Scale (SUS)

The System Usability Scale (SUS) scores for each participant in Groups A and B are presented in Figure 26, and their classification and interpretation are provided in Table 11. The results indicate that PhysioDrum achieved high usability ratings in both groups. Mean scores were consistently elevated and standard deviations were low, suggesting uniform performance across participants. The data also demonstrate compliance with established usability principles, including interface consistency, ease of operation, and reliability in task execution.



(a) Distribution of individual SUS scores for participants in Group A.



(b) Distribution of individual SUS scores for participants in Group B.

Figure 26: System Usability Scale (SUS) scores for Groups A and B.

Group	SUS Score	SD	Grade	Percentile	Adjective	Acceptability / NPS
General	85.0	9.44	A	∼95–97	Excellent	Acceptable / Promoter
A	84.5	9.12	A–	∼70–90	Excellent	Acceptable / Promoter
B	85.5	10.36	A	∼95–97	Excellent	Acceptable / Promoter

Table 11: System Usability Scale (SUS) results by experimental group.

Beyond the computation and classification of the SUS scores, an assessment of the internal consistency of participants' responses was conducted separately for each group. Subsequently, the item-level responses to the ten questions presented in this questionnaire were analyzed, aiming to identify inter-group discrepancies and to provide a more detailed understanding of how distinct haptic configurations might have influenced perceived system usability.

The distribution of responses for both groups is presented in Figure 27, while Table 12 reports the median values and standard deviations for each item. In Group A, participants' response patterns displayed a relatively high degree of dispersion across the ten items. This variability was reflected in the Cronbach's alpha coefficient ($\alpha = 0.5206$), indicative of moderate internal consistency. Such a reliability level suggests that participants did not exhibit fully uniform response behavior across all items, potentially reflecting the heterogeneity of user experiences with the uniform haptic feedback condition.

The analysis of Group B responses revealed a higher Cronbach's alpha coefficient ($\alpha = 0.7278$), indicating stronger internal consistency when compared to Group A. This suggests that participants in Group B exhibited a more coherent response pattern, potentially reflecting a clearer or more impactful interaction with the different haptic feedback conditions.

For Q1 (intention for frequent system use), both groups displayed similar response distributions, with higher frequencies in the "Agree" and "Strongly agree" categories. The median score for both was 4, indicating a consistently high intention for continued use. In Question 2 (perceived complexity), the dominant response in both groups was "Strongly disagree", with a median of 1, suggesting that the experimental manipulation did not influence perceived simplicity.

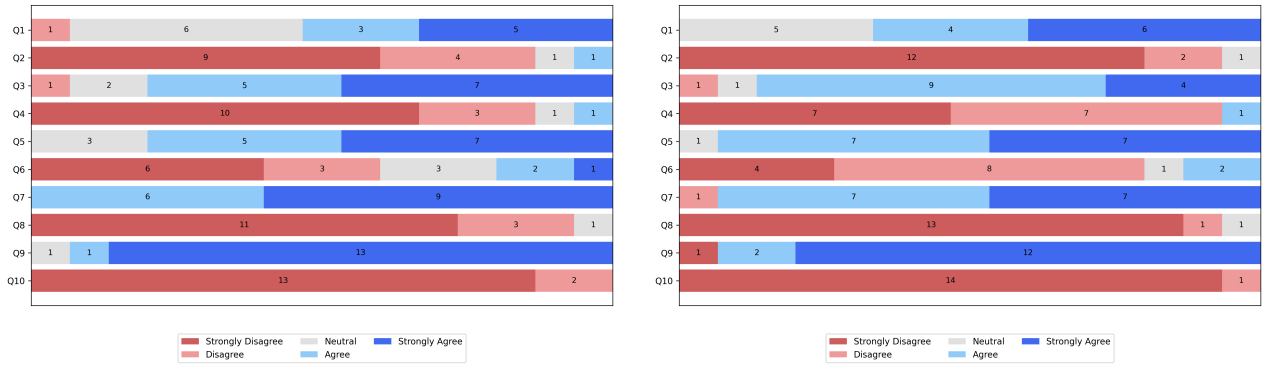
Question	Group A (Median; SD)	Group B (Median; SD)
Q1	4.0 (1.014)	4.0 (0.884)
Q2	1.0 (0.910)	1.0 (0.594)
Q3	4.0 (0.941)	4.0 (0.799)
Q4	1.0 (0.915)	2.0 (0.816)
Q5	4.0 (0.799)	4.0 (0.632)
Q6	2.0 (1.335)	3.0 (0.961)
Q7	5.0 (0.507)	5.0 (0.816)
Q8	1.0 (0.617)	2.0 (0.561)
Q9	5.0 (0.561)	5.0 (1.056)
Q10	1.0 (0.352)	1.0 (0.258)

Table 12: Median and standard deviation for each SUS questionnaire item across Groups A and B.

The evaluations for ease of use (Q3) and integration of functionalities (Q5) were equivalent across groups, both with medians of 4 and a prevalence of positive evaluations, confirming consistent perceptions of fundamental usability regardless of tactile experience mode.

Non-significant differences emerged in Q4, Q6, and Q8. In Q4 (need for technical support), Group A had a median of 1, whereas Group B presented a median of 2. Similar patterns were observed in Q6 (system inconsistency) and Q8 (confidence during use), with Group B reporting slightly higher medians. Although subtle, these variations may indicate that exposure to non-uniform haptic feedback requires greater perceptual or cognitive adaptation. However, they did not result in substantial usability impairments.

The remaining items — Q7 (ease of learning), Q9 (absence of prior knowledge requirements), and Q10 (overall simplicity) — showed no differences between the two experimental conditions. Participants



(a) SUS responses for Group A

(b) SUS responses for Group B

Figure 27: Comparison of item-level responses to the System Usability Scale (SUS) questionnaire across experimental groups.

in both cases reported highly positive evaluations, with medians of 4, 5, and 1, respectively. These findings underscore the robustness of the system's learning curve and structural clarity, regardless of the haptic feedback configuration.

This examination indicates substantial convergence between Groups A and B, with identical medians in seven out of ten questions. Even in cases with observable differences (Q4, Q6, Q8), median variations were modest and concentrated in Group B. The Mann–Whitney test (Table 13) confirmed the absence of statistically significant differences between groups. Collectively, these findings suggest a stable level of perceived usability, with no single aspect of the user experience being critically affected by the experimental manipulation of haptic feedback.

Question	**U Statistic**	**p-value**	**Significant ($p < 0.05$)**
Q1	96.0	0.48168	No
Q2	135.5	0.24682	No
Q3	127.0	0.52848	No
Q4	95.0	0.42509	No
Q5	104.5	0.73321	No
Q6	116.5	0.87904	No
Q7	130.5	0.40743	No
Q8	126.5	0.42179	No
Q9	120.0	0.65441	No
Q10	120.0	0.57653	No

Table 13: Mann–Whitney U test results for each question in SUS questionnaire.

In order to refine the assessment of internal consistency and to identify potential divergences in usability perception between the experimental conditions, the non-parametric effect size measure Cliff's Delta was employed, with the results presented in Table 14. While the Mann–Whitney test determines whether differences between the experimental conditions reach statistical significance, Cliff's Delta quantifies the practical magnitude of these differences, providing an estimate of the probability that a randomly selected score from Group A will exceed (or fall below) a randomly selected score from Group B.

The Cliff's Delta coefficients ranged from -0.15 to 0.20, with all values remaining below 0.33 — a threshold conventionally adopted as the lower bound for a medium effect size. This distribution indicates that, in addition to the absence of statistically significant differences between Groups A and B, the practical magnitude of these variations was also minimal.

Nevertheless, some patterns warrant further consideration. The highest $|\delta|$ values occurred in Q2 (ease of use, $\delta = 0.204$), Q4 (integration of functionalities, $\delta = -0.156$), and Q7 (ease of learning,

Question	Cliff's delta (δ)
Q1	-0.1467
Q2	0.2044
Q3	0.1289
Q4	-0.1556
Q5	-0.0711
Q6	0.0356
Q7	0.1600
Q8	0.1244
Q9	0.0667
Q10	0.0667

Table 14: Cliff's delta (δ) for each SUS questionnaire item comparing the responses from Groups A and B.

$\delta = 0.160$), which suggest a slightly more favorable perception of these aspects among participants in Group B.

The inclusion of Cliff's Delta in this analysis strengthens the assessment of internal consistency by providing a measure of practical relevance that complements the statistical significance testing. The low $|\delta|$ values are consistent with the Cronbach's Alpha coefficients (Group A: $\alpha = 0.52$; Group B: $\alpha = 0.72$), reflecting only minor variations in individual responses and indicating no substantive structural disruption in perceived system usability.

7.2.3.1 Comparison Between Musicians and Non-Musicians

The descriptive and inferential data obtained in this evaluation (see Table 40 - Appendix E) reveal a marked difference in perceived usability between participants with and without musical experience. Users with musical training achieved a mean score of 79.25 (SD = 8.90), whereas those without such skills reached a significantly higher mean of 87.88 (SD = 8.78). The median values mirrored this trend (81.25 vs. 87.50, respectively), and the mode further emphasized the disparity, with the most frequent score among musicians being 72.50 — notably lower than the corresponding mode for non-musicians (85.00). This distribution indicates a higher concentration of positive evaluations among participants without musical experience.

The Mann–Whitney test results confirmed that this difference is statistically significant ($U = 50.5$, $p = 0.0301$). The associated effect size ($r = 0.3975$) corresponds to a medium-to-large magnitude, underscoring the practical relevance of the finding. Complementarily, the Cliff's delta value ($\delta = -0.495$) corroborates the asymmetry in score distributions, suggesting an approximate 49.5% probability that a musician would rate the system's usability lower than a non-musician, a configuration that

represents a statistically and practically consistent trend within the subjective evaluation metrics.

Within the context of the PhysioDrum system, these results suggest that musically experienced users may apply higher expectations or more stringent evaluative criteria to interactive musical systems, particularly when interactions are mediated by immersive technologies and haptic feedback mechanisms. Such users are likely to be more sensitive to nuances in response latency, expressive control, and fine motor accuracy, dimensions that differentiate physical instruments from their virtual counterparts and may influence the overall user experience evaluation. In contrast, participants without musical training appear to prioritize general ease of use and intuitiveness, leading to more favorable assessments.

When conducting an intra-group examination, the data for Group A (see Table 41 - Appendix E) reveal a marked difference between participants with and without musical experience. Participants with musical training achieved a mean score of 78.33 (SD = 6.83), whereas those without such experience obtained a substantially higher mean score of 88.61 (SD = 8.30). This discrepancy is reflected in the medians (76.25 vs. 87.50) and modes (72.50 vs. 85.00), indicating a systematic tendency towards more critical evaluations among musicians.

The Mann–Whitney test indicated that this difference was statistically significant ($p = 0.037$), with an effect size of $r = 0.55$, classified as large, thereby underscoring the practical relevance of the observed divergence. These results support the conclusion that musical expertise exerted a measurable influence on usability perceptions specifically within Group A. One plausible explanation is that the uniform haptic feedback configuration employed in this condition may not have met the performance expectations of musically trained participants, who are generally more attentive to attributes such as expressivity, temporal precision, and realism in musical instrument interaction. By contrast, non-musicians may have perceived the uniform feedback as sufficiently informative and intuitive, resulting in more favorable evaluations. Consequently, musical background appears to heighten sensitivity to specific limitations in haptic interfaces, particularly in scenarios characterized by lower variability in sensory stimuli.

For Group B (see Table 42 - Appendix E), the comparison between musicians and non-musicians yielded no statistically significant difference ($p = 0.326$). Although non-musicians attained a slightly higher mean score (87.27; SD = 9.52) compared to musicians (80.63; SD = 12.48), the Mann–Whitney test did not reject the null hypothesis, and the effect size was small ($r < 0.3$).

The absence of statistical significance, combined with the low effect magnitude, suggests that the granular haptic feedback implemented in Group B was effective in attenuating the additional performance demands typically associated with musical training. Unlike Group A, where the uniform feedback configuration appeared to expose perceptual limitations for musicians, the tactile differentiation provided in Group B seems to have offered sufficient sensory cues to satisfy expectations related to expressivity and fine control. These findings demonstrate that, although musicians remain attuned to the characteristics of haptic feedback, such receptiveness can be attenuated through the provision of more nuanced tactile variation, thereby preserving usability perceptions at levels comparable to

those of non-musicians.

7.2.3.2 Comparison Between Musicians in Group A and Group B

The comparison between musicians allocated to Groups A and B (see Table 43 - Appendix E) revealed no statistically significant differences in the perceived usability of PhysioDrum, as indicated by the Mann–Whitney test ($U = 8.5$; $p = 0.516$). Despite this lack of significance, musicians in Group B achieved a slightly higher median score (85.0) compared to those in Group A (76.25), with corresponding means of 80.63 ($SD = 12.47$) and 78.33 ($SD = 6.83$), respectively. The effect size ($r = 0.235$), and Cliff's delta ($\delta = -0.29$) also indicated a small effect, suggesting a trend without statistical robustness.

Although not statistically substantiated, this trend points to a marginal preference among Group B participants, potentially attributable to the implementation of differentiated haptic feedback, which may benefit users with greater artistic–musical demands. Such findings highlight the potential practical value of employing more diverse haptic responses for this user profile.

7.2.3.3 Comparison Between VR Specialists and Non-Specialists

The comparison between participants with prior experience in immersive environments and those without such a background (see Table 44 - Appendix E) revealed no statistically significant differences in the perceived usability of PhysioDrum. Participants with VR experience obtained a slightly lower mean SUS score ($Mean = 80.41$; $SD = 11.00$) compared to participants without prior experience ($Mean = 86.14$; $SD = 9.11$). However, this difference was not statistically significant ($U = 51.50$; $p = 0.296$), and the calculated effect size was small ($r = 0.194$; $\delta = -0.284$), indicating a modest discrepancy between the two groups.

The examination of the impact of prior VR experience within Groups A and B indicates no differences between participants with and without such experience when considered within their respective groups, reflecting a pattern consistent with the findings from the broader analysis.

In Group A (see Table 45 - Appendix E), participants with VR experience achieved a mean SUS score of $Mean = 81.87$ ($SD = 6.57$), whereas those without prior experience reached $Mean = 85.45$ ($SD = 9.98$). This difference was not statistically significant ($p = 0.552$), with a small effect size ($r = 0.17$; $\delta = -0.227$).

Similarly, in Group B (see Table 46 & Table 47 - Appendix E), VR experts reported a mean score of $Mean = 77.50$ ($SD = 21.21$), which was lower than the mean obtained by non-experts ($Mean = 86.73$, $SD = 8.68$). This difference was also not statistically significant ($p = 0.609$), and the effect size remained modest ($r = 0.15$; $\delta = -0.269$). These findings indicate considerable variability in the responses of expert participants, while simultaneously confirming the absence of a consistent or systematic pattern across the groups.

The absence of statistical significance supports the conclusion that PhysioDrum provides an interface

that is sufficiently intuitive and accessible for novice VR users. At the same time, the fact that experienced VR users did not assign higher scores suggests that, although the system is effective in terms of overall usability, it may not deliver advanced interactive stimuli that meet the expectations of users more accustomed to immersive technologies. This observation identifies a clear opportunity for refinement, potentially through the integration of gesture personalization, dynamic multisensory feedback, and adaptive mechanics capable of responding in real time to user behavior in a more sophisticated manner.

7.2.3.4 Comparison Between VR Specialists in Group A and Group B

An analysis of VR experts exposed to distinct haptic feedback configurations (see Table 48 - Appendix E) revealed no measurable differences in the perceived usability of PhysioDrum. The Mann–Whitney test demonstrated identical results ($U = 4.0$, $p = 1.00$), with a null effect size ($r = 0.00$) and a Cliff's delta ($\delta = 0$), indicating complete overlap between the SUS score distributions. From a descriptive standpoint, Group A recorded a mean score of 81.88 ($SD = 6.57$), whereas Group B reported 77.50 ($SD = 21.21$). The substantial variability observed in Group B, combined with the limited number of cases, considerably weakens the statistical and inferential strength of these outcomes.

The small sample size, consisting of only four experts in Group A and two experts in Group B, significantly limits the statistical ability to detect even moderate effects. Additionally, the lack of any clear preference for one type of configuration indicates that, for users with extensive VR experience, the transition between uniform and differentiated haptic feedback is not, on its own, a key factor in assessing usability.

7.2.3.5 Synthesis and Implications of SUS Results

Analysis of SUS in the context of PhysioDrum indicated consistently high levels of perceived usability across all user profiles and experimental conditions. Mean SUS scores exceeded 84 points in all cases — classified as Excellent — and were accompanied by low standard deviations, suggesting that the system is accessible, reliable, and effectively integrated. These results are further supported by high percentile rankings, positive adjective ratings, and a predominance of "Promoter" classifications in recommendation rates.

From a psychometric standpoint, Group B demonstrated greater internal consistency in its responses compared to Group A, as well as slight advantages in items associated with reliability and functional integration. Item-level analysis revealed identical medians in 70% of the questions, reflecting strong convergence in perceived usability between the two experimental conditions. Although some items exhibited small tendencies favoring Group B (Q4, Q6, Q8), none of these differences reached statistical significance, and all effect sizes were classified as small.

Musical expertise emerged as the primary moderator of perceived usability. Musicians assigned lower SUS scores than non-musicians, particularly in Group A. This suggests that users with musical training

may be more attuned to nuances in expressiveness and tactile responsiveness, which could negatively influence evaluations when sensory diversity is limited. Notably, this difference disappeared in Group B, indicating that differentiated haptic feedback may attenuate perceptual demands among musically trained users, thereby enhancing satisfaction.

Prior experience with virtual reality did not produce significant effects on perceived usability. Both VR experts and novices rated the system equivalently from a statistical perspective, suggesting that PhysioDrum's interface is sufficiently intuitive and accessible for individuals with minimal prior exposure to immersive environments.

In summary, SUS outcomes demonstrate that PhysioDrum, in its current configuration, is usable, stable, and well-engineered, exhibiting consistent performance across diverse experimental conditions. The principal implication concerns the potential of tactile differentiation as a personalization and engagement strategy, particularly for users with refined technical and artistic repertoires, such as musicians.

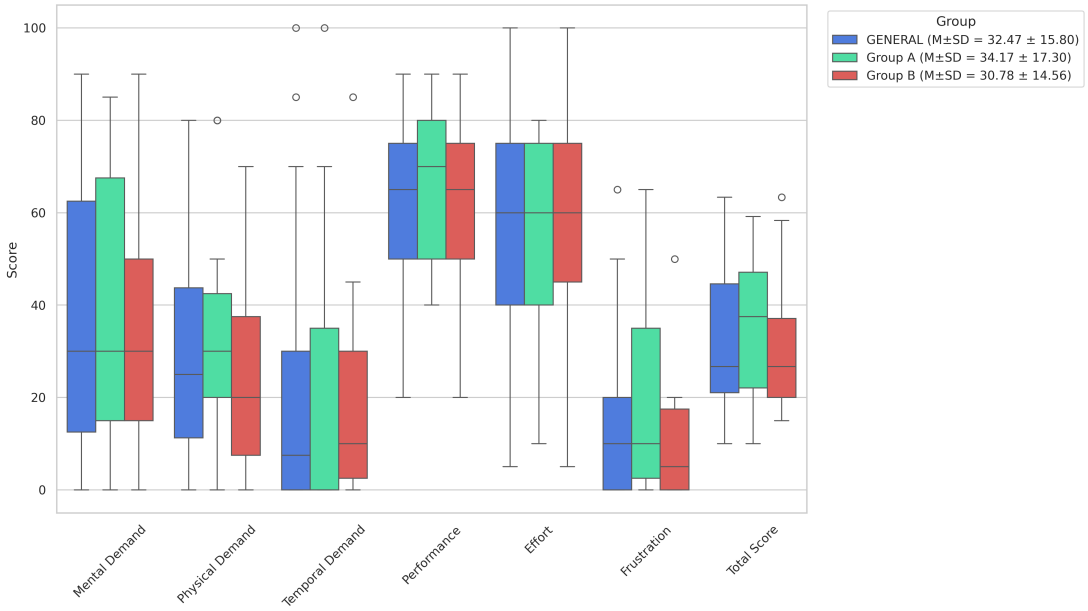
7.2.4 Assessment of NASA Task Load Index (TLX)

Figure 28 depicts the individual evaluation for each NASA-TLX metric, alongside a boxplot summarizing group-level performance. The analysis shows an overall mean score of 32.47 ($SD = 15.80$), with only minor variations between experimental conditions. Participants exposed to uniform haptic feedback achieved a mean of 34.17 ($SD = 17.30$), whereas those receiving differentiated feedback obtained 30.78 ($SD = 14.56$). The dimensions with the highest perceived workload were Effort ($M = 56.17$) and Mental Demand ($M = 37.67$), indicating that the PhysioDrum interaction was primarily characterized by physical exertion, consistent with the simulation of drumming motor skills, and by substantial cognitive engagement, given the rhythmic task requirements. The lowest scores were recorded for Frustration ($M = 15.50$) and Temporal Demand ($M = 20.17$), suggesting that participants generally experienced a sense of task accomplishment and did not perceive time constraints as a source of pressure.

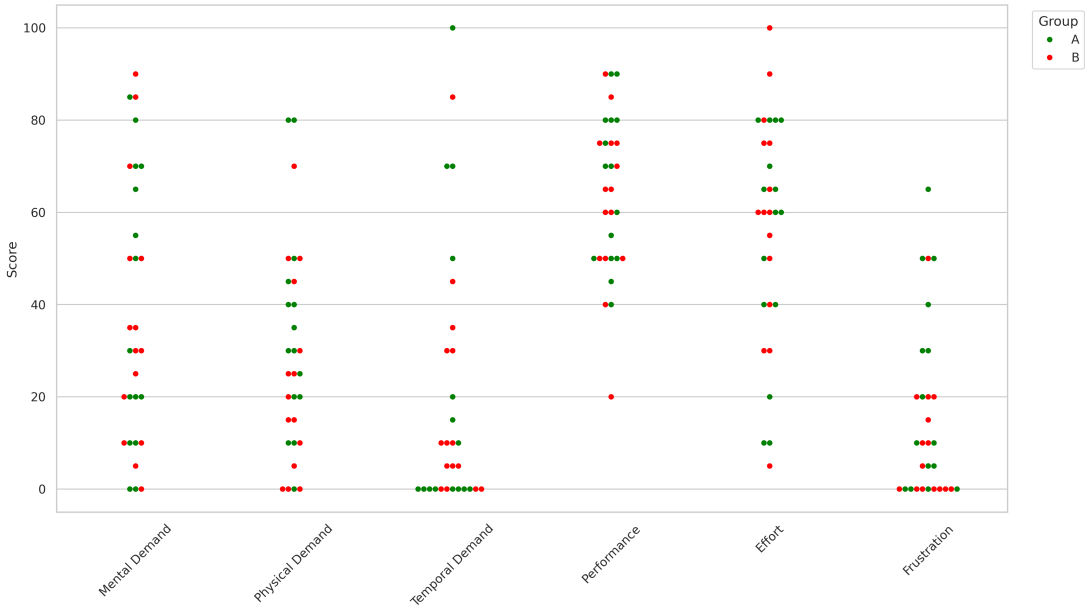
In the Mental Demand subscale, both groups exhibited identical medians (30.0) with highly similar score distributions, indicating that task execution did not require elevated cognitive effort, regardless of the feedback condition. This interpretation is supported by the Mann–Whitney test ($U = 114$; $p = 0.966$), which revealed no detectable difference between the groups.

The Physical Demand subscale showed a slight numerical variation between conditions: Group A reported a median of 30.0, whereas Group B indicated a lower workload, with a median of 20.0. Although this difference was not statistically significant ($U = 143$; $p = 0.211$), it may suggest that differentiated haptic feedback marginally reduced perceived physical exertion, albeit without conclusive evidence.

With respect to Temporal Demand, Group A exhibited a median score of 0.0, whereas Group B reported a median of 10.0. This outcome produced a non-significant result in the Mann–Whitney test ($U = 101.5$; $p = 0.652$), suggesting that the observed difference is attributable to random variation



(a) Distribution of NASA-TLX scores by group.



(b) Individual distribution of NASA-TLX scores.

Figure 28: Comparative visualization of NASA-TLX results between experimental conditions, emphasizing both group-level patterns and individual variations.

rather than a systematic effect. Accordingly, both experimental conditions were characterized by a perception of minimal time pressure.

In the Performance dimension, the medians were closely aligned across conditions, with Group A scoring 70.0 and Group B scoring 65.0. The Mann–Whitney test indicated no statistically significant difference ($U = 123.5$; $p = 0.660$), suggesting that participants perceived themselves as similarly effective in completing the tasks regardless of the haptic feedback configuration.

For the Effort construct, both groups reported identical medians of 60.0, reinforcing the interpretation that the physical and cognitive demands of operating the system were consistent and appropriately calibrated across different conditions. The corresponding statistical test ($U = 108.5$; $p = 0.883$) provided no evidence of an effect attributable to feedback type.

The Frustration parameter showed a slight numerical advantage for Group B, with a median score of 5.0 compared to 10.0 in Group A. While this difference may imply a marginal improvement in emotional comfort when differentiated haptic feedback was applied, the Mann–Whitney test results ($U = 145.5$; $p = 0.165$) fell short of statistical significance.

In light of the absence of statistically significant differences between the experimental groups, additional analyses were conducted using effect size metrics and Cliff's delta to provide a more granular interpretation of potential variations. All computed values are summarized in Table 50 - Appendix F.

The effect size assessment indicated predominantly small magnitudes, with the highest coefficients observed for the Frustration ($r = 0.249$) and Physical Demand ($r = 0.231$) factors. Although these results did not attain statistical significance, they suggest that differences in these components may possess practical relevance in the context of user experience. Specifically, the greater disparity in Frustration may reflect variations in adaptation or comfort with the system interface, with Group B ($M = 10.00$) reporting lower frustration than Group A (mean = 21.00). Such a difference may be associated with prior familiarity with immersive environments or with the system's learning curve, indicating opportunities to optimize the interaction design for specific user profiles. Likewise, the higher Physical Demand reported by Group A ($M = 34.33$), compared to Group B ($M = 24.00$), suggests a greater perception of physical exertion, potentially attributable to the bodily interaction style required by the PhysioDrum system.

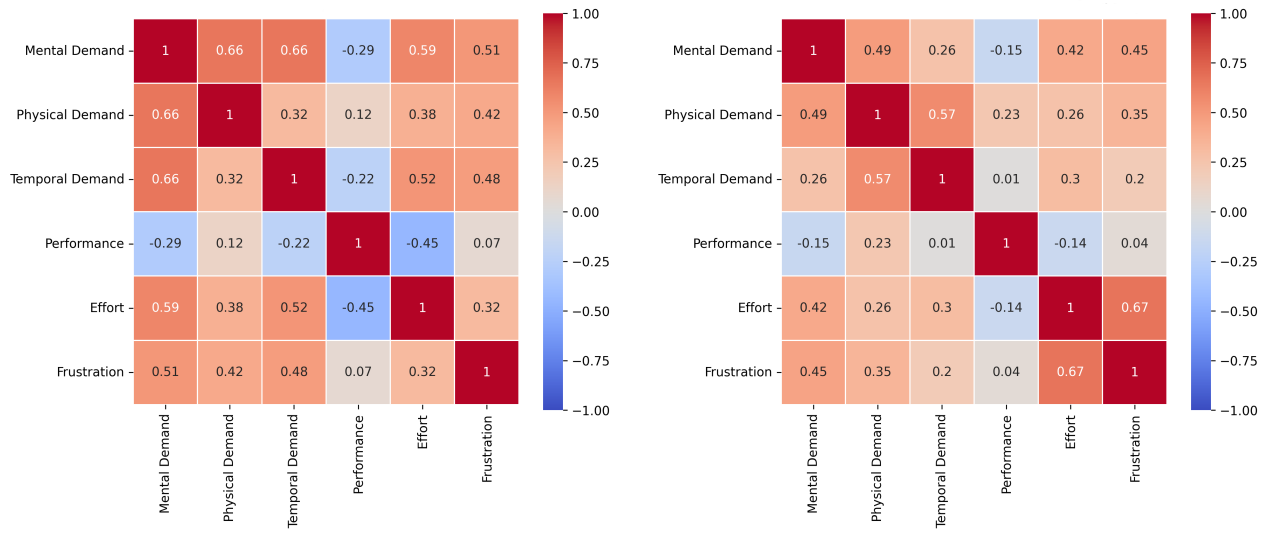
The Cliff's delta results corroborated these tendencies. The Frustration ($\delta = 0.293$) and Physical Demand ($\delta = 0.271$) subscales exhibited the largest absolute values, indicating a moderate probability of superiority, that is, the likelihood that a participant from one group perceives a greater workload than a participant from the other cohort. In contrast, the remaining subscales and the overall workload score presented deltas close to zero ($|\delta| < 0.15$), characterizing negligible effects and a high degree of similarity in workload perception across the experimental conditions.

To further refine the interpretation of the NASA-TLX outcomes, a Spearman correlation analysis was conducted for the subscales, as illustrated in Figure 29. In Group A, a strong association was observed among the Mental Demand, Physical Demand, and Temporal Demand subscales, with $\rho = 0.66$ for

all pairwise combinations. This robust interdependence suggests that participants perceiving a high workload in one domain tended to report similarly elevated levels in the others. Such a pattern may be interpreted as evidence that individuals receiving uniform haptic feedback experienced a more integrated workload profile, wherein cognitive effort, physical execution, and temporal urgency were simultaneously and interdependently perceived. This perceptual cohesion is indicative of heightened sensorimotor engagement and potentially greater attentional absorption during task execution.

Additional results within Group A reinforce this interpretation. The Effort subscale demonstrated positive correlations with Mental Demand ($\rho = 0.59$) and Temporal Demand ($\rho = 0.52$), suggesting that perceived exertion was not exclusively physical, but also shaped by cognitive load and time-related pressure. Furthermore, Frustration correlated positively with Mental Demand ($\rho = 0.51$), Temporal Demand ($\rho = 0.48$), and Effort ($\rho = 0.32$), indicating that negative affective states arose primarily from cognitive challenges and task complexity rather than physical limitations or performance deficits.

In Group B, correlations among subscales were systematically weaker compared to Group A. The most notable associations occurred between Effort and Frustration ($\rho = 0.67$), Physical Demand and Temporal Demand ($\rho = 0.57$), and Mental Demand and Physical Demand ($\rho = 0.49$). While these values still suggest interrelationships among certain workload components, the overall pattern reflects a more fragmented experience, with reduced integration of cognitive, physical, and emotional aspects. Notably, the Performance subscale displayed correlations close to zero with most other factors, suggesting a dissociation between self-assessed task success and perceived workload demands.



(a) Spearman correlation matrix for Group A participants in the NASA-TLX assessment.

(b) Spearman correlation matrix for Group B participants in the NASA-TLX assessment.

Figure 29: Spearman correlation analysis of NASA-TLX subscales for Group A and Group B.

7.2.4.1 Comparison Between Musicians and Non-Musicians

In the comparison between participants with and without prior musical experience (see Table 51 - Appendix F), musicians exhibited a higher mean perceived workload ($M = 37.83$, $SD = 15.00$) relative to non-musicians ($M = 29.79$, $SD = 15.87$), accompanied by greater median values (35.00 vs. 24.58) and mode scores (21.67 vs. 17.50). This pattern suggests that individuals with musical expertise tend to appraise the task as more demanding, whether in cognitive, physical, and/or affective terms.

Such differences may be attributable to the heightened engagement, concentration, and critical self-assessment often associated with trained musicians. These individuals may have invested greater attentional and expressive resources during task execution, which could amplify their subjective perception of mental load and physical effort. Moreover, their potentially greater sensitivity to latency, tactile inconsistencies, or minor audiovisual imperfections in the system's feedback could have contributed to this elevated demand perception.

Although the Mann–Whitney U test did not showed differences between the groups ($U = 135.0$, $p = 0.128$), the effect size metrics suggest a practically relevant tendency. The standardized effect size ($r = 0.281$) corresponds to a small-to-moderate magnitude, whereas Cliff's delta ($\delta = 0.35$) indicates a consistent directional difference, with musicians perceiving the experience as more cognitively demanding. Importantly, this perception should not be construed as inherently negative; rather, it may reflect greater immersion, aesthetic expectations, and performance-oriented engagement — factors that could be advantageous in advanced learning or high-level performance contexts. Alternatively, the comparatively reduced perception of workload among non-musicians may serve as a facilitating factor for their integration into immersive musical environments, offering a more accessible entry pathway for the progressive acquisition and refinement of musical skills.

An examination of the subgroups within Group A (see Table 52 - Appendix F) reveals differences in workload perception between musicians and non-musicians, with relevant implications for system design and adaptation. Musicians exhibited a mean NASA-TLX score of 36.39 ($SD = 15.62$), whereas non-musicians reported a slightly lower mean of 32.68 ($SD = 19.10$). The medians were nearly equivalent (35.00 for musicians and 37.00 for non-musicians), indicating a similar central tendency despite greater variability among non-musicians.

From a statistical perspective, the Mann–Whitney U test did not reveal a significant difference between the groups ($U = 32.5$; $p = 0.555$), a result that is consistent with expectations given the limited statistical power associated with the small sample size. Nevertheless, the effect size ($r = 0.167$) suggests a small-magnitude difference, corroborated by Cliff's delta ($\delta = 0.203$), which also indicates a modest but non-negligible effect (see Table 53 - Appendix F). These results suggest that, while workload perception remained relatively balanced, musicians still tended to report a slightly more demanding experience, even under uniform haptic feedback conditions.

This pattern may be interpreted as evidence that tactile feedback operated as a potential equalizing

factor. The close alignment of scores across profiles suggests that haptic cues contributed to a more consistent workload perception, regardless of musical expertise. Such findings support the hypothesis that sensory support in immersive environments can homogenize user experience by offering clear, structured tactile references that aid both musicians and non-musicians in maintaining task control and comprehension. From a design perspective, these results are particularly relevant as they highlight the system's potential for inclusivity and coherent performance across diverse user profiles.

In contrast, the within-group comparison for Group B (see Table 54 - Appendix F) revealed more pronounced differences in perceived workload. Musicians reported a mean NASA-TLX score of 40.00 ($SD = 16.06$), substantially higher than that of non-musicians ($M = 27.42$; $SD = 13.15$). This disparity was also evident in central tendency measures, with musicians presenting a median of 35.00 versus 21.00 for non-musicians, and a slightly higher modal score (26.00 vs. 17.00). These results indicate that, in a variable-feedback scenario, musicians perceived the experience as considerably more demanding.

Although the Mann–Whitney test did not reach conventional statistical significance ($U = 34.0$; $p = 0.131$), the effect size metrics are notable. The effect size ($r = 0.404$) denotes a moderate effect, and Cliff's delta ($\delta = 0.545$) indicates a medium-level difference, suggesting that, in most pairwise comparisons, musicians reported higher workload scores than non-musicians.

The heightened workload perception among musicians may be explained by greater engagement and situational awareness. Individuals with formal musical training are likely more sensitive to the absence of standardization in sensory feedback, which could have resulted in a less predictable and, consequently, more demanding experience. Inversely, non-musicians, lacking a strongly consolidated mental model of expected system responses, may have experienced the task with a lower degree of self-imposed demand.

The presence of a moderate difference in workload perception, modulated by musical expertise and feedback consistency, underscores the critical role of adapting system feedback to users' sensorimotor expectations. Specifically, experienced users may benefit from precise, coherent, and temporally synchronized tactile cues, while novice users may respond equally well to more exploratory and less conventional feedback strategies.

7.2.4.2 Comparison Between Musicians in Group A and Group B

The comparative analysis between musicians who evaluated the PhysioDrum system under distinct haptic feedback configurations revealed only modest differences between the groups (see Table 55 - Appendix F). Participants in Group B reported a slightly higher mean NASA-TLX score (40.00) compared to those in Group A (36.39). Despite the findings, the medians remained the same ($M = 35.00$) and the Mann–Whitney test showed no statistical significance ($U = 9.00$ and $p = 0.610$). Additionally, the effect size metrics indicated a small difference, with $r = 0.20$ and Cliff's delta at $\delta = -0.25$. Both measures suggest a limited magnitude of the difference observed.

The stability observed in workload assessments may be interpreted as evidence of the musicians' tolerance and adaptability to the sensory enrichment afforded by haptic feedback. This finding is particularly relevant in immersive musical performance contexts, where increased informational density — whether auditory, visual, or tactile — should not compromise execution fluency or perceived control. The ability of musicians to efficiently integrate multimodal sensory inputs underscores their aptitude for engaging in complex multisensory interactions and supports the feasibility of incorporating richly tactile interfaces into VR-based creation or improvisation settings.

These results suggest the viability of adopting differentiated haptic feedback as an expressive resource, even among users with elevated sensory demands such as experienced musicians. The absence of perceived overload indicates that such users may not only tolerate but also potentially benefit from more elaborate sensory interfaces. This opens promising avenues for the customization of immersive musical experiences, in which haptic feedback could be strategically deployed to reinforce dynamics, accentuate rhythmic contrasts, or enhance expressive nuance, without compromising cognitive comfort or performance fluidity.

7.2.4.3 Comparison Between VR Specialists and Non-Specialists

The comparison between participants with prior expertise in VR and those without such experience revealed notable differences in perceived workload when using the PhysioDrum system (see Table 56 - Appendix F). VR experts exhibited a higher mean NASA-TLX score (41.39; $SD = 13.53$) compared to non-experts (30.24; $SD = 15.78$). The median score was also higher among VR experts (44.16) relative to non-experts (25.00), with a smaller standard deviation in the former group indicating greater consistency in workload assessments. Although the modal value was slightly lower among experts (19.17) than non-experts (26.67), this isolated statistic does not detract from the overall trend of higher perceived workload among experienced users.

From a statistical perspective, the Mann–Whitney test indicated no significant difference between groups ($U = 102.0$; $p = 0.125$). The effect size ($r = 0.283$) corresponds to a small-to-moderate magnitude, suggesting potential practical relevance. Cliff's delta ($\delta = 0.416$) corroborates this interpretation, showing that in approximately 42% of pairwise comparisons, VR experts reported greater workload than non-experts.

Several interpretive hypotheses may contextualize these findings. First, experienced users tend to hold higher expectations regarding system responsiveness, sensory quality, and interaction coherence. Their heightened attentiveness to subtle nuances in system behavior can translate into increased cognitive effort. Second, this user profile may engage in more stringent self-assessment, which can elevate scores on the performance and mental demand subscales.

In the context of the PhysioDrum system, it is plausible that VR experts engaged more intensively with the technical dimensions of the application, such as the precision of drum strikes and rhythmic accuracy, thereby requiring greater cognitive and motor processing to achieve the intended level of

control. This engagement may also reflect a more strategic and goal-oriented interaction style, in which users actively seek to optimize their perceived musical performance. In opposition, non-experts may have approached the experience in a more exploratory or playful manner, placing less emphasis on precision and performance benchmarks, which could have mitigated perceptions of effort, demand, or frustration.

To examine whether prior experience with immersive environments influences perceived workload in the PhysioDrum context, subgroup analyses were conducted for each experimental condition (see Table 57 - Appendix F). Within Group A, participants classified as VR experts reported substantially higher NASA-TLX scores ($M = 48.75$; $SD = 7.15$) compared to their non-expert counterparts ($M = 28.86$; $SD = 16.95$). The median score for experts (46.25) was approximately twice that of non-experts (22.50), indicating a consistent difference between groups despite the limited sample size.

Although the p-value ($p = 0.078$) marginally exceeded the conventional significance threshold, it can be interpreted as borderline significant in exploratory contexts or when accompanied by effects of substantial magnitude. Indeed, the effect size ($r = 0.471$) denotes a strong effect, and Cliff's delta ($\delta \approx 0.64$) indicates a 64% probability that a VR expert in Group A would report a higher workload score than a non-expert user. These results indicate that, when exposed to equivalent haptic feedback conditions, individuals with extensive prior experience in immersive environments are more likely to exhibit heightened sensitivity to the sensory, cognitive, and motor elements underpinning the interaction. Summary statistics are provided in Table 58 - Appendix F.

In examining Group B, the results demonstrated a different configuration of outcomes compared to those identified in Group A (Table 59 - Appendix F). In this group, VR experts reported lower mean scores (26.67; $SD = 10.60$) than non-experts (31.41; $SD = 15.31$). This difference was not statistically significant ($U = 11.0$; $p = 0.798$), and effect size estimates indicated negligible impact ($r = 0.087$; $\delta = -0.1538$). The identical median values (26.67) reinforce the absence of a consistent difference between the subgroups.

The absence of substantial differences in Group B indicates that prior VR experience exerted minimal influence on workload evaluations under this specific condition. One possible interpretation is that the differentiated haptic feedback improved the perceptual clarity of the action–response relationship, thereby lowering cognitive demands. For VR-experienced participants, this may have attenuated the inclination to critically scrutinize interface performance, whereas for novice participants, it may have facilitated immersion without introducing additional cognitive strain. In both cases, the result was a relatively uniform perception of workload across varying levels of prior VR expertise.

7.2.4.4 Comparison Between VR Specialists in Group A and Group B

The comparison between VR experts assigned to different groups indicated that those who received identical haptic feedback reported substantially higher perceived workload scores ($M = 48.75$; $SD = 7.15$) than experts exposed to varied tactile feedback ($M = 26.67$; $SD = 10.61$).

Although the Mann–Whitney test ($U = 8.0$; $p = 0.133$) did not indicate statistical significance, the effect size metrics were high: $r = 0.755$ and $\delta = 1.00$. The latter value demonstrates that, in all pairwise comparisons, participants in Group A reported greater workload than those in Group B. This uniformity in responses, despite the limited sample size, suggests that the usage conditions associated with Group B were perceived as more favorable. These results are summarized in Table 60 - Appendix F.

Nonetheless, given the small sample size in this comparison, these findings cannot be generalized to the broader population of VR experts. Even so, the fact that all participants with prior VR experience assessed Group B as less demanding in terms of workload aligns with the broader trend observed in the macro-level analysis presented throughout this section.

7.2.4.5 Synthesis and Implications of NASA-TLX Results

The assessment of perceived workload during interaction with the PhysioDrum system, measured through the NASA–TLX questionnaire, indicated a favorable experience, characterized by low levels of frustration and time pressure. The most prominent dimensions were Effort and Mental Demand, consistent with the task's nature, which requires motor coordination, rhythmic attention, and controlled execution. Despite these expected peaks, both mean and median scores remained within a moderate range, suggesting an effective balance between challenge and usability, an essential condition for sustaining engagement without imposing excessive strain.

Although no statistically significant differences were observed between Groups A and B, complementary analyses revealed meaningful tendencies. Effect size and Cliff's delta estimates indicated slightly lower frustration and physical demand in Group B, suggesting that varied tactile feedback may contribute to a more comfortable and fluid interaction.

Subgroup analyses provided further insights. Participants with musical training tended to perceive the experience as more demanding, likely reflecting their higher engagement, aesthetic sensitivity, and self-critical standards, yet did not report increased frustration. This cohort maintained stable evaluations of performance and workload, indicating successful adaptation to the multisensory environment. Such tolerance suggests that PhysioDrum is capable of supporting advanced performance without compromising fluency. Among VR experts, a higher workload was observed in Group A, whereas this difference was absent in Group B, implying that the haptic feedback employed in the latter condition may have leveled the experience between expert and novice users, thereby enhancing both cognitive and physical accessibility.

From a design perspective, experienced users — whether in music or VR — tend to be more discerning and responsive to system nuances, potentially benefiting from more refined haptic responses. In an opposing manner, novice participants benefit from intuitive and cognitively lighter interactions that facilitate immersion while minimizing overload risk. The stability of performance ratings, even under conditions perceived as demanding in other dimensions, reinforces the system's robustness across user

profiles.

In conclusion, PhysioDrum demonstrated favorable outcomes regarding cognitive load during use, with the potential for adaptation to individual user profiles. The differentiated haptic feedback showed subtle benefits, particularly for participants with substantial VR experience. Moreover, the combination of subjective measures and statistical analysis indicates that, although the system performs robustly across varied user characteristics, its effectiveness is marginally enhanced when employing diverse feedback mechanisms.

7.2.5 Assessment of Haptic Questionnaire (HQ)

Figure 30 presents the mean scores for each subscale, as reported by the three groups under analysis, together with the overall score for the haptic experience. The comparative assessment between Groups A and B, based on the subscales, indicated a broadly similar response profile, with only minor variations across the evaluated dimensions. The Autotelic subscale exhibited virtually identical mean values (Group A: 3.47; Group B: 3.52) and closely aligned medians (≈ 3.4). The Mann–Whitney test confirmed the absence of a statistically significant difference ($U = 106.0$; $p = 0.801$), with a negligible effect size ($r = -0.0492$). Independent of the specific configuration of the tactile feedback, participants evaluated the experience as comparably engaging, particularly with respect to the inherent quality of the haptic effect itself (see Table 62 - Appendix G).

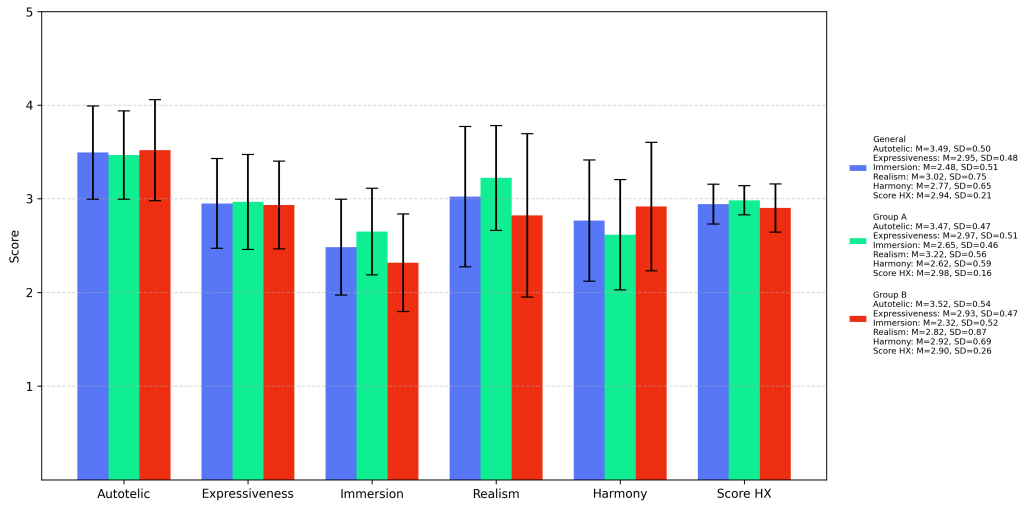


Figure 30: Mean scores per subscale of the Haptic Questionnaire (HQ) for Groups A and B.

In the Expressiveness construct, no meaningful differences were identified between the groups, with comparable mean values (Group A: 2.97; Group B: 2.93) and stable medians around 3.0. The Mann–Whitney test revealed no indication of dissociation ($U = 112.5$, $p = 1.000$), and the effect size was null ($r = 0.0$). These results indicate that both haptic feedback styles were perceived as sufficiently expressive, enabling users to convey sensory input through the system in a clear and subjectively satisfactory manner.

In contrast, the Immersion dimension revealed more pronounced differences. Although both groups

shared the same median score (2.5), Group A exhibited a higher mean (2.65) compared to Group B (2.32). While this difference was not statistically significant ($U = 153.0$; $p = 0.0879$), it was associated with a moderate effect size ($r = 0.306$), the highest among all subscales. The standardization of haptic feedback may have contributed to a more stable and continuous perception of immersion.

For the Realism dimension, Group A demonstrated a clear advantage ($M = 3.22$; $Median = 3.33$; $SD = 0.56$) over Group B ($M = 2.82$; $Median = 3.00$; $SD = 0.87$). Although the difference was not statistically significant ($U = 142.0$; $p = 0.218$), it was associated with a small effect size ($r = 0.223$), suggesting a more coherent integration between the haptic stimuli and the virtual environment for participants in Group A. The higher standard deviation observed in Group B may indicate that tactile feedback was perceived less consistently within this group, resulting in a more dispersed evaluation of realism.

For the Harmony factor, which assesses the integration of haptic feedback with other sensory modalities, Group B obtained a slightly higher mean score (2.92) compared to Group A (2.62), with medians remaining close (2.75 and 2.50, respectively). The Mann–Whitney test did not indicate a statistically significant difference ($U = 83.5$; $p = 0.231$), and the effect size was small and negative ($r = -0.2196$), suggesting a slight preference in Group B for the varied feedback condition in terms of synchrony between tactile stimulation, motor interaction, and auditory perception.

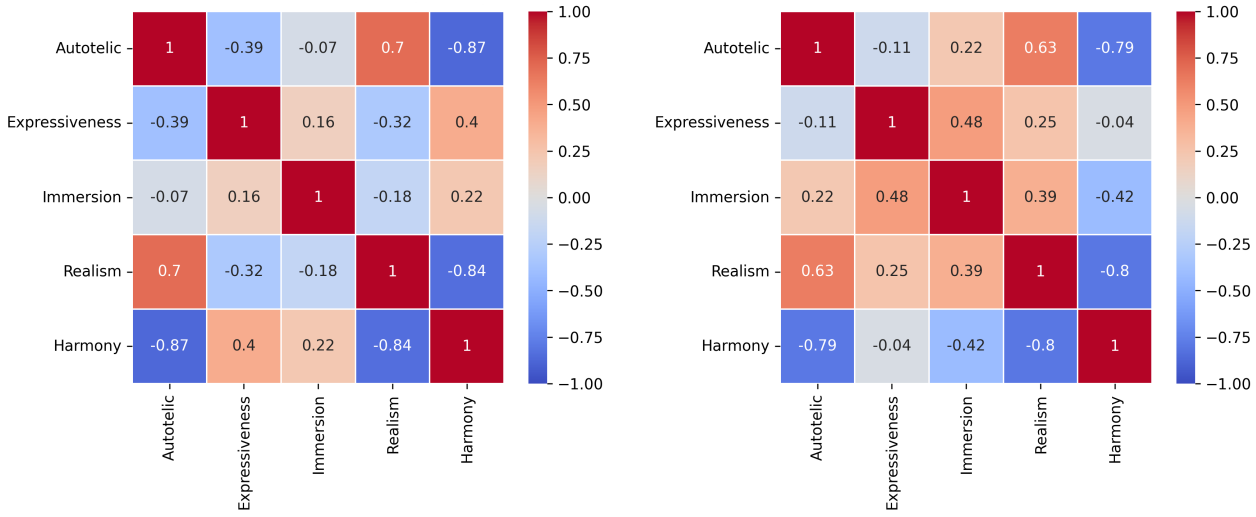
Regarding the overall score, mean values were highly similar between the groups (Group A: 2.98; Group B: 2.90), with both medians approximating 3.0. The Mann–Whitney test ($U = 32.0$; $p = 0.4301$) and the corresponding effect size ($r = 0.147$) corroborate the absence of a statistically significant difference. This aggregate measure reflects the subscale-level analysis, indicating that the quality of the haptic experience was consistently evaluated as positive and relatively stable across conditions. Minor advantages were observed for Group A in metrics related to immersion and realism, while Group B showed a slight advantage in Harmony and Autotelic results; however, neither trend exerted a substantial influence on the final overall scores.

Following this broader analysis, the correlation between the subscales was examined using Spearman's method, as presented in Figure 31, which reveals substantial differences in how participants integrated the tactile qualities of the PhysioDrum experience. In Group A, extremely strong and polarized correlations emerged, suggesting a perceptual architecture structured around two primary poles: pleasure and realism versus coherence and integration. The positive association between Autotelic and Realism ($\rho = 0.70$) indicates that, for these users, higher perceived tactile realism was accompanied by greater enjoyment and autonomy in the experience. Complementarily, the strong negative correlation between Autotelic and Harmony ($\rho = -0.87$) suggests that when tactile stimuli were perceived as more pleasurable, they were also regarded as less integrated with other sensory channels, highlighting the greater relative influence of haptic feedback on participants' perceptual judgments. This tendency is further reinforced by the negative correlation between Realism and Harmony ($\rho = -0.84$), which points to a dissociation between sensory fidelity and systemic cohesion: the closer the tactile stimuli approached the expected physical realism, the less harmoniously they were perceived to align with the

remainder of the immersive musical experience.

In Group B, the correlation pattern was more balanced and evenly distributed, reflecting a perceptual architecture that was comparatively more holistic and interconnected. The association between Autotelic and Realism remained high ($\rho = 0.63$), reinforcing the notion that hedonic value continued to be linked to the perception of tactile plausibility. However, negative associations with Harmony persisted ($\rho = -0.79$), suggesting that heightened sensory fidelity in haptic feedback could conflict with playful or representational elements of the experience. On the other hand, the more distributed correlation structure in Group B, with moderate positive coefficients among Immersion, Expressiveness, and Realism (ρ ranging from 0.25 to 0.48), indicates that the diversified feedback modality fostered a more integrated experiential framework, wherein expressivity, immersion, and realism were perceived as mutually reinforcing components.

From these observations, it can be inferred that Group A exhibited greater internal consistency in their responses, a pattern that may be attributed to the predictability of the haptic feedback. Such predictability tends to benefit users in virtual reality environments by reinforcing the perception of agency and control. To the contrary, the diversified feedback provided to Group B may have increased the sensory complexity of the experience, thereby enhancing perceptual richness and fostering expressive spontaneity for certain participants.



(a) Spearman correlation matrix for Group A participants in the Haptic Questionnaire assessment.

(b) Spearman correlation matrix for Group B participants in the Haptic Questionnaire assessment.

Figure 31: Spearman correlation analysis of Haptic Questionnaire subscales for Group A and Group B.

This polarization may indicate a perceptual mismatch between the haptic feedback, designed to emulate realistic sensations of playing an actual drum set, and the visual elements, which adopted a more playful and artistic approach, lacking motion cues or visual indicators corresponding to the activation of specific components. Moreover, the low performance of auditory realism, as highlighted by the PQ results presented in Section 7.2.2, may have further contributed to the advantage observed for the autotelic dimension of the haptic component.

A similar trend is observed in the weak correlations between Immersion and the other factors, suggesting that the sense of engagement in the virtual environment was not directly associated with Expressiveness or Harmony. The only statistically significant positive correlation involving Immersion was with Harmony ($\rho = 0.22$), although its magnitude remained low. This finding suggests that, in Group A, the experience was perceived in two distinct ways: either as realistic and pleasurable, or as integrated and coherent. This dichotomy highlights that a subset of participants focused primarily on the hedonic aspects of the system, whereas others prioritized its functional integration.

This perceptual cohesion can be interpreted as an indicator of the maturity of PhysioDrum's sensory design. Participants in Group B, despite being exposed to more complex stimuli, demonstrated an ability to integrate multiple dimensions of the haptic experience. Although conflicts between pleasure (Autotelic) and coherence (Harmony) persisted, their impact appears to have been mitigated by the presence of stronger interconnections among the remaining subscales. In this way, the tactile feedback was positively received by the participants.

A clear divergence was observed in the reception of the system by Group A. A subset of participants evaluated the system more objectively, thereby not associating the realistic haptic feedback with other sensory modalities, whereas another subset focused predominantly on the entertainment value and the playful aspects of the application. This pattern was not evident in Group B, indicating that the diversified feedback may have facilitated a more transparent integration between the experiential dimensions.

In light of these findings, future iterations of PhysioDrum should consider enhancing the realism of auditory and visual elements or, alternatively, adopting haptic components more closely aligned with hedonic features rather than purely functional attributes.

7.2.5.1 Comparison Between Musicians and Non-Musicians

The results indicate that there were no statistically significant differences between musicians and non-musicians in any of the assessed dimensions. The Mann–Whitney test confirmed the absence of significance across all subscales ($p > 0.30$ for all comparisons), while both effect size values and Cliff's delta further reinforced the very weak or negligible nature of the observed differences.

In the Autotelic dimension, both groups exhibited very similar mean scores (musicians: 3.44; non-musicians: 3.52), with nearly identical medians. The p-value ($p = 0.807$) and the negligible effect size ($r = 0.048$; $\delta = -0.06$) indicate that the intrinsic pleasure of engaging with the experience was perceived similarly, regardless of musical background. This finding reinforces that motivation and engagement with PhysioDrum's haptic system are not necessarily contingent upon prior experience with musical instruments.

In the Expressiveness construct, both groups again demonstrated highly similar means and medians, with scores hovering around 3.0. The statistical analysis yielded a p-value of 0.875 and a null effect size ($r = 0.032$, $\delta = 0.04$), indicating that musicians and non-musicians perceived the system as

equally expressive. PhysioDrum's haptic interface was able to convey communicative nuances that were consistently perceived across all participants, regardless of their prior musical training.

In the Immersion subscale, mean scores were virtually identical between groups ($M = 2.48$ for both), with a slightly higher standard deviation observed among musicians. The Mann–Whitney test once again indicated no statistically significant difference ($p = 0.70$), with a very small effect size ($r = 0.072$, $\delta = -0.09$).

The Realism measure was the only dimension to display a slight tendency toward group differentiation, with musicians reporting higher mean scores (3.13) than non-musicians (2.97). Musicians also had a higher median (3.5 vs. 3.0), and the larger standard deviation suggests greater variability within the group. Although the result did not reach statistical significance ($p = 0.301$), the effect size was relatively more pronounced compared to other dimensions ($r = 0.189$, $\delta = 0.235$), pointing toward a greater receptiveness of musicians to the system's sensory responses.

For the Harmony attribute, differences were minimal in both mean (2.80 vs. 2.75) and median values. The Mann–Whitney test ($p = 0.911$; $r = 0.024$; $\delta = -0.03$) confirmed a complete overlap in the distributions of responses between the two groups.

In the total score, both groups exhibited nearly identical mean values (musicians: 2.96; non-musicians: 2.93), accompanied by low standard deviations. The absence of statistical significance ($p = 0.741$) and the very small effect size ($r = 0.064$; $\delta = 0.08$) corroborate the general conclusion that the haptic experience was perceived similarly and positively across both profiles. This consistency between participant profiles, irrespective of musical expertise, suggests that the system's design effectively accommodates a diverse range of users without compromising experiential quality.

In the intra-group analysis, Group A exhibited a high degree of homogeneity in the perception of the haptic experience (see Table 63 - Appendix G). For the Autotelic category, both musicians and non-musicians reported identical mean scores (3.46), with a completely null statistical result ($U = 27.5$; $p = 1.000$; $r = 0.015$; $\delta = 0.019$), indicating an equally high perception of this metric across all participants. In Expressiveness, musicians reported a slightly higher mean score (3.16) compared to non-musicians (2.83), which may suggest a mild trend toward greater sensitivity among musicians to the expressive nuances of the haptic feedback. However, this difference did not reach statistical significance ($U = 37.0$; $p = 0.252$), although the effect size ($r = 0.304$) and Cliff's delta ($\delta = 0.370$) were both in the moderate range, indicating a tendency for higher Expressiveness scores among musically trained participants. A similar pattern was observed in the Realism dimension, where musicians assigned higher mean scores (3.39) than non-musicians (3.11), with identical effect size and p-value metrics to those reported for Expressiveness.

For the Immersion/Control and Harmony elements, the differences between the two groups were negligible. In Immersion/Control, the mean scores were virtually identical (2.62 for musicians and 2.66 for non-musicians), providing a completely null statistical result ($U = 27.0$; $p = 1.00$; $r = 0.00$; $\delta = 0.00$). For Harmony, the mean score among musicians was slightly lower (2.58) compared to non-

musicians (2.63); however, the effect direction was reversed and of negligible magnitude ($r = 0.030$, $\delta = -0.037$). In view of these results, the sense of immersion and the integration of tactile stimuli with motor actions were experienced similarly by both profiles, regardless of musical expertise.

For the HQ Total score, musicians reported a slightly higher mean (3.04) compared to non-musicians (2.94), suggesting a marginally more favorable overall perception among music experts, potentially driven by their slightly elevated ratings in this category. The p-value ($p = 0.261$) and the effect size metrics ($r = 0.304$; $\delta = 0.370$) maintain the pattern of a weak-to-moderate trend, without reaching statistical significance.

In Group B, the comparison between musicians and non-musicians revealed a pattern analogous to that observed in Group A, with no statistically significant differences in any of the HQ subscales (see Table 64 - Appendix G). Certain tendencies suggest subtle perceptual variations between the two profiles, particularly in attributes related to expressiveness and immersion.

In the Autotelic subscale, non-musicians reported a slightly higher mean score (3.56) compared to musicians (3.40). In spite of that, this difference was minimal and did not reach statistical significance ($U = 19.5$; $p = 0.793$), with a small effect size ($r = 0.084$) and a negative Cliff's delta ($\delta = -0.114$), indicating that this variation does not constitute a meaningful distinction between the groups under analysis.

Within the Expressiveness construct, the difference between groups was comparatively more pronounced, with non-musicians reporting a higher mean score (3.02) than musicians (2.68). Although this difference did not reach statistical significance ($U = 13.5$; $p = 0.279$), the effect size was moderate ($r = 0.287$; $\delta = -0.386$), indicating a tendency for non-musicians to perceive themselves as more capable of conveying nuances and emotions through the system. This tendency may reflect a greater openness to unfiltered sensory experiences, whereas musicians, owing to their formal training, may possess more sophisticated expectations regarding the system's expressive capabilities.

For the Immersion/Control feature, mean scores were closely aligned between groups (musicians = 2.25; non-musicians = 2.34), and the difference was not statistically significant ($U = 17.0$; $p = 0.542$), with a small effect size ($r = 0.169$; $\delta = -0.227$). This suggests a broadly comparable perception of immersion, albeit marginally higher among non-musicians.

Regarding Realism, results were convergent, with virtually no difference between groups (musicians = 2.75; non-musicians = 2.84), accompanied by the maximum p-value ($U = 22.5$; $p = 1.000$) and a negligible effect size ($r = 0.017$; $\delta = 0.023$). This convergence indicates a consistent interpretation of the haptic feedback across both participant profiles.

In the Harmony subscale, musicians reported higher mean scores (3.12) compared to non-musicians (2.84), accompanied by a positive Cliff's delta ($\delta = 0.091$), but the effect size was negligible ($r = 0.067$) and the difference was not statistically significant ($U = 24.0$; $p = 0.843$), suggesting only a marginal inclination of musicians to perceive stronger integration between multisensory stimuli.

About the HQ Total score, musicians exhibited a lower mean (2.84) than non-musicians (2.92), with

the Mann–Whitney test again indicating no statistical significance ($U = 16.5$; $p = 0.514$) and a small effect size ($r = 0.185$; $\delta = -0.250$). This decrease in musicians' overall scores may indicate stricter critical standards in assessing sensory stimuli, especially in contexts that involve more intricate haptic feedback.

On balance, perceptions between musicians and non-musicians in Group B were similar, although certain reversals were observed in subscales such as Expressiveness and Immersion, where non-musicians reported more positive experiences. Even when the application delivers differentiated and potentially richer tactile feedback, musical training does not constitute a determining factor for enhanced system enjoyment.

7.2.5.2 Comparison Between Musicians in Group A and Group B

In comparing musicians from Groups A and B (see Table 65 - Appendix G), the Autotelic dimension showed nearly identical mean values (Group A = 3.46; Group B = 3.40), with $U = 12.0$ and $p = 1.000$, resulting in null effect sizes ($r = 0.00$; $\delta = 0.00$). This outcome indicates that musicians found the haptic component to be intrinsically engaging, regardless of whether the feedback was uniform or varied. This result suggests that engagement driven by intrinsic motivation was unaffected by the characteristics of the sensory stimulus. It implies that both consistent and nuanced variations in haptic feedback can contribute positively to the overall user experience.

The Expressiveness dimension exhibited a more pronounced difference, with Group A ($M = 3.16$) surpassing Group B ($M = 2.68$). Although this difference did not reach statistical significance ($U = 17.5$; $p = 0.279$) and the effect size was moderate ($r = 0.371$; $\delta = 0.458$), musicians in Group A perceived the system as offering greater expressive capabilities.

In the Immersion metric, Group A also obtained higher mean scores (2.62 vs. 2.25), with a small-to-moderate effect size ($r = 0.270$; $\delta = 0.333$). This means that the sensation of "being there" in the experience was more pronounced among participants who received constant haptic feedback. Such an outcome may be attributed to the sensory predictability and stability afforded by uniform tactile stimulation, factors particularly valued by musicians, who often rely on rhythmic and sensory consistency to maintain focus and fluency during performance.

The evaluation of Realism revealed an even greater difference ($GroupA = 3.39$; $GroupB = 2.75$), with $r = 0.337$ and $\delta = 0.417$. This pattern shows that musicians in Group A perceived the haptic feedback as more realistic, possibly due to its closer alignment with the tactile effects typically expected in real instrumental contexts. Experienced musicians often develop specific expectations regarding sensorimotor congruence, and the variability of interaction modes in PhysioDrum may have compromised the perceived authenticity of the experience for those in Group B.

In Harmony, however, Group B outperformed Group A (3.12 vs. 2.58), although the effect was small ($r = 0.236$; $\delta = -0.292$). This may indicate that the sensory diversity in Group B enhanced the perceived integration between tactile stimuli and other performance components, such as auditory

cues, visual elements, timing, and musical gestures. These results suggest that multisensory richness may play a positive role in fostering a sense of functional cohesion, provided that it does not interfere with expressive clarity.

In the end, the total score was higher for Group A (3.04) than for Group B (2.84), with $U = 19.0$, $p = 0.171$, and a considerable effect size ($r = 0.472$; $\delta = 0.583$). This indicates that musicians attributed higher quality to the haptic experience when feedback was uniform — likely because it aligned more closely with their mental models of musical action and sensory response. The consistency of the stimulus appears to have provided a more predictable, stable, and interpretable environment, attributes highly valued by performers accustomed to rhythmic and functionally coherent tactile structures.

7.2.5.3 Comparison Between VR Specialists and Non-Specialists

The comparative analysis between VR experts and non-experts revealed differences in the HQ assessments (see Table 66 -Appendix G). In Autotelic metric, non-experts assigned higher scores ($M = 3.58$) than experts ($M = 3.13$), with a difference approaching statistical significance ($U = 35.5$; $p = 0.059$). The observed effect size was small-to-moderate ($r = 0.345$), and Cliff's delta was high ($\delta = -0.506$), indicating that in more than half of the paired comparisons, non-experts scored higher values. Accordingly, participants with little or no prior VR experience tended to perceive the haptic feedback as more enjoyable and intrinsically motivating, potentially due to the novelty of this approach to delivering such stimuli, thereby eliciting greater sensory interest. Experts users, in turn, may have compared the experience to more sophisticated VR applications, potentially lowering their intrinsic pleasure ratings.

In the Expressiveness dimension, experts presented higher mean scores (3.12) than non-experts (2.90), though the difference was not statistically significant ($U = 87.5$; $p = 0.428$), with a small effect size ($r = 0.146$) and a modest delta ($\delta = 0.215$). This minor differential may suggest that VR experts were slightly more adept at recognizing and leveraging the subtleties of tactile feedback to identify musical actions or intentions.

The Immersion factor showed similar mean values between groups (VR experts = 2.58; VR non-experts = 2.45), with a negligible effect size ($r = 0.033$) and a low Cliff's delta ($\delta = 0.048$), confirming that the sense of immersion was perceived equivalently. This result suggests that familiarity with VR environments was not a decisive factor in the perception of engagement within the immersive environment, which is encouraging for the system's applicability to broader audiences regardless of VR expertise.

In the Realism construct, non-experts reported higher scores ($M = 3.09$) than experts ($M = 2.72$), with a small-to-moderate effect size ($r = 0.194$; $\delta = -0.284$), though the difference was not statistically significant ($U = 51.5$; $p = 0.288$). This result is consistent with the Autotelic pattern, where users with less critical VR repertoires tended to find tactile stimuli more convincing and realistic, while experts likely maintained a more technical or critical perspective, benchmarking the system against

stricter standards of tactile realism.

About the Harmony, experts achieved higher outcomes (3.04 vs. 2.69), with a small effect size ($r = 0.175$; $\delta = 0.256$) and no statistical significance ($U = 90.5$; $p = 0.344$). This tendency suggests that the integration between tactile feedback and motor actions was perceived as more evident or functionally coherent by experienced users.

For the total HQ score, mean values were nearly identical (VR experts = 2.92; VR non-experts = 2.95), with no significant difference ($U = 53.0$; $p = 0.336$), a small effect size ($r = 0.179$), and a modest Cliff's delta ($\delta = -0.263$). Despite the lack of statistical significance, the trend of higher scores among non-experts reflects a consistent pattern of greater subjective impact of haptic feedback among participants with limited prior immersive experience, whereas experts scored better on functional scales addressing the integration of tactile elements with other system functionalities.

In this fashion, it possible to note that non-experts in VR tended to place greater value on the pleasurable and realistic aspects of the haptic experience, whereas experts demonstrated a slight advantage in the perception of expressiveness and sensorimotor harmony. Novice users may benefit from more intense or engaging stimuli, while advanced users require greater refinement, responsiveness, and multisensory integration to achieve full engagement.

When examining these same profiles exclusively within Group A, participants with VR expertise assigned lower scores in almost all subscales, with the exception of Expressiveness and Immersion, where mean values were only marginally higher. For instance, in the Autotelic subscale, experts reported a mean of 3.25, whereas non-experts scored higher ($M = 3.54$). Although not statistically significant ($U = 15.0$; $p = 0.390$), this difference presented a small-to-moderate effect size ($r = 0.236$; $\delta = -0.318$), suggesting a predisposition for more experienced users to perceive the tactile experience as less intrinsically rewarding. This pattern persisted in the total HQ score, with nearly identical means between experts (2.99) and non-experts (2.98) and a modest effect size ($r = 0.236$), indicating a broadly similar overall perception of tactile quality between the two subgroups (see Table 67 - Appendix G).

The scenario changes substantially in Group B (see Table 68 - Appendix G). In this case, the differences between experts and non-experts were more pronounced, with consistently lower scores among participants with prior VR experience. In the Autotelic subscale, experts assigned a mean of 2.90, in stark contrast to the 3.61 reported by non-experts. This difference was accompanied by a marginal p-value ($U = 2.5$; $p = 0.087$), a large effect size ($r = 0.460$), and a highly pronounced Cliff's delta ($\delta = -0.808$), indicating that in over 80% of paired comparisons, non-experts perceived the feedback as more autotelically positive than experts. A similar behavior was found in the Realism subscale (VR experts: 2.165; VR non-experts: 2.924), with $r = 0.197$ and $\delta = -0.346$, suggesting that experienced users regarded the haptic feedback as less plausible or coherent with the immersive experience. These findings suggest that, within Group B, tactile variety was particularly advantageous for less-experienced VR users, likely through the induction of surprise, novelty, and heightened sensory engagement. At the same time, experts may have evaluated the experience more critically or struggled

to adapt to feedback delivered directly through drumsticks rather than traditional controllers.

The Harmony subscale presented an exception. In this dimension, experts scored significantly higher ($M = 3.62$ vs. 2.80), with a medium effect size ($\delta = 0.50$), suggesting that they perceived greater synergy between tactile stimuli and motor interactions. This indicates that stimulus variation may have been functionally integrated and intelligible to more experienced users.

These results demonstrate that, for inexperienced users, varied haptic feedback can be highly engaging and enjoyable, contributing to a richer experience. For experienced VR users, the sophistication and consistency of sensory feedback become critical criteria for positive evaluation, making it essential to ensure realism, responsiveness, and temporal coherence with motor actions.

7.2.5.4 Comparison Between VR Specialists in Group A and Group B

A comparative analysis between VR-experienced participants in Groups A and B revealed nuanced differences across the evaluated subscales (see Table 69 - Appendix G). In the Autotelic dimension, participants exposed to uniform haptic feedback exhibited a higher mean score (3.25) compared to those receiving varied stimuli (2.90). Although the difference did not reach statistical significance ($p = 0.617$), the effect size was small ($r = 0.283$; $\delta = 0.375$). The positive sign of Cliff's delta suggests a tendency toward greater intrinsic motivation and enjoyment of interaction in Group A.

In the Expressiveness factor, both groups achieved identical mean scores (3.12), indicating equivalent perceived expressiveness regardless of haptic feedback type. The Mann–Whitney test confirmed this equality ($U = 4.0$; $p = 1.000$), with a null effect size ($r = 0.000$; $\delta = 0.000$), implying that, for VR-experienced users, expressiveness is consistently achieved independent of haptic modality.

Immersion facet demonstrated the greatest discrimination between groups. Group A scored substantially higher ($M = 2.81$) than Group B ($M = 2.12$). The p-value ($p = 0.100$) remained above the significance threshold and the effect size was very large ($r = 0.756$; $\delta = 1.000$), with no overlap in scores between groups. This finding reinforces the hypothesis that individuals familiar with immersive environments place greater value on sensory consistency, potentially enhancing absorption in the task and reinforcing presence and engagement.

For Realism, a moderate difference favored Group A ($M = 2.99$) over Group B ($M = 2.16$), even though the result lacked statistical significance ($p = 0.806$) and exhibited a small effect size ($r = 0.189$; $\delta = 0.250$). The trend nonetheless suggests that uniform haptic feedback was perceived as more realistic by VR-experienced participants, likely due to greater alignment with their internalized sensorimotor models developed through prior interaction with immersive systems.

In the Harmony subscale, Group B obtained a higher mean score (3.62) than Group A (2.75), suggesting a more favorable perception of the integration between haptic feedback and other sensory elements such as auditory and visual stimuli. Statistical significance was not reached ($U = 1.5$; $p = 0.348$) and the effect size was moderate ($r = 0.472$; $\delta = -0.625$), with the negative sign indicating a clear advantage for Group B. This may reflect the possibility that varied haptic stimuli enhance the perception

of multimodal synchrony, an aspect particularly valued by VR-experienced users who expect greater complexity and articulation in immersive systems.

The total score slightly favored Group A ($M = 2.99$) over Group B ($M = 2.79$), again without statistical significance ($p = 0.814$) and with a modest effect size ($r = 0.189$; $\delta = 0.250$). This trend aligns with the findings from the Autotelic, Immersion, and Realism subscales. For participants with VR experience, consistent haptic feedback tends to create a more cohesive and immersive experience. On the other hand, the sensory diversity characteristic of Group B may provide benefits for specific aspects, such as multimodal harmony.

7.2.5.5 Synthesis and Implications of HQ Results

This discussion indicates that the haptic system of PhysioDrum provides a strong sensory experience, receiving consistently positive evaluations across all HQ subscales, regardless of whether the feedback is uniform or varied. Uniform tactile feedback showed a stronger relationship with perceptions of realism and immersion, especially among experienced VR users and musicians, groups that generally value sensory consistency and predictability. This preference for more stable feedback can be seen as a pursuit of alignment between sensorimotor expectations and the system's responses, which is essential for tasks that require precision and rhythmic control, such as musical performance.

Varied tactile feedback seems to enhance aspects like multimodal harmony and contextual expressiveness. This operational mode can enrich the experience for less experienced users or those who are more exploratory and playful. Also, this approach is particularly effective in broadening the sensory experiences of novice users, helping them achieve a more integrated understanding of touch, sound, and movement.

Regarding the influence of prior musical experience, the findings point toward a form of "positive neutrality". Musicians and non-musicians evaluated the experience in similar ways, with only occasional, non-significant differences. This suggests that PhysioDrum is both inclusive and accessible, capable of eliciting pleasure, expressiveness, and immersion irrespective of users' technical background. In this context, musicians tended to slightly favor uniform stimuli, whereas non-musicians responded more positively to varied tactile feedback.

Prior VR experience emerged as an important moderating factor. Experts tended to be more critical of feedback realism, whereas non-experts demonstrated greater enthusiasm and hedonic engagement. This indicates that the impact of haptic feedback is partly determined by users' sensory repertoire, underscoring the need for adaptive modes of sensory delivery.

These results not only validate the relevance of the haptic dimension in immersive musical experiences but also provide a solid empirical foundation to guide future investigations in the fields of the IoMusT, interactive multisensory systems, and immersive virtual environments as a whole.

7.2.6 Qualitative Evaluation Through Semi-structured Interview

The qualitative analysis, grounded in data obtained from semi-structured interviews (available in Appendix H), elucidates key dimensions of interaction in immersive musical environments mediated by haptic technologies. The thematic approach employed enabled the identification of consistent patterns in engagement, usability, sensory perception, and improvement suggestions, thereby providing substantive evidence regarding both the efficacy and current limitations of PhysioDrum.

In General, participants' accounts indicate that PhysioDrum delivers a predominantly positive experience, frequently described as enjoyable, stimulating, and engaging. A recurrent theme was the sense of freedom and enjoyment during interaction, with several reports highlighting a progressive immersion as familiarity with the environment increased. This progression was exemplified by P03, who stated: "*At some point, you are just playing the drums without worrying about the surrounding elements*", suggesting a state of cognitive flow characteristic of activities that balance challenge and motor skills. Such experiences were corroborated by other participants (e.g., P04: "*It was fun. It was nice to have a drum kit in front of me, to play it, hear the sounds, and feel the vibration, especially in the guided part, feeling that I was playing the basics of a drum, with a simple beat structure*"; P07: "*The engagement I felt was very high and stimulated my curiosity, making me want to test different parts of the system. I wanted to try different musical options*"; P14: "*The system is dynamic, it captures your attention. You want to improve the next time you repeat the activity. I think it is impossible not to be entertained given the dynamism and attentional focus you employ to improve your performance. At the same time, I think it is a very technological, very modern system*"). These perspectives underscore not only the ludic aspect of the system but also its ability to sustain attention and foster curiosity and exploration.

The repeated association of the experience with gaming suggests that the system achieves engagement levels akin to gamified practices, which may serve as a strategic vector for educational and recreational applications (e.g., P02: "*The application reminded me of a game*"; P09: "*Having a virtual environment with a music game seems very cool. It is something I would like to play more often*"; P22: "*I thought the concepts behind this application were great. It seems like a game*"; P23: "*I kept thinking about Beat Saber*[3], *because it has a training session before the game. I found PhysioDrum similar, as if it were a training platform*"; P27: "*I really liked the environment, found it very comfortable, I was very relaxed. This game-like aspect, where you have to replicate movements, engaged me a lot. Having a specific goal to accomplish was better than just freely testing. That part didn't interest me much, but the other one I found very engaging*").

Despite the predominantly positive evaluations, technical limitations were also highlighted, particularly concerning the accuracy of strike detection and system ergonomics. Issues such as "ghost strikes[4]", double strike detection, and collision inaccuracies were noted (e.g., P01: "*Sometimes, I no-*

[3] https://beatsaber.com/

[4] Ghost strike can refer to several different game mechanics or techniques, depending on the context of the game being played. In general, it often involves a strike that is hidden or unexpected, or a strike that is delivered in a way that seems to defy the usual rules of the game.

ticed a duplicate strike. Once I got used to the environment, I understood I wasn't supposed to register those hits, so it was adaptable"; P06: "*I accidentally hit drum pieces I did not intend to. This issue is also common in acoustic drums, so I believe it's a feature encountered in drum practice*"; P08: "*I took some time to get used to the distance I needed to hit. Sometimes I thought I had hit but no sound was produced*"; P27: "*The main problem for me was that sometimes it registered a duplicate strike. That distracted me*"). These observations point to the need for refinements in physical modeling and spatial calibration.

Regarding the ergonomic assessment of the system, the drumsticks were predominantly perceived as intuitive input interfaces, with the additional components, such as the integrated electrical circuit and spheres, exerting no negative impact on usability. In this way, some participants reported that these modifications enhanced control and stability during interaction (e.g., P09: "*The spheres at the top felt strange at first glance, but then I realized they helped control the drumstick*"; P10: "*Since we had to hit the drum and then remove the hand, making the drumsticks heavier was actually helpful*").

In parallel, participants frequently described the foot pedals as both ergonomically adequate and easy to use (e.g., P00: "*Pedals are intuitive to use*"; P04: "*I found the pedals comfortable, with no major difficulties*"). These reports suggest that the system's physical interfaces achieved a favorable balance between functionality, comfort, and ease of learning.

Nevertheless, certain negative considerations were raised regarding both interfaces. For the drumsticks, several participants (P02, P10, P18, P29) emphasized their excessive weight, which was perceived as detrimental to prolonged use. Additionally, a discrepancy between the position of the physical and virtual drumsticks was reported, as illustrated by P03: "*Visually, the drumsticks appeared far apart, but when I performed certain movements, they would touch each other*".

Concerning the pedals, critical remarks focused primarily on the device noise generated upon activation, which was considered to disrupt immersion (P06: "*The pedal generates noise, and this interferes in the experience*"). Another recurrent point concerned their dissimilarity to traditional drum pedals, which imposed certain constraints on interaction. As noted by P04: "*The pedals differ from how we usually play. I am used to keeping the hi-hat pedal pressed, but here I could not do that. I also found them very light*". Similarly, P18 stated: "*I found the pedals uncomfortable, because in a real drum set the pedal is something you have to press hard, especially the bass drum pedal. Having to adapt to a more subtle motion was uncomfortable*". Furthermore, the absence of fixed positioning on the floor and the lack of virtual representation led some participants to lose spatial reference during use (P07: "*I could not see the pedal, so I had to touch around on the floor to find it*"; P29: "*I could not see the pedal. This disrupted my experience in some manner*"). Collectively, these findings indicate that, while the system's interfaces were broadly functional and intuitive, certain design aspects, particularly related to weight, sensory congruence, and spatial anchoring, warrant refinement to optimize ergonomics and immersion.

Towards the perception of haptic feedback, the participants' responses proved heterogeneous. While a subset of users acknowledged the value of vibration as a means of enhancing immersion and providing

sensory guidance (P00: "*I found the haptic feedback quite interesting. I did not expect that feeling a vibration would make such a difference. I think it was useful, because it helped me perceive whether I had actually hit the drum*"; P05: "*I found the feedback useful. I just wish it had been stronger*"; P13: "*I really enjoyed the haptic feedback. The vibration conveyed a sense that the hit had an effect, which greatly aided my understanding of the interaction*"; P18: "*I found the task with feedback more enjoyable. I was actually feeling something in my hand. It was not as if I was just moving randomly. The vibration provides a more engaging experience*"; P27: "*I thought it was really cool. I preferred the part with haptic feedback*"), others reported insensitivity, latency, or even indifference toward its presence (P04: "*I did not feel any difference between the sessions with or without feedback. I paid much more attention to the visual and audio parts (...) For me, honestly, having haptic feedback made no difference. If I had to choose, I would choose without, because it would make no difference at all. But, again, it neither hindered nor helped*"; P10: "*The haptic feedback did not make much difference*"; P22: "*I found the feedback annoying and prolonged. I think I performed better without it*").

This divergence suggests that the effectiveness of tactile responses depends on factors such as stimulus intensity, latency, and the ability to discriminate between different tactile events, factors that should be prioritized in future system optimizations. Nevertheless, accounts such as that of P14: "*The task with feedback ends up being a bit better, because you adapt the way you play to the feedback you receive*", corroborate the potential role of vibration as a sensorimotor reinforcer in the process of rhythmic skill acquisition.

Another important aspect observed in PhysioDrum concerns its accessibility and learning curve. The majority of participants perceived the system as intuitive, including those without prior experience with musical instruments or virtual reality technologies (P03, P05, P10, P16, P28). This perception of accessibility manifested across two complementary dimensions: i) technical, referring to the ease of operating the controls and the accurate recognition of user actions; and ii) cognitive, related to the clarity of the instructions provided and the logical progression of tasks.

Nonetheless, some participants with limited technological familiarity (P11, P27) encountered initial challenges regarding both control manipulation and spatial orientation. These observations suggest that the implementation of visual tutorials or interactive guidance tools could facilitate the early adaptation process, thereby reducing entry barriers and promoting a smoother integration into the system's interactive environment.

Beyond its ludic, educational, and artistic dimensions, PhysioDrum was also perceived by several participants as having potential applications in stress relief, functioning as a relaxing and therapeutic environment. For instance, P07 remarked: "*I felt relaxed and experienced several moments of catharsis*"; while P12 stated: "*If you are upset about something, you can vent your anger there. It generates relaxation*". Similarly, P17 noted: "*I felt energized! I experienced catharsis. You can play with energy, and express your emotions*". Participant P00 added: "*The tool is, in addition to being educational, an option for those who want to relax in a creative way*". Moreover, P05, who self-reported a diagnosis of Attention Deficit Hyperactivity Disorder (ADHD), emphasized that the application facilitated sus-

tained concentration, as the virtual environment minimized external distractions such as movement or conversation. These accounts suggest that PhysioDrum may serve not only as an interactive musical platform but also as a medium for emotional regulation and focused engagement, thereby expanding its potential applications into therapeutic and well-being contexts.

The improvement suggestions provided by participants delineate clear directions for the future development of the PhysioDrum system. Requests such as enhanced customization of the drum kit, fine adjustment of instrument position, timbral variations, and differentiated haptic feedback per component (P07, P12, P13, P24) underscore a demand for greater musical expressivity and fidelity. Further recommendations include the integration of background tracks, multiple task difficulty levels, performance feedback mechanisms (e.g., hit and miss counters), and the inclusion of musical scores (P23, P24, P26). These proposals reinforce the platform's potential to consolidate itself as a hybrid environment for practice, learning, and entertainment. Additionally, the adoption of game-inspired metaphors and the structuring of progressive challenges emerged as salient strategies to sustain user engagement and foster the gradual development of musical skills. Such enhancements, if implemented, may significantly increase both the pedagogical effectiveness and the long-term appeal of the system.

The qualitative analysis indicates that PhysioDrum enjoys a high degree of acceptance among users with diverse profiles, being consistently perceived as an innovative, accessible, and immersive experience. Simultaneously, the data highlight the need for targeted technical and functional refinements, particularly concerning sensorimotor accuracy, ergonomic optimization, and the responsiveness of the haptic feedback system. The insights gathered indicate a positive outlook for the ongoing development of the platform, highlighting its potential for wider use in educational, therapeutic, and recreational settings.

7.2.7 Quantitative Analysis of Performance Accuracy

This section provides a quantitative analysis of the accuracy of PhysioDrum, with a focus on comparing the expected number of hits to the actual number of hits performed by participants during the tasks. As previously mentioned, each session comprised four tasks, leading to a total of 272 expected hits per session. The analysis was conducted in a general manner, without differentiation between specific drum elements, and aimed to evaluate participant accuracy, instances of oversampling, and cases of underperformance.

Overall, tasks 1 and 2 achieved the highest accuracy rates, with average accuracy remaining at or above 90%, aside from a few instances of underperformance. However, there was slight oversampling in more sensitive elements, such as the snare. This was caused partly due to its central position, leading to involuntary hits when moving the hands, particularly when returning to a rest position. Additionally, alternating between drumsticks resulted in subtle asymmetries; the dominant hand tended to register more collisions than the non-dominant hand, especially to trigger hi-hat. In this sense, there was a tendency to trigger the hi-hat using drumsticks instead of the pedal, which also served to hit this element.

In tasks 3 and 4, the results exhibited greater dispersion. Both involved the pedal used to trigger the bass drum, which was precisely the element where most difficulties emerged. In several cases, partial failures in activation registration were observed, suggesting that for users without prior experience, the lack of awareness regarding the bass drum's position prevented them from seeing the visual cue, leading to missed timing when striking that element. Simultaneously, the snare and cymbal displayed peaks of oversampling, possibly due to the high speed and constant alternation between the elements to be played, as this was the most challenging task. This finding reinforces that the more demanding activities, those requiring higher coordination among arms and legs, were indeed the ones with the poorest performance.

When considering the participants' profiles, those with musical experience (P01, P06, P07, P08, P12, P13, P18, P19, P22, and P24) exhibited more consistent performance. These participants made fewer under-execution errors and maintained better rhythmic control during the initial tasks, yet they showed a higher incidence of oversampling on the snare. This behavior was possibly caused by their tendency to reproduce the same mechanics used when playing a traditional acoustic drum kit. Since there were no physical objects to constrain their movements in this context, involuntary hits occurred on drum elements as they attempted to return their hands to the resting position.

Participants with prior virtual reality experience (P02, P13, P20, P22, P23, and P28) demonstrated faster spatial adaptation, showing lower variability between the two sessions. Their accuracy rates improved as early as in the initial tasks, particularly in scenarios involving multiple targets. However, difficulties with the pedal persisted, indicating that familiarity with immersive environments does not necessarily translate into improved performance in musical tasks.

Conversely, participants without prior experience in either music or virtual reality exhibited a higher number of errors during the first session, particularly in the tasks involving the pedal. As they continued to use the application and perform the tasks, the number of executed hits increased, sometimes even exceeding the expected count, indicating a rapid yet not fully controlled learning process, as well as a lack of motor refinement. Both outcomes, in the initial tasks and after adaptation to the environment, were consistent with what would be expected for this participant profile.

The order of tasks also influenced performance. Participants who began without feedback and subsequently performed the session with feedback (P00, P03, P06, P07, P09, P11, P13, P15, P16, P18, P21, P23, P25, and P27) showed a marked improvement in the second session, with a significant reduction in missed hits (under-hits). In contrast, those who started with feedback and later proceeded without it (P01, P02, P04, P05, P08, P10, P12, P14, P17, P19, P20, P22, P24, P26, P28, and P29) maintained good accuracy but exhibited greater variability once the haptic support was removed. This suggests that vibrotactile feedback facilitates initial synchronization; however, its absence removes a sensory channel that likely aided participants' performance and guided them during task execution.

Overall, the analysis indicates that haptic feedback contributes to reducing omission errors; however, successful task execution is more closely associated with individual experience. The average behavioral trend shows that participants learn and adapt quickly, stabilizing their performance by the second half

of the experiment. Consequently, the most notable difficulties are concentrated in tasks with higher BPM and greater motor coordination demands (tasks 3 and 4), as well as with the snare specifically, suggesting that the position of this element within the virtual environment requires optimization.

These findings reinforce the importance of considering both user profile and feedback exposure order when evaluating performance in immersive musical systems. The intersection of these variables reveals that musicians and users experienced in VR adapt more rapidly to the environment and maintain high performance even without continuous feedback, whereas inexperienced participants rely more heavily on the initial sensory support to achieve a stable rhythm. Table 15 summarizes this discussion, presenting each participant's individual behavior, which further supports the analysis developed throughout this section.

Participant	Hits Session 1	Hits Session 2	Session Order	Musical Exp.	VR Exp.	Observations
P00	231/272	264/272	without feedback; with feedback	No	No	Marked improvement after feedback
P01	286/272	259/272	with feedback; without feedback	Yes	No	Slight drop without feedback; good control overall
P02	250/272	267/272	with feedback; without feedback	No	Yes	Increasing regularity; occasional pedal misses
P03	204/272	261/272	without feedback; with feedback	No	No	Feedback improved timing and coordination
P04	278/272	256/272	with feedback; without feedback	No	No	Mild oversampling; some reliance on feedback
P05	340/272	299/272	with feedback; without feedback	No	No	Strong hits; persistent oversampling
P06	259/272	285/272	without feedback; with feedback	Yes	No	Solid rhythm and coordination
P07	263/272	294/272	without feedback; with feedback	Yes	No	Precision improves with feedback
P08	272/272	267/272	with feedback; without feedback	Yes	No	Refined coordination; stable tempo
P09	218/272	259/272	without feedback; with feedback	No	No	Pedal issues corrected after feedback
P10	299/272	262/272	with feedback; without feedback	No	No	High precision in the second session
P11	223/272	264/272	without feedback; with feedback	No	No	Rhythm stabilized in second session
P12	305/272	272/272	with feedback; without feedback	Yes	No	Improved in the second session
P13	281/272	277/272	without feedback; with feedback	Yes	Yes	Consistent across sessions
P14	272/272	250/272	with feedback; without feedback	No	No	Solid performance; small drop without feedback
P15	239/272	269/272	without feedback; with feedback	No	No	Gradual improvement; better control in session 2
P16	254/272	272/272	without feedback; with feedback	No	No	Consistent and stable execution
P17	320/272	300/272	with feedback; without feedback	No	No	Recurrent snare oversampling
P18	231/272	277/272	without feedback; with feedback	Yes	No	Clearer pedal use and improvement with feedback
P19	283/272	266/272	with feedback; without feedback	Yes	No	Predictable performance
P20	261/272	256/272	with feedback; without feedback	No	Yes	Residual pedal faults
P21	212/272	264/272	without feedback; with feedback	No	No	Feedback improved timing and precision
P22	285/272	281/272	with feedback; without feedback	Yes	Yes	Consistently strong performance
P23	258/272	275/272	without feedback; with feedback	No	Yes	Clear improvement with feedback
P24	272/272	264/272	with feedback; without feedback	Yes	No	Strong performance
P25	245/272	266/272	without feedback; with feedback	No	No	Notable improvement; mild snare oversampling
P26	272/272	278/272	with feedback; without feedback	No	No	High precision and consistency
P27	231/272	264/272	without feedback; with feedback	No	No	Strong improvement
P28	327/272	299/272	with feedback; without feedback	No	Yes	High energy; typical VR oversampling
P29	239/272	269/272	with feedback; without feedback	No	No	Better control in session 2

Table 15: Aggregated hits per session, execution order, and participant experience profiles.

7.3 Analysis of Desirable Characteristics for the Io3MT Environment

This section analyzes the categories of the Io3MT reference model, emphasizing their correspondence with the system's key characteristics and their contribution to ensuring its technical and artistic dimensions.

7.3.1 General Characteristics of the Environment

Based on the defining characteristics of environments shaped by the Io3MT reference model, PhysioDrum demonstrates a loosely coupled architecture in which the operational components — namely the pairs of drumsticks (RemixDrum) and the physical pedals — operate autonomously and independently from each other. Each element functions as an isolated functional module whose input and corresponding sensory outputs (auditory, visual, and haptic) are directly linked to the contextual actions occurring in the virtual environment, without requiring synchronization with other devices in the system.

The system exhibits a high degree of scalability at both the software and hardware levels. Logically, the Unity-based application supports the integration of additional three-dimensional graphical elements, virtual musical instruments, and interactive functionalities. On the hardware side, the modular structure of RemixDrum accommodates the incorporation of additional sensors (e.g. biosensors) and alternative or complementary actuators to those already implemented (e.g., vibration motors with varied patterns, LEDs, thermal systems). This modularity broadens the range of functional and sensory possibilities. Such scalability is enabled by the use of low-cost, open-source technologies with modular architectures, facilitating both reconfiguration and component reusability.

This architectural flexibility also translates into ease of development and continuous system evolution. The virtual application ensures compatibility with advanced capabilities by supporting the integration of diverse libraries and functionalities widely adopted in industry and academia. This approach enables not only updates to existing modules but also the replacement of outdated methods with more modern solutions. Similarly, the physical drumsticks and pedals can be substituted, reconfigured, or expanded without requiring a complete system overhaul.

Service orchestration is achieved by capturing the physical gestures performed with the drumsticks, which trigger events in the virtual drum set. This configuration produce real-time auditory and visual responses, alongside the activation of vibration actuators to deliver haptic feedback. The pedals function as triggers for specific sounds (e.g., bass drum, hi-hat) and also activate visual cues in the virtual environment, thereby reinforcing the sensorimotor coherence of the experience. This clearly illustrates how a sequence of physical actions is transformed into the final multisensory output perceived by the user.

All devices integrate with the application without requiring advanced configurations, following a plug-

and-play strategy. This ensures transparent integration and promotes accessibility, even for users with no prior experience with technological devices or musical performance. Moreover, the system has a lightweight implementation, requiring minimal computational resources for both sensory data capture and control signal transmission. Although the Unity application inherently demands greater computational and graphical resources, empirical evaluation revealed no crashes, perceptible delays, or performance degradation during testing, supporting its operational viability even in resource-constrained environments.

It is important to note that the present investigation did not aim to conduct a detailed analysis of network metrics. Consequently, no specific measurements were collected regarding latency, jitter, bandwidth consumption, or fault-tolerance mechanisms in network communication.

7.3.2 Functional Requirements

Among the functional requirements addressed by PhysioDrum, behavioral interoperability stands out, as evidenced by the system's capacity to produce coherent and contextually appropriate responses based on diverse input signals. This operational capability is also linked to other requirements, such as the detection and tracking of entities in the physical environment and the collaborative processing of information, wherein multiple data sources cooperate to generate integrated and contextually relevant outputs.

Regarding network management, PhysioDrum employs an architecture grounded in reliable communication protocols, utilizing UDP for the low-latency transmission of network data between the physical devices and the Unity application. In terms of operational robustness, the system demonstrates minimal maintenance requirements and high energy efficiency. The drumsticks and pedals are encased in protective housings, safeguarding the electronic circuits against mechanical damage and thereby reducing the frequency of maintenance interventions. Furthermore, the devices are powered by batteries with an approximate autonomy of eight hours, ensuring uninterrupted operation even during extended usage sessions.

7.3.3 Non-Functional Requirements

With regard to the non-functional requirements, one of the most prominent aspects is the system's heterogeneity, as it integrates distinct devices and platforms. The virtual application was developed in Unity and executed on the Meta Quest 3 headset, whereas the physical devices are based on embedded hardware, most notably the ESP32 microcontroller board, in conjunction with touch sensors, accelerometers, and vibration motors. This heterogeneity is further reflected in the variety of programming languages and development tools employed: C# is used to control the behavior of the virtual application, Python supports drumstick tracking via computer vision, Pure Data can be employed for audio processing, while Processing may be used for multimedia content display.

In terms of reliability, PhysioDrum demonstrated consistency in delivering sensory data and motor re-

sponses, as observed during empirical experiments. In cases of minor errors, such as delayed or missed collision detection, the system maintained continuous operation, reflecting partial resilience and functional recovery support. High availability is further ensured by its local network-based architecture, which operates independently of external servers or specific network systems, thereby reducing the likelihood of downtime caused by external factors. Finally, the platform exhibits notable technological adaptability. PhysioDrum has been designed to accommodate new functionalities, incremental software improvements, and adaptation to different hardware configurations, thereby supporting long-term scalability and maintainability.

7.3.4 Musical and Multimedia Protocols and Data Types

The musical/multimedia protocol employed in the PhysioDrum system was the OSC, selected for its widespread adoption in interactive musical contexts and its capability to provide lightweight, flexible, and low-latency transmission of real-time musical control events. When a physical pedal was pressed, OSC messages were transmitted to the virtual application, triggering the bass drum and hi-hat sounds. Conversely, the same protocol was used to communicate back to the physical system, activating the vibration motors embedded in the drumsticks whenever a haptic response was required, thereby completing the interactive feedback loop.

For the playback of pre-recorded sound samples, such as the timbres of the virtual drum kit elements, the WAV format was adopted. This format was chosen due to its compatibility with the Unity platform and its ability to preserve audio fidelity without perceptual compression.

It is important to note that, given the operational characteristics of PhysioDrum and the nature of the data transmitted, there was no need to employ a specialized application-layer protocol. The methods described in this section were deemed adequate to meet the interactive requirements of the proposed system.

7.3.5 Artistic Requirements

Regarding the artistic requirements, the qualitative interviews showed a high degree of user satisfaction, with the experience frequently described as fun, intuitive, and enjoyable. The combination of visual, auditory, and tactile stimuli, enhanced by the ergonomic design of both drumsticks and pedals, resulted in a sensorially rich interface, promoting comfort during interaction and fostering spontaneous engagement with the musical environment.

Immersion and engagement form a central pillar of the system, achieved through the use of first-person virtual reality, precise multisensory integration, and accurate sensorimotor coupling. Data obtained via the Presence Questionnaire (PQ) confirmed high immersion levels, further reinforced by the perceived realism of haptic feedback and the quality of the visual interface.

The system also incorporates features that support users without prior artistic or technical background,

including guided tasks with simple rhythmic patterns, an intuitive interface based on natural gestures, and immediate feedback to reinforce trial-and-error learning. Several participants without previous VR or musical experience reported being able to interact autonomously and satisfactorily with PhysioDrum, thereby contributing to the accessibility of musical performance and enabling unrestricted participation.

In terms of creative stimulation, the system provides a free exploration and musical improvisation phase at the beginning of the test sessions. This unstructured stage encourages users to experiment with various sound, dynamic, and gestural combinations, fostering the emergence of individual expressive strategies. Qualitative accounts indicated high engagement levels and a desire to continue exploring beyond the proposed tasks.

The experience structure was designed to present information in a coherent flow, marked by a progressive increase in task complexity. This design fosters artistic confidence, namely the user's trust that their gestures will produce predictable and consistent aesthetic outcomes, encouraging continued performance. These characteristics contribute to the system's captivating nature, directly linked to the balance between challenge and reward.

7.3.6 Device Requirements

From the perspective of physical infrastructure, both the drumsticks and pedals integrate low-power embedded electronics. This structure enabled the development of autonomous, portable, and intelligent devices capable of continuous operation throughout the experimental sessions.

Communication between the physical devices and the virtual application was established over a wireless network, using the previously described UDP/IP protocols. This choice also allowed for the assignment of unique addresses to each module, ensuring unambiguous identification within the local network. Moreover, the efficient transmission of sensory data and control commands ensured a lightweight, responsive, and reliable communication structure.

Regarding human–computer interaction, the devices were positively evaluated in terms of usability, demonstrating a low learning curve, high flexibility, and efficiency in performing musical tasks. The design considered accessibility principles, enabling use by participants with varying degrees of technological familiarity, motor skills, and educational or social backgrounds. Communicability was also addressed, as the design logic was perceived as clear, and the objectives of each action were readily understood. This comprehension was reinforced by the consistent mapping between physical gestures and the multimodal effects produced, thus promoting predictability and expressive control. Also, PhysioDrum was purposefully designed to control a single performative process, focusing exclusively on the execution of virtual drumming.

7.3.7 Architecture Analysis

The PhysioDrum architecture demonstrates strong alignment with the Io3MT reference model, presented in 4. At the device layer, it employs a diverse set of physical components with multisensory sensing and actuation capabilities. The drumsticks capture data via touch sensors and accelerometers, while also delivering haptic output through vibration motors. The physical pedals operate solely as sensing units, serving as triggers for visual and auditory effects in the virtual drum kit. Collectively, these components support one of the core principles of Io3MT systems: the virtualization of physical actions, enabling the abstraction of physical data sources and their contextualized use in immersive environments.

Within the network layer, PhysioDrum leverages a local Wi-Fi–based network architecture. This design ensures logical addressing of devices, efficient packet routing, and traffic control among all communication nodes. Although it does not implement a formal QoS scheme, the network structure sustains interactive operations without degrading responsiveness, even under prolonged usage. Data routing is hard-coded, with specific IP addresses and ports assigned to each component, preventing route collisions and congestion. Connection stability is further reinforced by the devices' ability to attempt automatic reconnection when disconnected, eliminating the need for additional network-level resilience or fault-tolerance services.

At the application layer, the virtual environment serves as the central processing hub, orchestrating musical actions and sensory data. It coordinates the triggering of auditory, haptic, and visual responses, monitors device status, manages scene logic, and synchronously activates multiple output modules. This fulfills the core Io3MT application-layer services, including real-time management of physical objects, concurrency control, and the translation of physical actions into multimodal events.

In this way, PhysioDrum exhibits consistent adherence to the fundamental pillars of the Io3MT architecture, integrating sensory data acquisition and transformation, efficient and reliable information transmission, and real-time multimodal orchestration.

7.4 Comparative Analysis with Related Work

To enhance the applicability of this study for researchers, designers, artists, and composers, a comparative analysis was also conducted between the PhysioDrum and related air-drumming applications, presented in Section 3.6.

Regarding the input devices employed in these systems, there is a clear preference for computer vision techniques that track physical objects and, based on their spatial position, trigger corresponding audio tracks. Besides that, some applications, such as Paradiddle[5] and DigiDrum (WILLEMSEN; HORVATH; NASCIMBEN, 2020), leverage the VR headset's native controllers (joysticks) to operate the virtual drum kit. In industry-developed solutions, more individualistic methods are adopted. For

[5] https://paradiddleapp.com/

instance, Aerodrums[6] combines retroreflective markers with a high-speed camera to track drumstick motion, while Aeroband Pocket Drum 2[7] employs spatial mapping of drumstick position. PhysioDrum distinguishes itself from these models by applying computer vision techniques to detect the position of a SMI in the physical world and using its movements to trigger the virtual drum kit. This approach directly incorporates the phygital concept, wherein tangible elements from the physical environment have a direct and meaningful impact on the virtual application.

In both academic and commercial systems, pedal-based interactions have limited adoption. Among the examined examples, only the Aeroband Pocket Drum 2 and PhysioDrum implement such functionality. In the Aeroband application, the integrated pedals are used to trigger the bass drum and hi-hat. Similarly, PhysioDrum also employs pedals to control these same drum elements, but distinguishes itself by activating them through the OSC protocol, with messages exchanged over the network. The Paradiddle application supports the addition of external electronic pedals to control specific elements via MIDI protocol; however, this option is offered as an extension, and its standard operation does not rely on such devices.

Similar to pedal-based interaction, network integration is generally given secondary importance in air drumming applications. Only the Aeroband Pocket Drum 2, the system proposed by (YASEEN; CHAKRABORTY; TIMONEY, 2022), and PhysioDrum incorporate such functionality. In the first case, networking is used to transmit the drumstick's spatial position and trigger the corresponding drum element based on that position. In the second example, network communication enables multiple users to connect and interact within the same musical environment. Finally, PhysioDrum employs networking both to transmit input data from the physical devices to the VR application and to receive feedback from the virtual environment, thus establishing a bidirectional interactive loop.

In terms of audio synthesis, industrial applications primarily rely on proprietary software solutions, whereas academic systems tend to focus on triggering pre-recorded drum samples. VR-based systems leverage the built-in audio integration features of their respective platforms, enabling seamless synchronization of sounds with 3D drum kit models. A noteworthy example is the Aeroband Pocket Drum 2, which supports integration with commercial DAWs. PhysioDrum adopts a hybrid approach, combining the native audio synthesis capabilities of its VR environment with Pure Data for real-time sound processing.

Haptic feedback was implemented in three of the examined systems. The Aeroband Pocket Drum 2 delivers a uniform tactile response for all struck elements, serving as an auxiliary sensory channel to confirm the occurrence of a given action. The DigiDrum investigates how tactile feedback, provided through membranes of varying characteristics, influences musical expressivity. In contrast, PhysioDrum employs two distinct types of haptic feedback: one constant across all interactions, and another that varies according to the specific drum element triggered, thereby enhancing the multimodal differentiation of the performance.

[6] https://aerodrums.com/
[7] https://www.aeroband.net/products/pocketdrum2-plus

Among the tools that incorporate XR integration, Paradiddle stands out as it was specifically designed for such immersive environments. The most recent version of Aerodrums also offers a graphical VR representation of the drum kit, a feature likewise present in DigiDrum and PhysioDrum.

In terms of modularity and customizability, only a few systems meet these criteria: the Aeroband Pocket Drum 2, which allows customization of the drum kit; Paradiddle, which enables users to upload music score and adapt the kit to their preferences; and PhysioDrum, which supports both hardware and software modifications. These capabilities enhance adaptability and extend the potential use cases of the systems beyond their default configurations.

In respect of application purpose, industry-developed solutions primarily emphasize entertainment, with occasional extensions into live performance contexts. Academic initiatives, while also addressing entertainment, are predominantly oriented toward research and experimental studies. PhysioDrum distinguishes itself from these approaches by encompassing a broader spectrum of objectives. In addition to supporting performance and entertainment, it integrates elements specifically designed for academic research and scholarly inquiry. Moreover, its adaptability enables deployment in music education and therapeutic contexts, thereby extending its applicability beyond the typical scope of air drumming systems.

From the analysis of the most recurrent attributes in air drumming applications, it becomes evident that PhysioDrum differentiates from existing models through its unique input modality, its approach to audio synthesis, and the breadth of its intended applications. Furthermore, it is among the few systems to incorporate network integration while also supporting pedal-based control and offering both modularity and customization capabilities.

When comparisons are drawn with academic exemplars, the distinctions become even more pronounced. PhysioDrum stands out as the only system to incorporate the phygital concept and to integrate a SMI into its interactive environment. Furthermore, it uniquely combines haptic feedback capabilities with XR integration, two features whose joint implementation has not been addressed in previous projects.

The operational model of PhysioDrum, which integrates multiple functionalities and resources, endows the system with a set of advantages that position it as a highly promising platform for scientific investigations in the fields of Io3MT, multimodal musical interaction, and immersive environments. Its phygital-based approach enables the exploration of phenomena related to embodiment, presence, and musical expressivity in a more realistic and tangible manner, thereby broadening the scope of research on sensory–motor interaction in digital contexts.

The combination of differentiated haptic, visual, and auditory feedback for each drum kit element offers fertile ground for studies on multisensory perception and musical cognition. Furthermore, interoperability with protocols such as OSC, coupled with a modular architecture, makes PhysioDrum readily adaptable to a variety of experimental designs, facilitating replication, extension, and deployment across diverse contexts. Consequently, the platform is configured not merely as an interactive

artifact, but as a versatile research infrastructure for knowledge production in domains such as computer music, virtual reality, cognitive science, and the design of SMIs. The synthesis of this discussion is presented in Table 16.

Attribute	**Aerodrums**[8]	**Aeroband Pocket Drum 2**[9]	**Paradiddle**[10]	**Anytime Drumming (A2D)** (YADID et al., 2023)	**Augmented Virtual Drums** (ZAVERI et al., 2022)	**Air Drums** (TOLENTINO; UY; NAVAL, 2019)	**DigiDrum** (WILLEMSEN; HORVATH; NASCIMBEN, 2020)	**IoMusT Model** (YASEEN; CHAKRABORTY; TIMONEY, 2022)	**PhysioDrum**
Input Type	Reflective markers + camera capture	Spatial orientation	VR headset controllers	Computer vision	Computer vision	Computer vision	VR headset controllers	Computer vision	Phygital + Smart Musical Instrument + Computer vision
Pedal Control	No	Yes	Yes	No	No	No	No	No	Yes
Audio Synthesis	Proprietary software	Proprietary software; DAW integration	VR application	Drum samples	Drum samples	Drum samples	VR application	Drum samples	VR application + Pure Data
Haptic Feedback	No	Yes	No	No	No	No	Yes	No	Yes
Network Integration	No	Yes	No	No	No	No	No	Yes	Yes
XR Integration	Yes	No	Yes	No	No	No	Yes	No	Yes
Modularity and Customization	No	Yes	Yes	No	No	No	No	No	Yes
Main Purpose	Entertainment; performance	Entertainment; performance	Entertainment; performance	Academic studies; entertainment	Academic studies; entertainment	Academic studies; entertainment	Academic studies; entertainment	Academic studies; entertainment	Academic studies; entertainment; performance; music education; therapy
Initiative	Industry	Industry	Industry	Academia	Academia	Academia	Academia	Academia	Academia

Table 16: Comparative analysis of air drumming systems.

7.5 Final Remarks on PhysioDrum

This chapter introduced an immersive musical environment designed according to the principles of the Io3MT. To guide its development, a focus group was conducted with four experts in virtual reality, all of whom also had extensive experience in music practice. Their insights helped establish guidelines for creating such systems. As a proof of concept, PhysioDrum was developed, which integrates physical elements, multisensory stimuli, network communication, and a virtual drum set.

To examine the effects of haptic feedback on presence, immersion, usability, and cognitive load, a mixed-methods approach was employed, incorporating both quantitative and qualitative analyses. Thirty participants were randomly divided into two experimental groups: one group received uniform tactile feedback, while the other received differentiated haptic information that corresponded to distinct drum elements. The results showed that uniform feedback improved interaction naturalness, engagement, and perceived control over the system. In contrast, differentiated feedback resulted in superior outcomes in terms of experiential realism, significantly enhancing the perceptual quality of the virtual environment.

Qualitative interviews revealed that PhysioDrum was perceived as accessible, engaging, and promising in terms of fostering musical creativity. Participants highlighted its intuitive interface, playful nature, and applicability for both beginners and experienced musicians. Nonetheless, technical limitations were

[8] https://aerodrums.com/
[9] https://www.aeroband.net/products/pocketdrum2-plus
[10] https://paradiddleapp.com/

noted, including the precision of collision detection, minor ergonomic adjustments for pedal operation, and refinements to sound fidelity, issues that represent concrete opportunities for improvement.

The main contributions of this chapter are threefold: i) the formulation of a comprehensive and structured set of design guidelines for immersive Io3MT systems, with applicability extendable to XR-based musical applications; ii) the development of a testing protocol that integrates both objective and subjective assessment methods for the evaluation of immersive musical experiences; and iii) the development of the PhysioDrum system, which exhibited high technical performance, distinct operational and aesthetic attributes, and yielded robust empirical evidence on the role of tactile feedback in shaping perceptual and experiential dimensions of immersive musical interaction. Collectively, these contributions provide an empirical and methodological foundation to inform and guide future research in the domains of Io3MT and the Musical Metaverse.

Future developments will focus on expanding environment customization, enabling users to configure their own drum kits, adjust timbres, modify aesthetic configurations (both in environment and drum set), and reposition instruments according to personal preference, including ergonomic support for left-handed users. From a functional perspective, enhancements include the capability to record performance sessions, facilitate content sharing, incorporate pre-existing musical tracks and scores, and deploy novel interactive tutorial modules. Analytically, further testing with a larger cohort of music and VR specialists is envisaged, alongside in-the-wild experiments to validate operational aspects of PhysioDrum in real-world settings. These efforts aim to consolidate PhysioDrum as a versatile platform for educational, performative, and therapeutic purposes.

8 Conclusion

This thesis constitutes the first systematic exploration of the use of multisensory, multimedia, and musical information in artistic applications, within what has come to be termed Internet of Multisensory, Multimedia and Musical Things (Io3MT). The adoption of this nomenclature is intended to highlight the specific ways in which these elements can enrich artistic practices and related fields within IoS, with which it shares similar objectives.

This approach fosters interoperability among devices and enables continuous expansion through the addition of new hardware and software components. It also seeks to ensure the technical reproducibility of the proposed concepts, which can be interpreted in light of Walter Benjamin's reflections (BENJAMIN, 2008). Just as photography and cinema democratized access to aesthetic experience by breaking with the exclusive aura of the unique artwork, Io3MT creates possibilities for multiplying, distributing, and recombining musical and sensory experiences over networks, thereby reshaping both the modes of artistic production and reception.

As a result, general guidelines and a set of open tools were proposed to support the design and implementation of applications within this ecosystem. By establishing this reference domain, the proposal advances a conceptualization capable of describing the interrelationships among entities, providing a shared vocabulary for researchers and practitioners, accommodating multiple usage perspectives, and consolidating this emerging field of research. The proposed guidelines offer structured support for requirements elicitation, technology acquisition, architectural modeling, implementation, evaluation, maintenance, and continuous system evolution. Moreover, by explicitly specifying the relationships among components, these guiding principles enable more accurate adaptation decisions, thus optimizing both creative and technical processes.

To serve as proof of concept, the proposed guidelines were instantiated in two real-world use scenarios: a multimedia remixing system, and a virtual drum kit enhanced with haptic feedback. These experiments demonstrated the flexibility of the proposal in addressing diverse network capacities, heterogeneous computational demands, and distinct functionalities, thereby consolidating its suitability for multiple artistic and technical contexts.

The application of Io3MT principles indicates a departure from the traditional linear model of Western musical communication toward an interactive network in which participation and exchange occur through multiple pathways. This transformation facilitates the transposition of creative activities across domains, enabling interactive multimedia performances that are both responsive and non-

intrusive, with artists and audiences occupying active roles. Human perception of sensory stimuli varies in form and intensity, and the integration of technological elements provides mechanisms to accommodate this variability. Within this domain, technology should not be conceived merely as an infrastructure supporting the artwork but rather as an element integrated into its semiotic structure, functioning as a constitutive component of the work itself. Accordingly, Io3MT is not limited to the role of an enabling infrastructure; it constitutes a language embedded in the artwork, shaping it at its core. This reconfiguration aligns with *avant-garde* practices oriented toward the dissolution of boundaries and is situated within a distributed, interactive, and multisensory ecosystem.

8.1 Revisiting the Research Questions

The research questions that emerge from the gaps identified in the field of IoS, as outlined in Section 1.3, can be classified into three distinct dimensions: technological, evaluative, and artistic expression. Each of these dimensions is examined in depth below, highlighting the ways in which they have been addressed through the efforts undertaken in this thesis.

- **RQ1: What are the functional requirements for environments that enable the integration of multisensory and multimedia data within IoS-based systems?**

Existing literature on the IoS suggests that different input modalities have traditionally been treated in isolation, thereby exposing a major limitation of current systems: the absence of comprehensive integration across the heterogeneous modalities that constitute a musical environment. Within this context, the Io3MT is introduced as a domain designed to systematize such integration, consolidating musical, multimedia, and multisensory information into a unified environment. The underlying premise is that these three categories of data should not be conceptualized merely as secondary byproducts of user actions, but rather as mutually interdependent dimensions that exert reciprocal influence. This perspective marks a departure from conventional paradigms in which, for instance, haptic feedback is regarded solely as a derivative outcome of musical performance.

However, constructing an Io3MT environment involves far more than the mere selection of sophisticated devices and their subsequent interconnection. It requires addressing additional layers of complexity that arise from its interdisciplinary nature. This includes the absence of native support for multisensory devices and, even when standardized communication protocols are available, the challenges associated with their effective implementation continue to persist. Furthermore, enduring issues from the multimedia domain remain relevant, such as ensuring precise synchronization between content and sensory effects, mitigating masking issues, overcoming processing constraints, and managing network-related challenges, particularly those concerning latency and jitter.

This context indicates the necessity of addressing the core technical challenges of Io3MT and of formulating systematic guidelines for handling hardware and software heterogeneity in the integration of multimedia, multisensory, and musical elements. In response, this thesis introduces a reference model

and two prototype implementations grounded in Io3MT principles. The reference model specifies functional and non-functional requirements, data formats, interfaces, and functional modules essential for assembling such environments. Rather than being a closed or prescriptive solution, it is conceived as a flexible structure intended to guide researchers and developers in adapting its specifications to varied scenarios and application domains.

The services defined by Io3MT also affect the physical structure, general characteristics, and operational dynamics of the artifacts involved in such applications. Consequently, these devices must incorporate embedded electronics, support communication with networks and other devices, be uniquely identifiable and addressable, ensure hardware scalability, provide energy and operational reliability, and maintain loose coupling in order to facilitate integration. In addition, aesthetic and ergonomic aspects are essential to ensure usability during extended periods of performance and rehearsal. Functionalities such as packet routing, context awareness, and fault tolerance are equally relevant. These devices may be physical or digital, dedicated to the control or synthesis of specific processes, or designed to enable the simultaneous manipulation of multiple services.

- **RQ2: Which evaluation methodologies are most suitable for assessing Quality of Experience (QoE) in environments that combine auditory, multimedia, and multisensory stimuli?**

Although this work primarily addresses the technical aspects of Io3MT, it also acknowledges the importance of QoE assessment. Nevertheless, this dimension posed an additional challenge, as such evaluations tend to be diffuse and less formally structured. This is partly due to the specificities of each environment and its functionalities, which require different assessment methods, and partly because QoE has historically played a secondary role in the field of Computer Music. Furthermore, assessing QoE involves examining how users perceive multimedia, multisensory, and musical information, including their levels of engagement, satisfaction, or even boredom during system interaction. This process is inherently complex, as it encompasses not only technical factors (ergonomics of devices, content formats, network conditions) but also psychosocial dimensions (application context, users expectations, and emotional state).

Accordingly, two different evaluation approaches were adopted, each aligned with the characteristics of the proposed use cases. In the first scenario, assessment was conducted with a single expert through a semi-structured interview. Although this method has limitations in terms of generalizability, it is valuable for identifying critical issues and validating solutions directly with specialized user. The second case involved a mixed-method evaluation with 30 participants, combining quantitative and qualitative measures. This approach not only broadened the applicability of the findings but also enabled a detailed analysis of social, academic, and artistic profiles, reinforced by statistical triangulation that enhanced the validity and reliability of the results. Moreover, this study contributed to the establishment of constructs and analytical methods for immersive artistic applications, with potential applicability to broader activities within the field of Computer Music.

These approaches reflect the diversity of audiences and objectives, encompassing systems designed for experts, the general public, and individual artistic needs. In this sense, this research not only validates different evaluative proposals but also contributes to the development of a methodological foundation for future Io3MT applications.

- **RQ3: In what ways can the integration of sound, multimedia content, and multisensory feedback enhance artistic expressiveness, aesthetic experience, and immersion within IoS-based systems?**

Io3MT introduces new possibilities for composition and performance by integrating sound, multimedia, and multisensory information as interdependent dimensions of artistic creation. This convergence enables the development of novel narrative strategies, compositional paradigms, and feedback mechanisms that extend beyond practices focused exclusively on sound.

One of the central contributions lies in the expansion of artistic expressivity. The coexistence of auditory, visual, and tactile modalities fosters a richer semiotic interplay, wherein stimuli can reinforce, complement, or contrast with each other. This interweaving enhances the communicative capacity of musical gestures, enabling the transmission of subtle nuances and the creation of more impactful aesthetic experiences. In this respect, Io3MT does not merely incorporate additional modalities but redefines the grammar of artistic expression by situating it within an intermodal and interactive context.

Within the scope of aesthetic experience, immersion is no longer restricted to listening but becomes a holistic perceptual encounter in which participants can feel, see, and interact with the artwork. Virtual and augmented reality environments, supported by real-time synchronization across sensory modalities, intensify the sense of presence and embodiment. Furthermore, practices such as multimedia remix and virtual–digital integration expand the modes of authorship, allowing audiovisual and haptic materials to be dynamically recombined in response to performers and audiences actions.

These transformations give rise to a new generation of multimodal instruments and installations capable of expanding performative practice into an enriched sensory event. Such systems preserve the expressivity of traditional musical performance while extending it, providing creators with an expanded creative repertoire and offering audiences participatory and affective experiences.

8.2 Scientific Contributions

The contributions of this thesis can be organized into five main dimensions: theoretical advancements, the reference model, design principles for immersive environments, empirical findings, and technological innovations. Regarding the first topic, this work introduces a novel domain referred to as Io3MT, which extends the capabilities of the IoS by enabling multisensory, multimedia, and musical interactions within networked environments. This approach supports real-time interactive experiences

capable of integrating multiple modalities of information in such a way that they coexist and exert direct influence upon one another. Furthermore, the technologies proposed in this study provide insights that can be leveraged to enhance the design of applications across various domains, including XR, Musical Metaverse, IoMusT, among others.

In order to consolidate this emerging domain, a reference model was devised, representing the second central outcome of this thesis. Its formulation draws upon adjacent fields of Io3MT, as well as technical standards dedicated to the harmonization of services in heterogeneous ecosystems. The proposed model delineates services, network infrastructures, and device operations, while simultaneously functioning as a roadmap for the systematic development of applications within the Io3MT scope. Its design is informed by network performance and QoE metrics, thereby advancing the pursuit of more consistent, interoperable, and scalable systems. Furthermore, it demonstrates how aesthetic and multisensory dimensions may be formally embedded into artistic and interactive practices, offering methodological and technical support for designers, researchers, and practitioners engaged in the creation of complex, immersive, and sensorially rich ecologies.

In general, the proposed structure is not intended to be hermetic or definitive; rather, it seeks to organize the principles and technologies that underlie Io3MT. It is also aligned with contemporary trends in interactive media, particularly embodied interaction, multisensory design, and hybrid physical–digital spaces, holding the potential to shape the design of next-generation media experiences that are more immersive and inclusive of diverse sensory modalities.

The expansion of Io3MT concepts into immersive environments required the formulation of specific design principles for the development of such applications, constituting the third dimension of this research's contributions. Although conceived within the scope of Io3MT, these principles are also highly relevant to the broader XR field, particularly with regard to multisensory integration. Their pertinence to the establishment of the Musical Metaverse is equally noteworthy, as they provide a systematic method of formalization and guidance for the creation of immersive musical environments. In addition, this work introduces a set of constructs to be considered in such contexts, along with methodologies to assess QoE.

As a proof of concept for the proposed reference model, two distinct use-case scenarios were developed, each with specific artistic and technological objectives. The resulting empirical evidence made it possible to identify which requirements were fully implemented, the difficulties encountered during the process, and the challenges that remain unresolved. Moreover, the findings provide a foundational benchmark for subsequent investigations in the Io3MT domain, thereby delineating the fourth dimension of contributions.

The fifth and final dimension is defined by the technological advances resulting from the development of open toolkits, which encompass a range of hardware and software solutions designed to support Io3MT applications. The first prototype was an SMI named RemixDrum, distinguished from previous models by its capacity to remix multimedia content. Subsequently, the device was integrated into an immersive environment, enhanced with vibration motors to deliver haptic feedback to users. This

environment further introduced a series of innovations, including the use of pedals for the control of percussive elements, the implementation of networked communication in an air-drumming application, the first description and analysis of a haptic feedback system applied to immersive musical scenarios, as well as the exploration of the phygital concept.

Although broadly applicable and intentionally general in scope, the Io3MT reference model developed in this thesis is primarily oriented toward artistic performance, particularly the communicative, expressive, and embodied dimensions of multisensory, multimedia, and musical interaction. Rather than addressing audience reception directly, the model concentrates on organizing how artistic actions — gestures, sonic events, sensory mappings, and multimodal expressions — are coordinated and rendered within phygital environments. Its focus, therefore, lies in defining how performers engage with heterogeneous devices, modalities, and services, rather than prescribing how these experiences should be perceived or interpreted by spectators.

Even so, the conceptual and architectural foundations of Io3MT naturally create opportunities for extending multisensory and multimodal interactions to audience-oriented scenarios. For instance, the reference model can enable immersive multisensory spectatorship, allowing local or remote audiences to receive visual, auditory, and tactile cues that convey the expressive nuances of a performer's actions. Additionally, Io3MT environments may support co-present or distributed participatory experiences, empowering spectators to influence aspects of the artistic presentation through gestural input, mobile interfaces, or collective behavior.

The reference model also accommodates adaptive and personalized sensory reception, making it possible for audiences to experience artistic content through configurable multimodal profiles aligned with individual preferences, accessibility needs, or perceptual sensitivities. Finally, Io3MT opens pathways toward socially shared artistic ecologies, in which spectators no longer remain passive observers but instead become part of a dynamic, multisensory environment shaped by presence, embodiment, and shared agency.

8.3 Artistic Contributions

This thesis introduces a novel concept with the potential to significantly advance the state-of-the-art in IoT, paving the way for new applications across the domains of entertainment, healthcare, and education. Rather than conceiving these fields as linear connections between point A and point B, merely as transmission channels, the proposed approach shifts the perspective toward a more complex and relational understanding. Such environments are envisioned as multisensory and interconnected ecosystems, comparable to dynamic landscapes, sensitive atmospheres, or multidimensional fields, in which multiple modalities coexist and interact in real time, generating reciprocal effects and continuous transformations.

In the arts, Io3MT introduces new modes of aesthetic practice, breaking with established conventions and aligning itself with *avant-garde* approaches. The technical reproducibility (BENJAMIN, 2008)

of this domain extends beyond the mere replication of devices and services, fostering instead collaborative production, improvisation, and nonlinearity, thereby reconfiguring the relationships among artists, audiences, and technology. It constitutes a space in which constant ruptures challenge initial expectations, creating opportunities for novel forms of interaction, feedback, and artistic communication.

Followinf this idea, technological art generates engagement not as mere distraction but as a means of mobilizing sensibilities against oppressive structures, thereby democratizing modes of creation and circulation. Io3MT points to a model in which art ceases to be the privilege of a few and becomes a collective, interactive, and critical space. This convergence of codes and cultures, previously treated as disparate, may constitute a prelude to new forms of expression, abstraction, imagination, and artistic interaction, configuring what may be described as the space of all possible art.

The poetic dimension of this practice also resonates with Susan Sontag's reflections, according to which art should mobilize the senses and provoke the body, rather than being reduced to strictly intellectual interpretations (SONTAG, 2001). By integrating sounds, images, textures, and scents, Io3MT expands the aesthetic experience into a multisensory domain, eliciting adaptive responses and positioning participants as co-authors of the creative process. In this sense, the value of the artwork lies not solely in the object itself, but in the relational event established among subjects, devices, and environments.

This stance further resonates with Nicolas Bourriaud's concept of relational aesthetics, which conceives contemporary art as a practice oriented toward the creation of social interstices, where meaning emerges from the interactions between individuals and contexts (BOURRIAUD; SCHNEIDER; HERMAN, 2002). In this regard, Io3MT can be understood as a relational platform that organizes aesthetic encounters mediated by technology, enabling artists and participants to co-create complex sensory ecologies that remain open to improvisation.

Finally, the proposal also engages with Vilém Flusser's perspective on the nature of apparatuses. For the author, technical devices should not be regarded as neutral instruments but rather as programmable "black boxes", whose creative potential depends on the human capacity to exploit them against their original function in the search for new meanings (FLUSSER, V., 1985; FLUSSER, Vilém, 2012). Io3MT situates itself precisely within this space, as it transforms communication devices, sensors, and actuators into aesthetic agents, reprogramming their functions to compose musical, visual, and sensory experiences that subvert utilitarian use and open new expressive possibilities.

In this way, Io3MT shifts technological art from a paradigm centered on closed, rigid, and unidirectional works to a processual, open, and emergent model, in which aesthetics, technique, and interaction are deeply intertwined. The artistic experience thus becomes a networked event, continuously in transformation, whose legitimacy derives from the multiplicity of voices, gestures, and stimuli that converge within it. This approach establishes a creative ecology in which multiple agents — humans, devices, and media — participate interdependently, giving rise to hybrid, polyphonic and sensory environments that are potentially unlimited in their expressive capacity.

8.4 Limitations

Although the results achieved are consistent in terms of network performance and QoE, the proof of concept implementations were conducted in controlled environments, which naturally impose limitations on the generalizability of the findings. More specifically, Scenarios 1 and 3 were implemented on a small scale, involving a limited number of devices and participants. This characteristic constrains the extrapolation of results to broader contexts and underscores the need for further investigations in large-scale settings, where aspects such as scalability, latency in congested networks, synchronization across multiple devices, and interoperability within heterogeneous ecosystems can be more comprehensively examined.

Additionally, certain technical aspects were not addressed in the implementations. Among these, the lack of investigation into the influence of the spatial positioning of actuators relative to musicians and participants is particularly noteworthy. This limitation constrains a more precise analysis of the experience in scenarios involving movement within the environment, as sensory perception is modulated by distance, direction, and volumetric dispersion of stimuli. In the specific case of olfactory effects, the spatial propagation of scents was not examined. As a consequence, their irregular distribution in the environment may significantly affect how intensity and uniformity of the effect are perceived.

The effects of prolonged exposure on the perception of sensory elements were likewise not addressed. In the case of olfaction, for example, odors may persist in the olfactory epithelium for longer than anticipated, thereby altering perception and compromising the temporal precision required for performance. In the domain of thermosensory perception, the application of a cold stimulus may modulate subsequent sensitivity to heat, giving rise to unintended cross-modal interferences. Furthermore, the response time of actuators constitutes a critical factor for ensuring the quality of the experience. Any delays in the delivery or perception of stimuli disrupt synchronization with auditory and visual content, ultimately undermining the coherence and integrity of the multimodal experience.

Another limitation concerns the second scenario, in which variations in vibration intensity were not assessed, restricting the analysis to a binary dimension of stimulus presence or absence. Similarly, the potential impacts of prolonged exposure to this stimulus were not taken into consideration.

Furthermore, the PhysioDrum use case did not include an assessment of network performance, thereby precluding any analysis of whether participants' positive or negative impressions of the experiment were influenced by the network's operational conditions.

In both cases, communication relied exclusively on local networks. Consequently, typical characteristics of Internet-based connectivity—such as variable routing paths, fluctuating bandwidth availability, congestion episodes, and cross-domain latency—were not examined. Additional experiments are therefore required to determine how these factors influence the behavior and overall performance of Io3MT environments.

At the conceptual level, this work did not address ethical, legal, and security issues associated with

Io3MT. The discussion on the use of copyright-protected content within these environments remains unresolved, encompassing music, images, and other forms of media. Likewise, the implications of integrating artificial intelligence into such ecosystems were not examined in detail, particularly with regard to authorship, content curation, and potential algorithmic biases. The risks related to the misuse of artistic material, amplified by its circulation in distributed and networked environments, were also not subjected to systematic investigation. Cybersecurity and data privacy concerns similarly persist as open challenges.

These limitations, although they do not diminish the relevance of the results presented, underscore the need for further research developments. Future investigations should not only broaden the technical dimension through experiments conducted in scalable contexts, but also incorporate a critical reflection on the ethical, social, and legal implications of Io3MT within the contemporary ecosystem of media, art, and technology.

8.5 Open Challenges

Despite the advances achieved in consolidating Io3MT, several challenges still need to be addressed to enable its large-scale adoption. These challenges encompass issues of technological infrastructure, security, economy, society, and environment, reflecting the complex and transversal nature of this ecosystem.

In the technological domain, one of the main challenges lies in achieving high-quality audio capture and preserving real-time interactions in networked environments. Synchronizing audio and multimedia streams, often produced by devices that do not share a common clock, requires continuous resynchronization processes. In addition, devices must be capable of simultaneously processing music, multimedia, and multisensory information, which demands high performance in terms of memory, energy, and bandwidth. Issues of miniaturization, ergonomics, and energy autonomy also remain critical, as devices must be discreet and comfortable without compromising efficiency or accessibility.

In the area of privacy and security, the pervasive nature of Io3MT and its reliance on sensitive data require the development of protocols more robust than those currently employed in IoT. Security gaps remain in areas such as data collection and anonymization, network object identification, software vulnerabilities, and threats posed by malware targeting mobile and wearable devices. The architectural heterogeneity and computational constraints of these devices increase the complexity of implementing effective solutions.

Economic challenges are also identified. The music and cultural industries have historically adapted to technological change, from the phonograph to digital streaming, and Io3MT may extend this trajectory by introducing new mechanisms for creative control and monetization for artists and producers. Uncertainty remains regarding its impact on the production chain, including whether it will reduce costs and broaden access or, alternatively, constrain creative opportunities by replacing human functions with automated solutions, which could compromise artistic authenticity.

From a social perspective, Io3MT raises issues related to unequal access to technology, the intensification of disparities between urban and peripheral or rural regions, and the concentration of resources and power in specific economic actors. Risks include the potential deepening of social and cultural inequalities, leading to forms of digital exclusion. Developing alternatives such as solidarity-based economies, collaborative practices, and inclusive cultural projects is necessary to mitigate these impacts and to ensure a more equitable distribution of the benefits of Io3MT.

Finally, environmental challenges are linked to the life cycle of connected devices, encompassing issues ranging from raw material extraction to equipment disposal. These concerns include pollution, depletion of natural resources, electronic waste generation, and broader ecological impacts. The sustainable development of Io3MT requires the adoption of practices such as ecological design, the use of energy-efficient materials, virtualization of activities, recycling programs, and strategies to address planned obsolescence.

8.6 Future Work

As a continuation of the proposed reference model, one future research direction involves the definition of a common ontology for gestures, sound effects, multimedia information, and sensory events. Such an ontology aims to facilitate and accelerate the development of applications while reducing interpretative asymmetries among heterogeneous systems, thereby establishing more effective interoperability relations and contributing to the consolidation of Io3MT as a structured domain.

In parallel with this effort, another avenue for future investigation concerns the development of an evaluation methodology for the reference model. Such a framework would make it possible to examine the extent to which the model effectively supports researchers and developers in designing their own Io3MT environments. The insights gained from this process would, in turn, inform an iterative cycle of refinement, contributing to the progressive consolidation and maturation of the proposed model.

From a technical perspective, a future line of work involves the development of new modules and libraries for Pure Data, Unity, and other relevant tools, specifically designed to support the creation of Io3MT scenarios. Also, future work includes extending such environments to enable multi-user applications, thereby facilitating distributed interaction, shared multisensory experiences, and collaborative content generation across heterogeneous devices and networks.

Another avenue of advancement will be the incorporation of additional sensory effects, such as airflow and thermal stimuli, thereby broadening the diversity of inputs and enriching the multisensory experience.

In parallel, efforts will be directed toward the development of algorithms for the automatic extraction of multimedia and multisensory elements from musical tracks, which may lead to mechanisms for translation, mapping, alignment, and fusion of modalities. Furthermore, artificial intelligence–based services may play a central role in managing experiences, either by automatically synchronizing audio

tracks with multimedia and sensory elements or by detecting and notifying potential asynchronies between instrumental parts in specific segments. In addition, such services may contribute to the dynamic adjustment of network parameters, thereby mitigating latency and ensuring higher quality of service.

Besides that, the use of embedded AI emerges as a means of enabling devices to collect environmental information, interpret contexts, and perform autonomous actions in real time. These techniques can likewise be extended to big data management, ensuring heterogeneity among devices, accommodating both structured and unstructured data, addressing the high velocity of data generation and processing, and supporting continuous adaptation to changing system conditions.

With regard device management, future perspectives include the development of dashboards and APIs designed to simplify control, discovery, and communication among network elements. These APIs should provide mechanisms for authentication and access management, such as transaction tokens, thereby ensuring the regulated and secure use of critical functionalities, sensory resources, and data generated by devices. Discovery services incorporating monitoring, management, and reputation analysis of devices may also be implemented, enhancing the overall reliability of Io3MT ecologies.

Another future task will involve the creation of new use cases aimed at validating the reference model, alongside the development of additional device prototypes. These efforts will expand the range of examples to domains such as games, cinema, healthcare applications, education, and interactive artistic installations.

From an architectural perspective, the scope of Io3MT is expected to be expanded through integration with edge and cloud computing paradigms, thereby strengthening the connectivity infrastructure. This advancement also encompasses the practical implementation of the communication and application protocols discussed in the theoretical scope of this thesis but not yet applied in the presented use cases. To this end, equivalent scenarios will be developed operating under different protocols, in order to identify, through comparative analysis, which configurations demonstrate the greatest efficiency for each task.

In the artistic domain, a new form of musical notation is envisioned, one that incorporates sensory and multimedia elements. This would be accompanied by the development of libraries, data formats, multimedia encoders, and, eventually, programming languages specifically designed to ensure portability and functionality on devices with limited computational resources.

To facilitate the indexing and retrieval of these new resources, a key future task is the creation of an online repository dedicated to Io3MT, bringing together specifications, academic publications, device/environments templates, code snippets, and tutorials. This repository should function as an open knowledge base, supporting researchers, developers, designers, and artists in building their own applications, thereby accelerating the innovation cycle and ensuring the dissemination of the proposed paradigm.

REFERENCES

AKYILDIZ, Ian; MELODIA, Tommaso; CHOWDHURY, Kaushik. A survey on wireless multimedia sensor networks. **Computer networks**, Elsevier, v. 51, n. 4, p. 921–960, 2007.

_____. Wireless multimedia sensor networks: Applications and testbeds. **Proceedings of the IEEE**, IEEE, v. 96, n. 10, p. 1588–1605, 2008.

ALIPRANTIS, John; CARIDAKIS, George. A survey of augmented reality applications in cultural heritage. **International Journal of Computational Methods in Heritage Science (IJCMHS)**, IGI Global Scientific Publishing, v. 3, n. 2, p. 118–147, 2019.

ALMALKAWI, Islam et al. Wireless multimedia sensor networks: current trends and future directions. **Sensors**, Molecular Diversity Preservation International (MDPI), v. 10, n. 7, p. 6662–6717, 2010.

ALVI, Sheeraz et al. Internet of multimedia things: Vision and challenges. **Ad Hoc Networks**, v. 33, p. 87–111, 2015. ISSN 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2015.04.006. Available from: <https://www.sciencedirect.com/science/article/pii/S1570870515000876>.

AMAZON. **AWS IoT**. 2023. Available from: <https://aws.amazon.com/pt/iot/>. Visited on: 21 Aug. 2025.

ARAÚJO, João Teixeira et al. A technical approach of the audience participation in the performance 'O Chaos das 5'. In_____. **Proceedings of the 17th Brazilian Symposium on Computer Music**. São João del-Rei - MG - Brazil: Sociedade Brasileira de Computação, Sept. 2019. P. 28–34.

ASHTON, Kevin. That 'Internet of Things' Thing. **RFID Journal**, RFID Journal, v. 15, p. 1, 2009.

ATZORI, Luigi; IERA, Antonio; MORABITO, Giacomo. The Internet of Things: A Survey. **Computer Networks**, p. 2787–2805, Oct. 2010. DOI: 10.1016/j.comnet.2010.05.010.

AZMANDIAN, Mahdi et al. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In: PROCEEDINGS of the 2016 chi conference on human factors in computing systems. [S.l.: s.n.], 2016. P. 1968–1979.

AZUMA, Ronald. A survey of augmented reality. **Presence: teleoperators & virtual environments**, MIT press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . ., v. 6, n. 4, p. 355–385, 1997.

AZURE, Microsoft. **Azure IoT Central**. 2023. Available from: <https://azure.microsoft.com/pt-br/products/iot-central/>. Visited on: 21 Aug. 2023.

BALANDRA, Alfonso et al. Selective listening attention enhancement, using a simultaneous visual and haptic stimuli. **Journal of New Music Research**, Taylor & Francis, v. 48, n. 3, p. 294–308, 2019.

BALK, Stacy; BERTOLA, Dakota; INMAN, Vaughan. Simulator sickness questionnaire: twenty years later. In: UNIVERSITY OF IOWA, 2013. DRIVING assessment conference. [S.l.: s.n.], 2013. v. 7.

BARBOSA, Jerônimo et al. Illusio: A drawing-based digital music instrument. In: PROCEEDINGS of the International Conference on New Interfaces for Musical Expression. [S.l.: s.n.], 2013. P. 499–502.

BARGAS-AVILA, Javier; HORNBAEK, Kasper. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: PROCEEDINGS of the SIGCHI conference on human factors in computing systems. [S.l.: s.n.], 2011. P. 2689–2698.

BARRETO, Fábio. **Uma Proposta de Extensão do Middleware Ginga-NCL para Interação Multimodal e Suporte Multiusuário em Ambientes Hipermídia**. Dec. 2021. PhD thesis – Universidade Federal Fluminense, Niterói, RJ, Brasil.

BASSI, Alessandro et al. **Enabling Things to Talk: Designing IoT Solutions with TheIoT Architectural Reference Model**. [S.l.]: Springer Nature, 2013.

BEKELE, Mafkereseb Kassahun; CHAMPION, Erik. Redefining mixed reality: User-reality-virtuality and virtual heritage perspectives. In: PROCEEDINGS of the 24th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA). [S.l.: s.n.], 2019. v. 2, p. 675–684.

BENJAMIN, Walter. **The work of art in the age of its technological reproducibility, and other writings on media**. [S.l.]: Harvard University Press, 2008.

BERG, Leif; VANCE, Judy. Industry use of virtual reality in product design and manufacturing: A survey-virtual reality. **SpringerLink, Sep**, 2016.

BERGER, Christopher et al. The uncanny valley of haptics. **Science Robotics**, American Association for the Advancement of Science, v. 3, n. 17, eaar7010, 2018.

BERTHAUT, Florent. 3D interaction techniques for musical expression. **Journal of New Music Research**, Taylor & Francis, v. 49, n. 1, p. 60–72, 2020.

BERTHAUT, Florent; DESAINTE-CATHERINE, Myriam; HACHET, Martin. Interacting with 3D reactive widgets for musical performance. **Journal of New Music Research**, Taylor & Francis, v. 40, n. 3, p. 253–263, 2011.

BIMBERG, Pauline; WEISSKER, Tim; KULIK, Alexander. On the usage of the simulator sickness questionnaire for virtual reality research. In: IEEE. 2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW). [S.l.: s.n.], 2020. P. 464–467.

BIN, Xia et al. Activation and Transfer: The Application of Digital Media in Environmental Art Design. **Frontiers in Art Research**, Francis Academic Press, v. 5, n. 8, 2023.

BLACKING, John. **How musical is man?** [S.l.]: University of Washington Press, 1973.

BOEM, Alberto; TOMASETTI, Matteo, et al. User needs in the Musical Metaverse: a case study with electroacoustic musicians. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION. [S.l.: s.n.], 2024.

BOEM, Alberto; TURCHET, Luca. Musical Metaverse Playgrounds: exploring the design of shared virtual sonic experiences on web browsers. In: PROCEEDINGS of the 4th International Symposium on the Internet of Sounds. [S.l.: s.n.], 2023. P. 1–9. DOI: 10.1109/IEEECONF59510.2023.10335297.

BOLTON, Matthew; BILTEKOFF, Elliot; HUMPHREY, Laura. The mathematical meaninglessness of the NASA task load index: A level of measurement analysis. **IEEE Transactions on Human-Machine Systems**, IEEE, v. 53, n. 3, p. 590–599, 2023.

BOURRIAUD, Nicolas; SCHNEIDER, Caroline; HERMAN, Jeanine. **Postproduction: Culture as screenplay: How art reprograms the world**. [S.l.]: Lukas & Sternberg New York, 2002. v. 7.

BOWERS, Faubion. **Scriabin, a Biography**. [S.l.]: Dover, 1996. ISBN 9780486288970.

BRADLEY, Margaret; LANG, Peter. Measuring emotion: the self-assessment manikin and the semantic differential. **Journal of behavior therapy and experimental psychiatry**, Elsevier, v. 25, n. 1, p. 49–59, 1994.

BROOKE, John et al. SUS - A quick and dirty usability scale. **Usability evaluation in industry**, London, England, v. 189, n. 194, p. 4–7, 1996.

BROOKE, John. SUS: a retrospective. **Journal of usability studies**, v. 8, n. 2, 2013.

BROWN, Dom; NASH, Chris; MITCHELL, Tom. A user experience review of music interaction evaluations. **International Conference on New Interfaces for Musical Expression**, v. 23, p. 28, 2017.

CÁCERAS, Juan-Pablo; CHAFE, Chris. Jacktrip/Soundwire meets server farm. **Computer Music Journal**, JSTOR, v. 34, n. 3, p. 29–34, 2010.

CAPPELEN, Birgitta; ANDERSSON, Anders-Petter. Health Improving Multi-Sensorial and Musical Environments. In: PROCEEDINGS of the Audio Mostly 2016. Norrköping, Sweden: Association for Computing Machinery, 2016. (AM '16), p. 178–185. ISBN 9781450348225. DOI: 10.1145/2986416.2986427. Available from: <https://doi.org/10.1145/2986416.2986427>.

CARDOSO, António. Arte interativa: quem é o autor e onde está o espectador? **AVANCA| CINEMA**, p. 45–49, 2019.

CAROT, Alexander; WERNER, Christian. Fundamentals and Principles of Musical Telepresence. **Journal of Science and Technology of the Arts**, v. 1, May 2009. DOI: 10.7559/citarj.v1i1.6.

CARÔT, Alexander; WERNER, Christian. Network music performance-problems, approaches and perspectives. In: PROCEEDINGS of the "Music in the Global Village"-Conference, Budapest, Hungary. [S.l.: s.n.], 2007. v. 162, p. 23–10.

CARROLL, Erin A; LATULIPE, Celine. The creativity support index. In: CHI'09 Extended Abstracts on Human Factors in Computing Systems. [S.l.]: ACM Digital Library, 2009. P. 4009–4014.

CENTENARO, Marco; CASARI, Paolo; TURCHET, Luca. Towards a 5G Communication Architecture for the Internet of Musical Things. In: PROCEEDINGS of the 27th Conference of Open Innovations Association (FRUCT). [S.l.: s.n.], Sept. 2020. P. 38–45. DOI: 10.23919/FRUCT49677.2020.9210980.

CERRATTO PARGMAN, Teresa; ROSSITTO, Chiara; BARKHUUS, Louise. Understanding Audience Participation in an Interactive Theater Performance. In. DOI: 10.1145/2639189.2641213.

CHOI, Bumsuk; LEE, Eun-Seo; YOON, Kyoungro. Streaming Media with Sensory Effect. In: PROCEEDINGS of the 2011 International Conference on Information Science and Applications. [S.l.: s.n.], 2011. P. 1–6. DOI: 10.1109/ICISA.2011.5772390.

CICILIANI, Marko. Virtual 3D environments as composition and performance spaces. **Journal of New Music Research**, Taylor & Francis, v. 49, n. 1, p. 104–113, 2020.

CIOFI-SILVA, Caroline Lopes et al. Workload assessment: cross-cultural adaptation, content validity and instrument reliability. **Revista brasileira de enfermagem**, SciELO Brasil, v. 76, n. 3, e20220556, 2023.

CLARKE, Victoria; BRAUN, Virginia. Thematic analysis. In: ENCYCLOPEDIA of critical psychology. [S.l.]: Springer, 2014. P. 1947–1952.

COALLIER, François. IoT Standardization Strategies in ISO/IEC JTC 1/SC 41. In: IEEE. PROCEEDINGS of the IEEE 8th World Forum on Internet of Things (WF-IoT). [S.l.: s.n.], 2022. P. 1–6.

COOK, Perry. 2001: Principles for designing computer music controllers. **A NIME Reader: Fifteen years of new interfaces for musical expression**, Springer, p. 1–13, 2017.

_____. Re-Designing Principles for Computer Music Controllers: a Case Study of SqueezeVox Maggie. In: NIME. [S.l.: s.n.], 2009. v. 9, p. 218–221.

COVACI, Alexandra et al. Is multimedia multisensorial? A review of mulsemedia systems. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 51, n. 5, p. 1–35, 2018.

CRESWELL, John; CRESWELL, David. **Research design: Qualitative, quantitative, and mixed methods approaches**. [S.l.]: Sage publications, 2017.

CROSS, Ian. Music, cognition, culture, and evolution. **Annals of the New York Academy of sciences**, Wiley Online Library, v. 930, n. 1, p. 28–42, 2001.

CUNNINGHAM, Stuart; WEINEL, Jonathan. The Sound of the Smell (and Taste) of My Shoes Too: Mapping the Senses Using Emotion as a Medium. In: PROCEEDINGS of the Audio Mostly 2016. Norrköping, Sweden: Association for Computing Machinery, 2016. (AM '16), p. 28–33. ISBN 9781450348225. DOI: 10.1145/2986416.2986456. Available from: <https://doi.org/10.1145/2986416.2986456>.

DALSGAARD, Tor-Salve; SCHNEIDER, Oliver. A Unified Model for Haptic Experience. **ACM Transactions on Computer-Human Interaction**, ACM New York, NY, 2025.

DAVIDSON, Jack. **Apple Homekit: The Beginner's Guide**. [S.l.]: Van Helostein, 2017. v. 1.

DAVISON, Brian D. Predicting Web Actions from HTML Content. In: PROCEEDINGS of the 13th ACM Conference on Hypertext and Hypermedia. College Park, Maryland, USA: Association for Computing Machinery, 2002. (HYPERTEXT '02), p. 159–168. ISBN 1581134770. DOI: 10.1145/513338.513380. Available from: <https://doi.org/10.1145/513338.513380>.

DEGRAEN, Donald et al. Weirding haptics: In-situ prototyping of vibrotactile feedback in virtual reality through vocalization. In: PROCEEDINGS of the 34th Annual ACM symposium on user interface software and technology. New York, USA: ACM Digital Library, 2021. P. 936–953.

DENZIN, Norman K; LINCOLN, Yvonna S. **The Sage handbook of qualitative research**. [S.l.]: SAGE, 2011.

DIMOVA, Polina. **At the crossroads of the senses: The synaesthetic metaphor across the arts in European Modernism**. [S.l.]: Penn State Press, 2024.

DRASCIC, David; MILGRAM, Paul. Perceptual issues in augmented reality. In: SPIE. STEREOSCOPIC displays and virtual reality systems III. [S.l.: s.n.], 1996. v. 2653, p. 123–134.

DZIWIS, Damian; COLER, Henrik von; PORSCHMANN, Christoph. Live Coding in the Metaverse. In: PROCEEDINGS of the 4th International Symposium on the Internet of Sounds. [S.l.: s.n.], 2023. P. 1–8. DOI: 10.1109/IEEECONF59510.2023.10335358.

ECLIPSE. **Eclipse Kura**. 2023. Available from: <https://projects.eclipse.org/projects/iot.kura>. Visited on: 21 Aug. 2025.

EDMONDS, Ernest. The art of interaction. **Digital Creativity**, Taylor & Francis, v. 21, n. 4, p. 257–264, 2010.

EDMONDS, Ernest; TURNER, Greg; CANDY, Linda. Approaches to interactive art systems. In: PROCEEDINGS of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia. [S.l.: s.n.], 2004. P. 113–117.

FAUSTINO, Daniella Briotto et al. SensArt demo: A multisensory prototype for engaging with visual art. In: PROCEEDINGS of the 2017 ACM international conference on interactive surfaces and spaces. New York, USA: ACM Digital Library, 2017. P. 462–465.

FEICK, Martin et al. 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. [S.l.: s.n.], 2021. P. 1–13.

FEINER, Steven; MACINTYRE, Blair; SELIGMANN, Dorée. Knowledge-based augmented reality. **Communications of the ACM**, ACM New York, NY, USA, v. 36, n. 7, p. 53–62, 1993.

FELDMAN, Tony. **Multimedia**. [S.l.]: Psychology Press, 1994. v. 64.

FERTLEMAN, Caroline et al. A discussion of virtual reality as a new tool for training healthcare professionals. **Frontiers in public health**, Frontiers Media SA, v. 6, p. 44, 2018.

FILIMON, Rosina Caterina. Syncretism and synaesthesia in music: unification of arts and perceptions. **Artes. Journal of musicology**, Editura ARTES, n. 27-28, p. 167–184, 2023.

FITCH, Tecumseh. The biology and evolution of music: A comparative perspective. **Cognition**, Elsevier, v. 100, n. 1, p. 173–215, 2006.

FLORIS, Alessandro; ATZORI, Luigi. Managing the Quality of Experience in the Multimedia Internet of Things: A Layered-Based Approach. **Sensors**, v. 16, p. 2057, Dec. 2016. DOI: 10.3390/s16122057.

_____. Managing the quality of experience in the multimedia internet of things: A layered-based approach. **Sensors**, MDPI, v. 16, n. 12, p. 2057, 2016.

FLORIS, Alessandro; ATZORI, Luigi. Quality of Experience in the Multimedia Internet of Things: Definition and practical use-cases. In: PROCEEDINGS of the 2015 IEEE International Conference on Communication Workshop (ICCW). [S.l.: s.n.], 2015. P. 1747–1752. DOI: 10.1109/ICCW.2015.7247433.

FLUSSER, V. **Filosofia da caixa preta: ensaios para uma futura filosofia da fotografia**. [S.l.]: Ed. Hucitec, 1985. Available from: <https://books.google.com.br/books?id=mpp2LwAACAAJ>.

FLUSSER, Vilém. **O universo das imagens técnicas: elogio da superfcialidade**. [S.l.]: Imprensa da Universidade de Coimbra/Coimbra University Press, 2012.

FORNY, Leonardo. Arte e Interação: Nos Caminhos da Arte Interativa? **Razón y palabra, ISSN 1605-4806, Nº. 53, 2006**, http://www.razonypalabra.org.mx/anteriores/n53/lforny.html, Nov. 2006.

FREITAS, Sérgio Paulo Ribeiro de. Relação e sistema: duas palavras-chave na trajetória da teoria tonal. **Musica Theorica**, v. 3, n. 2, 2018.

AL-FUQAHA, Ala et al. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. **IEEE Communications Surveys & Tutorials**, v. 17, n. 4, p. 2347–2376, 2015. DOI: 10.1109/COMST.2015.2444095.

FURHT, Borko. **Handbook of augmented reality**. [S.l.]: Springer Science & Business Media, 2011.

_____. Multimedia systems: An overview. **IEEE MultiMedia**, IEEE, v. 1, n. 1, p. 47–59, 2002.

GABRIELLI, Leonardo; TURCHET, Luca. Towards a Sustainable Internet of Sounds. In: PROCEEDINGS of the 17th International Audio Mostly Conference. St. Pölten, Austria: Association for Computing Machinery, 2022. (AM '22), p. 231–238. ISBN 9781450397018. DOI: 10.1145/3561212.3561246. Available from: <https://doi.org/10.1145/3561212.3561246>.

GENG, Hwaiyu. **Internet of Things and Data Analytics Handbook**. [S.l.]: Wiley, 2017.

GERSHENFELD, Neil; KRIKORIAN, Raffi; COHEN, Danny. The Internet of Things. **Scientific American**, v. 291, p. 76–81, Nov. 2004. DOI: 10.1038/scientificamerican1004-76.

GHINEA, Gheorghita et al. Mulsemedia: State of the art, perspectives, and challenges. **ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)**, ACM New York, NY, USA, v. 11, 1s, p. 1–23, 2014.

GIBSON, David. **The Art of Mixing: A Visual Guide to Recording, Engineering, and Production, Second Edition**. [S.l.]: Thomson Course Technology, 2005. ISBN 9781598630602.

GIROUX, Félix et al. Haptic stimulation with high fidelity vibro-kinetic technology psychophysiologically enhances seated active music listening experience. In: IEEE. PROCEEDINGS of the 2019 IEEE World Haptics Conference (WHC). [S.l.: s.n.], 2019. P. 151–156.

GONSALEZ, Elisiane Crestani de Miranda; ALMEIDA, Kátia de. Cross-cultural adaptation of the speech, spatial and qualities of hearing scale (SSQ) to Brazilian Portuguese. **Audiology-Communication Research**, SciELO Brasil, v. 20, p. 215–224, 2015.

GOZDECKI, J.; JAJSZCZYK, A.; STANKIEWICZ, R. Quality of service terminology in IP networks. **IEEE Communications Magazine**, v. 41, n. 3, p. 153–159, 2003. DOI: 10.1109/MCOM.2003.1186560.

GRIER, Rebecca et al. The system usability scale: Beyond standard usability testing. In: SAGE PUBLICATIONS SAGE CA: LOS ANGELES, CA, 1. PROCEEDINGS of the human factors and ergonomics society annual meeting. [S.l.: s.n.], 2013. v. 57, p. 187–191.

GSÖLLPOINTNER, Katharina; SCHNELL, Ruth; SCHULER, Romana Karla. **Digital synesthesia: a model for the aesthetics of digital art**. [S.l.]: Walter de Gruyter GmbH & Co KG, 2016.

GU, Xiaoyuan et al. Network-centric music performance: practice and experiments. **IEEE Communications Magazine**, v. 43, n. 6, p. 86–93, 2005. DOI: 10.1109/MCOM.2005.1452835.

GUBBI, Jayavardhana et al. Internet of Things (IoT): A vision, architectural elements, and future directions. **Future Generation Computer Systems**, v. 29, n. 7, p. 1645–1660, 2013. Including Special sections: Cyber-enabled Distributed Computing for Ubiquitous Cloud and Network Services & Cloud Computing and Scientific Applications — Big Data, Scalable Analytics, and Beyond. ISSN 0167-739X. DOI: https://doi.org/10.1016/j.future.2013.01.010. Available from: <https://www.sciencedirect.com/science/article/pii/S0167739X13000241>.

GUEDES, Alan et al. Using NCL to synchronize media objects, sensors and actuators. In: SBC. EXTENDED Proceedings of the XXII Brazilian Symposium in Multimedia Systems and Web. [S.l.: s.n.], 2016. P. 184–189.

GUTH, Jasmin et al. Comparison of IoT platform architectures: A field study based on a reference architecture. In: IEEE. 2016 Cloudification of the Internet of Things (CIoT). [S.l.: s.n.], 2016. P. 1–6.

HALLER, Stephan. The Things in the Internet of Things. **Bern University**, Jan. 2010.

HALSALL, Fred. **Multimedia communications: Applications, networks, protocols and standards**. [S.l.]: Pearson education, 2001.

HARGREAVES, David J; NORTH, Adrian C. The functions of music in everyday life: Redefining the social in music psychology. **Psychology of music**, Sage Publications Sage CA: Thousand Oaks, CA, v. 27, n. 1, p. 71–83, 1999.

HARRISON, John E. **Synaesthesia: The strangest thing**. [S.l.]: Oxford University Press, 2001.

HART, Sandra; STAVELAND, Lowell. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: ADVANCES in psychology. [S.l.]: Elsevier, 1988. v. 52. P. 139–183.

HART, Sandra G. NASA-task load index (NASA-TLX); 20 years later. In: SAGE PUBLICATIONS SAGE CA: LOS ANGELES, CA, 9. PROCEEDINGS of the human factors and ergonomics society annual meeting. [S.l.: s.n.], 2006. v. 50, p. 904–908.

HAYES, Lauren. Vibrotactile Feedback-Assisted Performance. In: CITESEER. NIME. [S.l.: s.n.], 2011. P. 72–75.

HODGES, Bridger Scott. **The Effects of Haptics on Rhythm Dance Game Performance and Enjoyment**. [S.l.]: Brigham Young University, 2018.

HOLLÄNDER, Kai et al. Investigating the influence of external car displays on pedestrians' crossing behavior in virtual reality. In: PROCEEDINGS of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. [S.l.: s.n.], 2019. P. 1–11.

HU, Xiao et al. The MIREX grand challenge: A framework of holistic user-experience evaluation in music information retrieval. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 68, n. 1, p. 97–112, 2017.

HWANG, Inwook; SON, Hyungki; KIM, Jin Ryong. AirPiano: Enhancing music playing experience in virtual reality with mid-air haptic feedback. In: IEEE. PROCEEDINGS of the 2017 IEEE world haptics conference (WHC). [S.l.: s.n.], 2017. P. 213–218.

IAZZETTA, Fernando. O que é música (hoje). **Fórum Catarinense de Musicoterapia**, v. 1, p. 5–14, 2001.

IEEE 11073: Health informatics — Point-of-care medical device communication (PoCD). New York: [s.n.], 2004. (IEEE Standard 11073 family). Ongoing series, includes multiple parts such as IEEE 11073-10101, 11073-10201, 11073-20601, and device specializations (104xx).

IERUSALIMSCHY, Roberto; FIGUEIREDO, Luiz Henrique de; CELES, Waldemar. The evolution of Lua. In: PROCEEDINGS of the third ACM SIGPLAN conference on History of programming languages. [S.l.: s.n.], 2007. P. 2–1.

ILOVAN, Oana-Ramona; DOROFTEI, Iulia. **Qualitative research in regional geography: A methodological approach**. [S.l.]: Presa Universitară Clujeană Cluj-Napoca, 2017.

IORWERTH, Miriam; MOORE, David; KNOX, Don. Challenges of using Networked Music Performance in education. In: PROCEEDINGS of the Audio Engineering Society Conference. [S.l.: s.n.], Aug. 2015. Available from: <http://www.aes.org/e-lib/browse.cfm?elib=17857>.

IORWERTH, Miriam Anne; KNOX, Don. The application of Networked Music Performance to access ensemble activity for socially isolated musicians. English. In: PROCEEDINGS of the Web Audio Conference 2019 – Diversity in Web Audio. [S.l.: s.n.], Dec. 2019. Acceptance in SAN AAM: Unknown publisher policy - made file open and contacted publisher - ET 20-9-19 Event held at: Norwegian University of Science and Technology (NTNU), Trondheim, Norway. Dec 4-6 2019 Author confirmed publication date 7/2/20 ET.

ISO/IEC. **Information technology — Media context and control — Part 3: Sensory information**. Geneva: International Organization for Standardization, 2019. MPEG-V Part 3.

ISO/IEC 20924:2018: Information technology — Internet of Things (IoT) — Vocabulary. Geneva: [s.n.], Dec. 2018. Edition 1; provides definitions and a foundational terminology for the Internet of Things (IoT). Withdrawn in March 2021.

ISO/IEC 21823-1:2019: Internet of things (IoT) — Interoperability for IoT systems — Part 1: Framework. Geneva: [s.n.], Feb. 2019. Edition 1; provides a framework and overview for interoperability of IoT systems, enabling peer-to-peer information exchange.

ISO/IEC 23005-5:2019: Information technology — Media context and control — Part 5: Data formats for interaction devices. Geneva: [s.n.], Feb. 2019. Fourth edition; specifies syntax and semantics of data formats for interaction devices (actuators and sensors) using IIDL, Device Command Vocabulary and Sensed Information Vocabulary.

ISO/IEC 30118-1:2018: Information technology — Open Connectivity Foundation (OCF) Specification — Part 1: Core specification. Geneva: [s.n.], Nov. 2018. Edition 1; defines the OCF core architecture, interfaces, protocols, and services to enable implementation of IoT profiles; withdrawn and replaced by the 2021 edition.

ISO/IEC 30141:2024 – Internet of Things (IoT) — Reference architecture. 2. ed. Geneva: [s.n.], Aug. 2024. Supersedes the 2018 edition. Includes a technical revision, alignment with ISO/IEC/IEEE 42010:2022, enhanced usability, and extended support for implementation standards.

ISO/IEC TR 22417:2017: Information technology — Internet of things (IoT) use cases. Geneva: [s.n.], Nov. 2017. Edition 1; identifies IoT scenarios and use cases to support interoperability and standardization efforts.

ITU-T F.748.0: Common requirements for Internet of things (IoT) applications. Geneva: [s.n.], Oct. 2014. (ITU-T Recommendation F.748.0 (also Y.4103), F.748.0). Renumbered as ITU-T Y.4103 on 2016-02-05 without further modification.

ITU-T Y.2066: Common requirements of the Internet of things. Geneva: [s.n.], June 2014. (ITU-T Recommendation Y series, Y.2066). Approved 22 June 2014; renumbered as ITU-T Y.4100 on 5 February 2016 without further modification.

ITU-T Y.2068 (also Y.4401): Functional framework and capabilities of the Internet of Things. Geneva: [s.n.], Mar. 2015. (ITU-T Recommendation Y series, Y.2068 (Y.4401)). Approved 22 March 2015; renumbered as ITU-T Y.4401 on 5 February 2016 without modification.

ITU-T Y.4000/Y.2060: Overview of the Internet of things. Geneva: [s.n.], June 2012. (ITU-T Recommendation Y series, Y.4000/Y.2060). Approved 15 June 2012; Y.2060 renumbered as Y.4000 on 5 February 2016 without modification.

ITU-T Y.4100/Y.2066: COMMON REQUIREMENTS OF THE INTERNET OF THINGS. Geneva: [s.n.], June 2014. (ITU-T Recommendation Y series, Y.4100/Y.2066). Edition 1; approved on 22 June 2014; originally published as Y.2066 and renumbered to Y.4100 on 5 February 2016 without modification.

ITU-T Y.4111/Y.2076: SEMANTICS BASED REQUIREMENTS AND FRAMEWORK OF THE INTERNET OF THINGS. Geneva: [s.n.], Feb. 2016. (ITU-T Recommendation Y series, Y.4111/Y.2076). Edition 1; approved on 13 February 2016; originally published as Y.2076 and renumbered as Y.4111 on 5 February 2016 without modification.

ITU-T Y.4400/Y.2063: FRAMEWORK OF THE WEB OF THINGS. Geneva: [s.n.], July 2012. (ITU-T Recommendation Y Series, Y.4400/Y.2063). Approved 29 July 2012; originally published as Y.2063 and renumbered as Y.4400 on 5 February 2016 without modification.

ITU-T Y.4401/Y.2068: FUNCTIONAL FRAMEWORK AND CAPABILITIES OF THE INTERNET OF THINGS. Geneva: [s.n.], Mar. 2015. (ITU-T Recommendation Y series, Y.4401/Y.2068). Edition 1; approved March 2015; originally published as Y.2068 and renumbered to Y.4401 on 5 February 2016 without modification.

JACKSON, Susan A; MARSH, Herbert W. Development and validation of a scale to measure optimal experience: The Flow State Scale. **Journal of sport and exercise psychology**, Human Kinetics, Inc., v. 18, n. 1, p. 17–35, 1996.

JOHNSON, Daniel; GARDNER, John; PERRY, Ryan. Validation of two game experience scales: the player experience of need satisfaction (PENS) and game experience questionnaire (GEQ). **International Journal of Human-Computer Studies**, Elsevier, v. 118, p. 38–46, 2018.

JOHNSON, David. **MusE-XR: musical experiences in extended reality to enhance learning and performance**. Dec. 2019. PhD thesis – University of Victoria, Oak Bay, Canada.

JOSUÉ, Marina; ABREU, Raphael, et al. Modeling Sensory Effects as First-Class Entities in Multimedia Applications. In: PROCEEDINGS of the 9th ACM Multimedia Systems Conference. Amsterdam, Netherlands: Association for Computing Machinery, 2018. (MMSys '18), p. 225–236. ISBN 9781450351928. DOI: 10.1145/3204949.3204967. Available from: <https://doi.org/10.1145/3204949.3204967>.

JOSUÉ, Marina; MORENO, Marcelo; MUCHALUAT SAADE, Débora. Mulsemedia Preparation: A New Event Type for Preparing Media Object Presentation and Sensory Effect Rendering. In: PROCEEDINGS of the 10th ACM Multimedia Systems Conference. Amherst, Massachusetts: Association for Computing Machinery, 2019. (MMSys '19), p. 110–120. ISBN 9781450362979. DOI: 10.1145/3304109.3306230. Available from: <https://doi.org/10.1145/3304109.3306230>.

JOSUÉ, Marina Ivanov Pereira. **Preparação de Objetos de Mídia e Efeitos Sensoriais para Formatação de Documentos Mulsemídia**. Dec. 2021. PhD thesis – Universidade Federal Fluminense, Niterói, RJ, Brasil.

JULURI, Parikshit; TAMARAPALLI, Venkatesh; MEDHI, Deep. Measurement of Quality of Experience of Video-on-Demand Services: A Survey. **IEEE Communications Surveys & Tutorials**, v. 18, n. 1, p. 401–418, 2016. DOI: 10.1109/COMST.2015.2401424.

KENNEDY, Robert S et al. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. **The international journal of aviation psychology**, Taylor & Francis, v. 3, n. 3, p. 203–220, 1993.

KIM, Jaeha et al. Construction of a Haptic-Enabled Broadcasting System Based on the MPEG-V Standard. **Image Commun.**, Elsevier Science Inc., USA, v. 28, n. 2, p. 151–161, Feb. 2013. ISSN 0923-5965. DOI: 10.1016/j.image.2012.10.010. Available from: <https://doi.org/10.1016/j.image.2012.10.010>.

KIVY, Peter. **Introduction to a Philosophy of Music**. [S.l.]: Oxford University Press, 2002.

KOELSCH, Stefan. Brain correlates of music-evoked emotions. **Nature reviews neuroscience**, Nature Publishing Group UK London, v. 15, n. 3, p. 170–180, 2014.

KRAHMER, Emiel; UMMELEN, Nicole. Thinking about thinking aloud: A comparison of two verbal protocols for usability testing. **IEEE transactions on professional communication**, IEEE, v. 47, n. 2, p. 105–117, 2004.

KRASONIKOLAKIS, Ioannis et al. Store layout effects on consumer behavior in 3D online stores. **European Journal of Marketing**, Emerald Publishing Limited, v. 52, n. 5/6, p. 1223–1256, 2018.

KRUEGER, Richard A. **Focus groups: A practical guide for applied research**. [S.l.]: Sage publications, 2014.

LANGE, A. **Arthur Lange's Spectrotone System of Orchestration: A colorgraphic exposition of tone-color combinations & balance as practiced in modern orchestration**. [S.l.]: Co-Art, 1943. (Arthur Lange's Spectrotone System of Orchestration). Available from: <https://books.google.com.br/books?id=WH0FGwAACAAJ>.

LAZZARO, John; WAWRZYNEK, John. A Case for Network Musical Performance. In: PROCEEDINGS of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video. Port Jefferson, New York, USA: Association for Computing Machinery, 2001. (NOSSDAV '01), p. 157–166. ISBN 1581133707. DOI: 10.1145/378344.378367. Available from: <https://doi.org/10.1145/378344.378367>.

LEÃO, Lucia. O remix nos processos de criação de imagens e imaginários midiáticos. In: XXI Encontro da Compós. [S.l.: s.n.], June 2012.

LEE, Euijong et al. A Survey on Standards for Interoperability and Security in the Internet of Things. **IEEE Communications Surveys & Tutorials**, IEEE, v. 23, n. 2, p. 1020–1047, 2021.

LEE, Sueyoon; STRINER, Alina; CÉSAR, Pablo. Designing a vr lobby for remote opera social experiences. In: PROCEEDINGS of the 2022 ACM International Conference on Interactive Media Experiences. [S.l.: s.n.], 2022. P. 293–298.

LEE, Sueyoon et al. Designing and evaluating a vr lobby for a socially enriching remote opera watching experience. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 30, n. 5, p. 2055–2065, 2024.

LEVSTEK, Maruša et al. "It All Makes Us Feel Together": Young People's Experiences of Virtual Group Music-Making During the COVID-19 Pandemic. **Frontiers in Psychology**, Frontiers Media SA, v. 12, p. 703892, 2021.

LÉVY, Pierre. **Cibercultura**. 3rd. [S.l.]: Editora 34, 2010.

LEWIS, James R; SAURO, Jeff. Item benchmarks for the system usability scale. **Journal of Usability studies**, v. 13, n. 3, 2018.

LEWISOHN, Mark. **Revolution in the Head: The Beatles' Records and the Sixties**. [S.l.]: Hamlyn, 2021. ISBN 9780600637127.

LI, Jie; CÉSAR, Pablo. Social virtual reality (VR) applications and user experiences. In: IMMERSIVE video technologies. [S.l.]: Elsevier, 2023. P. 609–648.

LIN, Jie et al. A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications. **IEEE Internet of Things Journal**, v. 4, n. 5, p. 1125–1142, 2017. DOI: 10.1109/JIOT.2017.2683200.

LOURENÇO, Douglas Fabiano; CARMONA, Elenice Valentim;
MORAES LOPES, Maria Helena Baena de. Translation and cross-cultural adaptation of the System Usability Scale to Brazilian Portuguese. **Aquichan**, Universidad de La Sabana, v. 22, n. 2, 2022.

LOVERIDGE, Ben. Networked music performance in virtual reality: current perspectives. **Journal of Network Music and Arts**, v. 2, n. 1, p. 2, 2020.

MACDONALD, Ian. **Revolution in the Head: The Beatles' Records and the Sixties**. [S.l.]: Chicago Review Press, 2007. ISBN 9781556527333.

MADGWICK, Sebastian et al. Simple synchronisation for open sound control. In: PROCEEDINGS of the International Computer Music Conference. [S.l.: s.n.], 2015.

MANOVICH, Lev. Remixability and Modularity. unpublished. [S.l.], 2005.

_____. What Comes After Remix? unpublished. [S.l.], 2007.

MAROIS, René; IVANOFF, Jason. Capacity limits of information processing in the brain. **Trends in cognitive sciences**, Elsevier, v. 9, n. 6, p. 296–305, 2005.

MARSAN, Carolyn. **The Internet of Things: An Overview**. [S.l.]: Internet Society, 2015.

MARSHALL, Mark; WANDERLEY, Marcelo. Examining the effects of embedded vibrotactile feedback on the feel of a digital musical instrument. In: PROCEEDINGS of the International Conference on New Interfaces for Musical Expression. Oslo, Norway: NIME, 2011. P. 399–404.

MASSI, R Wood. **Music and Discourse: Toward a Semiology of Music**. [S.l.]: JSTOR, 1992.

MATTERN, Friedemann; FLOERKEMEIER, Christian. From the Internet of Computers to the Internet of Things. In: v. 33, p. 242–259. ISBN 978-3-642-17225-0. DOI: 10.1007/978-3-642-17226-7_15.

MATTOS, Douglas et al. Assessing Mulsemedia Authoring Application Based on Events With STEVE 2.0. **IEEE Access**, v. 13, p. 100970–100986, 2025. DOI: 10.1109/ACCESS.2025.3576167.

MATUSZEWSKI, Benjamin. A web-based framework for distributed music system research and creation. **AES-Journal of the Audio Engineering Society Audio-Accoustics-Application**, 2020.

MAYER, Richard E. Elements of a science of e-learning. **Journal of educational computing research**, SAGE Publications Sage CA: Los Angeles, CA, v. 29, n. 3, p. 297–313, 2003.

MCDOWELL, John; FURLONG, Dermont J. Haptic-Listening and the Classical Guitar. In: PROCEEDINGS of the International Conference on New Interfaces for Musical Expression. Virginia, USA: NIME, 2018. P. 293–298.

MCKNIGHT, Patrick E; NAJAB, Julius. Mann-whitney U test. **The Corsini encyclopedia of psychology**, Wiley Online Library, p. 1–1, 2010.

MED, Bohumil. **Teoria da música**. [S.l.]: BrasıĺIa: Musimed, 1996. v. 996.

MEHTA, Amardeep et al. Calvin constrained—A framework for IoT applications in heterogeneous environments. In: IEEE. PROCEEDINGS of the IEEE 37th International Conference on Distributed Computing Systems (ICDCS). [S.l.: s.n.], 2017. P. 1063–1073.

MEIXNER, Britta. Hypervideos and interactive multimedia presentations. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 50, n. 1, p. 1–34, 2017.

MEIXNER, Britta; EINSIEDLER, Christoph. Download and Cache Management for HTML5 Hypervideo Players. In: PROCEEDINGS of the 27th ACM Conference on Hypertext and Social Media. Halifax, Nova Scotia, Canada: Association for Computing Machinery, 2016. (HT '16), p. 125–136. ISBN 9781450342476. DOI: 10.1145/2914586.2914587. Available from: <https://doi.org/10.1145/2914586.2914587>.

MELE, Cristina et al. The phygital transformation: a systematic review and a research agenda. **Italian Journal of Marketing**, Springer, v. 2023, n. 3, p. 323–349, 2023.

MERENDINO, Nicoló; RODÀ, Antonio; MASU, Raul. "Below 58 BPM", involving real-time monitoring and self-medication practices in music performance through IoT technology. **Frontiers in Computer Science**, Frontiers Media SA, v. 6, p. 1187933, 2024.

MERRIAM, Alan P. Ethnomusicology discussion and definition of the field. **Ethnomusicology**, JSTOR, v. 4, n. 3, p. 107–114, 1960.

MEYER, Leonard B. **Emotion and meaning in music**. [S.l.]: University of chicago Press, 2008.

MILGRAM, Paul et al. Augmented reality: A class of displays on the reality-virtuality continuum. In: SPIE. TELEMANIPULATOR and telepresence technologies. [S.l.: s.n.], 1995. v. 2351, p. 282–292.

MOLINO, Jean. Facto musical e semiologia da música. **Semiologia da música**, p. 109–64, 1975.

MONKS, John et al. Quality of experience assessment of 3D video synchronised with multisensorial media components. In: IEEE. PROCEEDINGS of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). [S.l.: s.n.], 2017. P. 1–6.

MONTAGU, Jeremy. How music and instruments began: A brief overview of the origin and entire development of music, from its earliest stages. **Frontiers in Sociology**, Frontiers Media SA, v. 2, p. 8, 2017.

MORAES, Luan Carlos de Oliveira. **A Física da Música no Renascimento: Uma ABordagem Histórico-Epistemológica**. 2010. Master thesis – Universidade de São Paulo.

MULDER, Axel. Virtual musical instruments: Accessing the sound synthesis universe as a performer. In: PROCEEDINGS of the 1st Brazilian Symposium on Computer Music. [S.l.: s.n.], 1994. P. 243–250.

MYSTAKIDIS, Stylianos. Metaverse. **Encyclopedia**, MDPI, v. 2, n. 1, p. 486–497, 2022.

NANAYAKKARA, Suranga Chandima et al. Enhancing musical experience for the hearing-impaired using visual and haptic displays. **Human–Computer Interaction**, Taylor & Francis, v. 28, n. 2, p. 115–160, 2013.

NARDIM, Thaise. **Allan Kaprow, performance e colaboração: estrategias para abraçar a vida como potencia criativa**. Aug. 2009. S. 1–152. PhD thesis – Universidade de Campinas - UNICAMP.

NAUMAN, Ali et al. Multimedia Internet of Things: A comprehensive survey. **Ieee Access**, IEEE, v. 8, p. 8202–8250, 2020.

NAVAS, Eduardo. **Remix Theory: The Aesthetics of Sampling**. 1st. [S.l.]: Ambra Verlag, 2012.

NETTL, Bruno. **The study of ethnomusicology: Twenty-nine issues and concepts**. [S.l.]: University of Illinois Press, 1983.

NEUSTAEDTER, Carman; SENGERS, Phoebe. Autobiographical Design in HCI Research: Designing and Learning through Use-It-Yourself. In: PROCEEDINGS of the Designing Interactive Systems Conference. Newcastle Upon Tyne, United Kingdom: Association for Computing Machinery, 2012. (DIS '12), p. 514–523. ISBN 9781450312103. DOI: 10.1145/2317956.2318034.

NIELSEN, Jakob; MOLICH, Rolf. Heuristic evaluation of user interfaces. In: PROCEEDINGS of the SIGCHI conference on Human factors in computing systems. [S.l.: s.n.], 1990. P. 249–256.

O'MODHRAIN, Sile. A Framework for the Evaluation of Digital Musical Instruments. **Computer Music Journal**, v. 35, n. 1, p. 28–42, 2011. DOI: 10.1162/COMJ_a_00038.

O'NEILL, Eoghan; ORTIZ, Miguel. From Prototype to Performance Practice: Reflections on Iterative Instrument Design. In: PROCEEDINGS of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures. [S.l.: s.n.], 2024. P. 439–444.

O'DAIR, Marcus; BEAVEN, Zuleika. The networked record industry: How blockchain technology could transform the record industry. **Strategic Change**, Wiley Online Library, v. 26, n. 5, p. 471–480, 2017.

O'MODHRAIN, Sile; CHAFE, Chris. Incorporating haptic feedback into interfaces for music applications. In: PROCEEDINGS of the International Symposium on Robotics with Applications, World Automation Conference. Stanford, USA: Stanford University, 2000. P. 1–6.

OKADA, Koh et al. Virtual Drum: Ubiquitous and playful drum playing. In: IEEE. 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE). [S.l.: s.n.], 2014. P. 419–421.

OLIVEIRA, Andreia Machado. Potências de agir implicadas na Arte Interativa. **Visualidades**, v. 13, n. 2, 2015.

ONDERDIJK, Kelsey E; ACAR, Freya; VAN DYCK, Edith. Impact of lockdown measures on joint music making: playing online and physically together. **Frontiers in psychology**, Frontiers Media SA, v. 12, p. 642713, 2021.

ORSSO, Bianca et al. Pensar poético e fazer estético: a generalização dos processos de criação de uma instalação de arte híbrida e interativa. In. DOI: 10.5281/zenodo.7489812.

PARIKKA, Jussi. Ethologies of software art: What can a digital body of code do? **Deleuze and contemporary art**, Edinburgh University Press Edinburgh, p. 116–132, 2010.

PARSONS, Jim. Remix: Making Art and Commerce Thrive in the Hybrid Economy. **Journal of Teaching and Learning**, v. 7, Feb. 2010. DOI: 10.22329/jtl.v7i1.679.

PATEL, Aniruddh D. **Music, language, and the brain**. [S.l.]: Oxford university press, 2010.

PATRÍCIO, Pedro. MuDI-Multimedia Digital Instrument for Composing and Performing Digital Music for Films in Real-time. In: PROCEEDINGS of the International Conference on New Interfaces for Musical Expression. [S.l.: s.n.], 2012.

PEACOCK, Kenneth. Instruments to perform color-music: Two centuries of technological experimentation. **Leonardo**, The MIT Press, v. 21, n. 4, p. 397–406, 1988.

RAINER, Benjamin et al. Investigating the impact of sensory effects on the quality of experience and emotional response in web videos. In: IEEE. PROCEEDINGS of the 4th International Workshop on Quality of multimedia experience. [S.l.: s.n.], 2012. P. 278–283.

RAJ, Pethuru. Empowering digital twins with blockchain. In: ADVANCES in computers. [S.l.]: Elsevier, 2021. v. 121. P. 267–283.

RAUSCHNABEL, Philipp A et al. What is XR? Towards a framework for augmented and virtual reality. **Computers in human behavior**, Elsevier, v. 133, p. 107289, 2022.

RAZUMIEJCZYK, Eugenia; MACBETH, Guillermo. Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. **Universitas Psychologica**, 2011.

REFSUM JENSENIUS, Alexander; LYONS, Michael J. (Eds.). **A NIME Reader: Fifteen Years of New Interfaces for Musical Expression**. [S.l.]: Springer Cham, 2017. v. 3. (Current Research in Systematic Musicology). ISBN 978-3-319-47214-0. DOI: 10.1007/978-3-319-47214-0.

REPP, Bruno H. Sensorimotor synchronization: A review of the tapping literature. **Psychonomic bulletin & review**, Springer, v. 12, p. 969–992, 2005.

REY, Denise; NEUHÄUSER, Markus. Wilcoxon-signed-rank test. In: INTERNATIONAL encyclopedia of statistical science. [S.l.]: Springer, 2011. P. 1658–1659.

ROCHA, Carlos Gustavo Araújo da; SOUZA FILHO, Guido Lemos de. Um Framework para provisão de Qualidade de Serviço em redes IP. **UFRN, jun**, 2001.

RODRIGUES, Renato et al. A proposal for supporting sensory effect rendering in ginga-ncl. In: PROCEEDINGS of the 25th Brazillian Symposium on Multimedia and the Web. [S.l.: s.n.], 2019. P. 273–280.

ROMANI, Michele et al. BCHJam: a Brain-Computer Music Interface for Live Music Performance in Shared Mixed Reality Environments. In: IEEE. PROCEEDINGS of the 2024 IEEE 5th International Symposium on the Internet of Sounds (IS2). [S.l.: s.n.], 2024. P. 1–9.

ROSENTHAL, Robert; COOPER, Harris; HEDGES, Larry, et al. Parametric measures of effect size. **The handbook of research synthesis**, New York, v. 621, n. 2, p. 231–244, 1994.

ROSSI, Silvia. **Understanding user interactivity for the next-generation immersive communication: design, optimisation, and behavioural analysis**. 2022. PhD thesis – University College London (UCL).

ROTHSTEIN, Joseph. **MIDI: A comprehensive introduction**. [S.l.]: AR Editions, Inc., 1992. v. 7.

ROTTONDI, Cristina et al. An Overview on Networked Music Performance Technologies. **IEEE Access**, v. 4, p. 8823–8843, 2016. DOI: 10.1109/ACCESS.2016.2628440.

ROY, William G; DOWD, Timothy J. What is sociological about music? **Annual Review of Sociology**, Annual Reviews, v. 36, n. 1, p. 183–203, 2010.

SAID, Sadiq et al. Validation of the raw national aeronautics and space administration task load index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: pooled analysis study using mixed models. **Journal of medical Internet research**, JMIR Publications Toronto, Canada, v. 22, n. 9, e19472, 2020.

SALEME, Estevão; FALBO, Celso Santos Ricardo, et al. Towards a reference ontology on mulsemedia systems. In: PROCEEDINGS of the 10th International Conference on Management of Digital EcoSystems. [S.l.: s.n.], 2018. P. 23–30.

SALEME, Estevão; SANTOS, Celso; GHINEA, George. A mulsemedia framework for delivering sensory effects to heterogeneous systems. **Multimedia Systems**, Springer, v. 25, p. 421–447, 2019.

SALEME, Estevão; SANTOS, Celso, et al. MulseOnto: a Reference Ontology to Support the Design of Mulsemedia Systems. **J. Univers. Comput. Sci.**, v. 25, n. 13, p. 1761–1786, 2019.

SALEME, Estêvão; SANTOS, Celso; GHINEA, George. A Conceptual Architecture and a Framework for Dealing with Variability in Mulsemedia Systems. In: EXTENDED Proceedings of the Brazilian Symposyum on Multimedia Systems and Web. São Luís: SBC, 2020. P. 5–8. DOI: 10.5753/webmedia_estendido.2020.13052. Available from: <https://sol.sbc.org.br/index.php/webmedia%5C_estendido/article/view/13052>.

SANT'ANNA, Francisco; CERQUEIRA, Renato; SOARES, Luiz Fernando Gomes. NCLua: objetos imperativos lua na linguagem declarativa NCL. In: PROCEEDINGS of the 14th Brazilian Symposium on Multimedia and the Web. [S.l.: s.n.], 2008. P. 83–90.

SANTOS, Danilo; ALMEIDA, Hyggo; PERKUSICH, Angelo. Uma arquitetura para a internet das coisas aplicada a sistemas de saúde conectada. In: SBC. PROCEEDINGS of the Brazilian Symposium on Ubiquitous and Pervasive Computing. [S.l.: s.n.], 2013. P. 1952–1961.

SANTOS, Lidiane; SILVA, Eduardo, et al. An architectural style for internet of things systems. In: PROCEEDINGS of the 35th Annual ACM Symposium on applied computing. [S.l.: s.n.], 2020. P. 1488–1497.

SATHIYAMURTHY, Suji et al. Measuring Haptic Experience: Elaborating the HX model with scale development. In: IEEE. PROCEEDINGS of the 2021 IEEE World Haptics Conference (WHC). [S.l.: s.n.], 2021. P. 979–984.

SCHAFER, R Murray. **O ouvido pensante**. [S.l.]: Unesp, 1992.

SCHIAVONI, Flávio; QUEIROZ, Marcelo; IAZZETTA, Fernando. Medusa - A Distributed Sound Environment. In: UNIVERSIDADE de São Paulo. [S.l.: s.n.], Aug. 2011.

SCHIAVONI, Flávio; QUEIROZ, Marcelo; WANDERLEY, Marcelo. Network Music With Medusa: A Comparison Of Tempo Alignment In Existing MIDI APIs. In: PROCEEDINGS of the Sound and Music Conference. [S.l.: s.n.], Aug. 2013.

SCHMEDER, Andrew; FREED, Adrian; WESSEL, David. Best practices for open sound control. In: PROCEEDINGS of the Linux Audio Conference. [S.l.: s.n.], 2010. v. 10.

SCHNEIDER, Oliver; MACLEAN, Karon, et al. Haptic experience design: What hapticians do and where they need help. **International Journal of Human-Computer Studies**, Elsevier, v. 107, p. 5–21, 2017.

SCHNEIDER, Oliver S; MACLEAN, Karon E. Studying design process and example use with Macaron, a web-based vibrotactile effect editor. In: PROCEEDINGS of the 2016 IEEE Haptics Symposium (Haptics). New York, USA: IEEE, 2016. P. 52–58.

SEIFI, Hasti; MACLEAN, Karon E. Exploiting haptic facets: Users' sensemaking schemas as a path to design and personalization of experience. **International Journal of Human-Computer Studies**, Elsevier, v. 107, p. 38–61, 2017.

SERAFIN, Stefania; ERKUT, Cumhur; KOJS, Juraj; NILSSON, Niels C, et al. Virtual reality musical instruments: State of the art, design principles, and future directions. **Computer Music Journal**, MIT Press, v. 40, n. 3, p. 22–40, 2016.

SERAFIN, Stefania; ERKUT, Cumhur; KOJS, Juraj; NORDAHL, Rolf, et al. Virtual Reality Musical Instruments: Guidelines for Multisensory Interaction Design. In: PROCEEDINGS of the Audio Mostly 2016. Norrköping, Sweden: Association for Computing Machinery, 2016. P. 266–271. ISBN 9781450348225. DOI: 10.1145/2986416.2986431. Available from: <https://doi.org/10.1145/2986416.2986431>.

SEVINC, Volkan; BERKMAN, Mehmet Ilker. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. **Applied ergonomics**, Elsevier, v. 82, p. 102958, 2020.

SHAHAB, Muhammad Hamza; GHAZALI, Ezlika; MOHTAR, Mozard. The role of elaboration likelihood model in consumer behaviour research and its extension to new technologies: A review and future research agenda. **International Journal of Consumer Studies**, Wiley Online Library, v. 45, n. 4, p. 664–689, 2021.

SILVA, Gustavo et al. A questionnaire for measuring presence in virtual environments: factor analysis of the presence questionnaire and adaptation into Brazilian Portuguese. **Virtual Reality**, Springer, v. 20, n. 4, p. 237–242, 2016.

SILVA, Nishal Stanislaus. **Embedded Real-Time Musical Pattern Detection for Smart Musical Instruments**. Jan. 2025. PhD thesis – Università degli studi di Trento, Trento, Italy.

SLOBODA, John A. Everyday uses of music listening: A preliminary study. **Music, mind and science**, v. 354, p. 369, 1999.

SMITH, Wesley. Real-time multimedia composition using lua. **Proceedings of the Digital Art Weeks 2007**, 2007.

SOARES, Luiz Fernando G; MORENO, Marcio F; MARINHO, Rafael S. Ginga-NCL architecture for plug-ins. **Software: Practice and Experience**, Wiley Online Library, v. 43, n. 4, p. 449–463, 2013.

SOARES, Luiz Fernando Gomes. **Programando Em NCL**. [S.l.]: Campus, Jan. 2009. P. 360. ISBN 978-8535234572.

SOARES, Luiz Fernando Gomes; MORENO, Marcio Ferreira; NETO, Carlos de Salles Soares, et al. Ginga-NCL: Declarative middleware for multimedia IPTV services. **IEEE Communications Magazine**, IEEE, v. 48, n. 6, p. 74–81, 2010.

SONTAG, Susan. **Against interpretation: And other essays**. [S.l.]: Macmillan, 2001.

SOUSA JUNIOR, José Geraldo de et al. Estendendo NCL: objetos NCLua como exibidores para novos tipos de mídia. In: SBC. EXTENDED Proceedings of the Brazilian Symposium on Multimedia Systems and Web. [S.l.: s.n.], 2010. P. 214–219.

SPENCE, Jocelyn; FROHLICH, David; ANDREWS, Stuart. Performative experience design: where autobiographical performance and human–computer interaction meet. **Digital Creativity**, Taylor & Francis, v. 24, n. 2, p. 96–110, 2013.

STADLER, Sebastian et al. A tool, not a toy: using virtual reality to evaluate the communication between autonomous vehicles and pedestrians. In: AUGMENTED reality and virtual reality: the power of ar and vr for business. [S.l.]: Springer, 2019. P. 203–216.

STEINMETZ, Ralf; NAHRSTEDT, Klara. **Multimedia fundamentals, Volume 1: Media coding and content processing**. [S.l.]: Pearson Education, 2002.

STROHMEIER, Paul et al. BARefoot: Generating virtual materials using motion coupled vibration in shoes. In: PROCEEDINGS of the 33rd Annual ACM Symposium on User Interface Software and Technology. New York, USA: ACM Digital Library, 2020. P. 579–593.

SU, Zhong; YANG, Qiang; ZHANG, Hong-Jiang. A Prediction System for Multimedia Pre-Fetching in Internet. In: PROCEEDINGS of the Eighth ACM International Conference on Multimedia. Marina del Rey, California, USA: Association for Computing Machinery, 2000. (MULTIMEDIA '00), p. 3–11. ISBN 1581131984. DOI: 10.1145/354384.354394. Available from: <https://doi.org/10.1145/354384.354394>.

SUEN, Rax Chun Lung et al. 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. [S.l.: s.n.], 2015. P. 703–708.

SURASINGHE, Pabasara; HERATH, Pasan; THANIKASALAM, Kokul. An Efficient Real-Time Air Drumming Approach Using MediaPipe Hand Gesture Model. In: IEEE. PROCEEDINGS of the 5th International Conference on Advancements in Computing (ICAC). [S.l.: s.n.], 2023. P. 215–220.

TAMPLIN, Jeanette et al. Development and feasibility testing of an online virtual reality platform for delivering therapeutic group singing interventions for people living with spinal cord injury. **Journal of Telemedicine and Telecare**, v. 26, n. 6, p. 365–375, 2020. DOI: 10.1177/1357633X19828463.

TAVAKOL, Mohsen; DENNICK, Reg. Making sense of Cronbach's alpha. **International Journal of Medical Education**, v. 2, p. 53–55, 2011. DOI: 10.5116/ijme.4dfb.8dfd.

THALMANN, Florian et al. The Mobile Audio Ontology: Experiencing Dynamic Music Objects on Mobile Devices. In: ROCEEDINGS of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC). [S.l.: s.n.], 2016. P. 47–54. DOI: 10.1109/ICSC.2016.61.

TOLENTINO, Carl Timothy; UY, Agatha; NAVAL, Prospero. Air Drums: Playing Drums Using Computer Vision. In: IEEE. PROCEEDINGS of the 2019 International Symposium on Multimedia and Communication Technology (ISMAC). [S.l.: s.n.], 2019. P. 1–6.

TRAGTENBERG, João; ALBUQUERQUE, Gabriel; CALEGARIO, Filipe. Gambiarra and Techno-Vernacular Creativity in NIME Research. In: PROCEEDINGS of the International Conference on the New Interfaces for Musical Expression. [S.l.: s.n.], Apr. 2021. https://nime.pubpub.org/pub/aqm27581.

TURCHET, Luca. Musical Metaverse: vision, opportunities, and challenges. **Personal and Ubiquitous Computing**, Springer, v. 27, n. 5, p. 1811–1827, 2023.

_____. Smart Mandolin: Autobiographical Design, Implementation, Use Cases, and Lessons Learned. In: PROCEEDINGS of the Audio Mostly 2018 on Sound in Immersion and Emotion. Wrexham, United Kingdom: Association for Computing Machinery, 2018. (AM'18). ISBN 9781450366090. DOI: 10.1145/3243274.3243280. Available from: <https://doi.org/10.1145/3243274.3243280>.

_____. Smart Musical Instruments: vision, design principles, and future directions. **IEEE Access**, IEEE, v. 7, p. 8944–8963, 2018.

_____. Some reflections on the relation between augmented and smart musical instruments. In: PROCEEDINGS of the Audio Mostly 2018 on Sound in Immersion and Emotion. [S.l.]: ACM Digital Library, 2018. P. 1–7.

TURCHET, Luca; ANTONIAZZI, Francesco, et al. The Internet of Musical Things Ontology. **Journal of Web Semantics**, v. 60, p. 100548, 2020. ISSN 1570-8268. DOI: https://doi.org/10.1016/j.websem.2020.100548.

TURCHET, Luca; BARTHET, Mathieu. Jamming with a Smart Mandolin and Freesound-based Accompaniment. In: PROCEEDINGS of the 2018 23rd Conference of Open Innovations Association (FRUCT). [S.l.: s.n.], Nov. 2018. DOI: 10.23919/FRUCT.2018.8588110.

TURCHET, Luca; BENINCASO, Michele; FISCHIONE, Carlo. Examples of use cases with smart instruments. In: PROCEEDINGS of the 12th international audio mostly conference on augmented and participatory sound and music experiences. [S.l.: s.n.], 2017. P. 1–5.

TURCHET, Luca; CASARI, Paolo. The Internet of Musical Things Meets Satellites: Evaluating Starlink Support for Networked Music Performances in Rural Areas. In: IEEE. 2024 IEEE 5th International Symposium on the Internet of Sounds (IS2). [S.l.: s.n.], 2024. P. 1–8.

TURCHET, Luca; FAZEKAS, György, et al. The Internet of Audio Things: State of the Art, Vision, and Challenges. **IEEE Internet of Things Journal**, v. 7, n. 10, p. 10233–10249, 2020. DOI: 10.1109/JIOT.2020.2997047.

TURCHET, Luca; FISCHIONE, Carlo, et al. Internet of Musical Things: Vision and Challenges. **IEEE Access**, v. 6, p. 61994–62017, 2018. DOI: 10.1109/ACCESS.2018.2872625.

TURCHET, Luca; HAMILTON, Rob; ÇAMCI, Anil. Music in Extended Realities. **IEEE Access**, v. 9, p. 15810–15832, 2021. DOI: 10.1109/ACCESS.2021.3052931.

TURCHET, Luca; LAGRANGE, Mathieu, et al. The Internet of Sounds: Convergent Trends, Insights, and Future Directions. **IEEE Internet of Things Journal**, v. 10, n. 13, p. 11264–11292, 2023. DOI: 10.1109/JIOT.2023.3253602.

TURCHET, Luca; MCPHERSON, Andrew; BARTHET, Mathieu, et al. Co-design of a Smart Cajón. In: 4. AES. [S.l.: s.n.], 2018. v. 66, p. 220–230.

TURCHET, Luca; MCPHERSON, Andrew; BARTHET, Mathieu. Real-time hit classification in a Smart Cajón. **Frontiers in ICT**, Frontiers Media SA, v. 5, p. 16, 2018.

TURCHET, Luca; MCPHERSON, Andrew; FISCHIONE, Carlo, et al. Smart instruments: Towards an ecosystem of interoperable devices connecting performers and audiences. In: PROCEEDINGS of Sound and Music Computing Conference. [S.l.: s.n.], 2016. P. 498–505.

TURCHET, Luca; NGO, Chan Nam. Blockchain-based Internet of Musical Things. **Blockchain: Research and Applications**, v. 3, n. 3, p. 100083, 2022. ISSN 2096-7209. DOI: https://doi.org/10.1016/j.bcra.2022.100083.

TURCHET, Luca; ROSAIA, Raffaele, et al. Exposure to vibrotactile music improves audiometric performances in individuals with cochlear implants. **Scientific Reports**, Nature Publishing Group UK London, v. 15, n. 1, p. 20054, 2025.

TURCHET, Luca; ROTTONDI, Cristina. On the relation between the fields of Networked Music Performances, Ubiquitous Music, and Internet of Musical Things. **Personal and Ubiquitous Computing**, Oct. 2022. DOI: 10.1007/s00779-022-01691-z.

TURCHET, Luca; VIOLA, Fabio, et al. Towards a Semantic Architecture for the Internet of Musical Things. In: IEEE Open Innovations Association. [S.l.: s.n.], Nov. 2018. DOI: 10.23919/FRUCT.2018.8587917.

TURCHET, Luca; WEST, Travis; WANDERLEY, Marcelo M. Smart mandolin and musical haptic gilet: effects of vibro-tactile stimuli during live music performance. In: PROCEEDINGS of the 14th International Audio Mostly Conference: A Journey in Sound. New York, USA: ACM Digital Library, 2019. P. 168–175.

_____. Touching the audience: musical haptic wearables for augmented and participatory live music performances. **Personal and Ubiquitous Computing**, Springer, v. 25, n. 4, p. 749–769, 2021.

VALBOM, Leonel; MARCOS, Adérito. WAVE: Sound and music in an immersive environment. **Computers & Graphics**, Elsevier, v. 29, n. 6, p. 871–881, 2005.

VENKATESAN, Tara; WANG, Qian Janice. Feeling connected: the role of haptic feedback in VR concerts and the impact of haptic music players on the music listening experience. In: ARTS. Basel, Switzerland: MDPI, 2023. v. 12, p. 148.

VIEIRA, Daniel; PROVIDÊNCIA, Bernardo; CARVALHO, Helder. Design of a smart garment for fencing: measuring attractiveness using the AttrakDiff Mini method. **Human-Intelligent Systems Integration**, Springer, v. 5, n. 1, p. 1–9, 2023.

VIEIRA, Rômulo; BARTHET, Mathieu; SCHIAVONI, Flávio Luiz. Everyday Use of the Internet of Musical Things: Intersections with Ubiquitous Music. In: PROCEEDINGS of the Workshop on Ubiquitous Music 2020. Porto Seguro, BA, Brasil: Zenodo, Nov. 2020. P. 60–71. DOI: 10.5281/zenodo.4247759.

VIEIRA, Rômulo; GONÇALVES, Luan; SCHIAVONI, Flávio. The things of the Internet of Musical Things: defining the difficulties to standardize the behavior of these devices. In: 2020 X Brazilian Symposium on Computing Systems Engineering (SBESC). [S.l.: s.n.], 2020. P. 1–7. DOI: 10.1109/SBESC51047.2020.9277862.

VIEIRA, Rômulo; MUCHALUAT SAADE, Débora C.; ROCHA, Marcelo, et al. RemixDrum: A Smart Musical Instrument for Music and Visual Art Remix. In: 2023 IEEE 9th World Forum on Internet of Things (WF-IoT). [S.l.: s.n.], 2023. P. 1–7. DOI: 10.1109/WF-IoTXXX.

VIEIRA, Rômulo; MUCHALUAT SAADE, Débora Christina; CÉSAR, Pablo. PhysioDrum: Bridging Physical and Digital Realms in Immersive Musical Interaction. In: PROCEEDINGS of the 2025 ACM International Conference on Interactive Media Experiences. New York, USA: ACM Digital Library, 2025. P. 356–358.

VIEIRA, Rômulo; SAADE, Débora C. Muchaluat; CÉSAR, Pablo. Exploring Artificial Intelligence for Advancing Performance Processes and Events in Io3MT. In: SPRINGER. INTERNATIONAL Conference on Multimedia Modeling. [S.l.: s.n.], 2024. P. 234–248.

_____. Internet of Multisensory, Multimedia and Musical Things (Io3MT): Framework Design, Use Cases, and Analysis. In: PROCEEDINGS of the 2025 ACM International Conference on Interactive Media Experiences. [S.l.: s.n.], 2025. P. 484–487.

_____. Towards an Internet of Multisensory, Multimedia and Musical Things (Io3MT) Environment. In: PROCEEDINGS of the 4th International Symposium on the Internet of Sounds. Pisa, Italy: IEEE, 2023. (IS2 '23), p. 231–238. DOI: 10.1145/3561212.3561246. Available from: <https://doi.org/10.1145/3561212.3561246>.

VIEIRA, Rômulo; SCHIAVONI, Flávio; SAADE, Débora Muchaluat. Sunflower: a proposal for standardization on the Internet of Musical Things environments. In: PROCEEDINGS of the Brazilian Symposium on Computer Networks and Distributed Systems. Fortaleza/CE: SBC, 2022. P. 25–32. DOI: 10.5753/sbrc_estendido.2022.222447.

VIEIRA, Rômulo; WEI, Shu, et al. Immersive Io3MT Environments: Design Guidelines, Use Cases and Future Directions. In: IEEE. PROCEEDINGS of the IEEE 5th International Symposium on the Internet of Sounds (IS2). [S.l.: s.n.], 2024. P. 1–10.

VIRTANEN, Kai et al. Weight watchers: NASA-TLX weights revisited. **TheoreTical issues in ergonomics science**, Taylor & Francis, v. 23, n. 6, p. 725–748, 2022.

VUUST, Peter et al. Music in the brain. **Nature Reviews Neuroscience**, Nature Publishing Group UK London, v. 23, n. 5, p. 287–305, 2022.

WALTL, Markus; TIMMERER, Christian; HELLWAGNER, Hermann. Improving the quality of multimedia experience through sensory effects. In: IEEE. PROCEEDINGS of the 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX). [S.l.: s.n.], 2010. P. 124–129.

WALTL, Markus; TIMMERER, Christian; RAINER, Benjamin, et al. Sensory Effects for Ambient Experiences in the World Wide Web. **Multimedia Tools and Applications**, v. 70, Aug. 2011. DOI: 10.1007/s11042-012-1099-8.

WANG, Ge. Principles of visual design for computer music. In: PROCEEDINGS of the International Computer Music Conference (ICMC). [S.l.: s.n.], 2014.

WEINBERG, Gil. The aesthetics, history and future challenges of interconnected music networks. In: PROCEEDINGS of the International Computer Music Conference. [S.l.: s.n.], 2002.

WEINEL, Jonathan. Cyberdreams: visualizing music in extended reality. In: TECHNOLOGY, Design and the Arts-Opportunities and Challenges. [S.l.]: Springer International Publishing Cham, 2020. P. 209–227.

WILLEMSEN, Silvin; HORVATH, Anca-Simona; NASCIMBEN, Mauro. DigiDrum: a haptic-based virtual reality musical instrument and a case study. In: SMC NETWORK. PROCEEDINGS of the 17th Sound and Music Computing Conference. [S.l.: s.n.], 2020. P. 292–299.

WILSON, Chauncey. **Interview techniques for UX practitioners: A user-centered design method**. [S.l.]: Newnes, 2013.

WINCKEL, Fritz. **Music, sound and sensation: A modern exposition**. [S.l.]: Courier Corporation, 2014.

WITMER, Bob G; JEROME, Christian J; SINGER, Michael J. The factor structure of the presence questionnaire. **Presence: Teleoperators & Virtual Environments**, MIT Press, v. 14, n. 3, p. 298–312, 2005.

WITMER, Bob G; SINGER, Michael J. Measuring presence in virtual environments: A presence questionnaire. **Presence**, MIT Press, v. 7, n. 3, p. 225–240, 1998.

WRIGHT, Matthew. Open Sound Control: an enabling technology for musical networking. **Organised Sound**, Cambridge University Press, v. 10, n. 3, p. 193–200, 2005.

WRIGHT, Matthew et al. Managing complexity with explicit mapping of gestures to sound control with OSC. In: PROCEEDINGS of the International Computer Music Conference. [S.l.: s.n.], 2001.

YADID, Harel et al. A2d: Anywhere anytime drumming. In: IEEE. PROCEEDINGS of the 2023 IEEE Region 10 Symposium (TENSYMP). [S.l.: s.n.], 2023. P. 1–6.

YAOYUNEYONG, Gallayanee Starwind et al. Virtual dressing room media, buying intention and mediation. **Journal of Research in Interactive Marketing**, Emerald Publishing Limited, v. 12, n. 1, p. 125–144, 2018.

YASEEN, Azeema; CHAKRABORTY, Sutirtha; TIMONEY, Joseph. A cooperative and interactive gesture-based drumming interface with application to the Internet of Musical Things. In: SPRINGER. HCI International 2022 Posters: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part II. [S.l.: s.n.], 2022. P. 85–92.

YONKOV, Atanas. **Enhancing the Spectator Experience**. [S.l.: s.n.], 2024.

YOON, Kyoungro. End-to-End Framework for 4-D Broadcasting Based on MPEG-V Standard. **Image Commun.**, Elsevier Science Inc., USA, v. 28, n. 2, p. 127–135, Feb. 2013. ISSN 0923-5965. DOI: 10.1016/j.image.2012.10.008.

YOUNG, Gareth; MURPHY, David; WEETER, Jeffrey. A Qualitative Analysis of Haptic Feedback in Music Focused Exercises. In: PROCEEDINGS of the New Interfaces for Musical Expression. [S.l.: s.n.], May 2017.

YOUNG, Gareth W; O'DWYER, Néill, et al. Feel the music!—audience experiences of audio–tactile feedback in a novel virtual reality volumetric music video. In: MDPI, 4. ARTS. [S.l.: s.n.], 2023. v. 12, p. 156.

YUAN, Zhenhui; BI, Ting, et al. Perceived Synchronization of Mulsemedia Services. **IEEE Transactions on Multimedia**, v. 17, p. 1–1, July 2015. DOI: 10.1109/TMM.2015.2431915.

YUAN, Zhenhui; GHINEA, Gheorghita; MUNTEAN, Gabriel-Miro. Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery. **IEEE Transactions on Multimedia**, IEEE, v. 17, n. 1, p. 104–117, 2014.

ZAR, Jerrold H. Spearman rank correlation. **Encyclopedia of biostatistics**, Wiley Online Library, v. 7, 2005.

ZATORRE, Robert J; SALIMPOOR, Valorie N. From perception to pleasure: music and its neural substrates. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 110, supplement_2, p. 10430–10437, 2013.

ZAVERI, Dishant et al. Aero Drums-Augmented Virtual Drums. In: IEEE. PROCEEDINGS of the 2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA). [S.l.: s.n.], 2022. P. 516–520.

ZELLERBACH, Karitta Christina; ROBERTS, Charlie. A framework for the design and analysis of mixed reality musical instruments. In: PUBPUB. PROCEEDINGS of the International Conference on New Interfaces for Musical Expression. [S.l.: s.n.], 2022.

ZHENG, Lirong et al. Technologies, applications, and governance in the internet of things. In: INTERNET of things-Global technological and societal trends from smart environments and spaces to green ICT. [S.l.]: River Publishers, 2022. P. 143–177.

ZHU, Hua et al. A survey of quality of service in IEEE 802.11 networks. **IEEE Wireless Communications**, v. 11, n. 4, p. 6–14, 2004. DOI: 10.1109/MWC.2004.1325887.

ZHUO, Anjing; SIRIVESMAS, Veerawat; PUNYALIKIT, Ruenglada. Application Research of Installation Art Based on Digital Media. In: ATLANTIS PRESS. 2ND International Conference on Intelligent Design and Innovative Technology (ICIDIT 2023). [S.l.: s.n.], 2023. P. 289–297.

APPENDIX A – RemixDrum: Semi-structured Interview

1. **How would you describe the ease of activation of the RemixDrum during testing?**

   The activation process was straightforward, with sensors responding immediately to my touches and gestures. Minimal physical force was required, which I considered advantageous. However, I observed that in certain situations, particularly during very rapid intensity variations, there was a slight delay between my action and the corresponding auditory response. This latency was not noticeable in the visual component.

2. **Were there any gestures or movements that worked better or worse than you expected?**

   The more conventional and traditional gestures performed well. However, more subtle movements, such as ghost notes or minor variations, were not captured with the same degree of precision. I am accustomed to exploring fine nuances of attack and dynamics, and I felt that the system does not yet fully convey this richness. While this did not compromise the overall experience, it did limit the exploration of more sophisticated repertoires.

3. **How natural or intuitive was the initial interaction with the device?**

   The interaction was highly intuitive, as it preserved the fundamental logic of using drumsticks. What required adaptation was understanding which movements would trigger particular sonic effects. Over time, however, I became accustomed to these mappings and began adapting my playing style to the pre-recorded track.

4. **How expressive were the sounds and artistic outcomes within the environment?**

   The sounds produced were convincing. I perceived the system as a sort of "pedal stom for drummers", enabling the generation of new sounds through gestures and thereby expanding the sonic palette. This was the feature I appreciated the most.

5. **To what extent did the RemixDrum meet your creative or artistic needs?**

   The device substantially met my creative needs, particularly in terms of expanding sonic possibilities beyond those of a traditional acoustic drum kit. It enabled the exploration of novel timbres, the creation of hybrid effects, and even the simulation of situations that would be challenging to reproduce with conventional instruments.

6. **Did you feel that the outcomes stimulated your creativity?**

   Yes, no doubts. During use, I felt encouraged to explore new combinations of gestures and timbres—possibilities that would not typically emerge with an acoustic drum kit. This shift in perspective fostered creative expansion.

7. **How did you perceive the relationship between your gestures and the outcomes obtained?**

   Initially, it took me some time to clearly perceive the relationship between movement and sound. With practice, however, this relationship became more evident. Regarding the visual component, I was less concerned—even if it were somewhat random, I believe it would still be aesthetically interesting. My primary focus was on the auditory response.

8. **What musical or technical skills did you consider essential for using the device?**

   Fundamentally, the same skills required for playing a traditional drum kit. A basic degree of technological familiarity would be beneficial, particularly for understanding the behavior of the accelerometer and related features.

9. **What was your perception of ergonomics and physical comfort while using the device?**

   Overall, the ergonomics were satisfactory. However, certain discomfort arose from the physical arrangement of the circuitry and the presence of cables. In longer sessions, I believe this could become a significant issue.

10. **What would most motivate you to use this system in a real performance?**

    The primary motivation would be the possibility of expanding the sonic repertoire and enriching a performance with digital elements.

11. **Would you recommend the RemixDrum to other drummers? Why?**

    Yes. The system complements a drum kit effectively, and I particularly appreciated the ability to control visual animations, even when done somewhat randomly. I would also recommend it to musicians with an interest in technology.

12. **Which factors (effort required, compatibility with your repertoire, innovation) would weigh most in your decision to adopt this system?**

    The most important factors would be innovation and compatibility with my repertoire. If the device offered sounds and textures that aligned with the styles I perform, I would certainly integrate it into live performances. The effort required for use was minimal, which facilitates adoption, but the key to continued use lies in its artistic relevance and its capacity to add genuinely novel elements to my performances.

13. **Were there any frustrating or limiting aspects in your experience?**

The most frustrating aspect was the limitation imposed by the cables and circuit positioning, which reduced the naturalness of my movements and caused some discomfort, as I had to grip the drumstick higher than usual.

14. **What suggestions for improvement would you make to enhance the device?**

I would recommend investing in a more ergonomic and wireless design to increase gestural freedom. Additionally, greater flexibility in the customization of mappings and audio tracks would be beneficial.

APPENDIX B – Statistical Methods

The application of appropriate statistical methods in experimental research can ensure the reliability of findings and the validity of inferences. Such methods strengthen the methodological rigor of the study, allowing for a more precise evaluation of participants' subjective perceptions as captured through quantitative questionnaires.

In this context, the Mann–Whitney test (MCKNIGHT; NAJAB, 2010) was selected as a nonparametric statistical method suitable for comparing two independent samples when the assumptions of normality and homoscedasticity are not satisfied. This procedure determines whether the values in one sample tend to be systematically higher (or lower) than those in the other sample, based on the ranks assigned to the combined dataset. The calculations for this test are presented in Equation B.1, where n_1 and n_2 denote the sample sizes, and R_1 represents the sum of ranks for the first group after the data are jointly ordered. The test statistic corresponds to the smaller value between U_1 and U_2. A result with $p < 0.05$ provides statistical evidence to reject the null hypothesis that the two distributions are equivalent.

$$\begin{aligned} U_1 &= n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \\ U_2 &= n_1 n_2 - U_1 \end{aligned} \tag{B.1}$$

In turn, the Wilcoxon signed-rank test (REY; NEUHÄUSER, 2011) was applied for paired samples in cases where the distribution of differences between pairs did not meet the assumption of normality. This test, expressed in Equation B.2, is based on the differences $D_i = X_i - Y_i$ computed for each pair. Pairs for which $D_i = 0$ are excluded from the analysis. The absolute values $|D_i|$ are then ranked, after which the original signs are reapplied to the corresponding ranks. The positive (T^+) and negative (T^-) ranks are summed separately, and the test statistic W is defined as the smaller of these two sums. As with the Mann–Whitney test, a value of $p < 0.05$ indicates that the differences between conditions are statistically significant, thereby suggesting a systematic effect.

$$W = \min(T^+, T^-) \tag{B.2}$$

Complementing significance testing, effect size (ROSENTHAL; COOPER; HEDGES, et al., 1994) measures provide an estimate of the magnitude of the observed difference, independent of sample

size. The effect size coefficient r is frequently employed alongside the Mann–Whitney and Wilcoxon tests, and is calculated using Equation B.3, where Z denotes the standardized test statistic and N corresponds to the total number of observations. This metric expresses the strength of the association between the variables or conditions under analysis and is commonly interpreted according to the following reference values: $r \approx 0.1$ (small effect), $r \approx 0.3$ (moderate effect), and $r \geq 0.5$ (large effect).

$$r = \frac{Z}{\sqrt{N}} \tag{B.3}$$

Another nonparametric measure of effect size is Cliff's delta (δ) (RAZUMIEJCZYK; MACBETH, 2011), which quantifies the degree of stochastic dominance between two groups. Its computation involves comparing all possible pairwise combinations of elements from groups A and B, determining the proportion of cases in which a value from A is greater than a value from B, and subtracting the inverse proportion (i.e., when a value from B exceeds a value from A). Formally, the calculation is expressed by Equation B.4, where $|x_i > y_j|$ denotes the number of pairs (x_i, y_j) in which the value from group A is greater than that from group B, $|x_i < y_j|$ represents the number of pairs with the inverse relationship, and n_A and n_B correspond to the sample sizes of groups A and B, respectively. The result ranges from -1 to 1, where values near zero indicate substantial overlap between groups, while values approaching the extremes indicate complete dominance of one group over the other. Cliff's delta is particularly useful when data exhibit asymmetries or heterogeneous variability, and its interpretation follows the conventional thresholds: $|\delta| < 0.147$ (small effect), $|\delta| < 0.33$ (moderate effect), and $|\delta| \geq 0.474$ (large effect).

$$\delta = \frac{|x_i > y_j| - |x_i < y_j|}{n_A \times n_B} \tag{B.4}$$

Spearman's rank correlation coefficient (ρ) (ZAR, 2005) is a nonparametric measure used to evaluate the strength and direction of a monotonic association between two ordinal variables or between variables that do not meet the assumptions of normality. In contrast to Pearson's correlation, which measures linear associations and assumes normally distributed data, Spearman's method relies on the ranked values of the observations, evaluating the extent to which one variable exhibits a monotonic increase or decrease in relation to the other variable. The calculation is given in Equation B.5, where d_i denotes the difference between the ranks of paired observations and n represents the total number of pairs. Coefficients with values close to 1 or -1 indicate strong positive or negative associations, respectively, whereas values near 0 suggest the absence of a monotonic relationship.

$$\rho = 1 - \frac{6 \sum d_i^2}{n\,(n^2 - 1)} \tag{B.5}$$

Finally, Cronbach's alpha (α) (TAVAKOL; DENNICK, 2011) is a statistical coefficient used to assess the internal consistency of data collection instruments, assuming that all items measure the same latent construct. Its calculation is shown in Equation B.6, where k is the number of items, σ_i^2 the variance

of each item, and σ_t^2 the total variance of the scale. Values between 0.70 and 0.90 are considered satisfactory, while those below 0.60 suggest low reliability.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right) \tag{B.6}$$

APPENDIX C – Simulator Sickness Questionnaire (SSQ) Overview

C.1 Simulator Sickness Questionnaire Applied in the Research

Question	English Version	Brazilian Portuguese Version
Q1	General discomfort	Mal-estar generalizado
Q2	Fatigue	Cansaço
Q3	Headache	Dor de cabeça
Q4	Eye strain	Vista cansada
Q5	Difficulty focusing	Dificuldade de manter o foco
Q6	Salivation increasing	Aumento de salivação
Q7	Sweating	Sudorese (suor excessivo)
Q8	Nausea	Náusea
Q9	Difficulty concentrating	Dificuldade de concentração
Q10	Fullness of head	Cabeça pesada
Q11	Blurred vision	Visão embaçada
Q12	Dizziness with eyes open	Tontura com olhos abertos
Q13	Dizziness with eyes closed	Tontura com olhos fechados
Q14	Vertigo	Vertigem
Q15	Stomach awareness	Enjoo estomacal
Q16	Burping	Arroto

Table 17: Questions from the Simulator Sickness Questionnaire (SSQ) with Brazilian Portuguese translations.

C.2 Comparison Between Musicians and Non-Musicians

Subscale	Mean (Mus.)	Median (Mus.)	Mode (Mus.)	SD (Mus.)	Mean (Non-Mus.)	Median (Non-Mus.)	Mode (Non-Mus.)	SD (Non-Mus.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	14.31	9.54	0	21.68	8.59	0	0	13.46	114.5	0.50366	0.122	0.145
Oculomotor	16.68	11.37	7.58	17.06	11.75	7.58	0	14.01	122	0.33045	0.178	0.220
Disorientation	8.35	6.96	0	9.73	6.96	0	0	13.92	119	0.32923	0.178	0.190
Total SSQ	147.12	104.24	28.35	164.41	102.08	56.7	0	141.67	125.5	0.2641	0.204	0.255

Table 18: Comparison of SSQ subscale scores between musicians and non-musicians.

Subscale	Mean (Mus.)	Median (Mus.)	Mode (Mus.)	SD (Mus.)	Mean (Non-Mus.)	Median (Non-Mus.)	Mode (Non-Mus.)	SD (Non-Mus.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	11.13	9.54	9.54	14.04	5.30	0	0	12.72	38	0.15742	0.365	0.407
Oculomotor	12.63	11.37	7.58	9.18	11.79	7.58	0	17.41	33.5	0.46521	0.189	0.241
Disorientation	9.28	6.96	0	11.37	6.19	0	0	14.11	33.5	0.39742	0.218	0.241
Total SSQ	123.58	104.24	0	119.32	87.06	28.35	0	161.57	37	0.25247	0.295	0.370

Table 19: Comparison of SSQ subscale scores between musicians and non-musicians in Group A.

Subscale	Mean (Mus.)	Median (Mus.)	Mode (Mus.)	SD (Mus.)	Mean (Non-Mus.)	Median (Non-Mus.)	Mode (Non-Mus.)	SD (Non-Mus.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	19.08	4.77	0	32.12	11.27	9.54	9.54	14.03	21	0.94514	0.018	-0.045
Oculomotor	22.74	11.37	7.58	25.52	11.71	7.58	0	11.43	27.5	0.50422	0.172	0.25
Disorientation	6.96	6.96	0	8.04	7.59	0	0	14.42	25	0.69608	0.101	0.136
Total SSQ	182.44	86.40	56.70	233.66	114.38	99.71	0	129.93	25.5	0.69242	0.102	0.159

Table 20: Comparison of SSQ subscale scores between musicians and non-musicians in Group B.

Subscale	Mean (VR Exp.)	Median (VR Exp.)	Mode (VR Exp.)	SD (VR Exp.)	Mean (No VR Exp.)	Median (No VR Exp.)	Mode (No VR Exp.)	SD (No VR Exp.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	10.49	9.54	0.00	14.54	11.92	0.00	0.00	18.28	109.0	0.68331	0.074	0.09
Oculomotor	8.34	7.58	7.58	6.64	14.02	7.58	0.00	18.45	90.0	0.66186	0.080	-0.10
Disorientation	8.35	6.96	0.00	9.73	11.83	0.00	0.00	21.32	109.0	0.66400	0.079	0.09
Total SSQ	101.67	86.40	116.09	106.09	141.30	28.35	0.00	202.66	113.0	0.57742	0.102	0.130

Table 21: Comparison of SSQ scores between musicians and non-musicians post-experiment.

C.3 Comparison Between Musicians in Group A and Group B

Subscale	Mean (Group A)	Median (Group A)	SD (Group A)	Mean (Group B)	Median (Group B)	SD (Group B)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	11.13	9.54	14.04	19.08	4.77	32.12	12.5	1.00000	0.000	0.042
Oculomotor	12.63	11.37	9.18	22.74	11.37	25.52	10.5	0.82478	0.070	-0.125
Disorientation	9.28	6.96	11.37	6.96	6.96	8.04	13.0	0.90619	0.037	0.083
Total SSQ	123.58	104.24	119.32	182.44	86.40	233.66	11.0	0.91459	0.034	-0.083

Table 22: Comparison of SSQ scores between musicians in Group A and Group B - Pre-Experiment.

Subscale	Mean (Group A)	Median (Group A)	SD (Group A)	Mean (Group B)	Median (Group B)	SD (Group B)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	12.72	4.77	18.76	7.15	9.54	4.77	12	1.00000	0.000	0.000
Oculomotor	10.11	7.58	7.83	5.68	7.58	3.79	16	0.39871	0.267	0.333
Disorientation	9.28	6.96	11.37	6.96	6.96	8.04	13	0.90619	0.037	0.083
Total SSQ	120.08	86.40	133.51	74.05	75.89	48.63	13.5	0.82859	0.068	0.125

Table 23: Comparison of SSQ scores between musicians in Group A and Group B - Post-Experiment.

C.4 Comparison Between VR Specialists and Non-Specialists

Subscale	Mean (VR Spec.)	Median (VR Spec.)	Mode (VR Spec.)	SD (VR Spec.)	Mean (Non-VR Spec.)	Median (Non-VR Spec.)	Mode (Non-VR Spec.)	SD (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	11.13	9.54	9.54	14.04	10.34	0	0	17.32	82.5	0.57348	0.103	0.146
Oculomotor	17.69	7.58	7.58	20.15	12.32	7.58	0	13.74	81.5	0.63115	0.088	0.132
Disorientation	11.60	6.96	0	16.27	6.38	0	0	11.59	87	0.36747	0.165	0.208
Total SSQ	151.16	116.09	116.09	179.02	108.58	70.88	0	142.83	88.5	0.39962	0.154	0.229

Table 24: Comparison of SSQ scores between VR specialists and non-VR specialists - Pre-Experiment.

Subscale	Mean (VR Spec.)	Median (VR Spec.)	Mode (VR Spec.)	SD (VR Spec.)	Mean (Non-VR Spec.)	Median (Non-VR Spec.)	Mode (Non-VR Spec.)	SD (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	11.92	4.77	0	18.06	6.07	0	0	11.51	26	0.60161	0.135	0.182
Oculomotor	17.06	7.58	7.58	24.27	10.34	7.58	0	9.75	23	0.94625	0.017	0.045
Disorientation	13.92	6.96	0	19.69	5.06	0	0	9.38	28	0.39014	0.222	0.273
Total SSQ	160.45	72.22	0	229.97	80.30	28.35	0	103.60	26	0.64045	0.121	0.182

Table 25: Comparison of SSQ scores between VR and non-VR participants in Group A - Pre-Experiment.

Subscale	Mean (VR Spec.)	Median (VR Spec.)	Mode (VR Spec.)	SD (VR Spec.)	Mean (Non-VR Spec.)	Median (Non-VR Spec.)	Mode (Non-VR Spec.)	SD (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	9.54	9.54	9.54	0.00	13.94	9.54	0	20.83	15	0.78828	0.069	0.154
Oculomotor	18.95	18.95	7.58	16.08	13.99	7.58	0	16.62	17.5	0.48702	0.179	0.346
Disorientation	6.96	6.96	13.92	9.84	7.50	0	0	13.47	14.5	0.83893	0.052	0.115
Total SSQ	132.58	132.58	116.09	23.33	132.52	92.38	0	169.72	18.5	0.39107	0.221	0.423

Table 26: Comparison of SSQ scores between VR specialists and non-VR specialists in Group B - Pre-Experiment.

Subscale	Mean (VR Spec.)	Median (VR Spec.)	Mode (VR Spec.)	SD (VR Spec.)	Mean (Non-VR Spec.)	Median (Non-VR Spec.)	Mode (Non-VR Spec.)	SD (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	19.08	14.31	0.00	20.01	9.54	0.00	0.00	15.91	94.5	0.21337	0.227	0.312
Oculomotor	24.00	7.58	7.58	30.10	9.16	7.58	7.58	8.05	86.0	0.46390	0.134	0.194
Disorientation	25.52	13.92	0.00	29.75	6.96	0.00	0.00	12.31	102.0	0.07561	0.324	0.417
Total SSQ	256.58	133.93	0.00	291.06	95.97	32.02	0.00	122.43	97.0	0.19810	0.235	0.347

Table 27: Comparison of SSQ scores between VR specialists and Non-VR specialist - Post-Experiment.

C.5 Comparison Between VR Specialists in Group A and Group B

Subscale	Mean (Group A)	Median (Group A)	SD (Group A)	Mean (Group B)	Median (Group B)	SD (Group B)	U	p-value	Effect size (r)	Cliff's Delta (δ)
Nausea	11.92	4.77	18.06	9.54	9.54	0.00	3.0	0.80259	0.102	-0.250
Oculomotor	17.06	7.58	24.27	18.95	18.95	16.08	3.0	0.80573	0.100	-0.250
Disorientation	13.92	6.96	19.69	6.96	6.96	9.84	4.5	1.00000	0.000	0.125
Total SSQ	160.45	72.22	229.97	132.58	132.58	23.33	2.5	0.63859	0.192	-0.375

Table 28: Comparison of SSQ scores between Group A and Group B among participants with prior VR experience - Pre-experiment.

Subscale	**Mean (Group A)**	**Median (Group A)**	**SD (Group A)**	**Mean (Group B)**	**Median (Group B)**	**SD (Group B)**	**Cliff's Delta (δ)**
Nausea	16.70	9.54	22.54	23.85	23.85	20.24	0.583
Oculomotor	15.16	7.58	20.53	41.69	41.69	48.24	0.482
Disorientation	17.40	6.96	26.35	41.76	41.76	39.37	0.667
Total SSQ	184.22	90.06	256.94	401.30	401.30	403.35	0.667

Table 29: Comparison of SSQ scores between Group A and Group B among participants with prior VR experience - Post-experiment.

APPENDIX D – Presence Questionnaire (PQ) Overview

D.1 Presence Questionnaire version 3.0 Applied in the Research

Question	English Version	Brazilian Portuguese Version
Q1	How much were you able to control events?	O quanto você foi capaz de controlar os eventos?
Q2	How responsive was the environment to actions that you initiated (or performed)?	O quanto o ambiente foi responsivo às ações que você iniciou ou desempenhou?
Q3	How natural did your interactions with the environment seem?	Quão natural pareceram suas interações com o ambiente?
Q4	How much did the visual aspects of the environment involve you?	O quanto os aspectos visuais do ambiente envolveram você?
Q5	How much did the auditory aspects of the environment involve you?	O quanto os aspectos sonoros do ambiente envolveram você?
Q6	How natural was the mechanism which controlled movement through the environment?	Quão natural foi o mecanismo que controlava o movimento no ambiente?
Q7	How compelling was your sense of objects moving through space?	O quão convincente foi sua sensação sobre os objetos se movendo no espaço?
Q8	How much did your experiences in the virtual environment seem consistent with your real-world experiences?	O quanto as suas experiências no ambiente virtual se pareceram com suas experiências no mundo real?
Q9	Were you able to anticipate what would happen next in response to the actions that you performed?	Você foi capaz de antecipar o que aconteceria a seguir em resposta às ações que você desempenhou?
Q10	How completely were you able to actively survey or search the environment using vision?	O quão capaz você foi de, ativamente, explorar ou investigar o ambiente usando a visão?
Q11	How well could you identify sounds?	Quão bem você conseguiu identificar os sons?
Q12	How well could you localize sounds?	Quão bem você conseguiu localizar os sons?
Q13	How well could you actively survey or search the virtual environment using touch?	Quão bem você foi capaz de, ativamente, explorar o ambiente usando o tato?
Q14	How compelling was your sense of moving around inside the virtual environment?	Quão convincente foi sua sensação de mover-se dentro do ambiente virtual?
Q15	How closely were you able to examine objects?	O quão detalhadamente você foi capaz de examinar objetos?
Q16	How well could you examine objects from multiple viewpoints?	Quão capaz de observar objetos sob vários ângulos?
Q17	How well could you move or manipulate objects in the virtual environment?	Quão bem você pôde se mover ou manipular objetos no ambiente virtual?
Q18	How involved were you in the virtual environment experience?	O quão envolvido você estava na experiência do ambiente virtual?

Continued on next page

Question	English Version	Brazilian Portuguese Version
Q19	How much delay did you experience between your actions and expected outcomes?	Quanta demora você experienciou entre suas ações e os desfechos esperados?
Q20	How quickly did you adjust to the virtual environment experience?	Quão rápido você se adaptou à experiência no ambiente virtual?
Q21	How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	O quão proficiente em mover e interagir com o ambiente virtual, você se sentiu ao final da experiência?
Q22	How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?	O quanto a qualidade do dispositivo de visualização interferiu ou distraiu você na performance das tarefas designadas ou atividades requeridas?
Q23	How much did the control devices interfere with the performance of assigned tasks or with other activities?	O quanto os dispositivos de controle interferiram no desempenho das tarefas determinadas ou nas demais atividades?
Q24	How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?	Quão bem você pode se concentrar nas tarefas ou atividades exigidas ao invés de se concentrar nos mecanismos utilizados para realizar essas tarefas ou atividades?
Q25	How completely were your senses engaged in this experience?	Quão completamente os seus sentidos estavam envolvidos nessa experiência?
Q26	How easy was it to identify objects through physical interaction, like touching an object, walking over a surface, or bumping into a wall or object?	O quão fácil foi identificar objetos por meio da interação física, como tocar um objeto, caminhar sobre uma superfície ou esbarrar em uma parede ou objeto?
Q27	Were there moments during the virtual environment experience when you felt completely focused on the task or environment?	Houve momentos durante a experiência no ambiente virtual em que você se sentiu completamente focado na tarefa ou no ambiente?
Q28	How easily did you adjust to the control devices used to interact with the virtual environment?	O quão facilmente você se ajustou aos dispositivos de controle usados para interagir com o ambiente virtual?
Q29	Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?	A informação provida aos diferentes sentidos (ex: visão, audição, tato) pelo ambiente virtual foi consistente?

Table 30: Questions from the Presence Questionnaire (PQ) with Brazilian Portuguese translations.

D.2 Comparison Between Musicians and Non-Musicians

Subscale	Mean (Mus.)	Median (Mus.)	Mode (Mus.)	SD (Mus.)	Mean (Non-mus.)	Median (Non-mus.)	Mode (Non-mus.)	SD (Non-mus.)	U	p-value	Effect size (r)
Involvement/Control	4.89	4.86	4.27	0.82	5.47	5.68	5.36	0.84	53.0	0.04051	0.374
Naturalness	4.94	4.86	4.43	1.17	5.78	5.86	5.86	0.78	56.0	0.05467	0.354
Interface Quality	5.75	6.12	6.25	0.87	5.98	6.00	6.00	0.78	87.0	0.57850	0.919
Auditory Realism	3.50	3.67	3.67	0.79	3.43	3.33	4.00	0.73	109.0	0.70472	0.636
Haptic Realism	6.00	6.16	6.33	0.82	5.90	6.33	6.33	1.00	101.5	0.96439	0.106
Visual Fidelity	5.10	5.00	5.00	1.45	5.75	6.00	6.00	1.33	71.0	0.19198	0.239

Table 31: Comparison of PQ subscale scores between participants with and without musical experience.

Subscale	Mean (Mus.)	Mean (Non-mus.)	Median (Mus.)	Median (Non-mus.)	Mode (Mus.)	Mode (Non-mus.)	SD (Mus.)	SD (Non-mus.)	U	*p*-value	Effect size (*r*)
Involvement/Control	4.97	5.39	4.86	5.82	3.82	5.82	1.00	0.95	19.5	0.40856	0.884
Naturalness	5.14	5.81	5.00	5.86	3.71	5.86	1.20	0.97	18.0	0.31343	0.360
Interface Quality	5.58	5.92	5.50	6.25	4.75	6.50	1.04	1.05	21.5	0.55212	0.648
Auditory Realism	3.67	3.41	3.67	3.33	3.67	4.00	0.76	0.57	33.5	0.47024	0.766
Haptic Realism	5.94	5.52	6.00	6.00	7.00	6.00	1.00	1.01	33.5	0.47673	0.766
Visual Fidelity	5.00	5.33	5.00	6.00	5.00	6.00	1.41	1.41	22.0	0.58062	0.589

Table 32: Comparison of PQ subscale scores between participants with (Mus.) and without (Non-mus.) musical experience in Group A.

D.3 Comparison Between Musicians in Group A and Group B

Subscale	Mean (Mus.)	Mean (Non-mus.)	Median (Mus.)	Median (Non-mus.)	Mode (Mus.)	Mode (Non-mus.)	SD (Mus.)	SD (Non-mus.)	U	*p*-value	Effect size (*r*)
Involvement/Control	4.77	5.54	4.72	5.64	4.27	5.36	0.58	0.79	10.5	0.03032	0.480
Naturalness	4.64	5.75	4.86	5.86	3.00	5.71	1.24	0.63	9.0	0.10053	0.310
Interface Quality	6.00	6.02	6.12	6.00	5.25	6.00	0.54	0.53	22.5	1.00000	0.065
Auditory Realism	3.25	3.45	3.50	3.33	2.00	2.67	0.88	0.86	20.0	0.84321	0.261
Haptic Realism	6.08	6.21	6.16	6.33	5.33	6.33	0.57	0.92	16.5	0.50181	0.718
Visual Fidelity	5.25	6.09	5.50	6.00	3.00	7.00	1.71	1.22	14.5	0.33490	0.080

Table 33: Comparison of PQ subscale scores between participants with (Mus.) and without (Non-mus.) musical experience in Group B.

Subscale	Mean (Group A)	Mean (Group B)	Median (Group A)	Median (Group B)	Mode (Group A)	Mode (Group B)	SD (Group A)	SD (Group B)	U	*p*-value	Effect Size (*r*)
Involvement/Control	4.97	4.77	4.86	4.72	3.82	4.27	1.00	0.58	12.0	1.00000	0.107
Naturalness	5.14	4.64	5.00	4.86	3.71	3.00	1.20	1.24	14.5	0.66887	0.533
Interface Quality	5.58	6.00	5.50	6.12	4.75	5.25	1.04	0.54	9.0	0.58832	0.640
Auditory Realism	3.67	3.25	3.67	3.50	3.67	2.00	0.76	0.88	16.0	0.43988	0.398
Haptic Realism	5.94	6.08	6.00	6.16	7.00	5.33	1.00	0.57	11.5	1.00000	0.05
Visual Fidelity	5.00	5.25	5.00	5.50	5.00	3.00	1.41	1.71	10.5	0.82753	0.075

Table 34: Comparison between musicians in Group A and Group B for PQ subscales.

D.4 Comparison Between VR Specialists and Non-Specialists

Subscale	Mean (VR-spec.)	Mean (NonVR-spec.)	Median (VR-spec.)	Median (NonVR-spec.)	Mode (VR-spec.)	Mode (NonVR-spec.)	SD (VR-spec.)	SD (NonVR-spec.)	U	*p*-value	Effect Size (*r*)
Involvement/Control	4.67	5.43	4.36	5.50	3.82	5.36	0.89	0.81	40.0	0.10197	0.450
Naturalness	4.84	5.67	4.72	5.86	3.71	5.86	0.90	0.96	35.0	0.05743	0.490
Interface Quality	5.38	6.03	5.50	6.25	6.00	6.25	0.65	0.80	28.5	0.02430	0.550
Auditory Realism	3.28	3.50	3.50	3.67	3.67	4.00	0.71	0.75	59.5	0.52837	0.648
Haptic Realism	5.17	6.12	5.00	6.33	3.67	6.33	1.21	0.77	35.5	0.05821	0.490
Visual Fidelity	4.50	5.79	4.00	6.00	3.00	6.00	1.76	1.18	40.5	0.09442	0.450

Table 35: Comparison of PQ subscale scores between participants with (VR-spec) and without (NonVR-spec) prior virtual reality experience.

Subscale	Mean (VR-spec.)	Mean (NonVR-spec.)	Median (VR-spec.)	Median (NonVR-spec.)	Mode (VR-spec.)	Mode (NonVR-spec.	SD (VR-spec.)	SD (NonVR-spec.)	U	*p*-value	Effect Size (*r*)
Involvement/Control	4.50	5.49	4.18	5.64	3.82	4.18	0.92	0.86	8.5	0.08908	0.58300
Naturalness	4.89	5.78	4.79	5.86	3.71	6.86	1.14	1.00	10.5	0.14840	0.52877
Interface Quality	5.25	5.98	5.25	6.25	4.50	6.25	0.74	1.07	9.0	0.09961	0.56944
Auditory Realism	3.67	3.45	3.67	3.67	3.67	4.00	0.27	0.74	25.0	0.73890	0.13558
Haptic Realism	4.67	6.06	4.50	6.33	3.67	6.33	0.98	0.73	4.5	0.02552	0.69147
Visual Fidelity	4.25	5.55	4.00	6.00	3.00	6.00	1.50	1.21	10.5	0.13462	0.51521

Table 36: Comparison of PQ subscale scores between participants with (VR-spec) and without (NonVR-spec) prior virtual reality experience in Group A.

Subscale	Mean (VR-spec.)	Mean (NonVR-spec.)	Median (VR-spec.)	Median (NonVR-spec.)	Mode (VR-spec.)	Mode (NonVR-spec.)	SD (VR-spec.)	SD (NonVR-spec.)	U	*p*-value	Effect Size (*r*)
Involvement/Control	5.00	5.38	5.00	5.45	4.27	5.36	1.03	0.80	10.5	0.73271	0.16013
Naturalness	4.72	5.57	4.72	5.86	4.43	5.86	0.40	0.95	3.0	0.10445	0.56045
Interface Quality	5.62	6.08	5.62	6.00	5.25	6.00	0.53	0.50	6.0	0.25825	0.40032
Auditory Realism	2.50	3.54	2.50	3.67	2.00	3.33	0.71	0.79	3.5	0.12263	0.53376
Haptic Realism	6.16	6.18	6.16	6.33	5.33	6.33	1.18	0.82	13.0	1.00000	0.02669
Visual Fidelity	5.00	6.00	5.00	6.00	3.00	6.00	2.83	1.15	11.0	0.78808	0.13344

Table 37: Comparison of PQ subscale scores between participants with (VR-spec.) and without (NonVR-spec.) prior virtual reality experience in Group B.

D.5 Comparison Between VR Specialists in Group A and Group B

Subscale	Mean (Group A)	Mean (Group B)	Median (Group A)	Median (Group B)	Mode (Group A)	Mode (Group B)	SD (Group A)	SD (Group B)	U	*p*-value	Effect Size (*r*)
Involvement/Control	4.50	5.00	4.18	5.00	3.82	4.27	0.92	1.03	3.0	0.80000	0.520
Naturalness	4.89	4.72	4.79	4.72	3.71	4.43	1.14	0.40	4.0	1.00000	0.430
Interface Quality	5.25	5.62	5.25	5.62	4.50	5.25	0.74	0.53	2.5	0.63859	0.560
Auditory Realism	3.67	2.50	3.67	2.50	3.67	2.00	0.27	0.71	8.0	0.10021	0.090
Haptic Realism	4.67	6.16	4.50	6.16	3.67	5.33	0.98	1.18	1.0	0.26667	0.690
Visual Fidelity	4.25	5.00	4.00	5.00	3.00	3.00	1.50	2.83	3.0	0.80573	0.520

Table 38: Comparison between VR specialist in Group A and Group B.

APPENDIX E – System Usability Scale (SUS) Overview

E.1 System Usability Scale (SUS) Questionnaire Applied in the Research

Item	English Version	Brazilian Portuguese Version
Q1	I think I would like to use the PhysioDrum system frequently	Eu acho que gostaria de usar a PhysioDrum frequentemente
Q2	I found the PhysioDrum system unnecessarily complex	Eu achei a PhysioDrum desnecessariamente complexa
Q3	I found the PhysioDrum system easy to use	Eu achei a PhysioDrum fácil de usar
Q4	I think I would need the support of someone to use the PhysioDrum system	Eu achei que precisaria de ajuda de uma pessoa técnica para ser capaz de usar a PhysioDrum
Q5	I found the various functions of the PhysioDrum system well integrated	Eu achei que as várias funções da PhysioDrum foram bem integradas
Q6	I found too much inconsistency in the PhysioDrum system	Eu acho que a PhysioDrum apresenta muita inconsistência
Q7	I would imagine that most people would learn to use the PhysioDrum system very quickly	Eu imagino que a maioria das pessoas pode aprender a usar a PhysioDrum rapidamente
Q8	I found the PhysioDrum system very cumbersome to use	Eu achei a PhysioDrum muito complicada de usar
Q9	I felt confident using the PhysioDrum system	Eu me senti seguro(a) usando a PhysioDrum
Q10	I needed to learn a lot of things before I could get going with the PhysioDrum system	Eu precisei aprender muitas coisas antes que pudesse utilizar a PhysioDrum

Table 39: System Usability Scale (SUS) with translation into Brazilian Portuguese.

E.2 Comparison Between Musicians and Non-Musicians

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
79.25	87.87	81.25	87.50	72.50	85.00	8.90	8.78	50.5	0.0301	0.3975	-0.495

Table 40: SUS score comparison between musicians and non-musicians.

E.3 Comparison Between Musicians in Group A and Group B

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)
78.33	88.61	76.25	87.50	72.50	85.00	6.83	8.30

Table 41: Descriptive statistics for Group A, comparing musicians and non-musicians.

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)
80.62	87.27	85.00	90.00	62.50	72.50	12.47	9.51

Table 42: Descriptive statistics for Group B, comparing musicians and non-musicians.

Mean (Group A)	Mean (Group B)	Median (Group A)	Median (Group B)	Mode (Group A)	Mode (Group B)	SD (Group A)	SD (Group B)	Mann–Whitney U	p-value	Effect size (r)	Cliff's Delta (δ)
78.33	80.62	76.25	85.00	72.50	32.50	6.83	12.47	8.5	0.516	0.235	-0.291

Table 43: Descriptive statistics comparing musician participants in Groups A and B.

E.4 Comparison Between VR Specialists and Non-Specialists

Mean (VR Spec.)	Mean (Non-VR Spec.)	Median (VR Spec.)	Median (Non-VR Spec.)	Mode (VR Spec.)	Mode (Non-VR Spec.)	SD (VR Spec.)	SD (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's Delta (δ)
80.41	86.14	83.75	86.25	62.50	72.50	11.00	9.11	51.50	0.296947	0.194054	-0.284722

Table 44: Descriptive statistics comparing VR Specialists and Non-VR Specialists participants.

E.5 Comparison Between VR Specialists in Group A and Group B

Mean (VR Spec.)	Mean (Non-VR Spec.)	Median (VR Spec.)	Median (Non-VR Spec.)	Mode (VR Spec.)	Mode (Non-VR Spec.)	SD (VR Spec.)	SD (Non-VR Spec.)
81.87	85.45	83.75	85.00	72.50	72.50	6.57	9.98

Table 45: Descriptive statistics for Group A, comparing VR Specialists and Non-VR Specialists.

Mean (VR Spec.)	Mean (Non-VR Spec.)	Median (VR Spec.)	Median (Non-VR Spec.)	Mode (VR Spec.)	Mode (Non-VR Spec.)	SD (VR Spec.)	SD (Non-VR Spec.)
77.50	86.73	77.50	87.50	62.50	72.50	21.21	8.68

Table 46: Descriptive statistics for Group B, comparing VR Specialists and Non-VR Specialists.

Group	Mann–Whitney U	p-value	Effect size (r)	Cliff's Delta (δ)
A	17.0	0.552210	0.168550	-0.227273
B	9.5	0.609103	0.153485	-0.269231

Table 47: Descritive statistics comparing VR specialists between Groups A and B.

Mean (Group A)	Mean (Group B)	Median (Group A)	Median (Group B)	Mode (Group A)	Mode (Group B)	SD (Group A)	SD (Group B)	Mann–Whitney U	p-value	Effect size (r)	Cliff's Delta (δ)
81.88	77.50	83.75	77.50	72.50	62.50	6.57	21.21	4.00	1.0000	0.000	0.000

Table 48: Descriptive statistics for VR Users in Groups A and B.

APPENDIX F – NASA Task Load Index (TLX) Overview

F.1 NASA-TLX Questionnaire Applied in the Research

Question	English Version	Brazilian Portuguese Version
Q1	Mental Demand: How mentally demanding was the task?	Demanda Mental: Quão exigente mentalmente foi a tarefa?
Q2	Physical Demand: How physically demanding was the task?	Demanda Física: Quão exigente fisicamente foi a tarefa?
Q3	Temporal Demand: How hurried or rushed was the pace of the task?	Demanda Temporal: O quão pressionado pelo tempo você se sentiu durante a tarefa?
Q4	Performance: How successful were you in accomplishing what you were asked to do?	Desempenho: Quão bem-sucedido você acha que foi na realização da tarefa?
Q5	Effort: How hard did you have to work to accomplish your level of performance?	Esforço: O quanto você teve que se esforçar para atingir esse nível de desempenho?
Q6	Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?	Frustração: O quanto você se sentiu inseguro, desmotivado, irritado, estressado ou frustrado?

Table 49: NASA-TLX questions with translation into Brazilian Portuguese.

F.2 General Analysis

Subscale	U	p-value	Effect Size (r)	Cliff's Delta (δ)
Mental Demand	114.0	0.96677	0.0114	0.0133
Physical Demand	143.0	0.21164	0.2310	0.2711
Temporal Demand	101.5	0.65204	0.0833	-0.0978
Performance	123.5	0.66095	0.0833	0.0978
Effort	108.5	0.88391	0.0303	-0.0356
Frustration	145.5	0.16544	0.2499	0.2933
Total Score	126.5	0.57521	0.1060	0.1244

Table 50: Mann–Whitney U test results with effect sizes (r and Cliff's δ) for each NASA-TLX subscale and total score.

F.3 Comparison Between Musicians and Non-Musicians

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)	U	p-value	Effect size (r)	Cliff's δ
37.83	29.79	35.00	24.58	21.67	17.50	15.00	15.87	135.00	0.128	0.281	0.35

Table 51: Descriptive statistics for NASA-TLX scores comparing musicians and non-musicians.

F.4 Comparison Between Musicians in Group A and Group B

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)
36.39	32.68	35.00	37.00	21.67	10.00	15.62	19.10

Table 52: Descriptive statistics comparing musicians and non-musicians in Group A.

Group	Mann–Whitney U	p-value	Effect size (r)	Cliff's delta (δ)
A	32.50	0.555336	0.16736	0.203704
B	34.00	0.131512	0.40452	0.545455

Table 53: Statistical test results for NASA–TLX scores in Groups A and B.

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)
40.00	27.42	35.00	21.00	26.00	17.00	16.05	13.15

Table 54: Descriptive statistics comparing musicians and non-musicians in Group B.

Mean (Group A)	Mean (Group B)	Median (Group A)	Median (Group B)	Mode (Group A)	Mode (Group B)	SD (Group A)	SD (Group B)	U	p-value	Effect size (r)	Cliff's δ
36.39	40.00	35.00	35.00	21.67	26.67	15.62	16.05	9.00	0.609	0.202	-0.25

Table 55: Descriptive statistics comparing musicians and non-musicians in Groups A and B.

F.5 Comparison Between VR Specialists and Non-Specialists

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)	U	p-value	Effect size (r)	Cliff's delta (δ)
41.39	30.24	44.16	25.00	19.17	26.67	13.53	15.78	102.00	0.125	0.283	0.416

Table 56: Descriptive statistics comparing VR specialists and Non-VR specialist.

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)
48.75	28.86	46.25	22.50	43.33	22.50	7.15	16.95

Table 57: Descriptive statistics comparing VR specialists and Non-VR specialists in Group A.

Group	Mann–Whitney U	p-value	Effect size (r)	Cliff's δ
A	36.00	0.077713	0.471940	0.636364
B	11.00	0.798022	0.087706	-0.153846

Table 58: Statistical test results for VR specialists and Non-VR specialists for Groups A and B.

Mean (Mus.)	Mean (Non-Mus.)	Median (Mus.)	Median (Non-Mus.)	Mode (Mus.)	Mode (Non-Mus.)	SD (Mus.)	SD (Non-Mus.)
26.67	31.41	26.67	26.67	19.17	26.67	10.60	15.31

Table 59: Descriptive statistics comparing VR specialists and Non-VR specialists in Group B.

F.6 Comparison Between VR Specialists in Group A and Group B

Mean (Group A)	Mean (Group B)	Median (Group A)	Median (Group B)	Mode (Group A)	Mode (Group B)	SD (Group A)	SD (Group B)	U	p-value	Effect size (r)	Cliff's δ
48.75	26.67	46.25	26.67	43.33	19.17	7.15	10.60	8.00	0.133	0.755	1.000

Table 60: Descriptive statistics comparing VR specialists in Groups A and B.

APPENDIX G – Haptic Questionnaire (HQ) Overview

G.1 Haptic Questionnaire (HQ) Applied in the Research

Question	English Version	Brazilian Portuguese Version
Q1	The haptic feedback fits well with the other effects (e.g., sound effects, visual effects)	O feedback tátil combinou bem com os outros efeitos (ex.: efeitos sonoros, efeitos visuais)
Q2	I like having the haptic feedback as part of the experience	Eu gostei de ter o feedback tátil como parte da experiência
Q3	The haptic feedback felt disconnected from the rest of the experience	O feedback tátil pareceu desconectado do restante da experiência
Q4	The haptic feedback felt appropriate when and where I felt it	O feedback tátil pareceu apropriado quando e onde eu o senti
Q5	The haptic feedback felt out of place	O feedback tátil pareceu fora de lugar
Q6	The haptic feedback felt satisfying	O feedback tátil foi satisfatório
Q7	The haptic feedback would feel good by itself	O feedback tátil seria agradável mesmo sozinho
Q8	I disliked the haptic feedback	Eu não gostei do feedback tátil
Q9	I would prefer the system without the haptic feedback	Eu preferiria o sistema sem o feedback tátil
Q10	All the haptic feedback felt the same	Todo o feedback tátil parecia o mesmo
Q11	I felt adequate variations in the haptic feedback	Senti variações adequadas no feedback tátil
Q12	The haptic feedback helped me distinguish what was happening	O feedback tátil me ajudou a distinguir o que estava acontecendo
Q13	The haptic feedback changes depending on how things change in the system	O feedback tátil muda dependendo de como as coisas também mudam no sistema
Q14	The haptic feedback reflects varying inputs and events	O feedback tátil reflete entradas e eventos variados
Q15	The haptic feedback distracted me from the task	O feedback tátil me distraiu da tarefa
Q16	I felt engaged with the system due to the haptic feedback	Eu me senti engajado com o sistema devido ao feedback tátil
Q17	The haptic feedback helped me focus on the task	O feedback tátil me ajudou a focar na tarefa
Q18	The haptic feedback increased my involvement in the task	O feedback tátil aumentou meu envolvimento na tarefa
Q19	The haptic feedback was realistic	O feedback tátil foi realista
Q20	The haptic feedback was believable	O feedback foi crível
Q21	The haptic feedback was convincing	O feedback tátil foi convincente
Q22	The haptic feedback matched my expectations	O feedback tátil correspondeu à minha expectativa

Table 61: Haptic Feedback Questionnaire – English and Brazilian Portuguese Versions.

G.2 General Analysis

Subscale	U	p-value	Effect size (r)
Autotelic	106.0	0.8019	-0.0492
Expressiveness	112.5	1.0000	0.0000
Immersion	153.0	0.0879	0.3067
Realism	142.0	0.2186	0.2234
Harmony	83.5	0.2313	-0.2196
Score HX	132.0	0.4301	0.1477

Table 62: Statistical tests comparing Groups A and B for each subscale in Haptic Questionnaire.

G.3 Comparison Between Musicians and Non-Musicians

Subscale	Mean (Mus.)	Mean (Non-Mus.)	U	p-value	Effect size (r)	Cliff's delta (δ)
Autotelic	3.467	3.467	27.5	1.000	0.015	0.019
Expressiveness	3.167	2.833	37.0	0.252	0.304	0.370
Immersion/Control	2.625	2.667	27.0	1.000	0.000	0.000
Realism	3.390	3.111	37.0	0.251	0.304	0.370
Harmony	2.583	2.639	26.0	0.952	0.030	-0.037
HQ Total	3.045	2.943	37.0	0.261	0.304	0.370

Table 63: Comparison between musicians and non-musicians for Group A in each subscale and total HQ score.

Subscale	Mean (Mus.)	Mean (Non-Mus.)	U	p-value	Effect size (r)	Cliff's δ
Autotelic	3.400	3.564	19.5	0.793	0.084	-0.114
Expressiveness	2.688	3.023	13.5	0.279	0.287	-0.386
Immersion/Control	2.250	2.341	17.0	0.542	0.169	-0.227
Realism	2.750	2.849	22.5	1.000	0.017	0.023
Harmony	3.125	2.841	24.0	0.843	0.067	0.091
HQ Total	2.842	2.923	16.5	0.514	0.185	-0.250

Table 64: Comparison between musicians and non-musicians for Group B in each subscale and total HQ score.

G.4 Comparison Between Musicians in Group A and Group B

Subscale	Mean (Group A)	Mean (Group B)	U	p-value	Effect size (r)	Cliff's delta (δ)
Autotelic	3.467	3.400	12.0	1.000	0.000	0.000
Expressiveness	3.167	2.688	17.5	0.279	0.371	0.458
Immersion	2.625	2.250	16.0	0.443	0.270	0.333
Realism	3.390	2.750	17.0	0.321	0.337	0.417
Harmony	2.583	3.125	8.5	0.516	0.236	-0.292
HQ Total	3.045	2.842	19.0	0.171	0.472	0.583

Table 65: Comparison between musicians in Groups A and B for each subscale and total HQ score.

G.5 Comparison Between VR Specialists and Non-Specialists

Subscale	Mean (VR-spec.)	Mean (NonVR-spec.)	Median (VR-spec.)	Median (NonVR-spec.)	Mode (VR-spec.)	Mode (NonVR-spec.)	SD (VR-spec.)	SD (NonVR-spec.)	U	p-value	Effect size (r)	Cliff's δ
Autotelic	3.13	3.58	3.20	3.60	3.20	3.40	0.516	0.460	35.5	0.059884	0.34551	-0.506944
Expressiveness	3.12	2.90	3.12	3.00	3.00	3.25	0.564	0.459	87.5	0.428451	0.146723	0.215278
Immersion/Control	2.58	2.45	2.50	2.50	2.50	2.50	0.516	0.519	75.5	0.872880	0.033131	0.048611
Realism	2.72	3.09	3.16	3.16	3.33	3.67	0.927	0.699	51.5	0.288884	0.194054	-0.284722
Harmony	3.04	2.69	2.75	2.50	2.75	2.50	0.827	0.594	90.5	0.344605	0.175122	0.256944
HQ Total	2.92	2.94	2.89	2.98	2.89	2.98	0.222	0.214	53.0	0.336930	0.179855	-0.263889

Table 66: Comparison between participants with VR experience and without VR experience for each subscale and total HQ score.

G.6 Comparison Between VR Specialists in Group A and Group B

Subscale	Mean (VR Spec.)	Mean (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's delta (δ)
Autotelic	3.250	3.545	15.0	0.390	0.236	-0.318
Expressiveness	3.125	2.909	26.5	0.593	0.152	0.205
Immersion	2.812	2.591	26.5	0.593	0.152	0.205
Realism	2.998	3.305	13.0	0.255	0.303	-0.409
Harmony	2.750	2.568	27.0	0.549	0.169	0.227
HQ Total	2.990	2.982	15.0	0.394	0.236	-0.318

Table 67: Comparison between VR experts and non-VR users for Group A in each subscale and total HQ score.

Subscale	Mean (VR Spec.)	Mean (Non-VR Spec.)	U	p-value	Effect size (r)	Cliff's delta (δ)
Autotelic	2.900	3.615	2.5	0.087	0.460	-0.808
Expressiveness	3.125	2.904	16.0	0.660	0.132	0.231
Immersion	2.125	2.346	7.0	0.332	0.263	-0.462
Realism	2.165	2.924	8.5	0.489	0.197	-0.346
Harmony	3.625	2.808	19.5	0.304	0.285	0.500
HQ Total	2.790	2.918	7.0	0.350	0.263	-0.462

Table 68: Comparison between VR experts and non-VR users for Group B in each subscale and total HQ score.

Subscale	Mean (Group A)	Mean (Group B)	U	p-value	Effect size (r)	Cliff's δ
Autotelic	3.250	2.900	5.5	0.617	0.283	0.375
Expressiveness	3.125	3.125	4.0	1.000	0.000	0.000
Immersion	2.812	2.125	8.0	0.100	0.756	1.000
Realism	2.998	2.165	5.0	0.806	0.189	0.250
Harmony	2.750	3.625	1.5	0.348	0.472	-0.625
HQ Total	2.990	2.790	5.0	0.814	0.189	0.250

Table 69: Comparison between VR specialists and non-VR specialists in Groups A and B for each subscale and total HQ score.

APPENDIX H – PhysioDrum Semi-structured Interview

Question	Brazilian Portuguese Version	English Version
Q1	O que você achou da experiência com a PhysioDrum?	What did you think of the experience with the PhysioDrum?
Q2	Quão confortável você se sentiu ao usar os pedais e as baquetas para interagir com o sistema?	How comfortable did you feel using the pedals and drumsticks to interact with the system?
Q2-A	O que poderia ser melhorado?	What could be improved?
Q3	O que você achou do feedback háptico?	What did you think of the haptic feedback?
Q3-A	Você achou útil ou preferiria removê-lo?	Did you find it useful or would you prefer to remove it?
Q3-B	Como você descreveria o nível de conforto do feedback háptico?	How would you describe the comfort level of the haptic feedback?
Q3-C	Como você descreveria a naturalidade do feedback háptico?	How would you describe the naturalness of the haptic feedback?
Q3-D	Você acha que o feedback háptico influenciou ou não a sua sensação de imersão na aplicação?	Do you think the haptic feedback influenced your sense of immersion in the application?
Q4	Como você avalia a acessibilidade do PhysioDrum para indivíduos com diferentes níveis de habilidade musical ou familiaridade com tecnologia?	How do you assess the accessibility of the PhysioDrum for individuals with different levels of musical ability or familiarity with technology?
Q5	Como você descreveria seu nível de engajamento e criatividade ao usar a PhysioDrum?	How would you describe your level of engagement and creativity when using the PhysioDrum?
Q6	Há algo que você mudaria ou gostaria de ajustar na PhysioDrum?	Is there anything you would change or would like to adjust in the PhysioDrum?
Q7	Algum comentário ou sugestão que você queira deixar para nós?	Any comments or suggestions you would like to share with us?

Table 70: Semi-structured interview applied to user in PhysioDrum study.

APPENDIX I – Towards an Io3MT Live Performance Scenario

In addition to the scenarios that demonstrated the practical implementation of a custom device and an immersive environment grounded in Io3MT principles, a third use case was developed. This case focused on employing off-the-shelf technologies to validate the concepts proposed throughout this thesis using commercially available tools, thereby eliminating the need for a dedicated toolkit. However, this scenario diverges from the other examples presented in this work and is therefore included in the Appendix, serving as an exploratory reference and a source of inspiration for alternative implementations and future directions within the Io3MT domain.

To assess the system, measurements of both QoS and QoE are once again conducted. This evaluation employs an autobiographical approach, in which the author also assumes the role of system user and provides an analytical assessment of the environment's functional and artistic dimensions. Subsequently, an analysis is performed to identify which aspects predicted by the Io3MT framework have been fulfilled.

I.1 Io3MT Environment Design

This process was grounded in the use scenario related to live performances, as presented in Section 4.10.1. In general, the development is structured into three sequential phases (ORSSO et al., 2022). The first phase, referred to as Ideation, involves the formulation of broad and guiding abstractions concerning the artistic entity, with the aim of defining the purpose of the performance and identifying potential materials and technologies to be employed. The interaction among multisensory, multimedia, and musical elements, with the objective of imparting meaning to the work, is also addressed at this stage.

The subsequent phase is Experimentation, in which the formulations developed in the preceding stage are effectively implemented, thereby testing the proposed concepts and selected materials. This process enables the identification and execution of adjustments, allowing the originally conceived concept to be aligned with the constraints of the material context.

Thereafter, the Incorporation phase consolidates the outcomes of the previous stages, resulting in a functional environment that remains consistent with the artistic and expressive requirements defined

during the initial phase. Additional adjustments are introduced to address imperfections or to refine the artistic object. In summary, the conceptual and practical development of this scenario follows an iterative process, in which the knowledge accumulated in earlier stages informs and enhances the subsequent refinement of the work.

Upon the conclusion of this stages, the following practical scenario was defined (VIEIRA; SAADE; CÉSAR, 2023): Computer A displays a sequence of multimedia content organized into scenes. Each scene is associated with a luminous effect that reflects the predominant color palette of the video. In addition, a scent diffuser is incorporated into the environment and operates analogously to the lighting devices, releasing different essences in accordance with the visual content being presented. In parallel, the author/artist performs live music, with the audio data transmitted over the network from Computer A to Computer B, which is exclusively dedicated to processing this type of information. Finally, a graphic artwork is projected throughout the environment, with the properties of its geometric forms and execution speed also modulated via the network. The convergence and interaction of these elements support the storytelling defined during the Ideation phase.

The storytelling, in turn, draws inspiration from the album "*Clube da Esquina*" (Corner Club) and the novel "*Grande Sertão: Veredas*" (The Devil to Pay in the Backlands). The author of this thesis, who also assumed the role of performer in the artistic scenario presented, sought resonance between the creative journey of the musicians in "*Clube da Esquina*" recording process and his own personal trajectory. Both stories unfold from a departure: leaving the state of Minas Gerais, Brazil, toward new horizons, relocating to the city of Niterói in order to produce a work (this thesis in the case of the author, and the album in the case of the musicians). The visual composition evokes this path, presenting scenes from three cities that marked the author's life and studies, symbolizing the crossings that ultimately converged in the realization of this doctoral research. The musical piece selected for performance was "*O Trem Azul*" (Blue Train), featured in the aforementioned album, which references a train departing from the city of Amsterdam, a place where the author also lived and which exerted both scientific and artistic influence on the development of this work. The novel's inspiration emerges in its final word, *Travessia* (Journey), a term that also titles the first artistic offering of *Clube da Esquina*. In both contexts, it signifies the passage of a journey, with its struggles and discoveries, adventures and setbacks, culminating in the construction of an accomplished ending.

I.1.1 Technological Foundations

This section introduces the technological scope underlying the proposed environment. It outlines the physical devices employed for audio acquisition and multisensory rendering, together with the networking infrastructure that ensures stable communication. Furthermore, it examines the software stack, which provides the basis for the implementation and execution of this scenario.

I.1.1.1 Hardware Components

To enable this environment, the setup included an HP 256 laptop (Computer A) and an Acer Predator Helios 300 G3 (Computer B), both running the Linux Mint XFCE 20.04 operating system, a Yeelight E27 smart lamp[1], a condenser microphone, a six-string electric guitar, a Behringer UMC202HD audio interface — responsible for connecting the instruments to the computer — and a multimedia projector.

With regard to the physical devices responsible for rendering sensory effects (illustrated in Figure 32), the Moodo AIR[2] constitutes a commercial solution designed for home automation, built around an encapsulated system containing aromatic crystals. Communication with this device is established via HTTP requests through a RESTful API, which employs the JSON format for both input and output data exchange.

For lighting effects, the Yeelight lamp employs three distinct categories of network messages: COMMAND, RESULT, and NOTIFICATION. The first category, COMMAND, enables the transmission of instructions to switch the device on or off and to modify its color properties or light intensity. The second set, RESULT, represents the lamp's response to previously issued requests. The third class, NOTIFICATION, informs all devices connected to the lamp about state changes within the system. All information exchange is carried out through the TCP protocol.

(a) Moodo device. (b) Yeelight device.

Figure 32: Devices employed for the rendering of sensory effects.

In relation to the networking elements, a TP-Link Archer C5 router provides wireless connectivity to the system using the Wi-Fi IEEE 802.11n protocol, with a theoretical transmission limit of up to 100 Mbps. To improve performance and minimize latency and packet loss, a dedicated network was configured, and all security settings were disabled.

I.1.1.2 Software Components

In addition to the physical devices employed for audio capture and multisensory rendering, four virtual devices were implemented using Pure Data. The first two are responsible for capturing audio data from the microphone and the electric guitar and transmitting it over the network, while the third is

[1] https://en.yeelight.com/
[2] https://moodo.co/

dedicated to executing this content. The fourth component generates the visual effects, which are likewise controlled through remote channels. The transmission of audio and control data is carried out via the UDP protocol, whereas auditory properties, such as volume and reverb, and graphical attributes, like geometric form of the displayed objects, are managed through the OSC protocol.

The implementation of the multimedia and multisensory content was carried out using the Nested Context Language (NCL) (SOARES, 2009). This language was designed to specify hypermedia documents in a straightforward manner, without addressing the details inherent to execution semantics. In essence, NCL does not define media objects themselves but rather specifies their temporal and spatial relationships, thereby serving as a "glue" for multimedia applications. This declarative paradigm stands in clear contrast to procedural languages, which describe how a specific task is to be executed (SOARES; MORENO; MARINHO, 2013; BARRETO, 2021).

In this way, the language is capable of integrating a wide range of entities, such as image objects (JPEG, PNG, etc.), video objects (MPEG, AVI, MOV, etc.), audio objects (MP3, WAV, etc.), text objects (TXT, PDF, etc.), and declarative objects (HTML, SVG, etc.), among others. Although no particular type of media is restricted or prescribed, it is important to note that proper execution requires direct association with a player capable of supporting the specific content in question (SOARES; MORENO; MARINHO, 2013).

The current version of the NCL, version 4.0, enables developers to describe the temporal behavior of a multimedia presentation by associating hyperlinks (user interactions) with media objects and sensory effects such as light, scents, and wind (JOSUÉ, 2021). It also allows for the specification of application layouts across multiple devices, as well as interaction through voice and gesture commands.

Although NCL is relatively easy to learn for individuals without prior programming experience, it lacks flexibility and imperative support, limitations that directly affect mathematical processing, text manipulation, use of the interactivity channel, animations, and object collisions. Moreover, its applicability is advantageous only when the application relies on features explicitly covered within the scope of the language (JOSUÉ; ABREU, et al., 2018; SANT'ANNA; CERQUEIRA; SOARES, 2008).

To address this limitation and enhance the expressive capabilities of NCL, integration with a scripting language becomes essential. For this purpose, Lua was incorporated into the functional scope of NCL to extend its capabilities. Originally designed to be simple and easily embeddable within other applications and languages, Lua has since been widely adopted in diverse domains, including robotics, programming, distributed systems, image processing, web development, bioinformatics, and, most notably, gaming. Due to its flexible semantics and high execution performance, Lua has also proven to be well suited for real-time multimedia processing (IERUSALIMSCHY; FIGUEIREDO; CELES, 2007; SMITH, 2007).

Lua scripts are incorporated into NCL as if they were media elements. From this integration emerges a dialect known as NCLua, which enables functionalities such as issuing HTTP requests, communicating with external entities via TCP, performing graphical operations, collecting transmission data, and ac-

cessing primitives available in the operating system. Communication with the NCL document occurs bidirectionally through an event-driven API, which constitutes the primary characteristic distinguishing NCLua from traditional Lua scripts (SANT'ANNA; CERQUEIRA; SOARES, 2008; GUEDES et al., 2016; SOUSA JUNIOR et al., 2010). One of the advantages of NCLua is that it allows NCL applications to support new types of media that were not defined within the language's original scope.

For NCL and Lua applications to be executed independently of the receiving platform, a middleware layer is required to provide such support. As this is a multimedia application, the software layer must also be adaptable in order to enable the provisioning of services across heterogeneous multimedia networks for pervasive end users, support the rendering of diverse actions, and allow multi-user operation (ALVI et al., 2015).

A middleware that fulfills all these requirements is Ginga-NCL (SOARES; MORENO; NETO, et al., 2010). Its architecture is composed of subsystems that perform specific functions for the presentation of declarative applications and for ensuring interoperability across IPTV systems. In this context, the Presentation Environment, the Ginga Common Core (CC), the protocol stack, and the application and service layer are of particular relevance, as illustrated in Figure 33.

To incorporate sensory effects and support the execution of the most recent version of NCL, additional components were introduced, namely, the Presentation Orchestrator, Preparation Orchestrator, Sensory Device Calibrator, Sensory Effect Renderer Manager, and Interaction Manager. Furthermore, modifications were applied to existing components, such as the Player Manager and the NCL Context Manager, as depicted in Figure 34 (JOSUÉ, 2021; BARRETO, 2021).

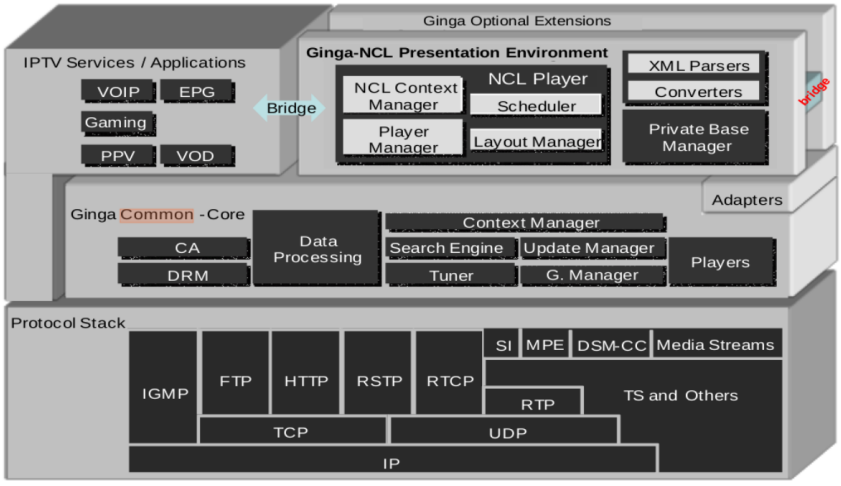


Figure 33: Basic architecture of the Ginga middleware (JOSUÉ, 2021).

In Ginga-NCL, communication between media players, sensory effect renderers, and the presentation engine is managed through a generic API. This API is divided into two parts: the external component, which defines the abstract class used to expose the middleware functionalities and support the implementation of graphical interfaces; and the internal component, which effectively implements the features prescribed by the middleware.

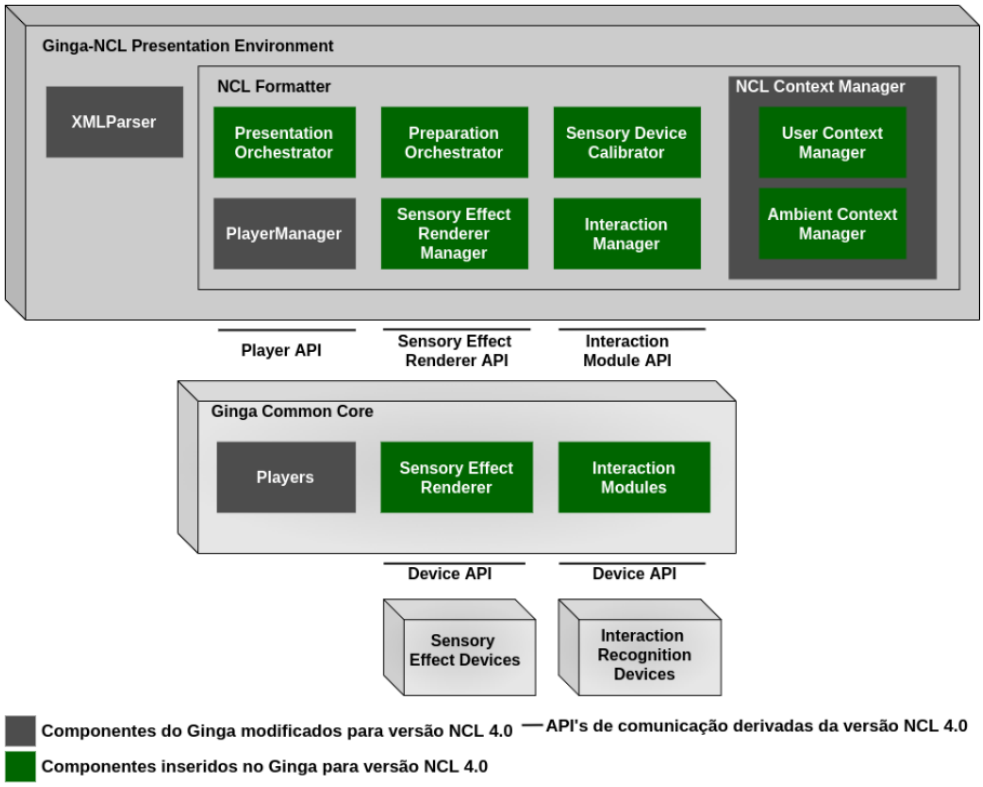


Figure 34: Architecture of Ginga-NCL adapted to support sensory effects (JOSUÉ, 2021).

This latter component is further characterized by a tripartite structure. The first subdivision is the Formatter, which is responsible for controlling the temporal and spatial synchronization of the presentation. The second is the Parser, tasked with performing the syntactic and semantic analysis of the content. The third and final element is the Document, which represents the object tree contained within the application (RODRIGUES et al., 2019).

In broad terms, Ginga operates by receiving an NCL document through its formatter component. The Parser then translates the specifications of this code into data structures that Ginga-NCL can process, thereby enabling control over the presentation. Subsequently, the scheduler component is activated to orchestrate this information, employing techniques such as content prefetching, link condition evaluation, and action scheduling. In addition, a dispatcher is initiated to instantiate the players responsible for handling specific types of media. Through this process, Ginga ensures accurate execution and synchronization of the information defined in the NCL document (SOARES; MORENO; NETO, et al., 2010).

In this specific scenario, Ginga-NCL functions as a centralized platform, providing capabilities for the integration of immersive content, management of media presentation intervals, synchronization of sensory effects with media elements, device description, and facilitation of interactivity within the defined ecosystem. However, since NCL does not provide native support for Pure Data, it was necessary to develop a dedicated player for this type of media object through NCLua. This implementation serves an analogous role to that of a native player within Ginga. Accordingly, functions for initialization, termination, pause, and resume were incorporated, as illustrated in Listing I.1. In the context of real-

time audio processing, the distinction between stopping and pausing the player becomes redundant, as both operations interrupt sound execution. Nevertheless, these four functionalities were implemented to ensure compliance with the established conventions of Ginga-NCL and to validate the chosen design strategies. It is also important to emphasize that this mechanism supports the transmission of information to the Pure Data environment, thereby establishing a bidirectional connection between the two systems.

```
-- creating actions for the Microphone Player
function handler(evt)
  if evt.class == 'ncl' and
    evt.type == 'presentation' and
    evt.action == 'start' then

    start = OpenMicPD()
  end

  if evt.class == 'ncl' and
    evt.type == 'presentation' and
    evt.action == 'stop' then

    start = CloseMicPD()
  end

  if evt.class == 'ncl' and
    evt.type == 'presentation' and
    evt.action == 'pause' then

    start = PauseMicPD()
  end

  if evt.class == 'ncl' and
    evt.type == 'presentation' and
    evt.action == 'resume' then

    start = ResumeMicPD()
  end
```

Listagem I.1: Pure Data player created in NCLua.

To enable this functionality, a library (also referred to as an external) was developed for Pure Data, with the capability of transmitting data to the code implemented in NCLua. Entitled GingaPD, its

development process was facilitated by the pd-lua tool[3], which was created to allow new Pure Data libraries to be implemented in Lua rather than in C, the language traditionally employed for such tasks. This approach makes it possible to generate components ranging from small-scale auxiliary objects to more complex modules, such as complete sequencers and algorithmic composition tools. As a result, the expressive capacity of the Pure Data environment is expanded, offering a broader spectrum of functionalities and potential applications.

Through this new component, it becomes possible to capture a variety of information, particularly numerical data, and direct it to the NCLua environment using pipes. This approach was deliberately adopted to avoid reliance on sockets or external Lua libraries, thereby simplifying both the communication with NCLua and the dissemination and use of this new element. Figure 35 illustrates the incorporation of this library within a functional Pure Data context.

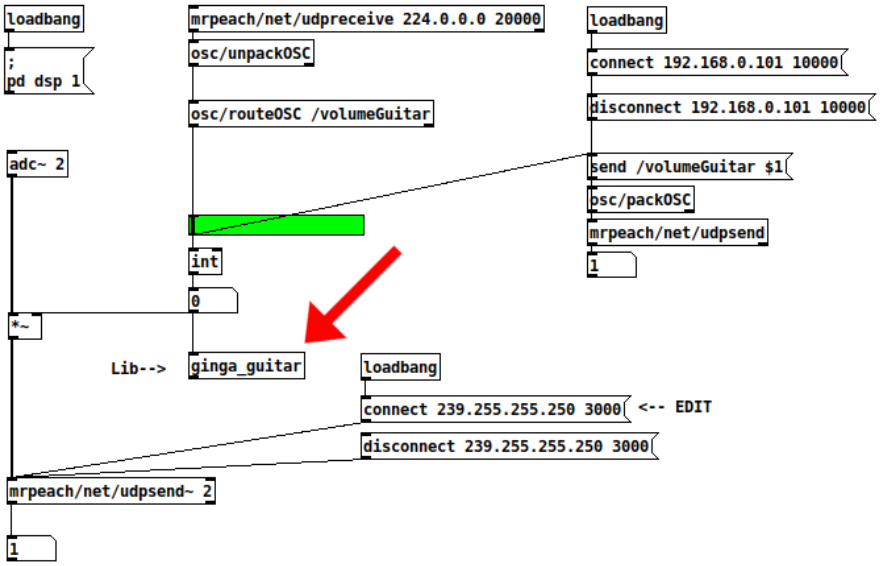


Figure 35: Use of the GingaPD library in a functional Pure Data patch.

Within the Io3MT framework, both the Pure Data media player and the GingaPD library facilitate bidirectional interaction, enabling audio data to modify the properties of sensory effects while, conversely, sensory effects influence audio parameters. This bidirectionality represents a core requirement for Io3MT environments. Moreover, the model extends to the manipulation of multimedia information in a similar manner. For instance, it is possible to envisage scenarios in which video playback is automatically interrupted once the microphone volume surpasses a predefined threshold.

Finally, to optimize the artist's control over these factors, an interface was developed using the TouchOSC tool, as illustrated in Figure 36. This interface centralizes the direct control of all the aforementioned properties, while its visual representation provides the artist with real-time feedback on the elements being modified through automated actions, thereby enabling precise adjustment of all parameters.

All system communication is illustrated in Figure 37, which depicts the interrelation between NCLua

[3] https://agraef.github.io/pd-lua/tutorial/pd-lua-intro.html

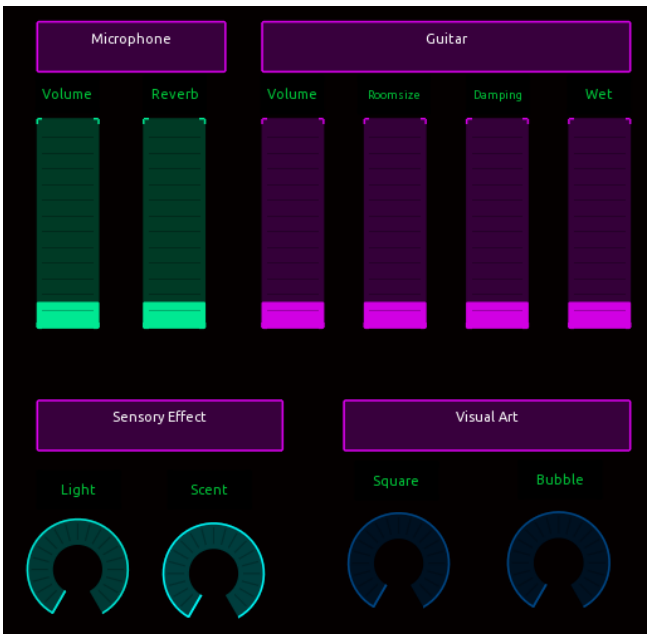


Figure 36: Control interface developed in TouchOSC.

and Pure Data mediated by the Lua language, implemented to extend the operational capabilities of both environments. Since the TouchOSC tool is closely integrated with the patches developed in Pure Data, it is also represented in the figure.

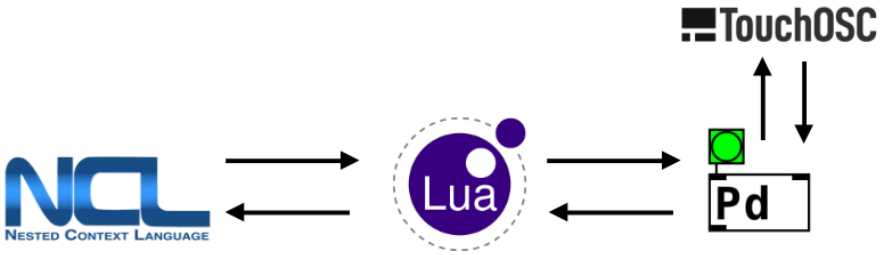


Figure 37: Communication flow among the technologies employed in the environment.

A synthesis of the complete setup employed in the practical experiment is presented in Figure 38, while a demonstration of the environment's operation can be observed in a video available on YouTube[4]. The corresponding source code is accessible in a GitHub repository[5].

I.2 Quality of Service (QoS) Analysis

To address this metric within the present context, analyses of latency, jitter, and packet throughput were once again conducted. Before delving further into this discussion, it is important to make a brief observation regarding the functioning of Pure Data in this scenario. In contrast to the RemixDrum application discussed in Chapter 5, where Pure Data was used to play a pre-recorded track and modify selected sound properties, in the tests reported here the tool was responsible for capturing

[4] https://youtu.be/PA2-n0PLyG4
[5] https://github.com/romulovieira-me/io3mt-environment

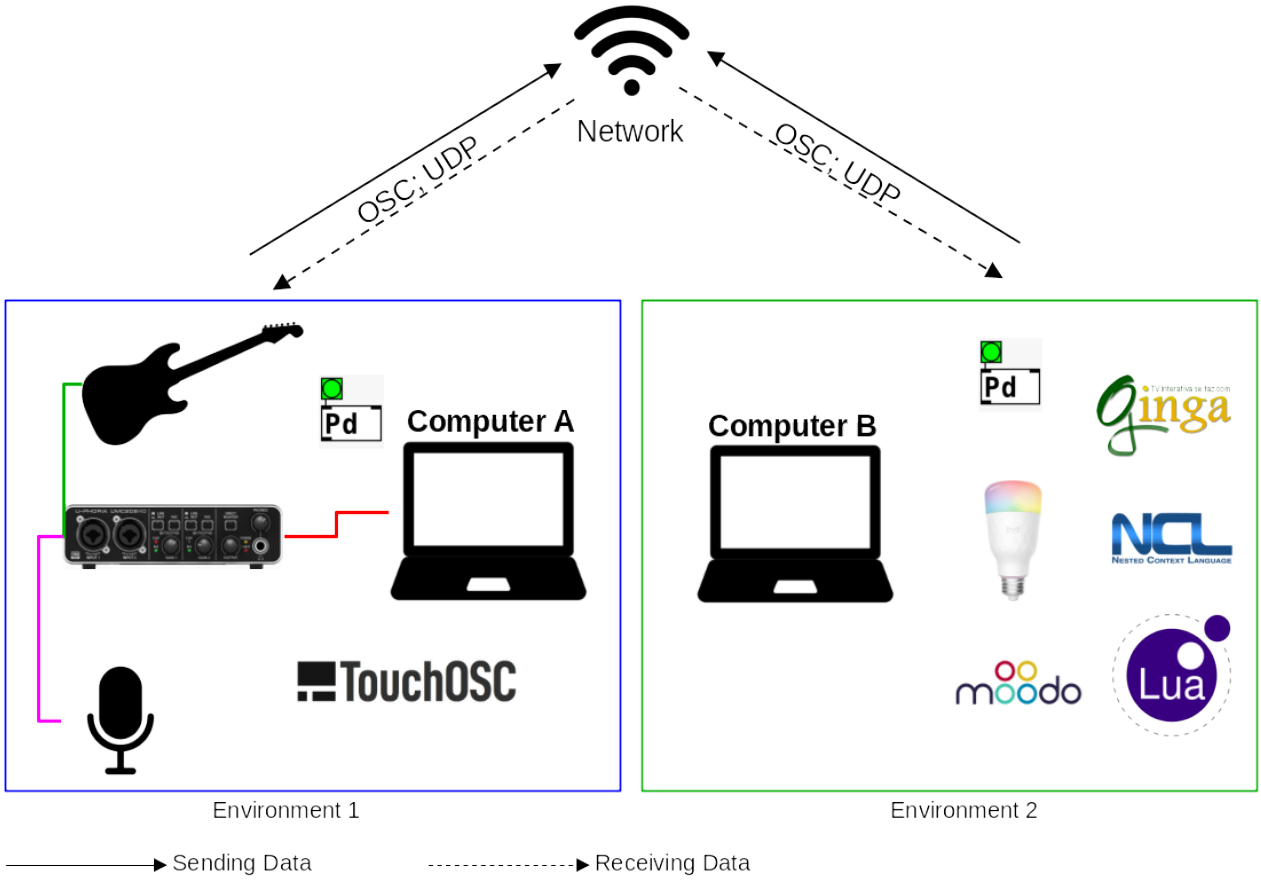


Figure 38: Overview of the experimental setup used in the study (VIEIRA; SAADE; CÉSAR, 2023).

and generating audio from the guitar and microphone, a task that introduced latency into the overall operation. This latency arises because audio information is processed in blocks of 64, 128, 256, 512, 1024, or 2048 samples. Once these blocks are formed and dispatched, they undergo the process of being copied into kernel space and subsequently returned to user space, with processing time depending directly on the block size. When processing occurs on the same sound card, the time required remains constant; however, when distributed across two or more machines, it depends on the block sizes accepted by each computer. This processing time constitutes the latency of Pure Data. In the environment described here, the audio samples were set to a sampling rate of 44,100 Hz, resulting in an average delay of 1.45 ms (VIEIRA; SCHIAVONI; SAADE, 2022).

This behavior is inherent to audio systems and accounts only for the processing time of the information blocks, without considering network-related delays. Potential solutions to this issue include the use of a single global clock to synchronize the oscillators of the sound cards across all participating devices, ensuring that they transmit information simultaneously. Another possible solution would be the adoption of local processing units, since increasing the amount of processing performed within a single block guarantees its execution at the same instant. However, as these topics extend beyond the scope of this work, they will not be discussed in further detail.

Accordingly, for latency measurements in tests of this nature, Equation 5.1 must be applied in con-

junction with the minimum latency introduced by Pure Data. Because this metric is characterized by variability and unpredictability, which may lead to temporal discrepancies, sequencing losses, and potential alterations in transmitted content, an appropriate management strategy is to recognize this condition as an inherent attribute of networked artistic practice. For this reason, the Laid-Back strategy was adopted in this scenario. The calculations for jitter and throughput follow the same procedures described previously.

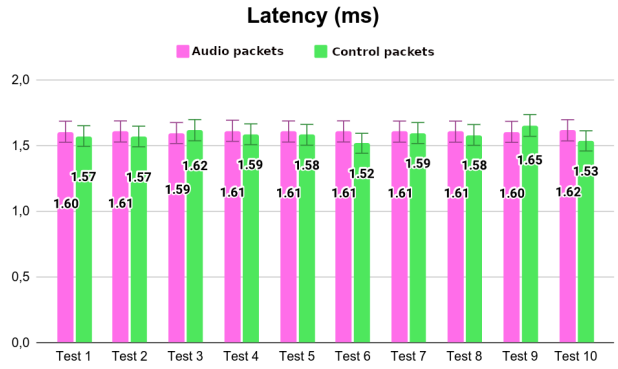
The measurement of these parameters also involved the execution of ten practical experiments. Due to the distinct nuances that characterize a performance, such as intensity, accentuation, and articulation, variations were observed in the number of packets transmitted across sessions. Table 71 presents a summary of this packet exchange, distinguishing once again between audio data and control data.

Test	**Audio Packets**	**Control Packets**
Test 1	10416	273
Test 2	11582	185
Test 3	10698	148
Test 4	10715	250
Test 5	11511	248
Test 6	11885	418
Test 7	11259	268
Test 8	12471	392
Test 9	12468	149
Test 10	12343	375

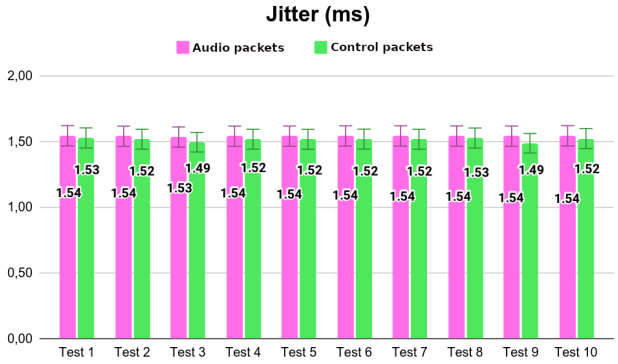
Table 71: Number of packets transmitted in each test (VIEIRA; SAADE; CÉSAR, 2023).

Figure 39 presents the measured values for each individual test, with a 95% confidence interval. The analysis shows that audio data latency ranged from 1.59 ms to 1.62 ms, with Test 3 yielding the most favorable results. For control data, latency varied between 1.53 ms and 1.65 ms. Since changes of this type of information occur in a linear manner (for instance, increasing the volume from 0 to 10 requires the transmission of 10 packets), musical expressiveness has a significant impact on this aspect, directly influencing latency. Consequently, Test 6 yielded the most favorable results with respect to control data latency.

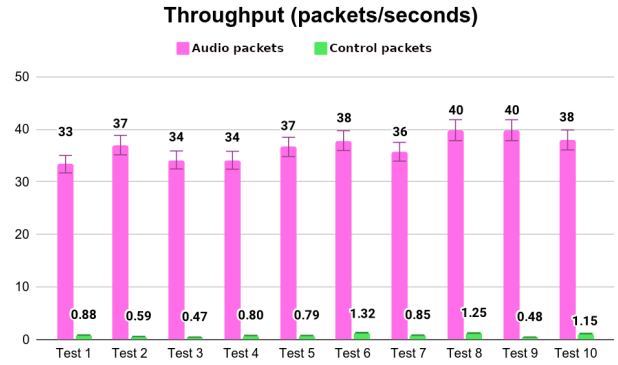
As discussed in Chapter 2.2, delays ranging from 10 ms to 24 ms provide a natural condition for performance, while delays between 25 ms and 60 ms still allow a certain degree of synchronization, though they already cause a noticeable slowdown in musical tempo. Beyond 60 ms, the system becomes impracticable (SCHIAVONI; QUEIROZ; IAZZETTA, 2011; ROTTONDI et al., 2016). In the present experiment, latency did not exceed 2 ms, remaining well within thresholds that do not compromise musical practice. This result can be attributed to the compact nature of the control packets combined



(a) Average latency across the 10 tests.



(b) Jitter across the 10 tests.



(c) Throughput across the 10 tests.

Figure 39: Network performance in the Io3MT environment (VIEIRA; SAADE; CÉSAR, 2023).

with a bandwidth adequate for audio transmission. Furthermore, the use of a dedicated router, the short physical distance between devices, and their respective processing capabilities contributed to these favorable outcomes. It is important to emphasize, however, that the network configuration was primarily designed to support communication and to demonstrate the underlying principles of the Io3MT architecture. To enable the scaling of the infrastructure for a larger number of users and operation in less constrained environments, further optimization of the network would be required.

The results concerning jitter reveal a profile comparable to that observed for latency, with minimal variation identified in the audio data and a broader amplitude recorded in the control data, ranging from 1.49 ms to 1.52 ms. In the case of audio data, Test 3 exhibited superior performance, whereas for control data, Tests 2 and 9 were the most prominent. In configurations of this nature, it is desirable for jitter values to remain below the 20 ms threshold (SCHIAVONI; QUEIROZ; IAZZETTA, 2011). Once this value is exceeded, users tend to perceive a substantial impact. It is noteworthy to emphasize that, once again, the results obtained remained well below the specified limits.

In terms of system throughput, the average audio transmission rate ranged from 33 to 40 packets per second, with Tests 8 and 9 accounting for the largest volume of transmitted data. In contrast, due to the sporadic nature of control information, an average of 0.5 to 1.2 packets per second was observed in this context. Within this scenario, Test 10 demonstrated the best performance with respect to control data throughput. This substantial discrepancy can be attributed to the predominance of audio data within the analyzed environment.

I.3 Quality of Experience (QoE) Analysis

The analysis of Quality of Experience (QoE) in this case diverges from the approaches previously presented, which relied on expert interviews or on mixed-method methodologies encompassing quantitative, qualitative, and statistical strategies. In this context, since the environment was both developed and tested by the same individual, an autobiographical evaluation was adopted. Such an approach is frequently employed in the design of novel interfaces for musical expression, where artists develop tools and techniques to address their own creative and performative requirements (NEUSTAEDTER; SENGERS, 2012; TURCHET, 2018a; REFSUM JENSENIUS; LYONS, 2017; O'NEILL; ORTIZ, 2024; BARBOSA et al., 2013; SPENCE; FROHLICH; ANDREWS, 2013).

I.3.1 Heuristic Evaluation (HE)

The assessment in this context is grounded in a set of guidelines through which the evaluator identifies the strengths and weaknesses of the system. These requirements are presented and explained below (JOHNSON, 2019; HU et al., 2017; NIELSEN; MOLICH, 1990).

- **Visibility of system status:** the system should consistently keep the user informed about ongoing processes through appropriate feedback delivered within a reasonable time frame;

- **Match between system and the real world:** the interface should employ concepts familiar to users, adhere to real-world conventions, and avoid unnecessary technical terminology;
- **User control and freedom:** the system should provide straightforward mechanisms for undoing and redoing actions;
- **Consistency and standards:** users should not be required to guess whether different words, situations, or actions share the same meaning. Adhering to conventions and established standards facilitates usability;
- **Error prevention:** the design should prioritize preventing errors rather than merely providing appropriate error messages after they occur;
- **Recognition rather than recall:** the system should minimize the user's memory load by ensuring that elements, actions, and options remain visible and easily accessible;
- **Flexibility and efficiency of use:** the interface should accommodate both novice and expert users, offering shortcuts and customization options that enhance the efficiency of frequent interactions;
- **Aesthetic and minimalist design:** the interface should avoid irrelevant or extraneous information that does not contribute to the user's task;
- **Help users recognize, diagnose, and recover from errors:** error messages should be clear, precisely indicate the problem, and suggest possible solutions, rather than presenting incomprehensible technical codes;
- **Help and documentation:** the system should provide clear, accessible, and well-structured documentation and support resources.

Based on these metrics, it was observed that, with regard to system status visibility, the devices and software employed provided immediate feedback to the performer during the execution. The interface developed in TouchOSC played a central role in this respect, displaying in real time the values associated with sound properties, as well as the intensity levels of both olfactory and lighting effects. The lighting component itself, in addition to contributing aesthetically and artistically to the multisensory narrative, also served as a feedback mechanism, since chromatic variations indicated changes in the multimedia content, while fluctuations in light intensity reflected corresponding auditory transformations. This continuous feedback ensured that the performer maintained full awareness of the performance's progression, thereby reducing uncertainty and providing greater confidence in controlling the environment.

Concerning the correspondence between the system and the real world, the interaction modalities followed those commonly found in traditional artistic performances, in which information input occurred through conventional musical instruments and the control of technical properties was mediated by sliders and knobs. This design choice proved effective, as it aligned familiar metaphors from the

musical domain with Io3MT systems. From an artistic perspective, the presence of a direct and perceptible causal relationship between multimedia and multisensory content further contributed to a clear interpretation of the interconnection between these elements.

The heuristic of user control and freedom was only partially satisfied. Although the multimedia application developed in NCL presented preprogrammed content without allowing modifications at runtime, the sound elements managed through Pure Data could be manipulated on the run, thereby also influencing the sensory effects. While the system did not include explicit undo or redo functions, the possibility of reinitializing the audio players proved to be an effective recovery mechanism in unexpected situations. In this way, the performer retained sufficient autonomy to preserve the flow of the performance without significant interruptions.

Consistency and adherence to standards were ensured both through the control interface developed in TouchOSC and through the adoption of well-established protocols such as OSC, TCP, and UDP. The use of familiar controls reinforced the uniformity of the experience and minimized the need for additional learning. In a similar manner, since the artistic application was preprogrammed, users were not required to interact with unfamiliar functionalities or manage unexpected behaviors. From a technical standpoint, consistency was further supported by the declarative nature of NCL, which organized the orchestration of media in a predictable and structured manner.

Error prevention was addressed primarily in the services responsible for network-related aspects. The use of a dedicated router, combined with the configuration of an exclusive communication channel, significantly reduced the likelihood of delays. The modular structure of the system, based on loosely coupled communication, ensured that local failures did not compromise the overall integrity of the environment. Furthermore, Pure Data, with its native reconnection functionalities, contributed to enhancing the resilience of the application. From a multimedia and sensory standpoint, the preprogrammed structure of the NCL application, combined with the prefetching mechanism provided by Ginga, contributed to maintaining both stability and predictability throughout execution.

The principle of recognition rather than recall was satisfied by centralizing control in TouchOSC, which displayed adjustable parameters in an explicit and visual manner. This reduced the need for the performer to memorize sequences of technical commands, allowing greater focus on the artistic dimension of the performance. Also, the declarative nature of NCL facilitated transparency in the relationship between media objects and effects, making these associations easily recognizable within the code.

Flexibility and efficiency of use constituted the least satisfied metric in this environment. Although the musical component operated within a plug-and-play model, NCL offered a relatively low learning curve and an intuitive structure, and Pure Data supported the exploration and customization of a wide range of sonic nuances, the programming and configuration of interactions among the different elements of the environment necessitated a substantial level of technological expertise. In this respect, the current setup proved to be more suitable for expert users with prior experience in multimedia programming than for novice users.

In terms of aesthetic and minimalist design, the system avoided informational overload by maintaining a control interface that was concise and focused on the parameters essential to performance. This design enhanced clarity in control, balancing visual simplicity with artistic expressiveness. As a result, the aesthetic integrity of the work was preserved without compromising the transparency of technological interaction.

The heuristic related to error recognition, diagnosis, and recovery was only partially satisfied. Although no textual error messages were provided, the system's behavior, such as the absence of sensory responses or the occurrence of sound distortions, served as sufficient indicators for the performer. Modularity also contributed in this regard, as localized failures could be bypassed without compromising the continuity of the performance.

Regarding help and documentation, the environment addressed this requirement primarily in its technical-scientific dimension. The source code and configuration instructions were made available in public repositories, accompanied by academic publications detailing the methodologies and results. This material supports replicability and facilitates continuity by other research groups.

I.3.2 Artistic Evaluation

The autobiographical analysis of purely artistic elements was conducted from the perspectives of design, performance, and composition (TURCHET, 2018a; O'MODHRAIN, 2011). For the first metric, the environment demonstrated robustness and reliability, as no data loss or technical failures in the devices were observed across the ten tests performed. This stability contributed to the development of creativity, expressiveness, and virtuosity within the practices explored.

The equipment utilized was lightweight and portable, and remained continuously connected to an external power supply, thereby avoiding complications related to battery autonomy. Even the Moodo device, which operated independently, exhibited no interruptions in power settings during its operation. Nevertheless, configuring and assembling the environment required considerable effort and imposed a significant cognitive load.

From the perspetive of sound design, the outcomes met the expectations initially envisioned by the performer. The pre-recorded tracks were carefully conceived and edited to ensure high quality, while the content generated in real time conformed to expectations, largely due to Pure Data's suitability for this type of activity. As anticipated, some processing overloads occurred, particularly in the data generated by the guitar. The quality of the external sound card and the computer responsible for processing these data were directly related to these minor faults. Furthermore, the possibility of balancing electronic sounds resulted in unique timbral nuances. From a functional perspective, the performer was able to carry out the artistic role freely, rather than merely mediating technological interactions or adapting to occasional failures.

In a broader scope, this analysis does not aim to classify artistic aspects as "effective" or "ineffective", but rather to highlight the discrepancy between the outcome anticipated by the performer and that

effectively achieved in practice. Within this context, it is important to emphasize the distinction between shortcomings associated with performance and/or technological choices and those arising from the intrinsic characteristics of the equipment employed. In any case, these considerations did not adversely affect the overall aesthetic integrity of the work.

The second parameter for evaluating QoE pertains to the performer's perspective, which constitutes the most relevant dimension of the environment, as the integration of all tools and protocols must consistently meet the artistic needs of the creator while also serving as a medium of expression. In this regard, the practical implementation proved capable of fulfilling these requirements through its precision, resolution, response time, and the spatial positioning of devices within the physical environment.

The integration of sensory and multimedia aspects into sound control was transparent, partly due to the use of TouchOSC and partly because the performer employed changes in these devices as a form of feedback, thereby extending the focus beyond exclusively auditory information.

Among the controllable properties, parameters such as volume, geometric shapes and their velocities, colors, aroma types, and corresponding intensities functioned as expected. However, the reverb effect applied to the microphone and guitar could be further improved, as it occasionally produced loop-back. This occurrence is intrinsically related to the characteristics of the sound effects rather than to any shortcomings in the techniques employed. Additionally, the physical proximity between the microphone and the loudspeaker was a contributing factor to this issue.

The last attribute to be analyzed concerns the composer's perspective. Within this dimension, the incorporation of multimedia and multisensory components increases the complexity of creative practices. This challenge results from the need to reconfigure compositional strategies, extending their scope beyond the strictly musical domain to include additional layers of information that shape the environment. Consequently, the time required to create a work under these conditions is extended, owing to the necessity of writing code in NCL and Pure Data, performing iterative testing, and integrating the corresponding media and files.

Besides that, composers must also establish meaningful connections with other media, code snippets, and functionalities of electronic devices, preserving both semantic and aesthetic consistency across these multiple dimensions.

Another aspect to be considered concerns the approaches for addressing the absence of an evident causal relationship between a specific multimedia or multisensory action and the corresponding sonic outcome. Historically, the manner in which musical instruments produce sound has been directly perceptible to audiences, but this paradigm began to shift with the introduction of digital elements into musical practice. In the present context, the issue becomes more pronounced, since even a simple media substitution or the activation of a sensory effect can trigger an auditory response. This creates significant challenges for ensuring transparency in communication and for enabling the performer to establish a meaningful connection between their actions and the sounds generated by the instrument.

Such considerations also extend to the production of multimedia and sensory content.

Despite the challenges, this activity remains versatile by not relying exclusively on specific tools or techniques, instead drawing on a technological plurality to achieve its objectives. Accordingly, depending on the artistic requirements, the compositional process can be organized in multiple ways, dynamically incorporating or removing devices and information as needed.

I.4 Analysis of Desirable Characteristics for the Io3MT Environment

This section investigates the extent to which the proposed environment satisfies the desirable characteristics outlined in the Io3MT framework. The analysis primarily emphasizes the system's functional behavior and its capacity to enable multisensory and multimedia integration, while also taking into account the artistic and performative context in which the evaluations were conducted. This discussion situates the assessment within both technical and creative dimensions.

I.4.1 General Characteristics of the Environment

Loosely coupled communication was achieved by ensuring the independence of both the devices and the software components involved. Each element, whether physical, such as the aroma diffusers, or digital, such as the Pure Data modules used to capture and transmit audio data over the network, operated according to its own internal logic and functioned autonomously, interacting only when required through network packets or specific requests. This architecture minimized network overhead and prevented local failures from affecting the overall functionality of the system. Moreover, this behavior fulfills the requirement of micro-systems, in which each device can be regarded as an "island of creation", operating autonomously in a distributed, modular, and sustainable manner, with the flexibility to expand or to function independently without full integration into the system.

Scalability, along with ease of development and evolution, was supported by the chosen software tools. Pure Data enables the integration of new libraries (externals) and plug-ins to extend its functionalities, whereas NCL allows the inclusion of additional media objects, albeit not at runtime, without compromising the existing structure. These properties facilitated the continuous evolution of the system, ensuring its openness to future updates and adaptations.

Service coordination was primarily managed through NCL, which was responsible for initiating, terminating, and synchronizing media and sensory effects, while establishing explicit relationships among them. This orchestration ensured transparent integration between multimedia and multisensory components, making their mutual influence evident. Furthermore, the use of this language enabled the achievement of one of the most prominent features of the environment: the synchronization of different types of data. This integration also enhanced content immersion, as videos, sounds, lights, and scents were presented in convergence to construct the artistic narrative.

From the networking perspective, the adopted configuration ensured low latency, minimal jitter, adequate bandwidth, and reliable data delivery. The use of a dedicated router with an exclusive transmission channel provided system stability. Audio and control packets transmitted via UDP exhibited no perceptible losses, while light effects delivered through TCP maintained complete reliability. These factors further reinforced the resilience of the environment, as execution was preserved even in the presence of minor incidents.

The lightweight implementation was enabled by the use of tools optimized for environments with hardware constraints. Pure Data ensured low computational overhead, allowing execution on both conventional computers and microprocessor boards. Complementarily, Ginga-NCL was designed to operate on set-top boxes and televisions with limited memory and processing capacity, ensuring efficiency and minimal consumption of computational resources.

The proposal also incorporated interactivity, as TouchOSC enabled real-time control of multiple parameters, including volume, reverberation, and sensory intensities. This functionality was further enhanced by an NCLua API, which facilitated remote and bidirectional control of heterogeneous devices.

Finally, both synchronous and asynchronous communication were integrated into the performance. While videos and sensory effects were preprogrammed (asynchronous), audio processing and control via TouchOSC occurred in real time (synchronous). This combination provided enhanced performative flexibility, balancing pre-planned elements with real-time execution.

I.4.2 Functional Requirements

The functional requirements were satisfactorily fulfilled. The system enabled remote access to devices through the network, as the control of sensory elements was carried out via communication protocols, allowing these devices to be operated and configured in real time regardless of their physical location. In addition, the sonic properties of both the microphone and the guitar were also accessible over the network, thereby expanding the possibilities for remote integration and manipulation.

Interoperability among different networks and compatibility between various data formats were successfully achieved. The environment integrated diverse equipment, utilizing NCL and Pure Data as abstraction layers. These languages managed media in multiple formats, such as audio, video, images, and sensory effects, without requiring users to interact directly with the execution logic or semantics of each individual device.

The environment also fostered collaborative information processing, once again, due to the interaction between Pure Data and NCL. Pure Data handled real-time sound data and transmitted control parameters, while NCL coordinated the orchestration of multimedia and multisensory effects. This collaboration between the two distinct systems allowed for the convergence of multiple expressive modalities into a unified performative flow.

Continuous operation with low maintenance requirements and efficient energy consumption was achieved thanks to the inherent functional characteristics of the equipment used. All devices were connected to the power supply, except for the Moodo, which relied on its battery and provided about seven hours of energy autonomy. Since no prototypes or handcrafted devices were created for this specific scenario, there was no need to address additional maintenance concerns.

The requirement for effective network management was only partially met. Each device and digital component involved in data generation and information exchange was assigned a specific network address, which required manual configuration to enable communication. Nevertheless, since the system was designed to accommodate a single user, no device management functionalities were implemented to support the dynamic integration of new equipment in real time. Furthermore, no dedicated support mechanisms were established to handle potential connection failures, limiting the system's adaptability and resilience.

Regarding device functionality, it was observed that each component served multiple purposes. For example, the Yeelight lamp allowed users to adjust both light intensity and color patterns. The Moodo enabled the selection and diffusion of various scents, while Pure Data functioned as a real-time sound processor and a control interface for other devices. This range of functions enhanced the expressiveness and adaptability of the environment.

The requirement of interoperability was fully achieved across its three dimensions. Syntactic interoperability was ensured through the adoption of formal formats and standardized protocols (JSON, OSC, TCP, UDP). Semantic interoperability was achieved by the capacity of NCL to establish explicit relationships among multimedia objects, sensory effects, and musical elements, thereby preserving the coherence of information within the artistic context. Behavioral interoperability was secured by the correct execution of operations resulting from information exchange, such as the use of audio parameters to control sensory devices, among other instances.

I.4.3 Non-Functional Requirements

Among the non-functional requirements, the completeness of heterogeneity is the most readily observable, as the system integrated a broad range of devices and platforms. These included two laptops, a guitar, a microphone, an audio interface, a smart bulb, the Moodo device, open-source software (Pure Data), different programming languages (NCL and Lua), and a middleware (Ginga-NCL). All these elements coexisted in a coordinated manner, preserving both the interoperability and the overall cohesion of the system.

Scalability was ensured through the modularity of the architecture, which allows the system to be expanded by incorporating new devices and media without compromising its integrity. Furthermore, the programming languages employed enable the integration of additional services and multimedia content, thereby supporting the system's adaptability across a wide spectrum of contexts, ranging from individual performances to large-scale collaborative scenarios.

Reliability was addressed at two distinct levels. With respect to data transmission, the dedicated network configuration, combined with the use of specific protocols, ensured consistency and the absence of perceptible losses. In terms of fault recovery, both Pure Data and Ginga-NCL incorporated internal mechanisms that preserved system execution even in the presence of minor incidents, such as temporary device failures. Consequently, the continuity of the performance was maintained without compromising either its artistic or technical quality.

When examining high availability, the environment was designed for continuous operation, with all devices permanently connected to the power supply, thereby avoiding limitations associated with battery autonomy. Practical tests demonstrated the absence of interruptions throughout the performances, ensuring stability and allowing the performer to focus exclusively on the artistic dimension of the presentation.

I.4.4 Musical and Multimedia Protocols; Message Protocols and Data Types

In the proposed Io3MT environment, message protocols played a pivotal role in enabling interoperability among heterogeneous devices, particularly in the control of sensory elements. HTTP, in conjunction with a RESTful API, was utilized to interface with the Moodo aroma diffuser, enabling the selection and activation of distinct essences through structured JSON messages. Among musical and multimedia protocols, OSC was adopted for the exchange of auditory, visual, and control parameters.

For audio types and data formats, WAV and MP3 files were adopted as pre-recorded tracks executed in NCL, while real-time processing relied on PCM, which encoded audio samples transmitted in blocks within Pure Data. Video data, used in the construction of different narrative scenes, was handled through AVI and MP4 formats.

Since NCL was in charge of coordinating the sensory effects, this approach dispensed with the adoption of specific formats for sensory elements, as well as semantic formats or vocabularies for media content.

I.4.5 Artistic Requirements

Among the artistic attributes incorporated, the system excelled in both usability and aesthetic coherence. The centralized control interface facilitated the manipulation of the system's core properties from a single access point, while the spatial arrangement of the sensory dispersers enhanced feedback reception and supported the automatic coordination of services through NCL. These features enabled the performer to focus primarily on the creative dimension of the performance, rather than on technical operational challenges. Furthermore, the convergence of multiple stimuli substantially reinforced the system's immersive quality and amplified its overall aesthetic impact.

Integration was facilitated through the adoption of open-source tools, such as Pure Data, and standardized technologies, including Ginga, NCL, and Lua, which enabled the consistent orchestration of

heterogeneous devices. The declarative nature of NCL, in particular, streamlined the aggregation of media and effects, thereby supporting a coherent and integrated system composition.

The proposal also fostered creativity, particularly in the compositional process, where the integration of the three primary categories of information underpinning the environment proved to be essential. The aggregation of these elements into a coherent flow ensured artistic consistency, resulting in a clear and transparent narrative progression while preventing perceptible discontinuities across the different media modalities.

I.4.6 Device Requirements

The experimental architecture designed for the Io3MT environment was structured through the integration of multiple heterogeneous devices. To varying extents, all these components rely on embedded electronics: whereas traditional musical instruments capture and process electrical pulses, the sensory renderer devices receive information through the network to perform specific actions, such as emitting light or dispersing scents.

In the scope of wireless communication, the sensory rendering devices operated over Wi-Fi, using protocols such as TCP and HTTP to receive real-time instructions. The electric guitar and microphone, despite relying on wired connections, were integrated into the digital environment through the audio interface, which transmitted the audio signals to the computer for subsequent dissemination across the network.

The system's capacity for sensing and actuation is likewise evident across the devices. The microphone and electric guitar can be classified as auditory sensors, responsible for capturing acoustic signals, whereas the sensory diffusers function as actuators, materializing visual and olfactory stimuli in synchrony with the multimedia and musical streams.

Device cooperation also emerged as a key feature of the environment. Multimedia actions executed on the laptops triggered responses in the sensory diffusers, thereby enabling the integrated articulation of audio, video, light, and scent within a coherent narrative. Furthermore, the devices operated simultaneously across multiple processes and were individually identifiable and addressable, either through IP (for networked devices) or through logical input and output ports (for the audio interface and Pure Data patches). This organizational structure ensured the precise routing of commands, avoiding ambiguities.

The solution demonstrated persistence and reliability, as all devices sustained continuous operation throughout the sessions without degradation or compromise of the artistic process. This robustness was further supported by the use of dedicated power supplies across all devices, which minimized the risk of unexpected interruptions. The architecture also incorporated the principle of loose coupling, as each device preserved autonomy in its operation and interacted only upon explicit request.

The system's communicability was ensured through the clear visual design of the control interfaces,

which conveyed the interaction logic in a transparent manner, thereby facilitating interpretation by both performers and technicians. Another important aspect was the consistent mapping of feedback, which reinforced the causal relationship between gestures and their outcomes. Adjustments in audio volume directly corresponded to changes in the sonic output, chromatic modifications were reflected in the ambient lighting, and scents were released in synchrony with the projected scenes. This coherence significantly strengthened the immersive experience.

The system enabled integration across sound, video, light, and scent. This multisensory and multimodal configuration was central to the artistic proposal, enhancing expressiveness and consolidating the environment as a creative platform for musical and performative experimentation.

I.4.7 Architecture Analysis

In this context, the device layer is defined by both physical elements, such as computers, guitar, microphone, lamp, and aroma diffuser, and digital components, including the patches responsible for integrating the musical instruments into the network and those dedicated to visual art generation. Among the three usage scenarios presented in this thesis, this one features the highest degree of device quantity and heterogeneity.

The network layer handled data addressing using the IP. In this setup, audio signals captured by the microphone were transmitted from computer A (192.168.0.20) to computer B (192.168.0.22) via port 3000, while guitar data were sent through port 3001. At the same time, control data were directed to port 10000. The audio played on computer B could be modified based on the media and sensory effects regulated by the NCL application. Consequently, these control messages were sent back to computer A through port 20000. The clear separation of these communication channels provided transparency within the system, explicitly indicating which data streams affected specific parameters. This clarity not only allowed for immediate understanding of causal relationships but also created a solid foundation for further analysis of these interactions.

I.5 Final Remarks on Io3MT Environment

The implementation and evaluation of the Io3MT environment presented in this chapter demonstrate the feasibility of integrating heterogeneous devices and multisensory, multimedia, and musical data streams within a coherent artistic ecosystem. The combination of technologies such as Pure Data, NCL/NCLua, the Ginga middleware, and off-the-shelf sensory devices enabled real-time communication and transparent interoperability, supporting both synchronous and asynchronous modes of execution. This configuration underscores the environment's capacity to function not only as an artistic tool but also as an experimental platform for advancing research in Io3MT systems.

The analysis of QoS parameters confirmed that the system achieves low latency, minimal jitter, and stable throughput, thereby fulfilling the fundamental requirements for networked musical practices.

The use of a dedicated router, lightweight communication protocols, and a modular design ensured temporal stability while minimizing the risks of packet loss or network congestion. These results position the environment as a robust prototype, capable of supporting real-time artistic practices with a high degree of reliability and scalability. Future enhancements should incorporate network management functionalities to enable the dynamic integration of new devices, as well as fault-tolerance mechanisms.

An examination of QoE through autobiographical heuristics revealed the effectiveness of the design in fostering creative immersion. The convergence of sound, video, light, and scent enabled a transparent mapping between performative gestures and artistic outcomes. Although certain aspects, like programming in NCL and Pure Data, still require a degree of technical proficiency, the system proved to be both expressive and functional.

In the artistic evaluation, the environment demonstrated significant potential by extending compositional and performative practices beyond the sonic domain, encompassing multimodal and multisensory dimensions. The integration of scents and lights effects, aligned with musical and visual content, contributed to narrative cohesion and intensified the immersive experience. At the same time, certain aspects warrant attention, particularly the cognitive and creative challenges arising from the heterogeneity of devices and media, which require performers to reconsider traditional techniques and adapt to new expressive paradigms.

From a functional perspective, the system fulfilled most of the requirements outlined in the Io3MT framework, including embedded sensing and actuation, distributed processing, interoperability, and real-time communication. Non-functional requirements such as reliability, adaptability, and scalability were also addressed, although the absence of device management mechanisms and context awareness indicates opportunities for further development. Owing to its modular and open architecture, the environment can be readily extended, supporting both large-scale artistic applications and integration into collaborative research initiatives

Among the areas identified as amenable to improvement, a key enhancement involves the incorporation of a management module capable of visualizing interconnected devices and exerting a certain degree of control over them. Such capabilities may range from modifying their functional properties to ensure system compliance, to establishing or terminating connections between nodes. At a more advanced level, a prospective refinement entails endowing participating elements with intelligence and adaptive capabilities, thereby enabling them to adjust dynamically to contextual demands.

To the best of the author's knowledge, no similar works have been documented in the literature that employ the Ginga-NCL system in artistic applications. Furthermore, this use case introduces the novelty of incorporating a new media and a new player into the system (Pure Data), enabling real-time audio manipulation, alongside the development of a new Pd library, GingaPD.

Since Ginga-NCL constitutes the official digital television standard in Brazil and has also been adopted by 16 other countries across Latin America and Africa, the system is assured of continued evolution

and long-term support, thereby guaranteeing the longevity of the technologies presented herein. Consequently, the proposal may enable the creation of musical systems and technologies designed for sustained and permanent operability. In addition, the environment exemplifies the possibilities and advantages of employing off-the-shelf components in artistic creation grounded in the principles of Io3MT.