"It Takes Two" - Shared and Collaborative Virtual Musical Instruments in the Musical Metaverse

Alberto Boem
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
alberto.boem@unitn.it

Damian Dziwis
KreativInstitut.OWL
Detmold University of Music
(TU Berlin / TH Köln)
Detmold. Germany
damian.dziwis@hfm-detmold.de

Matteo Tomasetti
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
matteo.tomasetti@unitn.it

Sascha Etezazi
KreativInstitut.OWL
Detmold University of Music
Detmold. Germany
sascha.etezazi@hfm-detmold.de

Luca Turchet
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
luca.turchet@unitn.it

2024 IEEE 5th International Symposium on the Internet of Sounds (IS2) | 979-8-3503-6652-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IS262782.2024.10704079

***Abstract*—The relevance and technical possibilities of Shared Virtual Environments (SVEs) are constantly growing as part of what is known as the Metaverse. This includes software and web platforms for creating SVEs and the availability of hardware for experiencing these environments. SVEs offer unique capabilities that have yet to be explored, especially in music. In this paper we explore the concept of networked Virtual Musical Instruments (VMIs) for the Musical Metaverse, where virtual spaces are specifically designed for musical collaboration and social interactions. We describe three prototypes for shared, collaborative VMIs that incorporate specific features of SVEs, such as spatial audio, data sonification, and embodied avatar-based interactions. We conducted a user study to investigate how these instruments can support creativity and usability and to what extent they can deliver a sense of social presence and mutual engagement between users. Finally, we discuss how the three implementations of the proposed shared and collaborative instruments provide novel avenues for music-making in the Metaverse. Our results show that the three instruments exhibit varying degrees of creativity and usability. However, instruments that employ symmetrical and embodied interactions better support social presence and interdependence among users.**

***Index Terms*—Musical Metaverse, virtual musical instruments, networked music performances.**

I. Introduction

Throughout history, musical instruments were designed not only to be played by a person alone, but some required two or more people to play them. One notable example related to Western music is the Organistrum, a precursor of the hurdy-gurdy from the IX century that must be used by two people: one to turn a crank and the other to press the keys [1]. Another is the pipe organ that - until the 19th century - was operated by one or more people whose provided air to the keyboard player by physically pressing a series of bellows [2]. In addition to these purely pragmatic aspects, such as the need for more than one person to operate the mechanics of a complex device, such kind of collaborative and shared instruments can facilitate the exploration of novel musical concepts through interactions among performers [3] [4] [5]. With internet-based music, shared and collaborative instruments emerged as systems that can facilitate the exploration of novel musical concepts by connecting geographically displaced users [6].

With the term "multi-user instruments", we refer to a particular instance of digital musical instruments that are performed simultaneously by multiple people at the same time [5]. However, compared to single-user instruments, such kind of many-people instruments have been relatively under explored since they pose a novel design challenge: they have not only to facilitate the interactions between performers and their instruments, but also between performers themselves [3] [5] . Moreover, they pose the issue of shared-control, since performers can simultaneously control the parameters of a single interface (e.g., [7] [8]).

One particular technological development that provides the scope for this work is that of the Musical Metaverse (MM) [9]. Consisting of networked, persistent, social, and immersive Virtual Environments (VEs) [10] , the Metaverse provides the possibility of supporting Shared VEs (SVEs) [11]. Here, displaced users can collaborate in a wide range of activities, including music. Thus, the Metaverse forms a new basis for collaborative music-making, in particular, for Networked Music Performances (NMPs) [12] [13] [14] where two or more players can perform together in the same virtual space where the audience can also participates in [15].

Being composed of virtual three-dimensional environments, the Metaverse offers countless possibilities for the visual de-

sign of environments such as stages and other virtual objects. Combined with spatial audio, they also enable the artistic application of spatial composition techniques [16]. Using such capabilities of VEs together with the possibilities of real-time digital sound synthesis and Virtual Reality (VR) hardware provides the basis for the development not only of novel Virtual Musical Instruments (VMIs) [17] [18] but also for live performances [19] and music-making in general [20].

In this paper, explore the concept of shared, collaborative instruments combined with the possibilities offered by the Musical Metaverse. For this purpose, we have developed three prototypes of VMIs to be used by two players simultaneously. Each of these prototypes corresponds to an instrument concept that incorporates specific unique capabilities of VEs, as also found in other shared VMIs (see II-B): *i)* the first incorporates the extended possibilities of virtual spaces and sound spatialization; *ii)* the second uses the extended abilities for sonification of the relations between users and virtual spaces; *iii)* the third explores embodied social interactions by using avatars as musical interfaces. Through such prototypes, we aim to explore MM's possibilities for creating unique, shared, and collaborative musical experiences. We evaluate these instruments in terms of usability and creativity support. Moreover, being collaborative and shared, we evaluate these in terms of how they afford social presence. This refers to the sense of being with others, and it encompasses the perception and experience of mutual awareness and real-time interaction with others, contributing to the authenticity of social interactions in virtual spaces [21].

The remainder of the paper is structured as follows. We first introduce the idea of shared collaborative instruments and the current state of VMIs in the context of multi-user immersive environments. Second, we discuss the design and conceptualization of the three VMIs prototypes and their technical implementation. Third, we present the evaluation of the implemented concepts for shared collaborative instruments in the Metaverse through the conducted user study. Finally, we provide a critical reflection on the achieved results and a discussion of future avenues.

II. Background

The Musical Metaverse represents a space for collaborative music-making since it allows group participation, social interactions, and real-time playing. We provide a survey of the most relevant works related to the topics addressed in the present study.

A. Collaborative and Shared Musical Instruments

Collaborative musical instruments allow more than one person to interact in a musical context. They are designed to be played by more than one player, with the goal of exploring communication and expression between players [22]. According to Jordà [5], multi-user instruments facilitate responsiveness and interaction not only between performers and their instruments, but mostly between performers. For Blaine and Fels [3], the quality of the experience of using a collaborative music instrument takes precedence over the music produced. Such instruments should be designed to be approachable by experts or novices [4]. Moreover, the key to a satisfactory user experience is social interactions since collaborative instruments can foster a sense of communication and connection with others. In the context of collaborative musical composition, getting in tune with others is a crucial component of creative engagement and group flow [23]. Existing examples of collaborative instruments are instruments where the same instrument is shared between two performers [24] or where several players can interact together through multi-touch and tangible surfaces [25] [7] [26]. At the same time, such instruments can resemble public installations and can depart from the canonical idea of an instrument and musical interface, like in the case of the SoundNet [27]. Here, performers can collaborate in creating a music improvisation by climbing a net made of sensorized ropes. A particular type of instrument is represented by the ones that use the internet through the use of web browsers and web-based applications that allow collaboration between geographically displaced users [28] [6] [29].

B. Shared Virtual Environments

Characteristics of VEs such as three-dimensional visual design, embodiment through avatars, data sonification, and sound spatialization enable novel possibilities for composition and music performances [20], [30]. Game engines have been explored to develop musical environments for multiple users, including concepts of player-based sonifications and AI-driven characters [31], [32]. Cerqueira et al. even used an unmodified game to sonify players' actions as musical material [33]. Further developments investigated SVEs in VR for musical applications [8], [34]. In "PatchWorld"[1], multiple users can create new VMIs in VEs and play them together.

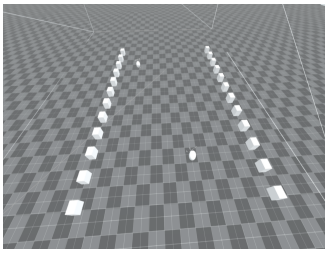
Apart from multi-user musical environments based on game engines, the idea of collaborative and shared instruments was also explored in the MM. A historical example comprises the instruments developed by the Avatar Metaverse Orchestra [35] in Second Life. Recent works have started to explore VMIs in web-based SVEs by using WebXR. "VERSNIZ" presents a collaborative audio-visual live coding system for generating virtual worlds in real-time by placing individual spatially distributed audio-visual fragments [36]. The "Musical Metaverse Playgrounds" explored the possibility of creating shared synthesizers in web-based Metaverse environments [37]. However, even though the aforementioned implementations provide SVEs with multi-user VMIs or VMIs that allow NMPs, the idea of collaborative music using VMIs in the MM is still under-explored [38].

C. Social Presence

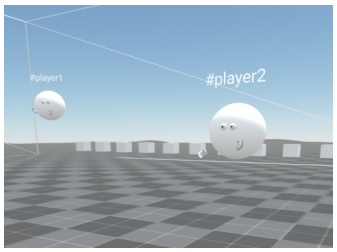
Social presence is a critical aspect of the experience within networked environments. It has been defined as the sense of *"being (somewhere) together"* with other people [39], and

[1]PatchWorld, https://patchxr.com/, accessed: 2024-07-08

it depends on a person's perception of having access to the thoughts and emotions of another [40]. Differently from other communication media, the Metaverse supports a variety of social cues through visual, audio, and haptic information channels [41]. Previous studies showed that in virtual environments several components of the experience can influence social presence, such as the representation of users, interactivity, the tasks, and the quality of visual display and audio (e.g., [42], [43]). However, a shared and collaborative virtual instrument is not a communication tool but a means of expression and co-creation. Therefore, it is necessary to understand how different technological features influence perceptions of social presence to inform the design of VR platforms. Social presence is an element largely under-explored not only in NMP contexts but also in multi-user VEs. Nevertheless, it is an essential element to explore, especially in the context of the MM.

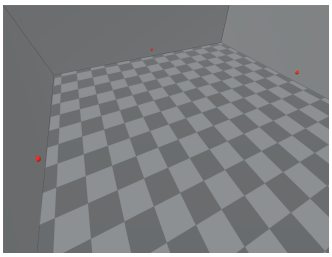


(a) The spatially distributed interface of the instrument.

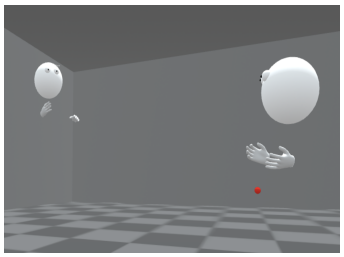


(b) Two participants interacting in the environment.

Fig. 1: The "Spatial Instrument."



(a) The environment and reference points of the instruments.



(b) Participants using the instrument.

Fig. 2: The "Sonification Instrument."

III. Design and Implementation

To explore the possibilities, and implications of shared, collaborative virtual instruments in Metaverse environments, we implemented three collaborative VMIs prototypes for multiple players[2]. To effectively evaluate the prototypes in a user study (see IV), we limited the prototypes to be used by two players In a NMP setting, both players can connect to a web-based Metaverse environment via a browser from different locations to play the instruments.

[2]The full code for the three instruments is available at https://github.com/CIMIL/It-Takes-Two.

Within the Metaverse environments, players are embodied as simple three-dimensional avatars. A neutral environment was implemented to avoid distractions. In general, these environments allow users to communicate via audio streaming. However, in this study, we did not use this feature, as communication was conducted through a Zoom video conference.

The three instruments are optimized for use with head-mounted displays (HMDs) and accompanying motion-based controllers. They are designed so that they can be played without an extensive introduction. In the conceptualization of the VMIs, three different unique capabilities of VEs were considered:

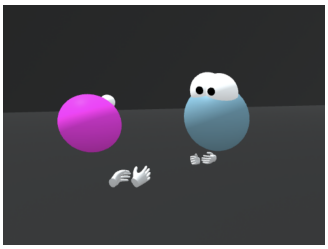
- **Spatial Instrument (SPI)**
  We developed this instrument to explore sound spatialization properties in the context of shared VMIs. We implemented a polyphonic ambient sound synthesizer that simultaneously generates high-range (C4-B4) and low-range (C2-B2) ambient sounds. The interface, resembling a two-part keyboard that divides the tonal ranges into semitone steps based on low and high registers, was positioned on two opposite, widely separated sides of the virtual space (see Fig. 1).
  The instrument emits the generated sound as a spatialized virtual sound source that has no fixed position in space. The first player ("#player1") plays tones on the spatially distributed interface, while the second player ("#player2") controls the spatial composition by moving the sound source. Therefore, the movement/position of the second player also determines the movement/position of the sound source.

- **Sonification Instrument (SOI)**
  This instrument uses a sonification approach to control a sound synthesizer. Such an approach is used to enable players to generate music through their own exploratory movements without relying on visual interfaces (see Fig. 2). To demonstrate this concept, we implemented a synthesizer that uses a single sawtooth oscillator to sonify different characteristics of the avatars and their relations with the virtual space:
  - The distance (D1) between the avatars of "#player1" (P1) and "#player2" (P2) is interpreted as the wavelength of the oscillator.
  - A first-order low-pass filter is applied to the oscillator. The cutoff frequency is controlled by the players' distances to the two red points on opposite walls.
  - The oscillator is also frequency modulated. The modulation frequency is determined by the distance of P2 from the third point on the wall.
  - The modulation depth, ranging from 0 to 1, is controlled by the height of P2 from the ground.
  - The oscillator's volume is controlled by the height of P1.

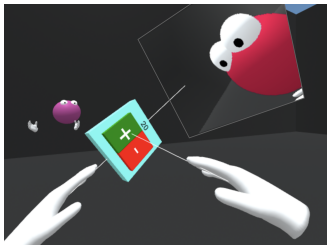
- **Body Instrument (BDI)**
  This instrument was developed to use the avatars as an

interface for collaborative music creation. The interaction is relatively simple: with its virtual hands, one of the users can point and select the avatar's head of the other. When this happens, a percussive-like sound is produced. However, each user can change the size of their avatar. The bigger the avatar, the lower the pitch of the generated sound becomes (see Figure 3a). To achieve this, each user can access a menu on the left hand composed of two buttons, one to increase and one to decrease the avatar's size. A handheld mirror was implemented since users cannot see themselves, as shown in Figure 3b. In addition to interact with the head, users can also touch the hands of the other avatar. When one hand of a user collides with the hand of the other, a sound of random pitch with a short envelope is generated.

In these prototypes, different degrees of interaction were implemented. In SPI, the two players are assigned different roles with different tasks and interaction possibilities; where in SOI, the type of interaction is identical for both players, but they control different parameters of the instruments' sound synthesis. With BDI, both players' roles and interactions are symmetrical. They can change parameters related to their own avatar, which is also reflected in the sound synthesis, but this requires the actions of one player to happen before the other.



(a) Two participants interacting with their avatars.



(b) The handheld menu with the mirror.

Fig. 3: The "Body Instrument."

A. Technical Implementation

All three prototypes were implemented in web-based Metaverse environments, developed using A-Frame[3] and the Networked-Aframe (NAF) library[4]. A-Frame abstracts the programming of 3D environments into a high-level markup language and integrates the WebXR Application Programming Interface (API)[5]. While the resulting VEs can be used with screen-based PCs or mobile devices, WebXR also enables the use of current VR/AR HMDs. The NAF library extents A-Frame for SVE development, by allowing the synchronization of the interactions between users and/or the environment. Therefore, NAF provides adapters for exchanging data via WebRTC or WebSockets. Attributes and states of shared A-Frame components, including user actions, are transferred between users in a peer-to-peer (P2P) network. WebRTC also allows audio and video streams to be transmitted with low latency.

The combination of open-source tools like A-Frame and NAF offers a suitable solution for developing Metaverse environments, providing a strong alternative to commercial platforms that restrict in-depth programming. A-Frame/NAF has already been proven in various MM applications such as [15], [36], [37], [44], [45].

For the sound synthesis part of the three prototypes we used the PdXR system developed by Dziwis [46]. PdXR is an implementation of the Pure Data (Pd) visual programming language [47] - widely used for DSP and music programming - adapted for A-Frame/NAF-compatible Metaverse environments. PdXR enabled us to first implement the sound synthesis algorithms for the VMI prototypes in the desktop version of Pd, and then to run them within the A-Frame/NAF Metaverse environments we developed. For sound spatialization, PdXR integrates the Resonance Audio spatializer for A-Frame[6], which is one of the proposed solutions for spatial audio in WebXR [48]. To implement the additional requirements imposed by the VMI prototypes, such as the control of the spatialized virtual sound source, the analysis and communication of data for sonification, as well as the extended avatar interfaces and interaction, additional components for A-Frame/NAF were programmed in JavaScript, which can communicate via interface functions with the Pd patch in PdXR.

IV. Evaluation

We evaluated the three applications using a user-centered design approach [49]. Our evaluation had two main goals. First, we aimed to gather preliminary feedback on the user experience and how these applications support creativity and usability. Second, we sought to understand how the applications facilitate social presence and interactions. Additionally, we collected suggestions for improvements.

A. Participants

We invited 12 participants to test the three instruments (10 males, 1 female, 1 preferred not to say, aged between 19 and 44 years old, mean = 30.46, standard deviation = 7.18). They were recruited through the personal network of the authors. Participants were located in Italy, Germany, and Canada. Participants are professionally involved in the field of music technology, with backgrounds in different types of music styles and genres. All participants provided informed consent. This study complies with the ethical standard of the NIME conference [50].

[3]A-Frame, https://aframe.io/, accessed: 2024-07-08
[4]NAF, https://github.com/networked-aframe/, accessed: 2024-07-08
[5]WebXR, https://immersive-web.github.io/webxr/, accessed: 2024-07-08

[6]Google Resonance for A-Frame https://github.com/mkungla/aframe-resonance-audio-component, accessed: 2024-07-08

B. Procedure

Participants were organized in pairs for each evaluation session, based on their availability and location. The sessions were guided and conducted by an experimenter. A total of six pairs were formed. Participants were required to wear a VR HMD provided by the experimenters or belonging to the participants. The HMDs used during the evaluation were the Meta Quest 2, Meta Quest 3, and Meta Quest Pro. Participants and experimenter were either placed in the same building (but separated into different rooms) or in different geographical locations (Italy, Canada, Germany). For testing the instruments, we adopted a methodology based on the *think-aloud protocol* [51], where users interact with the system while verbally describing its functions, and commenting on their experience and usability of each instrument.

At first, the participants and the experimenter joined a Zoom call on their laptops. Here, the experimenter provided a five-minute briefing about the system and explained the procedure and the study's goal. Zoom was used mainly to record each participant's voice during the evaluation sessions. Then, participants were asked to wear their HMD and connect to a Wi-Fi network. Subsequently, an URL for each instrument was sent to them by the experimenter. Participants had to input such URL on the Meta Quest Browser, directly available in their HMD. Afterward, participants were asked to join the VE of the instrument they had to evaluate. Then, they were asked to explore the instrument together with their partner. The experimenter was also present in the VE, but acted only as a facilitator. For each instrument, the evaluation session lasted approximately 10 minutes. The order of presentation of the three instruments was randomized between pairs.

At the end of each evaluation session, participants were asked to remove their HMD and fill out three questionnaires. Questionnaires were devised to investigate the level of creativity supported by the instruments, assess the instrument's usability, and assess the degree of social presence afforded by the instruments. Specifically, the questionnaires used were: *i)* the Creativity Support Index (CSI) [52], a tool used for evaluating the ability of a tool to support users' creativity; *ii)* the System Usability Scale (SUS) [53], a widely used questionnaire for assessing the usability of interactive systems; *iii)* the Networked Minds Social Presence Inventory (NMSPI) [54], used for understanding the dynamics of social presence and its impact on communication and collaboration in technology-mediated environments, such as VR. Finally, participants were asked by the experimenter to provide their thoughts and feedback on the instruments.

V. Results

A. Quantitative Data

Here we report the results of the three questionnaires administered to the participants: the Creativity Support Index, the System Usability Score, and the Networked Minds Social Presence Inventory. The results were analyzed by following the guidelines provided by the authors of the respective questionnaires [52] [55] [54]. We used generalized linear mixed effects models to assess differences among the systems evaluations. For each of the created models, the assumption of normally distributed residuals of the data was visually verified.

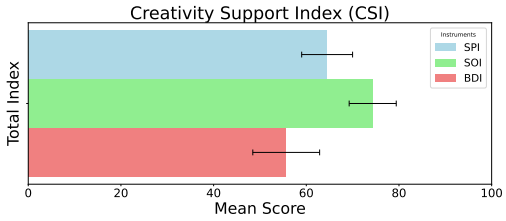


Fig. 4: Mean and standard error of the total Creativity Support Index for the three instruments.

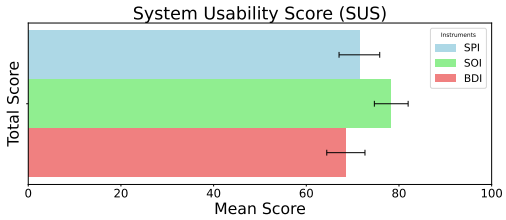


Fig. 5: Mean and standard error of the total System Usability Scale for the three instruments.

1) Creativity Support Index: The CSI metric, ranging from 0 to 100, is used for evaluating a tool's ability to support users' creativity. An aggregated CSI score below 50 indicates that the tool analyzed does not fully support creativity, while a score above 90 indicates excellent support for creativity. The resulting index was calculated according to the guidelines [52], and it is presented in Figure 4. SPI obtained an average CSI score of 64.5 (SD = 19.02), SOI obtained an average score of 74.33 (SD = 17.59), and BDI obtained an average CSI score of 57.72 (SD = 21.24). We performed an ANOVA on the results of a generalized linear mixed effects model (having subject as a random factor and instrument as a fixed factor). The results revealed that the difference between the three instruments was not statistically significant.

2) System Usability Scale: The SUS metric assesses the usability of a system using a scale from 0 to 100. An average SUS score of about 68 is considered to be a benchmark for usability [56]. Results were analyzed according to the guidelines [53]. Results are presented in Figure 5. SPI obtained an average SUS score of 71.45 (95% confidence interval: [62.85; 80.06]), which is slightly above average. SOI obtained an average SUS score of 78.33 (95% confidence interval: [71.18; 85.47]), which is above average. BDI obtained an average SUS score of 68.54 (95% confidence interval: [60.46; 76.62]), which is around average. We performed an ANOVA on the results of a generalized linear mixed effects model (with subject as a random factor and instrument as a fixed factor). The results revealed no statistical significance between

the three instruments.

3) Networked Minds Social Presence Inventory: The NM-SPI assesses the sense of social presence in mediated interactions. It includes two main dimensions: Perception of Self (PoS), which measures how individuals perceive their own presence and engagement within the interaction, and Perception of the Other (PoTO), which evaluates how individuals perceive the presence and engagement of others in the interaction. First, we calculated the total score for *Social Presence* by averaging the sum of all of the items of the questionnaires for each user. Then we calculated the average score for the three main scales of the inventory, defined as *Co-Presence*, *Perceived psychological engagement*, and *Perceived Behavioral Interdependence* for both dimensions of PoS and PoTO. We also calculated the *Subjective Symmetry* [54]. The results of the total score are shown in Figure 6. Figure 7 depicts the results of the three main scales.

Perception of Self: Overall, the mean *Social Presence* for SPI was 3.62 (SD = 0.559), for SOI was 4.287 (SD = 0.908), and for BDI was 4.394 (SD = 0.773). We performed an ANOVA on the results using a generalized linear mixed effects model (with subject as a random factor and instrument as a fixed factor). The analysis revealed a significant effect of the instrument, $\chi^2(2) = 12.535$, p = 0.001897. This suggests significant differences in the *Social Presence* scores across the different instruments. We then conducted post hoc tests using pairwise comparisons with Tukey correction. A statistical difference was found between BDI and SPI (p = 0.0095) and between SOI and SPI (p = 0.0263). For *Co-Presence*, the resulting averaged score of SPI was 3.68 (SD = 0.478), for SOI was 3.88 (SD = 0.547), and for BDI was 5.38 (SD = 0.63). For *Perceived psychological engagement* the averaged score for SPI was 3.7 (SD = 0.89), for SOI was 4 (SD = 1.09), for BDI was 4.18 (SD = 0.92). For *Perceived Behavioral Interdependence* the averaged score for SPI was 3.47 (SD = 1.26), SOI was 5.02 (SD = 1.64), BDI was 5.41 (SD = 1.12). We then performed an ANOVA on the generalized linear mixed effects models, one for each of the three sub-scales (with subject as a random factor and instrument as a fixed factor). The analysis revealed a significant effect of the instrument for *Perceived Behavioral Interdependence* $\chi^2(2) = 16.658$, p = 0.00024. This suggests that there are significant differences in scores across the three instruments. We then conducted post hoc tests using a pairwise comparisons with Tukey correction. A statistical difference was found between BDI and SPI (p = 0.0024) and between SOI and SPI (p = 0.0144).

Perception of the Other: The total average score for *Social Presence* regarding SPI was 3.78 (SD = 0.58), for SOI was 4.33 (SD = 0.85), and for BDI was 4.34 (SD = 0.76). We performed an ANOVA on the generalized linear mixed effects model (with subject as a random factor and instrument as a fixed factor). The analysis revealed no significant effect. For *Co-Presence*, the resulting averaged score of SPI was 3.86 (SD = 0.48), for SOI was 4 (SD = 0.68), and for BDI was 3.56 (SD = 0.63). For *Perceived psychological engagement* the averaged score for SPI was 3.76 (SD = 0.84), for SOI was 4.01 (SD = 1.1), for BDI was 4.15 (SD = 0.94). For *Perceived Behavioral Interdependence* the averaged score for SPI was 3.72 (SD = 1.3), SOI was 4.97 (SD = 1.67), BDI was 5.33 (SD = 1.13). We performed an ANOVA on the results, employing a generalized linear mixed effects model (with subject as a random factor and instrument as a fixed factor). The analysis revealed a significant effect of the instrument for *Perceived Behavioral Interdependence* $\chi^2(2) = 10.245$, p = 0.00596. This suggests that there are significant differences in scores across the different instruments. We then conducted post hoc tests using a pairwise comparisons with Tukey correction. A statistical difference was found between BDI and SPI (p = 0.0150).

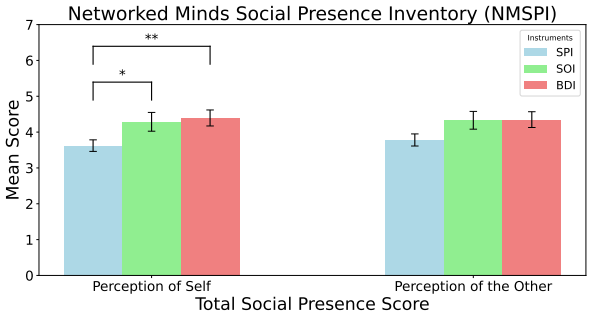


Fig. 6: Mean and standard error of the total Social Presence score for the NMSPI. On the left the results of the Perception of Self, on the right the results of the Perception of the Other. The statistical difference is presented with the symbol * corresponding to $p < 0.05$, and ** corresponding to $p < 0.01$.

Subjective Symmetry: Subjective symmetry measures social presence from the perspective of a user. For this, we computed a Pearson correlation coefficient that was calculated from the mean total scores of PoS and PoTO for each of the three dimensions. We then performed on the coefficients a Fischer Z-Transformation, to use the correlation values for significance testing. We calculated both the p-value and Bonferroni-adjusted p-value. Results are summarized in Table I. Regarding total *Social presence*, we found a strong correlation between PoS and PoTO for SOI only (*adjusted p-value* = 0.042). For *Perceived psychological engagement*, the correlations between the two perceptual dimensions of self and others revealed a strong correlation for SOI (p = 0.004) and BDI (p = 0.012). For *Perceived behavioral interdependence*, SOI shows a strong correlation between PoS and PoTO (p = 0.013).

B. Qualitative Data

Participants' comments were analyzed using an inductive thematic analysis [57] based on Grounded Theory [58]. Through this analysis, conducted by the authors, we generated codes that were further organized into the following themes that reflected shared patterns. We have identified two macro-themes: one including specific themes regarding each

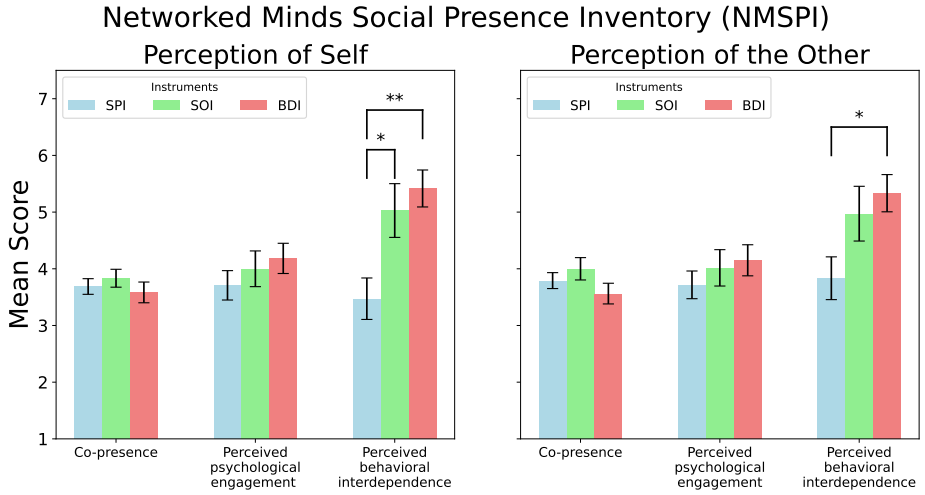


Fig. 7: Mean and standard error of the three main scales of the NMSPI: Co-presence, Perceived psychological engagement, and Perceived behavioral interdependence. On the top are the results for the Perception of Self, and on the bottom are the ones for the Perception of the Other. The statistical difference is presented with the symbols * corresponding to $p < 0.05$ and ** to $p < 0.01$.

individual instrument, and the other encompassing themes shared by all instruments.

Instrument-specific Themes

SPI - *Unclear difference in functionalities and roles*: Several users perceived the decoupling of roles as problematic in this instrument. Since one player selects the notes and the other moves the sound source within the VE, this results in a perceived imbalance between the players regarding their roles (e.g., *"I do not quite understand what my role is and what the other player's role is* [...] *"It seems confusing to me; this does not allow me to interact with each other and create something musical."*). In addition, the fact that only one player hears the spatialized binaural audio while the other does not, created a mismatch that some participants disliked.

SOI - *Mapping movement to sound promotes exploration:* In this instrument, the relationship between movement in space and sound positively influenced the participants' experience. For example, one participant noted: *"This is the one that is taking me the most because the player can try particular combinations based on where the player is in the space* [...] *it greatly allows me to explore this synth in space"*. Participants reported that the mapping used in this instrument enabled them to be more engaged in their activity, and motivated to explore the timbral possibilities offered by the instrument. While the mapping between the distance among users and pitch was immediately understood by most participants, some showed difficulties in understanding the other mappings and their functions. One participant expressed this confusion in the following terms: *"I do not understand the relationships between flying and sound."*.

BDI - *Embodied interactions:* Most participants noted that the gestural interactions of this instrument reminded them of drum circles. Since the body itself becomes the instrument, this facilitated more direct and engaging collaborations among musicians (e.g., *"You do the kick drum, I do the snare"*, *"I like that fact that by changing the size of the body I change the pitch of the instrument [...] is very engaging"*).

Perceived Latency: Participants reported experiencing some latency between gestures and sound. Even minimal, this latency negatively affected the instrument's usability, particularly hindering their ability to create complex rhythms (e.g., *"creating a tempo is difficult"*, *"I was synchronizing with my gestures, not with the sound"*).

Cross-instrument Themes

The importance of timbre and sound design: Most participants reported that to enhance the use of the systems, the timbral qualities of the instruments should be carefully considered, as they influence the experience as much as the interactions do. For instance, when referring to SPI, one participant explained that *"the continuous sound of this instrument is bothering me, even if the other changes the pitch"*).

Importance of visual feedback: Some participants reported that in SVEs, it is essential to have more detailed visual feedback, which could increase engagement and the overall user experience. For example, one participant said, *"I would like the cubes to change when I touch or interact with them so that I have a visual response to the interaction I am making."* Additionally, some participants noticed that locating the avatars in the environment was challenging when their partner moved too far away. Well-designed visual feedback should provide users with such spatial information.

Interactions and relationships with space: Some participants reported that in the three instruments, the relationship between their virtual bodies with the space and the sound is extremely intertwined that more references and guides are needed. Therefore, each interaction needs to be better contextualized within

the purpose of the developed instrument (e.g., *"Different combinations of actions and social interactions really has an impact on the final timbre"*, *"If my partner does something wrong (going too far from me) the complexity and quality of the sound diminish, therefore we need to really discuss and cooperate together to make the instrument sound good"*).

TABLE I: Pearson Correlation Coefficient, z-score, and p-value for correlations between Perception of self and Perception of the other, for Total Social Presence, Co-Presence, Perceived Psychological Engagement, and Perceived Behavioral Interdependence. Statistically significant correlations (with *adjusted p-value* < 0.05) are highlighted in yellow.

Total Social Presence				
Instruments	***Pearson Corr. Coeff.***	***z-score***	***p-value***	***adj. p-value***
SPI	**0.876**	**1.359**	**0.174**	**0.523**
SOI	**0.985**	**2.458**	**0.014**	**0.042**
TMI	**0.902**	**1.482**	**0.138**	**0.415**
Co-presence				
Instruments	***Pearson Corr. Coeff.***	***z-score***	***p-value***	***adj. p-value***
SPI	**0.769**	**1.018**	**0.309**	**0.926**
SOI	**0.761**	**0.998**	**0.318**	**0.955**
TMI	**0.417**	**0.444**	**0.657**	**1.000**
Perceived Psychological Engagement				
Instruments	***Pearson Corr. Coeff.***	***z-score***	***p-value***	***adj. p-value***
SPI	**0.932**	**1.672**	**0.095**	**0.284**
SOI	**0.997**	**3.178**	**0.001**	**0.004**
TMI	**0.994**	**2.888**	**0.004**	**0.012**
Perceived Behavioral Interdependence				
Instruments	***Pearson Corr. Coeff.***	***z-score***	***p-value***	***adj. p-value***
SPI	**0.882**	**1.384**	**0.167**	**0.500**
SOI	**0.993**	**2.857**	**0.004**	**0.013**
TMI	**0.936**	**1.706**	**0.088**	**0.264**

VI. Discussion

The Creativity Support Index (CSI), System Usability Scale (SUS), and Networked Minds Social Presence Inventory (NMSPI) revealed key performance differences among the instruments. These insights contribute to refining the design and usage of collaborative and shared virtual instruments in the Musical Metaverse.

The CSI scores show varying levels of support for creativity across the three instruments, with SOI's scoring higher than SPI and BDI. Regarding usability, the SUS score reveals moderate differences among the instruments. All scores are above the average usability benchmark threshold, with SOI exhibiting the highest usability of the three, followed by SPI and BDI. However, these differences between instruments' scores for CSI and SUS are not substantial enough to conclusively determine one instrument's superiority over the others.

Qualitative analysis further elucidates such aspects of creativity and usability. For SOI, the overall mapping between the avatars' position in space and the sound was considered intuitive and positively influenced the overall participants' creative experience. However, some users struggled to understand all the different mapping strategies used precisely. For SPI, participants reported that confusion over their roles and functionalities hindered their creative process. For BDI, users found the embodied interactions engaging and intuitive; however, latency issues negatively impacted these interactions, affecting the overall user experience. These issues primarily disrupted synchronization between users, crucial for the rhythm-based music creation that BDI facilitates. The timbral quality of the sound generated by the instruments also limited participants' creativity. In BDI, participants noted that the sound generated by the two users was too similar. In SPI, the sound was unengaging and difficult to control. Conversely, SOI's timbral complexity received positive evaluations, enhancing engagement and promoting a sense of flow during the task.

The NMSPI results offer another perspective on the characteristics of the three instruments, highlighting how different social interactions impact the experience of making music together within the evaluation task. The overall *Social presence* scores reveal significant differences among the instruments, indicating participants' varied evaluation of their own sense of presence. BDI is characterized by the strongest sense of *Social presence*, especially in terms of *Perceived Behavioral Interdependence*, with SOI following closely. The analysis of the symmetry between the PoS and PoTO shows that SOI strongly correlates not only in terms of *Social presence* but also regarding *Perceived Psychological Engagement* and *Perceived Behavioral Interdependence*. With SOI, users felt more mutually aware and behaviorally interdependent, leading towards a more balanced perception of presence.

By looking at the aggregated results, we can observe that SOI excelled across multiple metrics, establishing itself as a versatile instrument that can enhance creativity, usability, and social presence. BDI stood out in terms of social presence, effectively promoting mutual awareness and behavioral interdependence. Conversely, while satisfactory, SPI did not meet the levels of SOI and BDI, particularly in fostering social presence.

The relationship between users and their roles critically influenced the experience of the three musical instruments. Analysis of social presence showed a clear distinction between SPI and the other two instruments. In SOI and BDI, each user's functions are equal, and their roles are symmetrical. In SOI, the spatial relationships (such as position) existing between the users determine the sound. In BDI, their roles adhere to a musical metaphor reminiscent of a "*drum circle*" fostering a "*call and response*" type of interaction. Furthermore, controls are integrated into the avatars' physical characteristics, such as their size. Thus, these instruments encourage a high level of interdependence between users, as the sound heavily relies on their relational dynamics, that is, spatial positioning in SOI and avatar size in BDI. Conversely, SPI disrupts this symmetry as it clearly divides functions between users: one is responsible for creating melodies while the other handles sound spatialization. Although this division did not significantly affect creativity and usability, it did foster a sense of separation and hindered mutual understanding between users.

This is an aspect that should be considered in the design of shared and collaborative virtual instruments. Since in the

Metaverse relationships are inherently social, the roles that users assume and how the functionalities are shared impact how the instruments are experienced and used and how users relate with each others.

Based on participants' feedback and observed behaviors, we recommend that instruments similar to SPI be used for pre-composed or choreographed musical pieces, where the relations and roles that users assume must be defined and known in advance. Such instruments may better suited for more experienced players. Conversely, systems like those of SOI and BDI appear to be more ideal for novices. Their embodied and situated interactions, along with symmetrical user relationships, appear to facilitate impromptu improvisation and exploratory activities, where a preliminary understanding of roles is not necessary. The difference between these instruments showed how interdependence between users plays a vital role in the use of shared and collaborative virtual instruments. However, collaborative and shared instruments in the Metaverse should not omit the importance of mapping between actions and sound, as well as the importance of multimodal feedback to enhance social presence and increase the understanding of the instrument. However, these topics have not been fully explored in the MM and in multi-user immersive musical environments in general. Therefore, they deserve further attention that goes beyond the scope of our work.

However, our study presents some limitations. First, we surveyed only 12 participants, divided into 6 pairs. The limited testing time and the exploratory nature of the tasks might explain the results for CSI and SUS. Participants were predominantly male. While this exploratory study does not aim for generalizability, future research should ensure more balanced demographics [59]. Second, using questionnaires to measure social presence, particularly in musical tasks within VR settings, presents constraints, as discussed in [60]. Future studies should consider a mixed-methods approach. Third, we did not measure or account for system and network latency, since the aim of our work was to implement and test a specific concept. We found that only for BDI the performance of few participants was effected by latency. Therefore, future work should better characterize network latency for VMIs based on NAF and PdXR, especially when using Wifi connections. This might provide a more complete understandings of the limitations and possibilities offered by these tools in the context of the MM.

VII. Conclusion & Future Work

We presented the design and evaluation of three prototypes of shared and collaborative VMIs exploring different aspects of the Musical Metaverse, such as spatialization, sonification, embodied and social interactions. Our study revealed that while different instruments exhibit different degrees of support for creativity and usability, social presence was influenced by the role and the shared functionalities that characterize each instruments. Among the three VMIs, SOI performed reasonably well regarding creativity support, usability, and social presence, while BDI stood out in promoting mutual awareness and behavioral interdependence. In contrast, SPI revealed that dividing an instrument's functionalities between users hinders social presence. This suggests that designs promoting symmetrical user roles and embodied interactions are more effective for collaborative and shared VMIs in the Musical Metaverse.

Regarding the technical implementation, we were able to realize all conceptual ideas of the instrument prototypes using the A-Frame/NAF and PdXR systems. However, the need for an interface concept for the communication of scripts/elements of the Metaverse environment with the Pd patch, as well as an optimization of the latency during the interaction, became apparent. These improvements are planned for the future development of PdXR.

Acknowledgment

We would like to thank all the participants in the user study and Samuel Johnstone for proofreading this paper. The authors Alberto Boem, Matteo Tomasetti, and Luca Turchet received support from the MUR PNRR PRIN 2022 grant, prot. no. 2022CZWWKP, funded by Next Generation EU. The work by Damian Dziwis and Sascha Etezazi was realized as a part of KreativInstitut.OWL, funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia, Germany.

References

[1] G. Severini, A. Orlando *et al.*, "Organistrum in santiago de compostela: Symphonia coelestis," *Mediterranean Archaeology and Archaeometry*, vol. 18, no. 4, pp. 345–345, 2018.

[2] S. Jeans, "The pedal clavichord and other practice instruments of organists," in *Proceedings of the Royal Musical Association*, vol. 77. Cambridge University Press, 1950, pp. 1–15.

[3] T. Blaine and S. Fels, "Contexts of collaborative musical experiences," in *Proceedings of the 2003 conference on New interfaces for musical expression*, 2003, pp. 129–134.

[4] S. Fels, "Designing for intimacy: Creating new interfaces for musical expression," *Proceedings of the IEEE*, vol. 92, no. 4, pp. 672–685, 2004.

[5] S. Jordà, "Multi-user instruments: models, examples and promises," in *Proceedings of the 2005 conference on New interfaces for musical expression*, 2005, pp. 23–26.

[6] A. Barbosa, "Public sound objects: a shared environment for networked music practice on the web," *Organised Sound*, vol. 10, no. 3, pp. 233–242, 2005.

[7] M. Kaltenbrunner, S. Jorda, G. Geiger, and M. Alonso, "The reactable*: A collaborative musical instrument," in *15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06)*. IEEE, 2006, pp. 406–411.

[8] L. Men and N. Bryan-Kinns, "LeMo: exploring virtual space for collaborative creativity," in *Proceedings of the 2019 on Creativity and Cognition*, 2019, pp. 71–82.

[9] L. Turchet, "Musical Metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, 2023.

[10] S. Mystakidis, "Metaverse," *Encyclopedia*, vol. 2, no. 1, pp. 486–497, 2022.

[11] R. Waters and J. Barrus, "The rise of shared virtual environments," *IEEE Spectrum*, vol. 34, no. 3, pp. 20–25, 1997.

[12] P. Oliveros, S. Weaver, M. Dresser, J. Pitcher, J. Braasch, and C. Chafe, "Telematic music: Six perspectives," *Leonardo Music Journal*, vol. 19, pp. 95–96, 2009.

[13] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[14] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Springer, 2016.

[15] D. Dziwis and H. von Coler, "The Entanglement – Volumetric Music Performances in a Virtual Metaverse Environment," *Journal of Network Music and Arts*, vol. 5, no. 1, pp. 1–12, 2023.

[16] M. A. Baalman, "Spatial composition techniques and sound spatialisation technologies," *Organised Sound*, vol. 15, no. 3, pp. 209–218, 2010.

[17] T. Mäki-Patola, J. Laitinen, A. Kanerva, and T. Takala, "Experiments with virtual reality instruments," in *Proceedings of the 2005 conference on New interfaces for musical expression*, 2005, pp. 11–16.

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

[19] B. Loveridge, "An Overview of Immersive Virtual Reality Music Experiences in Online Platforms," *Journal of Network Music and Arts*, vol. 5, no. 1, 2023.

[20] L. Turchet, R. Hamilton, and A. Camci, "Music in Extended Realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[21] F. Biocca and M. R. Levy, *Communication in the age of virtual reality*. Routledge, 2013.

[22] E. R. Miranda and M. M. Wanderley, *New digital musical instruments: control and interaction beyond the keyboard*. AR Editions, Inc., 2006, vol. 21.

[23] N. Bryan-Kinns and F. Hamilton, "Identifying mutual engagement," *Behaviour & Information Technology*, vol. 31, no. 2, pp. 101–125, 2012.

[24] S. Fels and F. Vogt, "Tooka: Explorations of two person instruments," in *Proceedings of the 2002 conference on New interfaces for musical expression*, 2002, pp. 1–6.

[25] T. Blaine and T. Perkis, "The jam-o-drum interactive music system: a study in interaction design," in *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, 2000, pp. 165–173.

[26] R. Laney, C. Dobbyn, A. Xambó, M. Schirosa, D. Miell, K. Littleton, and N. Dalton, "Issues and techniques for collaborative music making on multi-touch surfaces," 2010.

[27] B. Bongers, "Physical interfaces in the electronic arts," *Trends in gestural control of music*, pp. 41–70, 2000.

[28] L. Dahl, J. Herrera, and C. Wilkerson, "Tweetdreams: Making music with the audience and the world using real-time twitter data." in *NIME*, 2011, pp. 272–275.

[29] Á. Barbosa, "Displaced soundscapes: A survey of network systems for music and sonic art creation," *Leonardo Music Journal*, vol. 13, pp. 53–59, 2003.

[30] M. Ciciliani, "Virtual 3D environments as composition and performance spaces*," *Journal of New Music Research*, vol. 49, pp. 104–113, 1 2020.

[31] R. Hamilton, "Q3osc or: How i learned to stop worrying and love the bomb game," *Proceedings of the international computer music association conference.*, 2008.

[32] R. Hamilton, "The Procedural Sounds and Music of ECHO::Canyon," *11th Sound and Music Computing Conference and 40th International Computer Music Conference (SMC/ICMC2014).*, 2017.

[33] M. Cerqueira, S. Salazar, and G. Wang, "Soundcraft: Transducing starcraft 2," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Daejeon, Republic of Korea: Graduate School of Culture Technology, KAIST, May 2013, pp. 243–247.

[34] R. Hamilton and C. Platz, "Gesture-based Collaborative Virtual Reality Performance in Carillon," *Proceedings of the international computer music association conference.*, 2016.

[35] G. Martín, "Social and psychological impact of musical collective creative processes in virtual environments; Te Avatar Orchestra Metaverse in Second Life. Musica/Tecnologia Music." *Technology*, vol. 75, pp. 75–87, 2018.

[36] D. Dziwis, "VERSNIZ – Audiovisual Worldbuilding through Live Coding as a Performance Practice in the Metaverse," in *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research*, Nov. 2023, p. 289–300.

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

[38] A. Boem, M. Tomasetti, and L. Turchet, "Harmonizing the musical metaverse: unveiling needs, tools, and challenges from experts' point of view," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2024.

[39] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Frontiers in Robotics and AI*, vol. 5, p. 409295, 2018.

[40] F. Biocca, "The cyborg's dilemma: Progressive embodiment in virtual environments," *Journal of computer-mediated communication*, vol. 3, no. 2, p. JCMC324, 1997.

[41] T. Hennig-Thurau, D. N. Aliman, A. M. Herting, G. P. Cziehso, M. Linder, and R. V. Kübler, "Social interactions in the metaverse: Framework, initial evidence, and research roadmap," *Journal of the Academy of Marketing Science*, vol. 51, no. 4, pp. 889–913, 2023.

[42] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, pp. 481–494, 2003.

[43] E.-L. Sallnäs, "Effects of communication mode on social presence, virtual presence, and performance in collaborative virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 14, no. 4, pp. 434–449, 2005.

[44] D. Dziwis, H. Von Coler, and C. Pörschmann, "Orchestra: a Toolbox for Live Music Performances in a web-based Metaverse," *Journal of the Audio Engineering Society*, vol. 71, no. 11, pp. 802–812, 2023.

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

[46] D. Dziwis, "PdXR - the Evolution of Pure Data into the Metaverse," in *ICMC 2023 : Proceedings of the International Computer Music Conference*. ICMC, 2023, pp. 52–58.

[47] M. Puckette, "Pure Data : another integrated computer music environment," in *Second intercollege computer music concerts*, 1997, pp. 37–41.

[48] M. Tomasetti, A. Boem, and L. Turchet, "How to spatial audio with the webxr api: a comparison of the tools and techniques for creating immersive sonic experiences on the browser," in *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2023, pp. 1–9.

[49] S. Kujala, "User involvement: a review of the benefits and challenges," *Behaviour & information technology*, vol. 22, no. 1, pp. 1–16, 2003.

[50] F. Morreale, N. Gold, C. Chevalier, and R. Masu, "Nime principles & code of practice on ethical research," 2023.

[51] J. Nielsen, "Estimating the number of subjects needed for a thinking aloud test," *International journal of human-computer studies*, vol. 41, no. 3, pp. 385–397, 1994.

[52] E. Cherry and C. Latulipe, "Quantifying the creativity support of digital tools through the creativity support index," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 21, no. 4, pp. 1–25, 2014.

[53] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[54] F. Biocca and C. Harms, "Networked minds social presence inventory:—(scales only, version 1.2) measures of co-presence, social presence, subjective symmetry, and intersubjective symmetry," 2003.

[55] J. R. Lewis, "The system usability scale: past, present, and future," *International Journal of Human–Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.

[56] J. R. Lewis and J. Sauro, "Item benchmarks for the system usability scale." *Journal of Usability Studies*, vol. 13, no. 3, 2018.

[57] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[58] B. Glaser and A. Strauss, *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[59] T. C. Peck, L. E. Sockol, and S. M. Hancock, "Mind the gap: The underrepresentation of female participants and authors in virtual reality research," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 1945–1954, 2020.

[60] B. Van Kerrebroeck, G. Caruso, and P.-J. Maes, "A methodological framework for assessing social presence in music interactions in virtual reality," *Frontiers in Psychology*, vol. 12, p. 663725, 2021.