

MDPI

Article

# Banging Interaction: A Ubimus-Design Strategy for the Musical Internet

Damián Keller <sup>1,\*,†,‡</sup>, Azeema Yaseen <sup>2,‡</sup>, Joseph Timoney <sup>2,‡</sup>, Sutirtha Chakraborty <sup>2,3,‡</sup> and Victor Lazzarini <sup>3</sup>

- <sup>1</sup> Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA 94305, USA
- Department of Computer Science, Maynooth University, W23 F2H6 Maynooth, Co. Kildare, Ireland; azeema.yaseen.2020@mumail.ie (A.Y.); joseph.timoney@mu.ie (J.T.); sutirtha.chakraborty.2019@mumail.ie (S.C.)
- Department of music, Maynooth University, W23 F2H6 Maynooth, Co. Kildare, Ireland; victor.lazzarini@mu.ie
- \* Correspondence: dkeller@ccrma.stanford.edu
- † Current address: Amazon Center for Music Research (NAP), Federal University of Acre, Rio Branco 69920-900, Brazil.
- ‡ These authors contributed equally to this work.

Abstract: We introduce a new perspective for musical interaction tailored to a specific class of sonic resources: impact sounds. Our work is informed by the field of ubiquitous music (ubimus) and engages with the demands of artistic practices. Through a series of deployments of a low-cost and highly flexible network-based prototype, the Dynamic Drum Collective, we exemplify the limitations and specific contributions of banging interaction. Three components of this new design strategy—adaptive interaction, mid-air techniques and timbre-led design—target the development of creative-action metaphors that make use of resources available in everyday settings. The techniques involving the use of sonic gridworks yielded positive outcomes. The subjects tended to choose sonic materials that—when combined with their actions on the prototype—approached a full rendition of the proposed soundtrack. The results of the study highlighted the subjects' reliance on visual feedback as a non-exclusive strategy to handle both temporal organization and collaboration. The results show a methodological shift from device-centric and instrumental-centric methods to designs that target the dynamic relational properties of ubimus ecosystems.

Keywords: ubiquitous music; mid-air interaction; percussion; adaptive HCI



Citation: Keller, D.; Yaseen, A.; Timoney, J.; Chakraborty, S.; Lazzarini, V. Banging Interaction: A Ubimus-Design Strategy for the Musical Internet. *Future Internet* **2023**, 15, 125. https://doi.org/10.3390/ fi15040125

Academic Editor: Yang Wang

Received: 12 February 2023 Revised: 18 March 2023 Accepted: 22 March 2023 Published: 27 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

How to interact with musical materials has been a persistent and complex issue that emerged with the use of analog equipment, gaining traction as soon as digital systems were able to support synchronous interaction [1]. On the one hand, early acoustic compilers such as MUSIC IVF [2] and MUSIC V [3] furnished a fairly flexible means of dealing with musical decision-making. A general-purpose language (FORTRAN in this case) and a custom data representation could be used to furnish computational instructions that the compiler interpreted to yield a sonic output. Inheriting the vocabulary of acoustic instrumental music, this set of instructions came to be known as a "score". This approach entailed a cumbersome process of either programming or score-writing, code compilation, and finally sonic rendering that could take from several hours to several weeks. On the other hand, the immediacy of the interaction with analog hardware typical of the early electronic music studios took at least two decades to be embraced as a model for computer-based musical interaction. Initial attempts included MUSYS and GROOVE. But viable synchronous interaction was only enabled when both the interface and the synthesis engine became compact and specialized [4].

Future Internet 2023, 15, 125 2 of 17

These two tendencies in musical interaction were described by [1,5] as a paradox of generality vs. strength. Apparently effective musical interaction for creative aims may only be achieved by systems that target a reduced set of sonic techniques. In contrast, tools that target versatility—such as acoustic compilers and music-programming languages—tend to demand strategies of interaction that are not efficient when deployed for synchronous usage. Buxton [1] suggested that this caveat is caused by the lack of a musical theory that can support the representation of any computable musical knowledge. This is an interesting challenge because it is not tied to a particular technology. It implies that if we find representations of musical knowledge that are general enough, we may be able to materialize them by computational means. Some will be doable with current technology. Others will demand specific implementation efforts. Later developments tended to confirm Buxton and Truax's insights. More specialized tools are often easier to deploy in diverse contexts and platforms. For instance, an acoustic compiler that targets almost exclusively audio synthesis and processing—like Faust—bypasses the complications introduced by the instrumental approaches to interaction. Consequently, Faust-based tools can be deployed in mobile, web-based and embedded platforms without the overhead involved in generalpurpose languages [6].

Despite the exponential expansion of tools for music-making triggered by the introduction of personal computers during the early 1980s, representations of musical knowledge that are general and deployable are still rare. Note, instrument, melody, rhythm, harmony and other notions derived from the acoustic instrumental paradigm form the basis for the design of both 'general' and 'specialized' tools (see [7–9] for critical perspectives on the acoustic instrumental views on music-making). These notions are justified when the target is a specific musical genre or the support for musical knowledge linked to a specific instrument or set of instruments, such as the orchestra. However, they make little sense as general representations of a diverse and growing body of musical knowledge aimed at venues such as domestic spaces, transitional spaces, social networks or multimodal formats. The disappearing computer envisioned by [10] has a poignant musical corollary: the disappearance of the acoustic instrument as a metaphor for musical interaction.

Within the realm of the computing methods designed to handle visual, gestural (Our usage of gesture in this paper is strictly tied to the movements of the human body. Particularly during the 1990s and 2000s there was a profusion of proposals that treated musical gestures as semiotic entities. This approach may be applicable to ethnomusicological endeavors or to some discussions in musicology. So far, its application for interaction design is doubtful. Hence, we adopt the current term mid-air to describe the computational processing of contactless human gestures) and sonic information tailored for networked environments, we focus on the emerging issues faced by prototyping for casual musical interaction. Among the targets of ubimus design, casual interaction is an extreme case because it features a large array of unforeseen factors and it tends to preclude the screening of subjects. In this paper, we tackle the development of support for musical activities in everyday settings entailing a change in approach to interaction design. Rather than taking as a point of departure the affordances of acoustic instruments or their emulations, we explore the deployment of adaptive interaction techniques. Furthermore, instead of incorporating a fixed set of instrumental gestures, we propose an open-ended strategy for interaction design.

The research carried out within the field of ubiquitous music (ubimus) (Ubiquitous music is a research field that targets creative activities enabled by technological ecosystems, involving human stakeholders and material resources) [11] indicates that sustainable and accessible metaphors for creative action demand a change of perspective in interaction design. First, ubimus prototyping needs to target low-cost, widely available hardware and open-source tools. Custom-built devices, e.g., data gloves or bodysuits [12], or off-the-shelf equipment such as Kinect and Leap motion are often used for data capture in musical interaction endeavors [13,14]. Although useful for initial deployments and testing tasks, some of these devices tend to have higher costs and shorter life cycles than consumer-level,

Future Internet 2023, 15, 125 3 of 17

standard portable-computer equipment. As a result, these special-purpose devices tend to lose technical support and eventually become discontinued.

Second, ubimus designs need to tackle flexible approaches to interaction without relying on device or music-genre specificities. This strategy has pushed ubimus research toward the implementation of creative-action metaphors that have been successfully applied in everyday settings, engaging with stakeholders of varied profiles including musically untrained subjects [15]. Creative-action metaphors are design strategies that are not linked to any particular musical aesthetic. Thus, they apply to diverse sonic resources and can be tailored for various creative styles. Nevertheless, it is not yet clear whether and how the widely adopted paradigm of acoustic instrumental thinking—involving a strict division between active and passive musical stakeholders—can be replaced. In other words, if the point of departure of the design is not an acoustic instrument, how should we deal with the variability of multiple potential sonic resources and their outcomes?

Third, the restrictions imposed by the COVID-19 pandemic have expanded the awareness of distributed creativity, configuring new forms of musical practices [16,17]. Although the implications for music-making and human–computer interaction are very slowly starting to be acknowledged, these restrictions were immediately felt by the artistic communities, triggering various concerns. Can music-making be supported without increased health risks? Should potentially hazardous cultural rituals—such as instrumental performances in enclosed spaces involving a large number of people—still be encouraged? The ongoing waves of contagion and the reduced access to vaccination in several parts of the world point to the need to invest in remote engagements as a complement to in-place artistic initiatives. How to enable these hybrid forms of participation and resource-sharing is an open avenue of research.

This paper explores casual interaction in transitional settings in the context of music-making. We present the results of deploying a creative-action metaphor based on whole-body, mid-air interaction, the Dynamic Drum Collective [18]. A feature of this new prototype is its ability to adapt to the particularities of its usage. In a sense, the software is tuned by deploying it. By keeping track of the spatial relationships between the parts of the body, the parametric adjustments are kept aligned with the specific demands of the participant's body characteristic. Furthermore, the Dynamic Drum Collective prototype provides an example of a new approach to design, Banging Interaction. We describe the motivations behind this design proposal, highlighting its contributions within the context of the developments that target network technologies for music-making. Instead of focusing exclusively on the technological aspects, we also consider the implications of the musical activities and their influence on the design process. This perspective, grounded on ubimus research, points to a definition of the musical Internet that avoids the pitfalls of genre-oriented designs. Furthermore, our banging interaction proposal may contribute to representing musical knowledge without relying on genre-specific traits.

# 2. A Musical Internet

The initial investigations in music computing involved the use of mainframe hardware [19]. Access to computing facilities was expensive and usage by artists was fairly limited. Thus, the artistic exploration of networked infrastructure only emerged after the appearance of the personal computer. In 1977, The League of Automatic Music Composers gathered several Commodore KIM-1 machines to write interdependent computer compositions across a local network [20] (http://www.sothismedias.com/home/the-league-of-automatic-music-composers, accessed on 11 February 2023). They relied on hardware and knowledge gathered over two decades, including affordable stationary devices and digital—analog converters, data-exchange protocols and acoustic compilers, configuring what later came to be known as a DIY approach (Do It Yourself) to technology. Strongly influenced by the analog—electronic practice of the late 1960s, their artistic approach was mostly improvisational. Their target was not the performance of structured musical works but an ongoing collaborative activity in which each member contributed to a shared sonic

Future Internet 2023, 15, 125 4 of 17

outcome. The League was active until 1986. They quit one year after the debut of a remotenetworking effort, involving two sites and featuring communication via telephone lines.

During the early 1990s, the availability of the Internet and the emergence of standard musical data-exchange protocols stimulated a wider interest in the development of network infrastructure for musical usage. According to [21], evolving technology implies alterations to any taxonomical or theoretical framework applied to interconnectivity, data-exchange architectures or devices for musical interaction. Thus, the notions related to a musical Internet tend to remain as work-in-progress. Networked music performance is one of the approaches that has been applied in this field [22]. A caveat of a device-centric view of music-making is the assumption that technological changes must drive musical activity. This perspective tends to be aligned with the proposals on telematic art which—during the 1980s—pushed for widespread adoption of any technology, regardless of its social or cognitive impacts. An alternative view was implied by the work in ubiquitous computing. A target of ubiquitous computing is to reduce the centrality given to the device. As expressed by [10], this involves designing for a disappearing computer. Rather than dealing with isolated computational resources, Weiser's vision emphasizes the situated relationships between computing and humans in the context of utilitarian activities. Whether these proposals are applicable to non-utilitarian creative goals is one of the threads of investigation carried out in ubimus [23].

The artistic practices of the late 1990s found a suitable platform in the convergent technologies of the first decade of the 21st century. Although some researchers stuck to acoustic instrumental notions to target virtuosic performance [24] and to emulate instrumental chamber music built around the concept of the orchestra, a group of researchers in Brazil proposed loosely structured artistic interventions involving web-oriented technologies, embedded platforms and the deployment of musical activities on consumer-level mobile devices [25–27]. These proposals engage with network technologies as enablers for the exploration of spaces not amenable to the mainstream, acoustic instrumental perspective on musical interaction. They highlight the opportunities for the participation of non-musicians in artistic endeavors, fostering the development of everyday musical creativity [17].

Thus, ubiquitous music not only deals with the infrastructure required to sustain established musical practices, it also probes the constraints and opportunities for the expansion of music-making. Consequently, two opposed socio-technological tendencies can be traced [15]. One tendency entails the support of extant musical genres, involving the adaptation of current methods to inherited forms of musical thinking and of legacy technologies (cf. [28]). Another tendency involves diversifying the musical knowledge base and exploring interaction designs that foster the emergence of new configurations and relational properties among resources and stakeholders. In short, it involves the development and deployment of ubimus ecosystems [11].

Taking into account the pace of change in technological resources and the impact of social, economic and health factors on everyday activities, we can trace emerging tendencies in the deployment of assets labeled as the musical Internet. First of all, computational infrastructure is not fixed, it evolves as a result of ongoing negotiations among multiple stakeholders [29]. Thus, treating the Internet as if it were an acoustic musical instrument or a homogeneous and persistent resource is prone to generate brittle art, i.e., ephemeral cultural products that are highly dependent on legacy technology and which tend to become extinct. Several of the systems reviewed in this paper demand special-purpose hardware and time-consuming configuration or calibration procedures. Dedicated hardware or specialized knowledge may be enforced when the focus is the laboratory or the professional-artist studio. Nevertheless, when the targets are everyday settings and casual participation, the implementation of support is faced with a different set of challenges.

There are, of course, niches for the use of legacy tools, as seen in the DIY practice of recycling personal computers, video-game consoles and more recently handheld controllers. For instance, a ubimus proposal laid out by [30] incorporates vintage analog synthesizers. Rather than attempting to replicate historic artistic practices, the authors adapt the legacy

Future Internet 2023, 15, 125 5 of 17

hardware for network-based usage, hence enabling usage by a large community of users of an expensive and hard-to-maintain asset.

This community-oriented perspective involves new challenges linked to a highly diverse pool of stakeholders. Some participants may share common interests and backgrounds, but more often than not, each one carries her own specific set of cultural backgrounds, demands and expectations. Consequently, any assumption on a fixed aesthetic target tends to bias the design toward a pre-established musical genre. If on the one hand infrastructure evolves as a result of a process of negotiation, on the other hand the constraints and opportunities embedded in the legacy technology tend to reduce the potential for musical diversity. For instance, by adopting the acoustic instrumental chamber ensemble as an ideal model for musical interaction, the musical Internet presents serious constraints linked to network latency and jitter [31,32]. Despite being the key targets of several initiatives in networked music performance, these issues tend to lose relevance as soon as the stakeholders adopt aesthetically pliable approaches to music-making. Given the potential breadth of musical activities supported by the network infrastructure, linking the musical Internet exclusively to the synchronous usage of instruments does not seem to encourage the expansion of current musical practices (This is a key difference between the approaches centered on musical things and those involving musical stuff (see [33])).

In summary, we have identified aesthetic pliability, sustainability and the exploration of new musical horizons as factors to be considered in ubimus interaction-design endeavors. Taking into account the dual nature of technological convergence in music-making, we may define a musical Internet as a collection of persistent resources used to support distributed musical activities involving two or more nodes. This definition stresses the requirement of a fairly stable infrastructure, which may be complemented with volatile resources [34]. Colocated activities are not excluded, though the existence of several nodes highlights the centrality of collaborative activities involving more than a single stakeholder. No assumptions are made about the characteristics of the stakeholders or on the musical genres. Thus, a ubimus ecosystem employing the musical Internet may entail the participation of musicians, of laypeople, of autonomous robots or of any entity capable of making aesthetic decisions.

The next section will explore related works that have the potential to contribute toward a more flexible conceptualization of designs targeting percussive sounds. There is, of course, a vast literature on computer support for percussion. We will only address proposals that employ these sonic sources within the context of network-based collaborative usage. As highlighted by the previous discussion, we are interested in strategies that are applicable to non-musicians or to artists engaged in casual interaction. Hence, extensive training, complex calibration procedures or closed designs are requirements to be avoided. Furthermore, the usage of widely deployed infrastructure is not an aim in itself but it is an important factor when peripheral countries are considered (Many ubimus projects have been deployed in Brazil. But the ubimus community is also present in India and it is starting to incorporate partners from Africa).

# 3. Handling Percussive Sonic Resources

Within the context of the evolving artistic usage of the network and considering the specific case of percussion-oriented interactions, a variety of applications have been proposed over the years. An early one, written on the Java platform, is Networked Drumsteps [35]. This is an educational tool that supports up to six people collaborating. Its results emphasize how collaboration can have a significant impact on learning. In a similar vein, Ref. [36] employ the network as an ensemble in the context of school and community-based case studies of their networked improvisation educational tool, jam2jam. Their prototype is designed to be used either over a local network or the Internet and highlights the application of the meaningful-engagement framework within a ubiquitous-music perspective [37].

Tailored for genre-centric practices, two drumming prototypes target Internet-based deployments: MMODM [38] and GroupLoop [39]. Based on the Twitter streaming API,

Future Internet 2023, 15, 125 6 of 17

MMODM is an online drum machine [38]. Using plain-text tweets from around the world, participants create and perform musical sequences together. To set rhythms, filters and blends of sequences, the users interact with a graphical interface, featuring up to 16-beat note sequences across 26 different presets. Tome and coauthors do not report a formal evaluation, but occasional feedback was gathered after the performances.

GroupLoop is a browser-based, collaborative audio looping and mixing system for synchronous usage [39]. This prototype supports a peer-to-peer connected mesh with a configurable topology which—given the right conditions—may achieve low latency. Participants send their microphone stream to their partners while simultaneously controlling the mix of the audio streams. According to the authors, the collaborative design of GroupLoop affords the creation, alteration, and interaction of multiple simultaneous feedback paths that may encompass multiple acoustic spaces. The synchronous and distributed control of this evolving mix allows the users to selectively turn on or off the incoming raw streams while applying audio processing. The evaluations indicated that GroupLoop demands a high level of skill and sonic expertise, with narrow margins for error. How to render this system accessible for novice users and how to handle the temporal variability intrinsic to faulty networks remain open questions.

Another thread relevant to the expansion of usage of network infrastructure involves the design and implementation of mid-air musical-interaction metaphors. Through independent but complementary approaches, Refs. [40,41] propose mid-air interaction techniques that are applicable to the musical Internet. Handy is a creative-action metaphor that features contactless bimanual interaction deployed on embedded devices and on personal computers, using low-cost hardware. A prototype relies on ultrasonic movement-tracking techniques and another prototype employs video-based motion-tracking to enable network-based interaction with audio-synthesis engines. Both prototypes were deployed in everyday settings yielding musically relevant results and pointing to limitations to be addressed by future projects. A complementary approach is described by [41] focused on solo usage of an emulated drum set. The authors implemented a movement-tracking system tailored for usage with two sticks. The proposed activity targeted a user sitting on a chair with two drum sticks featuring markers of different colors. Eight different sonic events would be triggered by a downward motion of a stick into each of the eight regions of the visual display.

Tolentino and coauthors proposed a simple task involving periodic sequences of hits. This activity allowed them to come up with a figure of maximum sensitivity, around nine hits per second. This is a good technical detail but not enough to characterize the system as musically effective. Thirty non-musicians answered a survey after the activity, furnishing interesting information. For instance, the authors report that "for the inexperienced drummers, the problems they encountered were [related to] playing the drums. This led to wild gesticulations with the sticks, [not] hitting the inside of the bounding box at all, [and sticks] frequently outside of the frames." Two questions targeted a comparison between the implemented support for tracking knee movements as an addition to the use of the sticks versus the sticks-only condition. Unsurprisingly, the participants tended to prefer just using two sticks rather than having to deal with the complexity of coordinating upper and lower limb movements.

These observations are useful for future developments in mid-air musical interaction. Furnishing support for calibration and configuration seems to be a requirement for mid-air systems. As stated by [41], "a more flexible system [allows for] the usage of markers with different colors, shapes, and sizes". Nevertheless, the introduction of sticks and the enforcement of "playing while sitting" seem to be unwarranted. Tracking the movements of both upper and lower limbs is a highlight of this proposal. This contribution would be enhanced by allowing the participants to stand and walk while playing. But more refined mechanisms need to be in place to enable tracking without restricting the movements to small regions of the visual display.

Future Internet 2023, 15, 125 7 of 17

Following a similar approach as [40,41], Ref. [42] propose a mid-air device to control synthesizers via real-time finger tracking, using input from a consumer-grade camera. A contribution of this proposal is handling multiple mid-air controllers in conjunction with a networked infrastructure to enable—within a limited distance—collaborations among geographically displaced users. Once an event has been triggered, an Open Sound Control (OSC) message is transmitted to both a local synthesizer or DAW and the network. The system was positively evaluated by non-musicians. Within a few minutes, the participants engaged in collaborative actions. Possibly because the interactions required less physical and mental effort, the evaluation was described as an enjoyable experience. These outcomes align well with the philosophy of ubimus.

Summing up, supporting interaction with impact sounds on the musical Internet presents specific challenges and opportunities particularly when involving casual interaction by musicians and untrained subjects. The research reviewed in this section highlights a trend toward the reduction of hardware requirements, increasing the investments in flexible movement-tracking systems that unveil the possibility of whole-body interaction, bypassing obtrusive and expensive equipment. Several proposals make use of standard consumer-level devices that are already available to a large segment of the population in both central and peripheral countries. Despite these advances, some caveats need to be addressed. The existing systems demand a fixed spatial reference. The subjects need to adapt their positions and behaviors to the constraints of the visual tracking method. Furthermore, previous studies have focused on performative measurements without providing a specific musical context. Our methods incorporate techniques to ensure musically situated actions and focus on exploratory and qualitative aspects of interaction at an early stage of design. The next section describes the implementation and deployment of the Dynamic Drum Collective prototype, and the materialization of our proposal for banging interaction. We describe the architecture, the experimental strategies and the outcomes obtained throughout three deployments.

# 4. Materials and Methods: The Dynamic Drum Collective

The Dynamic Drum Collective (DDC) is a camera-based gesture recognition system that enables individual and collaborative control of events aiming at both colocated and distributed collective music-making with percussive sounds. The prototype tracks the presence of the human body, captures limbs movements and maps the motions to audio-synthesis and processing parameters. To promote remote collaborations, a client–server architecture was implemented. A socket server handles multiple clients using multi-threading. Thus, the prototype is akin to a virtual, multi-user and multi-timbral percussion system featuring a fairly low demand of bandwidth.

Calibration. Participants use the square 'Calibration box' to position the adaptive visual tokens. One hand is employed to select a kit and the other arm serves to drag the token position to the desired location. In the first two prototypes, this functionality allowed the subjects to set up the visual referents using both hands. Expanding gesture-triggered sonic resources would require a new mapping strategy for calibration. In this study, we do not tackle the gestural mappings for calibration and thus this functionality was not available to the participants.

Adaptive techniques. The Dynamic Drum Collective employs the computer-vision technology Mediapipe Pose Estimation (MPE) model for synchronous human-pose estimation [43,44]. Pose is defined as an arrangement of human joints and human-pose estimation is the localization of the human joints from a set of landmarks in a sequence of captured images [45]. The system tracks the human body continuously to calculate the distance from the limbs to the camera. At a higher level of abstraction it furnishes data to detect gestures. Distance is measured to adjust the visual referents according to the movements, thus allowing a 'walk and play mode'. When a difference in body position between continuous video frames is found, the system attempts to map the motion to a gesture. Then a sound is triggered.

Future Internet 2023, 15, 125 8 of 17

Network infrastructure. To make remote interaction possible, socket programming based on a client–server architecture was implemented in Python. The devices were connected to a local network. In the implemented bi-directional socket client–server architecture, the listener node supports the read and write processes. One node listens on a port using an IP while the other node sends a request to establish a connection. The server functions as a listener node and the client requests communication.

*Sonic resources*. Depending on where a hit is detected, a MIDI event is triggered. The server receives a MIDI message and sends it to the other client. Then the message is sent to an audio-synthesis engine (MixCraft9 in this case) to render the audio. The available drum sounds in this prototype are Kick, Hi-Hat, Crash Cymbal and Snare (left to right in Figure 1). In this version of the prototype, the participants can hear the audio during the collaborative interaction and can read the name of the percussive sound triggered.

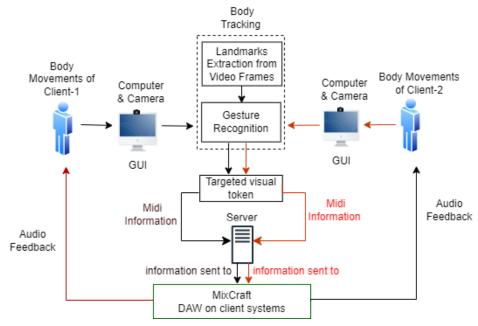


Figure 1. Components and flow of the Dynamic Drum Collective.

*Limitations*. If the limb position changes rapidly during five consecutive frames, all the values in the speed buffer increase, indicating a faster movement. The system performance, as it currently stands, is constrained by the camera resolution (with a typical frame rate between 30 and 150 fps) and by the processing capabilities of the computer.

In camera-based pose estimation techniques, camera and light conditions can influence the quality of pose estimation and associated targets. Documentation on minimum or maximum light conditions to improve the quality of pose detection using MediaPipe is still pending.

# 5. Deployments

First Prototype. The first prototype of the Dynamic Drum Collective was presented at the Ircam Workshop 2022 [18]. This version featured: (i) A calibration mode supporting the action of grasping or pinching to handle the visual referents of a drum kit, (ii) Adaptive support for movement-tracking to enable consistent sonic feedback during whole-body interaction, and (iii) Multi-user collaboration through network connectivity.

A professional drummer and a non-musician participated voluntarily in a live demonstration. Overall, the system performed well with respect to latency. At various moments of the exploratory session, varying light conditions caused visual jitter and informational noise impacting the image-processing cycle. This caused longer delays than the average obtained during the laboratory trials, compromising the temporal alignment between the visual and the sonic events.

Future Internet 2023, 15, 125 9 of 17

*Second Prototype*. A second version of the Dynamic Drum Collective incorporates colors as visual tokens, replacing the images of instruments. This prototype was used by an experienced percussionist. We studied the relationship between the interaction design and the body movements to identify major flaws and potentially useful features.

The study featured two sessions: (i) free exploration and (ii) semi-structured tasks. The objective was to focus on the limitations of the support for casual interaction by tracking the time needed to reach a basic understanding of the functionalities. The aim was to allow usage without any previous preparation.

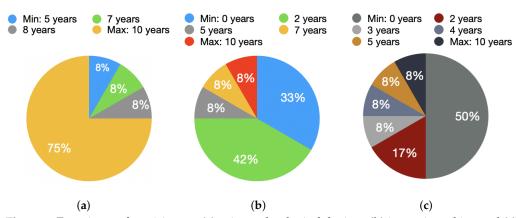
During the exploration phase, the participant was given no visual feedback of his movements (no-visuals condition). He could only see the bounded regions for triggering sound. The instruction was: "a 'hit event' occurs by moving the index finger inside the visual tokens". The lack of spatial referents made it hard for the subject to align the visual tokens in relation to his body position. According to his report, this condition had a negative impact on the accuracy of his actions. Our analysis of the footage corroborated these observations. The exploratory session also helped us to understand how a musician deals with latency, pointing to the need to employ a musical context to check how the technical issues impact the musical outcomes.

Subsequently, we observed whether the subject was comfortable doing targeted activities with the DDC prototype. This stage allowed us to gather information on the specific design dimensions and demands, hinting at problems to be encountered in field deployments.

*Third Prototype*. This case study was aimed at identifying the possibilities offered by the DDC system, the creative exploration of the available sonic resources and the interaction among participants during musical collaborations.

## 5.1. Participants

Twelve participants with at least five years of experience using technological devices volunteered for this study. The majority had used devices for at least a decade. Among the participants, 33 % had no prior experience in music-making. A percentage of 42% had been involved in music-making activities for no more than 2 years. One quarter of the participants had instrumental training ranging between 5 and 10 years. Thus, almost a quarter of the subjects had music lessons, having played acoustic instruments during their school time. Sixty-seven percent had less than two years of training. Figure 2 shows the participants' profile of musical experience.



**Figure 2.** Experience of participants: (a) using technological devices, (b) in music-making and (c) in formal music training.

# 5.2. Materials

Two iMac 14.3 computers, with 16 GB RAM and 3.1 GHz Intel Core i7 processors, running canonical Ubuntu 22.04 LTS were used. The camera resolution was  $1280 \times 720$  (16:9) pixels. According to a valley benchmark, the frame rate was 38 fps. This equipment was donated to the experimenters. Within the spirit of ubimus, we chose to employ machines that were not state-of-the-art technology. To ensure better control of the sound

Future Internet 2023, 15, 125 10 of 17

mix, a Yamaha MW12 mixing desk and a pair of Yamaha HS7 active speakers were used (Figure 3).

A walkable space was used by the participants during this experiment. It was 3.0 m long and 2.40 m wide. As the DDC prototype supports "walk and play" mode, the subjects were moving within the space while playing. The measurements were recorded based on the positioning of the subjects during the sessions. We used built-in Mac cameras as mentioned above, without additional calibration. MediaPipe requires good lighting. Therefore, the experimental sessions were conducted in a standard laboratory with LED louver lights with Osram LEDs. We used the lux meter (luminance meter) to measure the level of light reflected on the subjects within the space.

The minimum distance from the camera was 2.2 to 2.3 m. The tracking algorithm searches for the lower body to display the percussion objects (e.g., Kick or Hi-hat). At a distance of 2.2 m from the camera, the object has a luminance of 294 lux and remains stable while changing the vertical position. But the luminance value fluctuates between 216 lux to 217 lux for horizontal movements. With these luminance values, the algorithm performance for pose detection and the response time was stable.

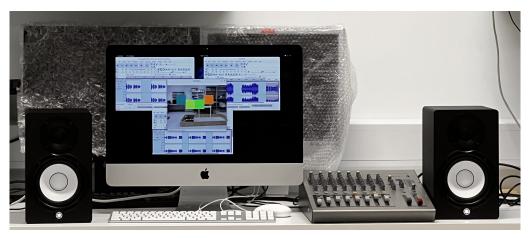


Figure 3. Experiment hardware setup.

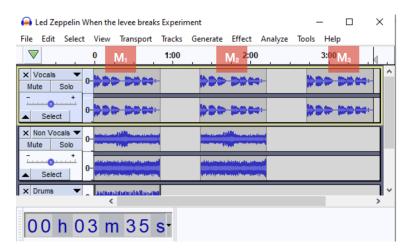
# 5.3. Session Setup

During this case study, the interactive drum performances were accompanied by sonic gridworks derived from three different soundtracks (Table 1 and Figure 4). The tracks selected were Anti-Hero (97 bpm) by Taylor Swift (clip-I), When the Levee Breaks (70 bpm) by Led Zeppelin (clip-II) and Piano Man (178 bpm) by Billy Joel (clip-III). The choice of materials was based on their different treatment of instrumentation and tempo. A verse and chorus from each of these tracks was played in three combinations: M1: vocals, drums and instruments, M2: vocals and instruments only and M3: vocals only. The subjects were prompted to perform: (1) with the original drums, (2) without the drums. In the latter condition, the rhythmic cues had to be inferred from the other instruments and the vocals. The duration of each accompaniment track varied according to the song's structure. Clip-I was presented for 4:21 min, clip-II for 3:35 min and 5:30 min for clip-II. The time for each track includes a 30-s pause given to the participants between three modes (M1, M2, M3). Considering a few minor adjustments between these three sound clips, the experiment duration was between 15 and 20 min, excluding free exploration and discussion.

**Table 1.** Sonic gridworks: sources and durations.

Sonic Gridwork	Author	Tempo	Start Time	End Time	Clip Duration
Anti-Hero	Taylor Swift	90 bpm	0:00	1:09	4:21
When the Levee Breaks	Led Zeppelin	70 bpm	1:24	2:16	3:35
Piano Man	Billy Joel	178 bpm	1:57	3:26	05:30

Future Internet 2023, 15, 125



**Figure 4.** Audacity window with three types of sonic gridworks (Leave Breaks by Led Zeppelin). M1: Vocals with drums and other instruments, M2: Vocals and instruments without drums, and M3: Vocals only.

# 5.4. Session Deployment

Session Deployment. Before the session, the individuals and pairs of participants were identified. They were given instructions about the complete session procedures including the individual and the collaborative aspects of the experiment (Figure 5). They were not interrupted during the tasks. The first five minutes of each session were allocated to a free exploration of the resources, allowing the participants to understand the mechanics of usage and to discover their own strategies to interact with the system. As soon as they were ready, they were asked to start playing with the sonic gridworks.

After their individual tasks, each subject played with a fellow participant. This activity involved the three modes of playing (M1, M2 or M3). The client–server mode was activated after the free exploration. In this condition, our primary focus was not to encourage mimicry of the given musical materials but to foster creative behaviors prompted by the musical context.

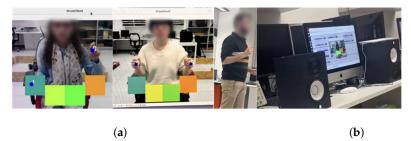
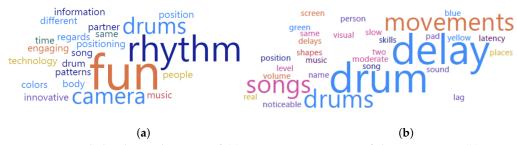


Figure 5. (a) Sitting usage. (b) Standing usage.

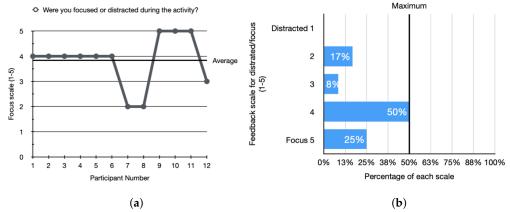
## 5.5. Results

Our approach to design targets the analysis of the users' experiences within the context of musical activities. We explore the collected data to expand the scope of ubiquitous-music strategies for design purposes. Another goal is to obtain feedback and opinions from differently skilled subjects to unveil how they deal with and assess the technical limitations. Data about the user experience and system performance was gathered through an experience assessment form at the end of each session and also via direct observation (Figures 6 and 7).

Future Internet 2023, 15, 125



**Figure 6.** Word-cloud visualizations of (a) positive assessments of the experience, (b) negative assessments of the experience.



**Figure 7.** (a) Profile of level of focus during the activities for all participants, (b) Distribution of participants according to their ability to focus on the target modalities.

The use of the accompanying tracks provided the participants with an initial scaffold to engage creatively. Their reactions to the proposed rhythmic patterns were varied. Most comments indicated a positive attitude toward the experience (described as "fun (8 %)") and some concerns related to the technical limitations (expressed in the negative comments as "delay") (Figure 6a,b). Other than "delay" (4 %), some subjects used the terms "lag/latency" to describe the system behavior.

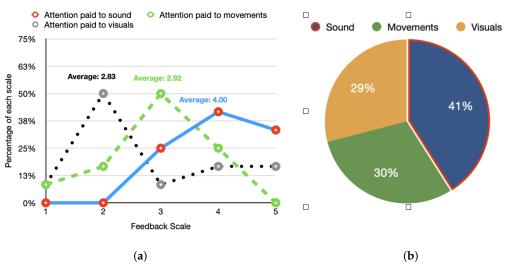
Overall, there was a balanced split between the attention paid to the sounds and the gestures, in some cases leaning toward the visual resources (Figure 8). The survey did not include any item describing the experience as feeling or being "engaged" and we did not use these terms during the discussion. Thus, the occurrence of "engaging (4%)" shows a positive assessment of percussion-oriented mid-air interaction by non-musicians (Figure 6a). Assessing the subjective quality of the sonic results and the support for specific creative factors are challenges that will be tackled in the next stages of design. The early-stage techniques applied in this study provide us with useful information on four factors: (1) tempo, (2) size of gestures, (3) sonic gridworks and (4) collaboration.

# 5.6. *Tempo*

Almost all the participants reported a delay between the gestures and the sonic feedback. This issue emerged when they played music with faster tempi and when they were encouraged to try faster arm movements (for instance, by employing sonic gridworks that exceeded 97 bpm). Interestingly, the evaluations pointing to "too much" delay were directly correlated with the subjects' level of musical training. Half of the subjects that had 5 years of musical experience marked delay as being "too much". The same answer was given by 17% that were trained for 2 years. Contrastingly, only eight percent of the non-musicians ranked delay as "too much". Fifty percent of the untrained subjects stated that the delay was "moderate" and said they could play fairly well. During the discussion, participants mentioned they perceived the system responsiveness as closely related to the tempo of the music. Except for one subject, all the participants highlighted that their

Future Internet 2023, 15, 125 13 of 17

actions were well aligned when they adopted slow to moderate tempi. We can conclude that regarding this dimension the user experience was significantly affected by the selection of the musical materials and by the profile of the subjects.



**Figure 8.** (a) Level of attention according to the target modalities (sounds, visuals, movements), (b) Average self-assessments by subjects according to the three targets.

# 5.7. Size of Gestures

The scale of the strokes is based on the distance between the projection of the subject's hands to the targeted visual tokens. The impact of this parameter was confirmed by differences in performance between small- and large-scale gestures. Large movements take more frames for accurate sampling, increasing the size of the sample proportionally to the accuracy of the outcome. The subtlety of small and fast movements tended to be lost. Thus, when trying to align their actions with the soundtrack the subjects frequently adopted large movements.

# 5.8. Sonic Gridworks

The use of sonic gridworks constitutes a key contribution of our method and was prompted by the caveats observed during the first two sessions. Of the three types of sonic gridworks proposed, most subjects preferred mode 3, i.e., sonic materials that only featured vocals. There were contradictions in how the participants justified their choices of mode. Some of them mentioned having more space to create their own rhythms when employing mode 3 (vocals only). Others mentioned that this condition was the most demanding. The number of participants who selected the vocals-only gridwork (37%) was slightly larger than those who preferred the complete soundtrack (34%) and than those who chose the vocals and instruments mode (29%).

These results, when considered in tandem with the impact of the type of musical material, indicate a variable to be addressed during the later stages of design: the sonic context. Sonic gridworks are flexible tools to deal with the contextualization of the creative activity because they are configurable. We observed that partial renditions of the sonic resources (in this case, the sections of vocals without percussive or instrumental parts) were enough to prompt musically significant participation. This means that remote stakeholders could be provided with simplified sonic gridworks without a negative impact on their engagement. Nevertheless, the amount of reduction of sonic information involved in the production and sharing of sonic gridworks is still an open question.

#### 5.9. Collaboration

Collaborative interactions highlighted that participants managed to find a musical common ground despite dealing with fairly simple musical resources. A leader-follower

Future Internet 2023, 15, 125 14 of 17

relationship was often observed. (They were not prompted for this strategy and they did not refer to it while explaining their approaches). The majority of the rhythmic patterns featured two sounds, mostly using the Kick drum and the Snare and less often using the Cymbal. The subjects tended to keep their performance periodic and aligned with the soundtrack, instead of attempting a free exploration of the available sonic resources.

#### 6. Discussion

An element to be considered in the implementation and deployment of technical support is the role of the musical materials incorporated as resources. Although networked music performance and telematic approaches to musical interaction use mechanisms aligned with acoustic instrumental chamber music practices, biased toward virtuosic performance or top-down social arrangements such as the orchestra [22,24], ubimus frameworks strive for aesthetic pliability and consequently avoid the use of acoustic instrumental notions. Open-ended, creativity-oriented and adaptive strategies are preferred. These designs are aligned to the stakeholders' profiles and to the available resources, highlighting: 1. The mechanisms for knowledge-sharing (to avoid the enforcement of common-practice notation or other score-oriented formats), 2. The temporal and spatial characteristics of the activity (avoiding the master-slave social structure of centralized and synchronous network usage of the orchestral formats), 3. The flexible role of the stakeholders (beyond the composer-performer-audience division of labor).

Casual musical interactions present specific challenges, given that the subjects only have a few minutes to familiarize themselves with the tools and the musical resources. The techniques adopted in this series of studies, involving the use of sonic gridworks, yielded useful outcomes. The participants' frequent usage of "fun" to describe the experience indicates that despite the intense focus required by the activity, the experience was pleasant. Interestingly, the negative evaluations were mostly related to the technical limitations, expressed as "delay". It is not surprising that the subjects preferred the gridworks that complemented but not those that replaced the percussive sounds. In other words, they avoided the materials with percussive characteristics (such as the recordings featuring the original drums) and tended to choose materials that opened spaces to be complemented by their participation.

These choices point to two parameters to be explored in future studies: timbral richness and timbral complementarity. The former can be implemented using resynthesis. Rich sonic gridworks can be created by combining multiple layers of simple sounds or by selectively filtering complex sonic scenes. These techniques apply to any source. Therefore, they do not imply a bias toward acoustic instrumental sources. Timbral complementarity can be explored by creating multiple sonic gridworks, some closely aligned to the percussive sounds being used by the subjects, and others that feature highly contrasting materials. Especially interesting cases are probably situated in the range of closely related but still distinct timbral classes.

Another insight yielded by this series of studies is the reliance on visual feedback as a non-exclusive strategy to deal with temporal organization in collaborative tasks. Although it has been argued that presence is an important factor in network-based musical experiences, the specific dimensions of this concept and its impact demand field research. The three targets of attention (sounds, visuals and movements) indicated a slight tendency to concentrate on the aural dimension, although all three modalities of interaction were considered important. We also observed interactions organized around patterns of leader–follower behaviors. Were these behaviors prompted by the periodic qualities of the sonic gridworks? Were they induced by the body images of the stakeholders or by the fact that the activities were colocated? Or are these behaviors an emergent property of any group-based interaction, regardless of the type of multimodal support provided? The experimental results indicate that visual feedback serves as a complement to other modalities in mid-air interaction, opening a range of experimental paths to be explored in future deployments.

Future Internet 2023, 15, 125 15 of 17

The creative-action metaphor featured in this paper supports sonic sources with temporal profiles that can be roughly classified as impact sounds. It incorporates either audio-synthesis or audio-processing techniques that do not demand a high computational cost and that are easily deployable through the network infrastructure. Locally, the Dynamic Drum Collective prototype supports synchronous usage. It may be employed in the context of individual or collective musical activities. Given its reliance on local resources, it tends to be scalable and may be integrated into the Internet of Things and the Internet of Musical Stuff [33], expanding the functionality of the musical Internet. As long as the musical activities employ impact sounds, the metaphor remains fully transparent to various aesthetic perspectives. Most importantly, it does not demand domain-specific knowledge or specialized intensive training and it is not limited to a predefined set of gestures or to a fixed body posture, the participants can explore the whole gamut of whole-body interactions.

Our research on banging interaction highlights how computational networks can be integrated with platforms for artistic practice. As a point of departure for future research, we have proposed a viable definition of a musical Internet: a collection of persistent resources used to support distributed musical activities involving two or more nodes. This definition stresses the multiplicity of musical applications, highlights the need for sustainable strategies to support open-access resources and leaves room for the incorporation of volatile resources that characterize many artistic practices.

Three components of banging interaction—adaptive interaction, mid-air techniques and timbre-led design—target the development of creative-action metaphors with an emphasis on percussive sources and audio-processing tools which are suited for everyday settings. These settings demand a flexible approach to deal with fast-changing conditions and varied participant profiles [18,46]. Two ubimus threads provide partial answers to current music-making demands: 1. Deployments on networked and embedded platforms, and 2. Contactless, mid-air, whole-body interaction. Gesture-based, contactless interaction may be applied in transitional settings involving casual participation. Prototypes based on custom equipment may also be employed in some contexts [13], though as previously discussed recent initiatives tend to explore standard stationary equipment and embedded platforms. Despite its flexibility, mid-air interaction presents various challenges due to the drastic reduction of tactile feedback, pointing toward an emerging area of research [47].

A perspective adopted in banging interaction is timbre-led design. A recurring issue in ubimus deployments is how to enlarge the creative potential of the prototypes without increasing the cognitive demands of the activity. This is particularly problematic in transitional settings because the participants tend to apply preconceived notions of sound-making when faced with open-ended activities and resources. Rather than starting the design process from a "musical instrument", timbre-based design engages with the characteristics of the possibilities afforded by the musical materials. As exemplified in the methods presented in this paper and in previous ubimus projects, human-based heuristics, machine-based decision-making processes involving adaptive techniques and hybrid forms of sonic analysis and annotation may be combined during the interaction-design process.

Adaptive design, another component of banging interaction, entails the usage of computational strategies to adjust the tools to the behavioral and ergonomic characteristics of the subjects [48]. An adaptive tool is the behavioral opposite of the instrument. An instrument, depending on the complexity of its functionality, frequently demands reshaping human attitudes and behaviors. Complex instruments, such as the acoustic string ensemble, involve a large amount of documentation and specialized knowledge to achieve minimally usable results. Contrastingly, adaptive tools change their own behaviors as a result of human-tool interactions. Thus, their computational components are shaped by the needs of the subjects and avoid hardwired body postures, highly constrained movements or awkward cognitive patterns.

Let us, then, restate the challenge laid out by Bill Buxton more than four decades ago: Would it be possible to conceive and implement a musical theory that can support the representation of any computable musical knowledge? Banging interaction hints that

Future Internet 2023, 15, 125

musical experiences cannot be limited to explicit musical knowledge or its representations. Thus, the design targets change from representing knowledge to enacting experiences. We believe that the banging interaction approaches exemplified in this paper provide a good starting point for this research path.

**Author Contributions:** Software, S.C.; Investigation, A.Y. and V.L.; Writing—original draft, D.K.; Writing—review and editing, J.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the Brazilian National Council for the Development of Science and Technology, grants [CNPq 202559/2020-3, 308790/2021-9] to D.K.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: https://github.com/SutirthaChakraborty/Banging-interaction, (accessed on 11 February 2023).

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Buxton, W.A. A composer's introduction to computer music. *Interface* 1977, 6, 57–71. [CrossRef]
- 2. Roberts, A. An All-FORTRAN Music-Generating Computer Program. J. Audio Eng. Soc. 1966, 14, 17–22.
- 3. Mathews, M.V.; Miller, J.E.; Moore, F.R.; Pierce, J.R.; Risset, J.C. *The Technology of Computer Music*; MIT Press: Cambridge, MA, USA, 1969; Volume 5.
- 4. Truax, B. For Otto Laske: A communicational approach to computer sound programs. *J. Music Theory* **1976**, 20, 227–300. [CrossRef]
- 5. Truax, B. The inverse relation between generality and strength in computer music programs. *J. New Music Res.* **1980**, *9*, 49–57. [CrossRef]
- 6. Orlarey, Y.; Fober, D.; Letz, S. Syntactical and semantical aspects of Faust. Soft Comput. 2004, 8, 623–632. [CrossRef]
- 7. Bhagwati, S. Towards interactive onscreen notations for comprovisation in large ensembles. In *Sound & Score: Essays on Sound, Score and Notation*; Leuven University Press: Brussels, Belgium, 2013; pp. 143–177.
- 8. Bown, O.; Eldridge, A.; McCormack, J. Understanding interaction in contemporary digital music: From instruments to behavioural objects. *Organised Sound* **2009**, *14*, 188–196.
- 9. Keller, D. Compositional processes from an ecological perspective. Leonardo Music J. 2000, 10, 55–60. [CrossRef]
- 10. Weiser, M. The Computer for the 21 st Century. Sci. Am. 1991, 265, 94–105. [CrossRef]
- 11. Lazzarini, V.; Keller, D.; Otero, N.; Turchet, L. Ubiquitous Music Ecologies; Routledge: London, UK, 2020.
- 12. Torre, G.; Andersen, K.; Baldé, F. The Hands: The making of a digital musical instrument. *Comput. Music J.* **2016**, 40, 22–34. [CrossRef]
- 13. Bernardo, F.; Pestana, P.; Martins, L. The Smart Stage: Designing 3D interaction metaphors for immersive and ubiquitous music systems. In Proceedings of the International Conference on New Music Concepts (ICNMC 2015), Treviso, Italy, 7–8 March 2015.
- 14. Hopkins, T.; Suibi, C.C.; Vanukuru, R.; Wenzel, E.A.; Banic, A.; Yi-Luen, E. AR Drum Circle: Real-Time Collaborative Drumming in AR. *Front. Virtual Real.* **2022**, *3*, 91. [CrossRef]
- 15. Keller, D.; Barreiro, D.L. Editorial—Ubiquitous Music Section: Forces of attraction and challenges in ubimus research (Editorial—Seção Temática Música Ubíqua: Forças de atração e desafios na pesquisa ubimus). *Rev. V'ortex* **2018**, *6*, 1–14.
- 16. Fink, L.K.; Warrenburg, L.A.; Howlin, C.; Randall, W.M.; Hansen, N.C.; Wald-Fuhrmann, M. Viral tunes: Changes in musical behaviours and interest in coronamusic predict socio-emotional coping during COVID-19 lockdown. *Humanit. Soc. Sci. Commun.* **2021**, *8*, 180. [CrossRef]
- 17. Keller, D. Everyday musical creativity. In *Ubiquitous Music Ecologies*, 1st ed.; Taylor & Francis, Howick Place: London, UK, 2020; pp. 23–51.
- 18. Chakraborty, S.; Yaseen, A.; Timoney, J.; Lazzarini, V.; Keller, D. Adaptive touchless whole-body interaction for casual ubiquitous musicalactivities. In Proceedings of the International Computer Music Conference (ICMC2022), Limerick, Ireland, 3–9 July 2022; pp. 132–138.
- 19. Hiller, L.; Isaacson, L.M. Experimental Music: Composition with an Electronic Computer; McGraw-Hill: New York, NY, USA, 1959.
- 20. Bischoff, J.; Gold, R.; Horton, J. Music for an interactive network of microcomputers. Comput. Music J. 1978, 2, 24–29. [CrossRef]
- 21. Weinberg, G. Interconnected musical networks: Toward a theoretical framework. Comput. Music J. 2005, 29, 23–39. [CrossRef]
- 22. Lazzaro, J.; Wawrzynek, J. 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, NY, USA, 25–26 June 2001; Association for Computing Machinery: New York, NY, USA, 2001; NOSSDAV '01, pp. 157–166. [CrossRef]
- 23. Pimenta, M.S.; Keller, D.; Flores, L.V.; de Lima, M.H.; Lazzarini, V. Methods in creativity-centred design for ubiquitous musical activities. In *Ubiquitous Music*; Springer: Cham, Switzerland, 2014; pp. 25–48.

Future Internet 2023, 15, 125 17 of 17

24. Wessel, D.; Wright, M. Problems and prospects for intimate musical control of computers. *Comput. Music J.* **2002**, *26*, 11–22. [CrossRef]

- 25. Barbosa, A. Performance musical em rede (Network music performance). In *Musical Creation and Technologies: Interdisciplinary Theory and Practice*; 2010; Volume 2, pp. 180–200. Available online: https://repositorio.ucp.pt/bitstream/10400.14/5008/1/2010 -Book%20Chapter%20(Anpoon).pdf (accessed on 11 February 2023).
- 26. Keller, D.; Flores, L.V.; Pimenta, M.S.; Capasso, A.; Tinajero, P. Convergent trends toward ubiquitous music. *J. New Music Res.* **2011**, *40*, 265–276. [CrossRef]
- 27. Miletto, E.M.; Pimenta, M.S.; Bouchet, F.; Sansonnet, J.P.; Keller, D. Principles for music creation by novices in networked music environments. *J. New Music Res.* **2011**, *40*, 205–216.
- 28. Lazzarini, V.; Keller, D. Towards a Ubimus Archaeology. In *Workshop on Ubiquitous Music (UbiMus 2021)*; Ubiquitous Music Group: Porto, Portugal, 2021. [CrossRef]
- 29. Dourish, P.; Bell, G. Divining a Digital Future: Mess and Mythology in Ubiquitous Computing; Mit Press: Cambridge, MA, USA, 2011.
- 30. Zawacki, L.F.; Johann, M.O. Analogue audio recording using remote servers. In *Ubiquitous Music*; Keller, D., Lazzartini, V., Pimenta, M.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 83–107. [CrossRef]
- 31. Frid, E. Accessible Digital Musical Instruments: A Review of Musical Interfaces in Inclusive Music Practice. *Multimodal Technol. Interact.* **2019**, *3*, 57. [CrossRef]
- 32. Rottondi, C.; Chafe, C.; Allocchio, C.; Sarti, A. An overview on networked music performance technologies. *IEEE Access* **2016**, 4,8823–8843. [CrossRef]
- 33. Messina, M.; Filho, M.C.; Mejía, C.M.G.; Keller, D.; Aliel, L.; Simurra, I.; Aliel, L.; Simurra, I. A Internet do Bagulho Musical (Internet of Musical Stuff)—IoMuSt. In *Ubiquitous Music Symposium (UbiMus* 2022); Ubiquitous Music Group: Curitiba, Brazil, 2022; pp. 85–93.
- 34. Keller, D.; Otero, N.; Lazzarini, V.; Pimenta, M.S.; Lima, M.H.; Johann, M.; Costalonga, L. Relational properties in interaction aesthetics: The ubiquitous music turn. In Proceedings of the Electronic Visualisation and the Arts Conference (EVA 2014), London, UK, 8–10 July 2014; Ng, K., Bowen, J.P., McDaid, S., Eds.; Computer Arts Society Specialist Group: London, UK, 2014.
- 35. McCarthy, C.; Bligh, J.; Jennings, K.; Tangney, B. Virtual collaborative learning environments for music: Networked drumsteps. *Comput. Educ.* **2005**, *44*, 173–195. [CrossRef]
- Brown, A.; Dillon, S. Networked improvisational musical environments: Learning through on-line collaborative music making. In Music Education with Digital Technology; Continuum International Publishing: London, UK, 2007; pp. 95–106.
- 37. Brown, A.R.; Stewart, D.; Hansen, A.; Stewart, A. Making meaningful musical experiences accessible using the iPad. In *Ubiquitous Music*; Springer: Heidelberg, Germany; New York, NY, USA, 2014; pp. 65–81.
- 38. Tome, B.; Haddad, D.D.; Machover, T.; Paradiso, J.A. MMODM: Massively multiplayer online drum machine. In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), Vancouver, BC, Canada, 31 May–3 June 2015; pp. 285–288.
- 39. Ramsay, D.B.; Paradiso, J.A. Grouploop: A collaborative, network-enabled audio feedback instrument. In Proceedings of the International Conference on New Interfaces for Musical Expression, Baton Rouge, LA, USA, 31 May–3 June 2015; pp. 251–254.
- 40. Keller, D.; Gomes, C.; Aliel, L. The Handy Metaphor: Bimanual, touchless interaction for the internet of musical things. *J. New Music. Res.* **2019**, *48*, 385–396.
- 41. Tolentino, C.T.; Uy, A.; Naval, P. Air Drums: Playing Drums Using Computer Vision. In Proceedings of the 2019 International Symposium on Multimedia and Communication Technology (ISMAC), Quezon City, Philippines, 19–21 August 2019; pp. 1–6. [CrossRef]
- 42. Cocchiara, D.; Turchet, L. Democratizing access to collaborative music making over the network using air instruments. In Proceedings of the 17th International Audio Mostly Conference, New York, NY, USA, 6–9 September 2022; pp. 211–218.
- 43. Chung, J.L.; Ong, L.Y.; Leow, M.C. Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet* **2022**, 14, 380. [CrossRef]
- 44. Zhang, Y. *Applications of Google MediaPipe Pose Estimation Using a Single Camera*; California State Polytechnic University Pomona: Pomona, CA, USA, 2022.
- 45. Josyula, R.; Ostadabbas, S. A Review on Human Pose Estimation. arXiv 2021, arXiv:2110.06877.
- 46. Keller, D.; Simurra, I.; Messina, M.; Neiva, T.M.; Tedesco, S.; Mesz, B. Domestic ubimus. *EAI Endorsed Trans. Creat. Technol.* **2022**, 9. [CrossRef]
- 47. Subramanian, S.; Seah, S.A.; Shinoda, H.; Hoggan, E.; Corenthy, L. Mid-air haptics and displays: Systems for un-instrumented mid-air interactions. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 3446–3452.
- 48. Graham-Knight, K.; Tzanetakis, G. Adaptive Music Technology:History and Future Perspectives. In Proceedings of the Internationc Computer Music Conference (ICMC2015), Denton, TX, USA, 25 September–1 October 2015; pp. 416–419.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.