

A design space for viewpoint-decoupling in mixed-reality musical ensembles

Florent Berthaut

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL

F-59000 Lille, France

florent.berthaut@univ-lille.fr

Pierrick Uro

Munster Technological University

Cork, Ireland

pierrick.uro@gmail.com

Abstract—Mixed-reality musical instruments allow for rich co-located collective music making, with novel expressive capabilities. They however tend to limit spatial exploration, due to the required alignment between physical and virtual spaces. One solution is viewpoint-decoupling, where musicians exist as both physical persons and as virtual avatars that can freely navigate virtual musical scenes. This approach has however only been implemented and analysed for pairs of musicians. In this paper we extend this idea to ensembles with three or more musicians, and propose a design space for viewpoint-decoupling in mixed-reality ensembles. The design space is composed of five dimensions which describe the way musicians will manipulate and visualise viewpoints. We illustrate the design space with a working implementation and a series of mockup instruments. Our goal is that this design space will help instrument designers and musical ensembles implement advanced collaboration features.

Index Terms—extended reality musical instruments, mixed-reality musical ensembles, viewpoint-decoupling

I. INTRODUCTION

Extended reality musical instruments (XRMI) [1] offer novel opportunities for musical ensembles: building and sharing large virtual instruments, sharing the control of virtual components without risk of physical conflicts, providing alternative views and gestures with the same instrument (*e.g.*, adapting to expertise [2]).

These virtual ensembles can be broadly divided in ones where musicians only coexist in a virtual environment (*i.e.*, in Virtual Reality - VR) and ones where they share virtual components within the same physical space (*i.e.*, in Mixed Reality - MR). They can also be separated into co-located or networked, in which case they can be designated as a musical metaverse [3]. A mix of co-located and networked musicians can also play together, either in VR or MR [4], [5].

VR ensembles provide specific advantages such as the ability to navigate large musical structures and to define shared and private musical expression spaces [6], [7]. However, they may suffer from usability issues due to network latency, from a lack of awareness of other musicians due to inaccuracies in avatars and from reduced co-presence.

Conversely, MR ensembles provide a clear (because physical) awareness of other musicians and familiarity in physical sharing and collaborating around virtual instruments, through non-verbal communication such as proxemics [8]. But by coupling the physical position of each performer with a

shared virtual scene, the interaction potential for building and playing the instrument is severely limited, as any navigation or positional manipulation would affect everyone, even in the middle of their own interaction. It is therefore not possible for musicians to zoom-in and focus on a specific, potentially simpler, part of an instrument, or to zoom-out and access more components simultaneously, or simply to transform the instrument to make it more playable from their position in space.

In this paper, we investigate an approach that reintroduces in MR ensembles the freedom of point-of-view manipulation enabled by VR ensembles: *viewpoint-decoupling*.

This technique was proposed by Uro *et al.* [9], in a collaborative autoethnographic study investigating interaction strategies emerging on a collaborative mixed-reality musical instrument. In their two-musician MR instrument *Grainy-Cloud*, musicians are able to choose between three spatial transformations of the instrument displayed in the physical space, which in fact constitute three views on the instrument. The FIXED view serves as a base state, a guaranteed and fixed initial transformation which performers can go back to. A performer goes into their SELF automatically whenever they manipulate the spatial transformation of the instrument, and can jump back to it at anytime when in another view. Finally, the OTHER view allows a performer to see the instrument the way their partner has transformed it. Whenever the two musicians are seeing the instrument differently, their respective viewpoints are visualised as avatars (headsets and controllers). This approach therefore allows for free individual exploration of a virtual instrument while preserving both a physical and virtual awareness of each other.

However, in this first proposal only two performers interacted together, which allowed for using very simple interactions for viewpoint selection. Scaling up the model to 3 or more performers will not suit with this initial implementation, and alternatives need to be considered. Along with viewpoint selection, other design choices that were made also require discussion, such as viewpoint transitions and visual feedback.

II. A DESIGN SPACE FOR VIEWPOINT-DECOUPLING

In this section, we present the generalized approach behind viewpoint-decoupling and propose a design space composed of 5 dimensions in order to help instrument designers and

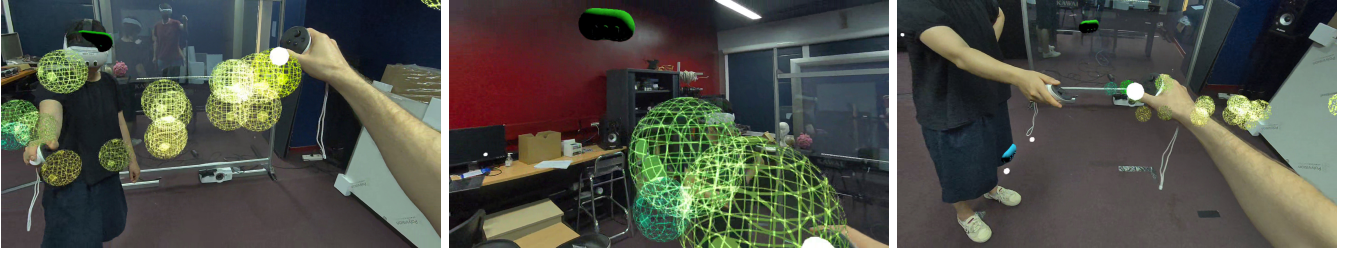


Fig. 1. Viewpoint-decoupling used in a corpus based concatenative synthesis instrument played by two musicians, seen by the first of them. Left: The musicians share the same view/transformation (green HMD avatar aligned with the other physical musician). Middle: The first musician zooms in to focus on one specific grain, the green avatar reflects the other musician’s viewpoint on the transformed instrument (large green HMD at the top). Right: Then the first musician zooms out to play multiple grains, the green HMD avatar corresponds now to the scale at which the other musician sees the instrument.

musicians integrate this approach in their own practices. The values for each dimension can be chosen in order to fit the specific instrument, ensemble scale or design constraints.

A. Generalized approach

The generalized viewpoint-decoupling technique applies to musical ensembles in mixed-reality with two or more musicians. The viewpoint manipulation is in fact a transformation of the virtual scene: instead of translating and rotating the camera within to the environment, one transforms the virtual scene with respect to the camera. This falls into the common “world-manipulation” 3D navigation techniques [10]. It can also be seen as similar to virtual workspaces [11] used to realign virtual avatars with the co-located physical users when exploring larger virtual environments.

Changes in position, rotation and scale of the virtual scene allow musicians to individually explore musical spaces, to access various levels of interaction accuracy or various musical controls. However because each musician can perform this on their end, from the others perspectives, their positions related to the virtual environment ends up being different from their physical position in the room.

This can be seen in Figure 1, in a corpus-based concatenative instrument, where avatars are simply virtual head-mounted displays (HMD). In this case, the musician starts with the default viewpoint (shared with the other musician, so that their avatar is aligned with their physical headset), then increase the scale of the scene (which results in seeing the viewpoint of the other as a large HMD avatar) and then scales down the scene.

In an ensemble with N musicians, there are $N + 1$ viewpoints: one for each musician, and the *default* viewpoint. Each musician is then present twice in the ensemble, as a physical musician and as a virtual avatar. They are perceived aligned by other musicians sharing the same viewpoint.

In order to preserve collective engagement and musical collaboration, our approach then allows each musician to select between their own viewpoint, the viewpoints of others and the default viewpoint. This can be seen in Figure 2 where one musician selects the other to access their viewpoint, using one selection possibility from our design space.

B. Dimensions

Implementing viewpoint-decoupling in a musical ensemble leads to a number of design questions that have not been addressed in the literature. We propose a set of 5 dimensions that will help designers of XR musical ensembles implement viewpoint-decoupling in their own instrument.

This set of dimensions was chosen following the practice sessions of the initial paper [9] and discussions between the authors. They however remain conceptual and were not evaluated beyond the implementation of a few features. We provide perspectives in the discussion on how they could be evaluated and refined.

Viewpoint selection The most direct way of selecting a viewpoint is with buttons like in the initial study by Uro *et al.*. This however might prove inefficient if one need to select from a large number of viewpoints. Alternatively, a menu could be brought up to select a viewpoint from a list. This does however bring up a need to identify each performer so they can be named in the menu.

A more embodied approach could rely on a virtual ray selection technique to directly select the subject of the desired viewpoint. This also opens up considerations for using either a simple virtual ray technique, or more elaborate techniques from the literature addressing inherent limitations such as RayCursor [12] or BubbleRay [13]. The actual targets of this selection can then be either only the virtual representation of the person, their physical body, or any of the two.

We refine further this design space dimension in the following potential values:

- **Button selection:** one or several buttons on the controller allow for switching between viewpoints
- **Graphical menu selection:** a popup 2D or 3D menu with names of all musicians and direct or indirect (*e.g.*, virtual ray) selection
- **Raycast to physical musicians** (or room center): a virtual ray allows for changing viewpoint by selecting directly the musician in the physical space (see Figure 2)
- **Raycast to virtual musicians:** by selecting the avatars directly (which can be much larger or smaller than the physical musician)

These can in addition be combined, for example to enable both fast (*e.g.*, jump to the main view with a button) and more flexible (*e.g.*, choose between other musicians while interacting).

Viewpoint transition The second dimension pertains to how one transition from one (origin) viewpoint to another (target) viewpoint. It is akin to the targeted movement category of 3D navigation [10].

Here are the four values that we envision.

- **Instantaneous:** akin to a 3D teleportation technique with respect to the instrument, *i.e.*, one jumps from the origin to target.
- **Transform automatic:** akin to predefined camera transformations, there is an interpolation between origin and target viewpoints across a fixed duration.
- **Transform controlled:** the transition is manually controlled, *e.g.*, by "grabbing the air" and pulling to interpolate until one reaches the target viewpoint.
- **Portal controlled:** a preview of the other viewpoint is displayed as a window in the virtual environment.

The choice for this dimension should consider continuity of context awareness and motion sickness, but a musical application also raises a need for minimised disruption of interaction, *e.g.*, allowing for transitioning while playing.

Viewpoint association This dimension describes how the association between virtual avatars (which represent musicians' viewpoints) and physical avatars/musicians is visualised. It is an essential aspect as it allows one to perceive where each musician is in the virtual scene.

- **Labels:** text labels with the name of each musicians are attached to the virtual and physical avatars.
- **Colours:** the same colour is used for the virtual and physical avatars.
- **Virtual links:** a 3D line connects the physical and virtual avatars at all time.

The choice on this dimension relates to the visual complexity of the virtual scene, *e.g.*, links might visually overload the physical space.

Viewpoint visualisation This dimension describes how the decoupled viewpoint is visualised within the virtual scene. This has an effect on how one perceives the viewpoints currently used by others and the overall awareness. We foresee the following values for this dimension of the design space.

- **Viewpoint source:** As shown in Figure 1, the first possibility is to display only the transformation of the virtual avatar, *e.g.*, the HMD or body.
- **Viewpoint frustum:** A more complex possibility is to represent the full frustum perceived by one musician.
- **Viewpoint attribution:** Finally one can display both the frustum and the source of the viewpoint (is it one's own or are they using someone else's viewpoint).

The choice for this dimension might also have an impact on the visual overload of the scene, and should therefore be decided as a compromise between a gain in awareness and a loss in readability.

Interaction Level-of-Detail Finally, the opportunity of changing the scale of the instrument through viewpoint manipulation enables changes in the gestures and interactions. In fact, when scaling viewpoint, interaction with the instrument changes scale, therefore precision in control gesture might vary. The simplest example might be our first example instrument shown in Figure 1. When the user zooms in, interacting with the musical spheres is easier and leads to more accurate control of the position of the control within the spheres. On the other hand, when zoomed-out the small spheres do not allow for precisely controlling the position within them.

We therefore envision three possible values.

- **Scaled:** the size of the virtual scene reduces or increases the gesture precision.
- **Cut:** the precision is preserved by reducing the control values interval within the virtual instruments (*e.g.*, a smaller fader only gives access to a subset of the values).
- **Changed:** depending on the scale of the virtual scene, the behaviour of the instrument changes completely (*e.g.*, continuous controls become discrete triggers of sounds)

III. IMPLEMENTATION

In order to implement a prototype of our design space, we relied on the IVMI-builder framework [14] that combines Godot for immersive visual rendering and 3D interaction, and PureData for sound synthesis and analysis. We used Quest 3 mixed-reality headsets.

A. Spatial alignment

The implementation first requires aligning the origin of the tracked space for all XR headsets. To that extent, we rely on OpenXR spatial anchors, implemented in Godot. Each musician aligns a small virtual plane to the same physical panel placed in the room. Once this is done, the anchor is retrieved, and the application assigns its transform to the root of the virtual environment, below which are the physical avatars, the instruments and so on. This alignment only needs to be performed once in a given space. Each client then continuously transmits the transformation of the headset and controllers to other clients, so that the virtual scene is spatially aligned between musicians.

B. Avatars

As mentioned earlier, each musician is represented by a physical avatar, always attached to their HMD in the physical space, and a virtual avatar, that represents their potentially decoupled viewpoint within the virtual scene. Physical avatars are children of the scene anchor, so that they are spatially aligned with the physical musicians after spatial synchronisation. For a first attempt, we chose to use simple models of the Quest 3 headset as avatars and simple spheres to represent

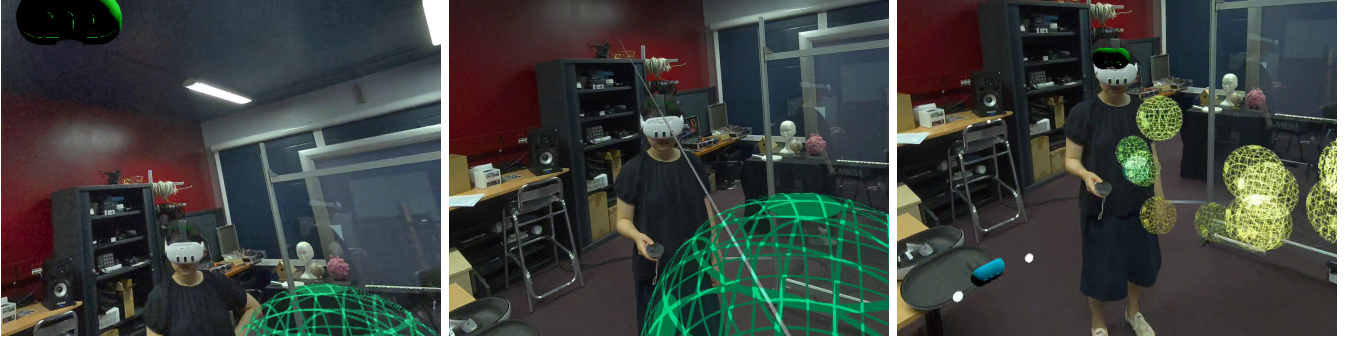


Fig. 2. Example technique for Viewpoint Selection: From a personal zoomed-in viewpoint (left), the musician uses a virtual ray to another physical musician (middle) to jump to their viewpoint (right)

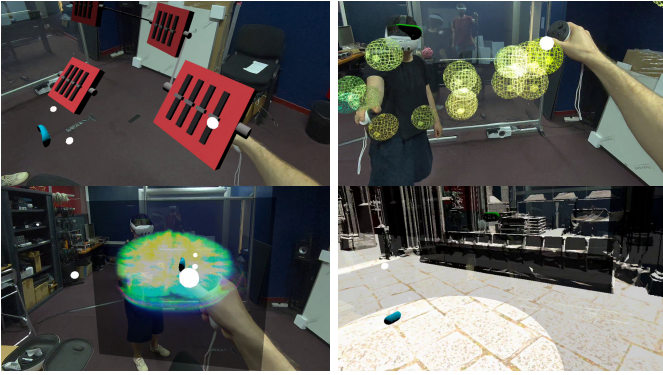


Fig. 3. Example envisioned collaborative mixed-reality instruments with integrated viewpoint-decoupling (left to right, top to bottom): Modular synthesiser with two levels of interaction accessed simultaneously, Corpus-based concatenative synthesis where one can travel to different parts of the corpus, Sonified Volumetric texture with one musician inside and the other outside, Large scene (here a church) with spatialized sound sources

the controllers. Physical avatars are half transparent, so that the physical musician remains visible enough. Virtual avatars are also 3D HMD models but more opaque so that they are clearly visible in the virtual scene.

C. Viewpoint transformation and selection

In our current implementation, the viewpoint transformation is done through the manipulation (translation, rotation, scaling) of the virtual scene, through a "grabbing the air" metaphor (or Point-Tugging [15]), similar to what was proposed by Uro *et al.*. Grabbing (with the grasp button on the Quest 3 controllers) with one hand allows for translating the virtual scene, and grabbing with two hands for translating, scaling (with distance between hands) and rotating (with respective position between hands).

IV. EXAMPLE INSTRUMENTS

In this section, we envision the integration of viewpoint-decoupling in four XRMIs, based on more or less common paradigms for musical expression in extended reality. They can be seen in Figure 3, although only the corpus-based system is working, the others remain mockups.

A. Corpus-based concatenative instrument

Similar to GrainyCloud [9], we extract a corpus of grains, created by analysing audio features of a recording using the PureData implementation of Flucoma [16]. Grains are then represented as spheres in the virtual scene, and can be played by passing the controller inside them. The position of the controller controls the tempo at which the grain is played, a bandpass filter on the grain, and the grain duration. Using viewpoint decoupling, musicians can then explore various parts of the corpus, zoom-in to focus on specific grains, or zoom-out to quickly play multiple grains.

B. Patching environment / Virtual Modular Synthesiser

Another potential application are patching environments and virtual modular synthesisers, which are common in the musical expression literature [17], [18] but also in commercial applications such as Patchworld.

Viewpoint-decoupling here enables ensembles to distribute roles between a patching shared view and individual viewpoints to focus on each module for fine grained interactions.

C. Volumetric texture / Pointcloud sonification

Another potential application is through the musical sonification of pointclouds or of volumetric textures. In this context, viewpoint-decoupling could allow for combining full body exploration of large volumetric textures or activation of smaller textures, as proposed by Berthaut [19].

D. Musical manipulation of Virtual Acoustic spaces

Finally, viewpoint decoupling can be used for musical 3D navigation [20], in the exploration of spatialized sound sources, by alternating between individual and shared exploration phases. It could also provide the possibility of manipulating spatial sound sources and / or virtual architecture (materials, geometries, propagation medium) in order to affect the overall sonic response of a virtual building or landscape.

V. PERSPECTIVES

One major limitation of this paper is the lack of formal evaluation, which prevents from validating the dimensions of the design space and its generative power [21]. We envision this evaluation in the form of sessions with groups of

musicians with the same mixed-reality musical instrument. The instrument will give access to a range of values for each dimension of our design space, *e.g.*, viewpoint selection with buttons, menu or virtual ray. At the beginning of each experiment, participants will learn how to use the instrument. The experiment will then be divided into multiple practice blocks, each with a different set of sounds and combination of our design space dimensions. For instance, one block will rely on menu selection of viewpoints, viewpoint frustum visualisation, and label viewpoint association. Between sessions, we will vary the size of the groups (from 2 to 5), in order to evaluate the scalability of each combination. We will rely on both behavioral analysis (frequency and nature of changes in viewpoints, proxemics), reflexive thematic analysis of interviews, and usability and social presence questionnaires. This will allow us to understand the specific impact of dimensions of our design space.

The second limitation is the testing within a single instrument. As proposed in the previous section, we will conduct a long-term evaluation, *i.e.*, across multiple sessions, of the use of viewpoint-decoupling in multiple instruments with one ensemble dedicated to each. This will allow us to understand the applicability of each dimension in various musical contexts.

VI. CONCLUSION

In this paper, we presented a design space for viewpoint-decoupling in Extended Reality Musical Instruments. We generalize the idea proposed in [9], with ensembles of more than two musicians, and we explore the potential of this technique in various potential instruments. We believe that it enables the combination of benefits from VR and MR ensembles, and that it can help instrument designers and ensembles explore novel collaboration strategies, both in composed and improvised performances.

However, this paper merely proposes the conceptual bases and an example implementation that can support this novel technique. Future work will need to evaluate our approach through a series of experiments, in order to refine the set of dimensions and propose a full implementation that will be released as part of the IVMI-builder open-source framework [14].

REFERENCES

- [1] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.
- [2] F. Berthaut, M. Desainte-Catherine, and M. Hachet, "Drile: an immersive environment for hierarchical live-looping," in *New interfaces for musical expression*, 2010.
- [3] C. Rinaldi and C. Centofanti, "The musical metaverse: Advancements and applications in networked immersive audio," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 2024, pp. 1–7.
- [4] R. Schlagowski, D. Nazarenko, Y. Can, K. Gupta, S. Mertes, M. Billingham, and E. André, "Wish you were here: Mental and physiological effects of remote music collaboration in mixed reality," in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 2023, pp. 1–16.
- [5] A. Boem, M. Tomasetti, and L. Turchet, "Between immersion and usability: A comparative study of 2d and mixed reality interfaces for remote music making," *International Journal of Human-Computer Studies*, 2025.
- [6] A. Boem, S. Skouras, G. B. Hinkis, M. Mousset, and L. Turchet, "Co-designing virtual reality musical instruments and spatial layouts for collaborative music-making," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, B. Gaster, J. Tragtenberg, A. Xambó, and T. Mitchell, Eds., London, United Kingdom, June 2026, pp. 820–831.
- [7] L. Men and N. Bryan-Kinns, "LeMo: Exploring Virtual Space for Collaborative Creativity," in *Proceedings of the Conference on Creativity and Cognition*. San Diego, CA, USA: ACM, Jun. 2019, pp. 71–82.
- [8] J. A. Harrigan, "Proxemics, kinesics, and gaze," *The new handbook of methods in nonverbal behavior research*, pp. 137–198, 2005.
- [9] P. Uro, F. Berthaut, T. Pietrzak, and M. Wanderley, "Decoupling Physical and Virtual Spaces for New Collaboration Strategies in Co-Located Mixed Reality Instruments," in *NIME 2025 - International Conference on New Interfaces for Musical Expression*, Canberra, Australia, Jun. 2025. [Online]. Available: <https://inria.hal.science/hal-05126240>
- [10] J. Jankowski and M. Hachet, "A survey of interaction techniques for interactive 3d environments," in *Eurographics 2013-STAR*, 2013.
- [11] Y. Zhang, H. Nguyen, N. Ladevèze, C. Fleury, and P. Bourdot, "Virtual workspace positioning techniques during teleoperation for co-located collaboration in virtual reality using hmds," in *2022 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 2022, pp. 674–682.
- [12] M. Baloup, T. Pietrzak, and G. Casiez, "Raycursor: A 3d pointing facilitation technique based on raycasting," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3290605.3300331>
- [13] Y. Lu, C. Yu, and Y. Shi, "Investigating bubble mechanism for raycasting to improve 3d target acquisition in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 35–43.
- [14] F. Berthaut, "GDPD / IVMI-BUILDER: A Libre Software Framework for Extended Reality Musical Instruments and Sonic Installations," in *Proceedings of the 19th Linux Audio Conference*, Lyon, France, Jun. 2025. [Online]. Available: <https://hal.science/hal-05095979>
- [15] N. Coomer, S. Bullard, W. Clinton, and B. Williams-Sanders, "Evaluating the effects of four VR locomotion methods: joystick, arm-cycling, point-tugging, and teleporting," in *Proceedings of the 15th ACM Symposium on Applied Perception*, ser. SAP '18, New York, NY, USA, 2018, pp. 1–8.
- [16] P. A. Tremblay, O. Green, G. Roma, J. Bradbury, T. Moore, J. Hart, and A. Harker, "The fluid corpus manipulation toolbox," Jul. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.6834643>
- [17] J. Bell, "Networked Music Performance in PatchXR and FluCoMa," in *Proceedings of the International Computer Music Conference*. Shenzhen, China: CUHK-Shenzhen, Oct. 2023.
- [18] M. Buffa, D. Girard, and A. Hofr, "Using Web Audio Modules for Immersive Audio Collaboration in the Musical Metaverse," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–10.
- [19] F. Berthaut, "Musical Exploration of Volumetric Textures in Mixed and Virtual Reality," in *2021 NIME Proceedings*, ser. 2021 NIME Proceedings, Shanghai, China, Jun. 2021. [Online]. Available: <https://hal.science/hal-03260466>
- [20] —, "3D interaction techniques for musical expression," *Journal of New Music Research*, vol. 49, no. 1, pp. 60–72, Jan. 2020.
- [21] M. Beaudouin-Lafon, "Designing interaction, not interfaces," in *Proceedings of the working conference on Advanced visual interfaces*, 2004, pp. 15–22.