

# First Steps Towards Augmented Reality Interactive Electronic Music Production

Valentin Bauer\*

LISN, CNRS  
Université Paris-Saclay

Tifanie Bouchara

CEDRIC (EA4626), CNAM  
HeSam Université

## ABSTRACT

Immersive music production starts to integrate Virtual Reality (VR) and Augmented Reality (AR) technology. VR creates a virtual environment separated from real-life, often based on the performer-audience relationship. AR goes further by expanding the usual user listening experience over headphones but remains unexplored. This paper presents a work in progress of an AR compositional platform which allows to create interactive music experiences, through the creation process of an electronic piece.

**Keywords:** Spatialized music production, Augmented reality, New interface for musical expression.

**Index Terms:** Applied computing—Arts and humanities—Media arts; Human-centered computing—Human computer interaction (HCI) —Interaction paradigms—Mixed / augmented reality

## 1 INTRODUCTION

Recent trends in the development of immersive spatial audio technologies have opened new artistic possibilities for audio composers, by considering the space dimension in the same way as the other four audio parameters, i.e. intensity, duration, frequency, and timbre [3]. Spatial music compositions often concern the contemporary music field [20], but a recent democratization of the spatial audio tools, particularly with the advent of spatial audio over headphones (i.e. binaural audio), has led projects to emerge in others music fields, e.g. electronic music with for instance the *Sequence* project [26].

While audio often has an ancillary role compared with visuals in Virtual Reality (VR) and Augmented Reality (AR) experiences, audio-first VR/AR experiences, i.e. VR/AR experiences with audio as a primary focus, can allow to create new forms of expressivity by adding a multimodal immersive aspect to spatial audio compositions. In particular, adding visuals to immersive audio media can allow not only to visualize it, but to create a meaningful relationship with the music content. Hence, it can support its understanding or add a complementary layer of meaning. The recent development of portable and affordable technology such as Head Mounted Displays (HMDs) has allowed interdisciplinary audio-first VR experiences to be developed, between art and sciences, with five main focuses [9], in particular Immersive Virtual Musical Instruments (IVMI) [22] and compositional frameworks and live performance practices [8], but also auditory localization and perception, cinema and installation, and pedagogy.

The “Don’t port” design principle for creating artful new media posits that considering the possibilities that a new medium has to offer is necessary to generate ideas and design projects for this new medium, and avoid copying elements from an old medium [1]. Though, IVMIs and VR compositional frameworks often remain close to the old medium, by extending already existing music practices. For instance, many IMSIs have been developed, such as virtual xylophones or air guitar in room-scale environments [17], using three main interaction techniques, i.e. selection, manipulation and navigation [5]. About VR compositional frameworks, some of them allow to procedurally generate sounds with modular interfaces, enabling collaborative performances and improvisation [21,27]. This paper considers the VR multisensory and interactive potentials to present a project oriented both for the listener and composer, drawing from IMSIs and compositional frameworks.

Audio-first VR experiments allow the listener to experience spatial audio through a multimodal interactive and immersive artwork, drawing from interactive and multimodal art installations. The listener could then be considered as a performer who experiences the spatial audio piece, by navigating it using natural gesture interaction (i.e. without controllers), with some musical events being already planned by the composer and others being triggered by his/her behavior. In comparison with art installations, such audio-first VR performances are directly accessible to the user, through HMDs and ubiquitous technologies such as smartphones, but remain unexplored. The work in progress here presented explores these aspects through an AR multimodal audiovisual spatialized music production.

To generate new designs and ideas taking advantage of the multimodal VR interaction potential, Atherton et Wang [1] suggest that audio-first VR experiences could focus more on the being, i.e. inhabiting the world around us, rather than on the doing, i.e. taking action with an intended outcome. Some projects already adopted this approach. The *Plumage* project consists in discrete granular sonic representation in VR using visual feathers, that the user can navigate to explore the music environment [15]. Then, Buckley et Carlson [8] focus on VR compositional frameworks to deliver music-driven performances, and suggest that composers could combine predetermined interactive elements to make the music affordances of a VR soundscape intuitively discoverable by the user being in the virtual space. At last, focusing more on the being can also imply offering clear action/reaction interactions along with some unexpected environmental reactions to engage the user [14].

Most IMSIs experiments consider musical performances through the lens of the relationship between a performer and an audience, inspired from the context of a stage [6]. Yet, users often listen to music in the private sphere over headphones, while seeing their usual environment. Since AR adds information on top of reality, it can extend this usual music listening experience by making it multimodal and focusing on the being of the experience. Though, most audio-driven experiments are developed in VR, leaving AR

---

\*e-mail: valentin.bauer@limsi.fr

quite under-explored. An AR user oriented scenography could be developed, intertwining audio and visuals to generate creative artworks based on first-person user's gaze [9]. This complementary role of visuals also draws from previously addressed considerations in audiovisual live electronic performances [12]. Hence, AR can bring interactivity for the user through multimodal interactive and immersive performances connecting with the real space, which could renew spatial music experiences.

This paper presents a proof of concept of such an AR multimodal and interactive audiovisual spatialized music production. After presenting the technological platform, we will address the AR music production methods through the use case of the creation process of *REM*, i.e. an AR electronic music piece. The discussion will then detail the remaining work so that the system can be used by other composers and researchers, and raise future perspectives.

## 2 THE SYSTEM

The main requirements were to connect with the traditional listening experience over headphones, augment the real environment with immersive spatial electronic music, and finally supplement it with visuals and natural embodied interaction, i.e. without controllers. Hence, an AR see-through platform was used (see figure 1), which means that a live feed of the real space is captured by a stereoscopic camera plugged onto a HMD, then sent inside it, and augmented by virtual objects. To that respect, the Zed Mini stereoscopic camera was used on top of the HTC Vive Pro headset, to get a non-perceivable latency of around 60ms, visual resolution of 720p per camera, and wide field of view of 90°horizontal 60°vertical. The front cameras of the HTC Vive Pro were previously tried but abandoned because achieved lower performances in terms of resolution (480p per camera) and latency (100 to 200ms). The AR system *Hololens 2* was also not considered because had a narrower horizontal field of view of 52° [28].

The system was developed using Unity3D software version 2019.4.3f1, and Google Resonance Audio SDK version 1.2.1 for the spatial audio rendering due to its high-quality with 3rd-order ambisonics, to enhance the binaural audio externalization cues over headphones [13].

The music system adopts an Object-Based Audio perspective, drawing from current immersive approaches for spatialized music production [24]. Thus, each musical source corresponds to an audio object in Unity, which is visually displayed depending on the source. Objects can be spanned vertically and/or horizontally, and follow different trajectories (ellipses, square, etc.) previously planned by the composer. The common metaphor of a cursor moving on a line was used to symbolize the audiovisual objects, as in other art projects [29], or spatial trajectory software, e.g. *Iannix* [16], and *Sound Trajectories* [19]. To prevent graphical objects from being associated with silence, they only appear when the loudness of the corresponding musical source is high enough. Though, the visibility threshold can be changed if the composer previously chooses to visualize some silent audio objects.

The system allows static and dynamic events. Static events are planned by the composer, whereas dynamic events are triggered by the user, based on his/her head elevation, body position in space, voice input. The architecture uses four main scripts. A master handles the timeline and changes the events that are available depending on time. It communicates with a script attached to every

sound object which controls its specificities (shape, movement etc.) to trigger the corresponding events. Two other scripts inform the master about user actions: the first one checks for voice input based on the user voice input loudness using the HTC Vive microphone, the second checks for the user gestures, i.e. head and body.

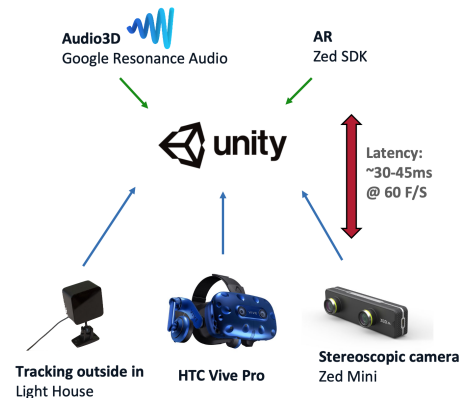


Figure 1 System architecture

## 3 AR MUSIC PRODUCTION: USE CASE OF THE MUSIC PIECE *REM*

Since the article presents a new way to produce AR spatialized music, the composition process requires different techniques not commonly used by sound artists. An attempt was done to formalize them through the use case of *REM*, an AR electronic music piece.

### 3.1 Pre-production

As a composer and audio-engineer, the first author chose an electronic music composition style with some high-pitched percussive elements. He was inspired by previous audio-only electronic immersive music experiences from the *Sequences* project [26], and previous spatialized music projects [4,25], where these types of elements were well externalized over headphones.

During the pre-production period, he used his usual stereo workflow so that to keep his benchmarks, using Logic Pro X software. He also took many notes about which visuals and movements to use for which audio content. At the end, a track-by-track export including audio effects was made to then import each channel output as an audio object in unity. A stereo mix of the pre-production was also made, which is available at this link<sup>1</sup> [https://drive.google.com/drive/folders/1XIBZtPm54ZEJ0Ho\\_NfzXWf1PV\\_7JNSH2?usp=sharing](https://drive.google.com/drive/folders/1XIBZtPm54ZEJ0Ho_NfzXWf1PV_7JNSH2?usp=sharing).

### 3.2 Production

The production phase was entirely conducted in Unity3D software and focused on four aspects hereafter detailed, i.e. musical intent, audiovisual representations, audio mixing, and interactivity. A video of the project is available at the link aforementioned.

#### 3.2.1 Musical intent

*REM* stands for Rapid Eye Movement, i.e. a stage of sleep where most recalled dreams occur. The piece questions the boundary between real, virtual, and mental spaces, by visually displaying spatial sounds matching the mental visual representations of the composer. These audiovisual representations, usually only imagined by the listener, give an additional layer of meaning to the music content. By doing so, the composer starts questioning the role of internal representations in everyday embodied perceptions.

<sup>1</sup> When the paper will be accepted, it will be hosted on a permanent webpage.

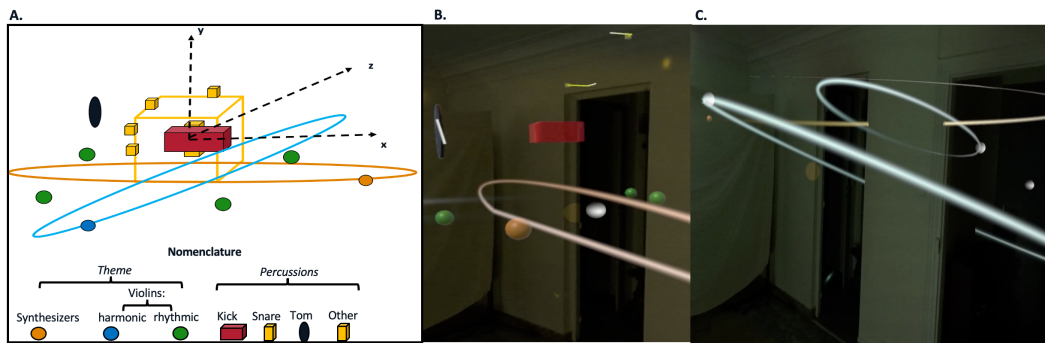


Figure 2. Schematic of visuals (A), User view of the first section (B), User view of the outro (C)

### 3.2.2 Audiovisual representations

Musical objects were represented by simple geometric shapes (e.g. cubes, spheres etc.), using two simple metaphors: the more low-pitch the sound, the bigger the visuals; the further the sound, the smaller the visuals. Object visuals (shapes, colors, etc.) depended on their musical roles, based on esthetic choices made by the composer. For instance, resonant objects left particular rays, and red and yellow color tones represented the drums' elements (see Figure 2B). The color of the environmental lighting was modulated to highlight the different sections of the music piece, i.e. blue tones for the introduction, outro and middle sections, orange tones for the first and last sections. Figure 2B and 2C respectively display the user view in the first section and the outro.

### 3.2.3 Audio mixing

The audio mixing concerned the intensity and position of the audio objects, since binaural audio intelligibility did not require more audio processing (e.g. compression, side chain), and adding a reverberation. Slight doppler effects and increased first reflections were used to strengthen the 3D audio cues. Stereo was only used to mix the kick and snare so that to convey the cohesive effect of electronic music, according to recent works about binaural mixing in the hip-hop music field [25]. By giving the user the possibility to be physically surrounded by sounds which are visually represented, the project raised questions about how to handle the music dynamic aspects and reverberation specificities. Indeed, the visuals and embodied navigation possibilities modified the auditory perception of these aspects by the audio engineer, in comparison with his spatial audio mixing experience with no visuals. The platform was currently limiting to handle these considerations, but next versions may allow to dynamically adjust the sound characteristics (reverberations, intensity, etc.) depending on user actions.

### 3.2.4 User Interactions

Three natural user interactions were chosen: looking at an audio object to change its trajectory; stooping or standing up to shift the cutting frequency of a low-pass filter; navigating in space to trigger audio effects (e.g. delays). Voice interaction, nodding the head, and gesture speed were abandoned because deemed unnatural by the composer to connect with a music experience over headphones.

## 4 DISCUSSION AND CONCLUSION

This section first presents the current work in progress and possibilities to extend the AR music production project that was presented. It then outlines perspectives which appeared.

Seven people were contacted during the production phase to get insights about this work in progress, with various expertise in

different artistic and technical fields: music, spatial audio, photography, cinema, VR. They were contacted by phone or email and gave their advice after looking at videos of the music production, since the Covid pandemic situation did not allow to undertake user testing. Little explanation was given when the video was sent to avoid influencing the participants, mostly about the fact that the spatial audio piece was normally experienced in AR through an HMD.

All participants enjoyed discovering the project. They suggested to explore more the audiovisual relationship to create more meaningful representations. Hence, further research is needed to find "meaningful" AR audio descriptors, so that to visually represent sounds, be it in a congruent, or non-congruent way so that to give an additional layer of meaning to the listener. Drawing from previous research [2,7], this could help to prevent the confusion and cognitive load sometimes experienced by participants, due to seeing too many elements at the same time without inferring a clear link between them. Furthermore, this raises questions about which visuals to use for which sounds, and in relation with which interaction parameters. This question remains unexplored to our knowledge, since only the opposite aspect of adding audio to visuals was explored [11]. Then, further examining the interactions between real and virtual objects was proposed, in particular by considering the role of the body. This focus could help to address the role of expressivity and embodiment in AR music, building from previous works about embodied musical interaction [23]. It could also lead to dynamically adjust AR music depending on the real space. This last possibility would then renew with a long tradition of creating music for specific spaces, for instance with the audiovisual music piece called *Poème Electronique* created by Le Corbusier, Xenakis, and Varèse at and for Philips Pavilion in 1958.

Several iterations will be needed to refine the AR platform, depending on the aspects aforementioned. To precise them, real testing has to be conducted with various user, i.e. music performers, music composers and non-musicians. For instance, comparing the audiovisual rendering through the AR platform with two-dimensional videos could help to evaluate the role of the medium in the listener's understanding of the musical piece proposed and compare the respective cognitive loads of listeners depending on the system. Comparing these systems would also help to evaluate if audio descriptors proposed in non-AR audiovisual research could be used in AR. After that, comparing different interaction possibilities in AR for a selection of audio effects could help to assess which musical interactions are most engaging for the end-users. At last, engaging in a collaboration with visual artist would help to investigate ways to add visuals to the audio content in order to offer complementary layers of meaning.

So far, the AR platform developed cannot be used as a comprehensive compositional environment but was an integrator which allowed to raise many questions.

Further research needs to focus on building immersive AR/VR music production tools, taking into account the entire spatialized music production pipeline, and drawing from recent works about sound spatialization [18]. Indeed, the platform described in this paper required the composer to connect with his usual stereo workflow, but it immediately appears that this can hinder and slow down the creation process. New tools better connected to this new medium have to appear so that to be able to think for this medium, and not only adapt old practices, in line with the “Don’t port” design principle already mentioned [1]. Furthermore, this lack of tools is surprising, since many VR/AR visual authoring tools already exist, some being already professionally used by animators in the 3D animation field [10]. About immersive music production tools, recent projects have emerged, such as the modular project called PatchXR [27], but more research has to be conducted.

At last, this project came from a wider project regarding how to use interactive spatial music and audiovisual conception for considering creative tools used in a therapeutic context. This calls for participative projects including patients in the creative design process via intuitive immersive music production tools.

## REFERENCES

- [1] J. Atherton and G. Wang, “Doing vs. Being: A philosophy of design for artful VR,” *Journal of New Music Research*, vol. 49, no. 1, pp. 35–59, Jan. 2020.
- [2] E. Audry and J. Garcia, “Towards congruent cross-modal audiovisual alarms for supervision tasks,” presented at the International Workshop on Haptic and Audio Interaction Design - HAID2019, Lille, France, 2019, p. 5.
- [3] M. A. J. Baalman, “Spatial Composition Techniques and Sound Spatialisation Technologies,” *Org. Sound*, vol. 15, no. 03, pp. 209–218, Dec. 2010.
- [4] V. Bauer, H. Déjardin, and A. Pras, “Musicians’ binaural headphone monitoring for studio,” *Audio Engineering Society Convention 144*, p. 10, 2018.
- [5] F. Berthaut, “3D interaction techniques for musical expression,” *Journal of New Music Research*, vol. 49, no. 1, pp. 60–72, Jan. 2020.
- [6] F. Berthaut, V. Zappi, and D. Mazzanti, “Scenography of Immersive Virtual Musical Instruments,” p. 6, 2014.
- [7] T. Bouchara, C. Jacquemin, and B. F. G. Katz, “Cueing multimedia search with audiovisual blur,” *ACM Trans. Appl. Percept.*, vol. 10, no. 2, pp. 1–21, May 2013.
- [8] Z. Buckley and K. Carlson, “Towards a Framework for Composition Design for Music-Led Virtual Reality Experiences,” in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, 2019, pp. 1497–1499.
- [9] A. Çamcı and R. Hamilton, “Audio-first VR: New perspectives on musical experiences in virtual environments,” *Journal of New Music Research*, vol. 49, no. 1, pp. 1–7, Jan. 2020.
- [10] A. Cannavo, C. Demartini, L. Morra, and F. Lamberti, “Immersive Virtual Reality-Based Interfaces for Character Animation,” *IEEE Access*, vol. 7, pp. 125463–125480, 2019.
- [11] M. Chion, *La Musique au cinéma*. Paris: Fayard, 1995.
- [12] N. N. Correia, D. Castro, and A. Tanaka, “The Role of Live Visuals in Audience Understanding of Electronic Music Performances,” in *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, London United Kingdom, 2017, pp. 1–8.
- [13] M. Gorzel, A. Allen, I. Kelly, J. Kammerl, A. Gungormusler, and H. Yeh, “Efficient Encoding and Decoding of Binaural Sound with Resonance Audio,” *Audio Eng. Soc. Convention*, p. 12, 2019.
- [14] V. Isnard and T. Nguyen, “L’étrangeté perceptive en réalité virtuelle,” IRCAM, 2020.
- [15] C. Jacquemin, R. Ajaj, R. Cahen, Y. Olivier, and D. Schwarz, “Plumage: design d’une interface 3D pour le parcours d’échantillons sonores granularisés,” in *Proceedings of the 19th International Conference of the Association Francophone d’Interaction Homme-Machine - IHM ’07*, Paris, France, 2007, p. 71.
- [16] G. Jacquemin, T. Coduys, and M. Ranc, “IANNIX 0.8,” presented at the Journées d’Informatique Musicale, Mons, France, 2012, p. 10.
- [17] T. Mäki-Patola, J. Laitinen, A. Kanerva, and T. Takala, “Experiments with Virtual Reality Instruments,” p. 6, 2005.
- [18] J. D. Mathew, S. Huot, and B. F. G. Katz, “Comparison of spatial and temporal interaction techniques for 3D audio trajectory authoring,” *J Multimodal User Interfaces*, vol. 14, no. 1, pp. 83–100, Mar. 2020.
- [19] M. Mercier, “Sound Trajectories.” [Online]. Available: <https://www.tripinlab.com/>. [Accessed: 20-Jan-2021].
- [20] F. Otondo, “Contemporary trends in the use of space in electroacoustic music,” *Org. Sound*, vol. 13, no. 1, pp. 77–81, Apr. 2008.
- [21] M. Palumbo, A. Zonta, and G. Wakefield, “Modular reality: Analogues of patching in immersive space,” *Journal of New Music Research*, vol. 49, no. 1, pp. 8–23, Jan. 2020.
- [22] S. Serafin, C. Erkut, J. Kojas, N. C. Nilsson, and R. Nordahl, “Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions,” *Computer Music Journal*, vol. 40, no. 3, pp. 22–40, Sep. 2016.
- [23] A. Tanaka, “Embodied Musical Interaction: Body Physiology, Cross Modality, and Sonic Experience,” in *New Directions in Music and Human-Computer Interaction*, S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson, and M. M. Wanderley, Eds. Cham: Springer International Publishing, 2019, pp. 135–154.
- [24] N. Tsingos, “Object-Based Audio,” in *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*, Focal Press., 2017, pp. 258–289.
- [25] K. Turner and A. Pras, “Is Binaural Spatialization the Future of Hip-Hop?,” *New York*, p. 9, 2019.
- [26] “Sequences.” [Online]. Available: <https://www.france.tv/spectacles-et-culture/festivals/sequences-electronic-live-sessions/>. [Accessed: 07-Jan-2021].
- [27] “PatchXR.” [Online]. Available: <https://www.patchxr.com/>. [Accessed: 18-Dec-2020].
- [28] “Hololens2.” [Online]. Available: <https://www.microsoft.com/fr-fr/hololens>. [Accessed: 07-Jan-2021].
- [29] “Alacran group project.” [Online]. Available: <https://alacrangroup.com/>. [Accessed: 07-Jan-2021].