



HAL
open science

Empty Room, exploring the plasticity of electroacoustic music spatialization in VR with Ambisonic 3D and Binaural techniques

Christine Webster, Anne Sèdes

► To cite this version:

Christine Webster, Anne Sèdes. Empty Room, exploring the plasticity of electroacoustic music spatialization in VR with Ambisonic 3D and Binaural techniques. The Global Composition 2018 Conference on Sound, Ecology, and Media Culture, Oct 2018, Darmstadt-Dieburg, Germany. hal-02276901

HAL Id: hal-02276901

<https://hal.science/hal-02276901v1>

Submitted on 3 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Global Composition 2018 Conference on Sound, Ecology, and Media Culture
Darmstadt-Dieburg/Germany, October 04 - 07, 2018

Empty Room, exploring the plasticity of electroacoustic music spatialization in VR with Ambisonic 3D and Binaural techniques

Christine Webster and Anne Sedes

(webster@sfr.fr) & (anne.sedes@univ-paris8.fr)

<http://www.edesta.univ-paris8.fr>

MUSIDANSE/CICM – EA 1572, University of Paris 8 - St Denis
France

ABSTRACT

Our paper present two 3D spatialization experiments in VR achieved during the creative process of the Empty Room project. The project shows how Virtual Reality can be considered as a good host for actual and future multi dimensional spatialization paradigms especially when we experiment electroacoustic music in VR, using Ambisonic and Binaural techniques.

INTRODUCTION

The Empty Room project is an electroacoustic composition spatialized in 3D in VR created by Christine Webster. The project is at the core part of Christine Webster's actual work in progress thesis : Electroacoustic Composition experimented in Digital 3D space. Exploration and study of emerging compositional techniques in VR. Directed by Anne Sedes, CICM, Paris 8 and co-directed by François Garnier Spatial Media Ensadlab.

An overall presentation of the Empty Room project can be found on HAL archives [1]. In the present paper we will focus on Empty Room's two first spatialization experiments. The Alpha version spatialization using Unity's version 5 native 3D sound properties and the Beta version spatialization using a third part Ambisonic 3D layer in Unity provided by French company 3D Sound Labs, inspired from the code of HOA Library developped at CICM [2].

CONTEXT AND MOTIVATION

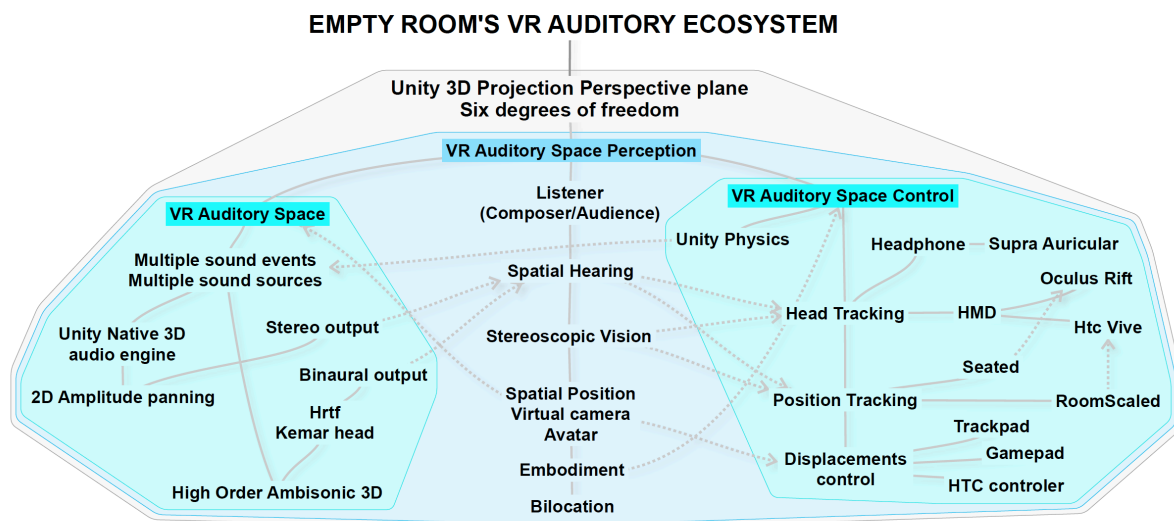
Empty Room was created during our collaboration with EnsadLab Spatial Media [3] in Paris (2013/2016). Our former experiments associating Electroacoustic Music with Virtual Reality were done on Second Life [4] since 2007. The distribution of the first development kit for Oculus Rift DK1 in 2013 transformed the whole VR ecosystem possibilities and opened a worldwide new era of researches and innovations. To produce Empty Room we abandoned Second Life for Unity 3D [5] we considered far more efficient and flexible regarding our future investigations.

OUR HYPOTHESIS

We consider the Virtual 3D scene as a new compositional space to explore for experimental electronic music. We want to detach from the linear process commonly in use in Digital Audio Workstation as well as we want to detach from the Audio Middleware used with game

engines, both optimized for Audio Visual and MultiMedia Post-production techniques for cinema and video games. Starting from this perspective how do we spatialize our composition? How do the sounds behave each other in the composition? How will the embedded 3D tools influence the creative process? How will this sound space be sensed and how will the listening be transformed? The purpose of our research is to confront the practice of *electroacoustic* music with a virtual 3 dimensional geometrical and graphic space we consider as a genuine place for 3D sound composition, fixation and diffusion.

EMPTY ROOM'S VR AUDITORY ECOSYSTEM



Picture 1

To start our experiments we assembled an operating environment we can compare to an ecosystem we call our VR Auditory Ecosystem. Picture 1. This system is made up with digital components accessible through Unity 3D (multi media files, GUI, 2d screen, code), one human in immersion and the necessary external controllers, like HMD, tactile control, and tracking system. The VRAE is a sensible environment divided into 3 groups of components which are synchronized in real time with interdependencies and capacity of interaction and control :

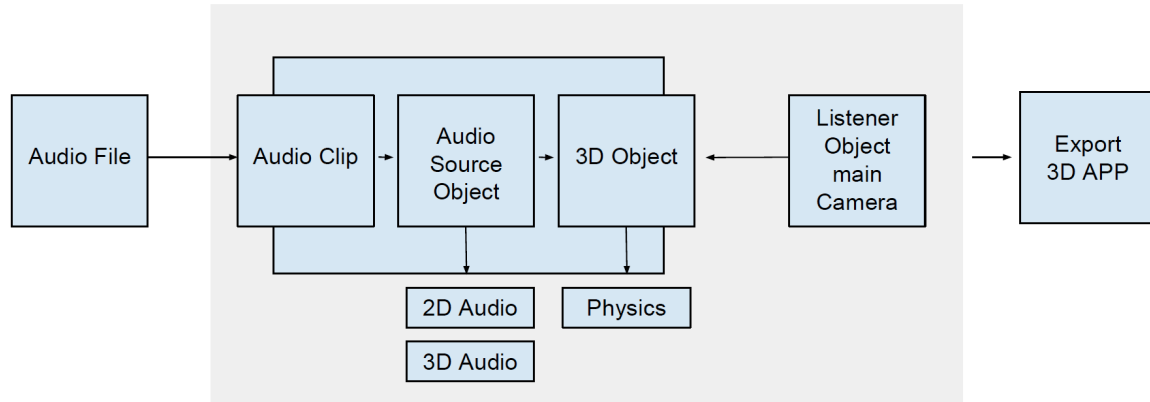
The VR Auditory Space Perception group. This is where we are confronted with the notion of Embodiment, (Lelièvre 2012) Bilocation and Instanciation (Amato 2008) of the Listener/Composer in VR immersion. There, the human senses navigate between two qualities of space, Reality and Virtuality, with the following distribution : the human body is bilocated in sync in the space of reality and in the virtual space, while hearing and vision disconnect naturally from reality and focus to sense only the virtual world. This kind of partial dissociation/association of the senses is very specific to human immersion in VR.

The VR Auditory Space group. Here we handle all the aspects concerning auditory spatial awareness in our Unity 3D scene. Notion such as 3D sound for Virtual Reality (Durant R. Begault 2000), VAS, Virtual Auditory Space (Nicol 2010), Ambisonic in multi media (Daniel 2002), HOA Library [6], Hrtfs and Binaural rendering through headphones (Blauert 1993) take part of the process.

The VR Auditory Space Control group. Here we use two types of controls, the control applied in the digital domain through coding like physical properties assigned to a 3D object and the

external hardware controllers, which are electronic devices working in relationship with the digital domain like headphones, stereoscopic Hmd, tactile controller and tracking systems. The VR Auditory Ecosystem provides all the necessary components to a complete human immersion in a VR Auditory Space.

EMPTY ROOM'S AUDIO WORKFLOW IN UNITY3D



Picture 2

Here we describe Empty Room's audio workflow. Picture 2. For both Alpha and Beta version we used Unity 3D version 5 providing enhanced audio features and compatibility with Oculus Rift and HTC Vive stereoscopic headsets. The audio materials were recorded from a modular synthesizer and mixed in a DAW. We made a stereo premix in Pro Tools for the Alpha version and a Bipan-Transpan mix in Reaper for the additional experiments included in the Beta version. The audio files were individually bounced from the DAW's in 48khz 24bits wav format then imported and compressed in Unity in high quality Vorbis format. Then we deployed each sounds in the 3D scene attaching the audio clips to virtual sound source objects, Unity's AudioSource object. These sound objects can manage 2D sound or 3D sound properties and can be linked to a physical 3D object. In this case the AudioSource object can take advantage of the physical behaviors assigned to the 3D Object such as trajectories, bouncing, hovering or collide effects. In an audio 3D context the audio source object is in constant relationship with the AudioListener object attached to the main virtual camera. The distance, angle and movement between these two objects create the illusion of an auditory space perceived in a synthetic 3 dimensional plane, our VR Auditory Space. In the end of the process the spatial composition can be distributed as a standalone app.

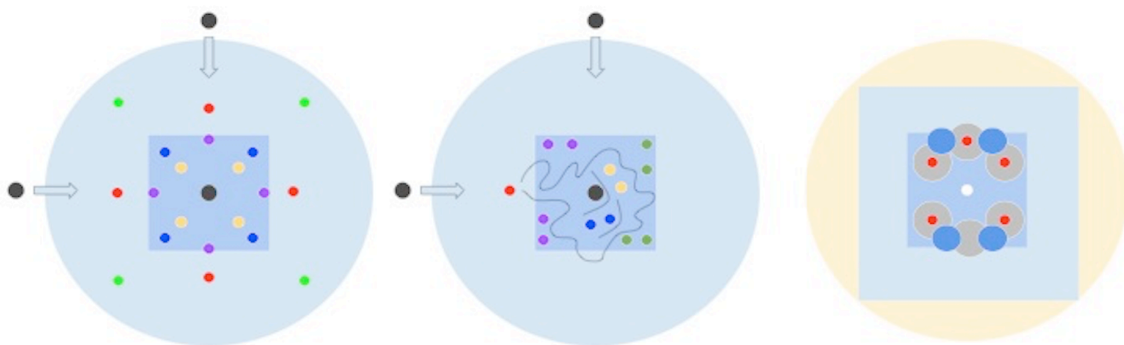
EMPTY ROOM ALPHA VERSION

The intention of Christine Webster, composer of Empty Room is to validate that full human body immersion in VR can be compatible with the sensing of a complex VR Auditory Space. If the VRAS is sensed correctly from the composer's cognitive perspective as well as from the Audience one, we can extend classical 2D and 3D VAS to the VR domain and imagine a new kind of compositional environment. To do so we imagined a musical composition producing different qualities of space to be sensed operating in relation with the graphical 3D space. We defined a musical structure divided in three sequences which form a 10' composition. The composition and visual environment are linked : Seq 1 - 4' - Large and aerial - environment

color Blue. Seq 2 - 4' - Closed, shifted perspective, lost of spatial references - environment color Red. Seq 3 - 2' - generative panic room, confinement - environment color Grey-White. She released the *Alpha version* of Empty Room in September 2015 and developed it on Unity 3D Version 5 for Mac OSX. This version is playable only in seated mode with a gamepad used for the displacements, with Oculus Rift DK2 VR headset providing a stereoscopic view and a supra auricular headphone for a binaural hearing.

Unity's 3D sound features

Unity version 5 provided new sound features among them an integrated mixer working in conjunction with their 3D positional audio pan. This type of pan is in use in the gaming industry since the 1990s and can be more or less sophisticated. In Unity the Spatial Blend parameter 3D position of the AudioSource object takes the source sound and regulates the gains of the left and right ears contributions based on the distance and angle between the AudioListener and the AudioSource object. Providing simple directional cues for the player on the horizontal plane with an Azimuth angle of 180°. No elevation or rear perception is possible. Additional 3D settings permit to adjust the roll off curve of the Volume with a minimum and maximum attenuation distance. The sounds radiate at 360° in the distance zone. The 3D position pan creates a dynamic immersive stereo soundfield from where we can manipulate the spatial field depth to obtain different qualities of perspective. We bypassed additional integrated features like Pitch and Doppler effects or Unity's Reverb Zone and audio FX. But we used the mixer and the mixer groups.



Picture 3 - Spatialization maps of Empty Room's Seq 1, 2 and 3 (beta)

The spatialization map

Webster has created a 2D spatialization map originally designed to perform with 64 discrete virtual channels plus 10 mobile channels. She had to down sized it to 56 channels to preserve a feeling of spatial clarity. Picture 3 shows how we organized the channel distribution for each of the 3 Sequences.

Seq 1 - the sound is distributed through 4 groups : Group 0 plays without directivity (ambient air sound). Group 1 plays the sound perceived from the central platform (blue square). Group 2 plays the sounds perceived beyond the platform limits (blue circle) Group 3 plays 3 mobile channels translating in the x, y and z axis, attached to a black monolith (black dots). To stay

close to the spatial dramaturgy of Seq 1 described as *large and aerial*, the spatialization is designed as a structure made of 5 quadraphonic layers stretching from the center to the external limits. The three central layers form a quadraphonic-octaphonic structure projecting dotted sound events responding each other in a question-answer mode. Combined with the two external quadraphonic layers (distant in red, very distant in green) we obtain a homogeneous sound field crossed by 8 perspective lines providing different depth of field. Seq 2 - *Closed, shifted perspective, lost of spatial references* - we have a rumble ambient sound in Group 0, in Group 1 we played with two pair of shifting stereo (yellow and blue) and four pairs of mono clusters (green and violet). Group 2 is disabled and Group 3 got a new sound element : the hectic cube, travelling trough the platform in random mode at fast pace. Seq 3 – In the Alpha version this sequence got temporary sounds. The final spatialization was set in the Beta version.

EMPTY ROOM BETA VERSION

Integration of the VR Audio Kit plugin

Unity allows replacing the standard 3D panner by more advanced ones with an open interface, the Audio Spatializer SDK, which is an extension on top of the Native Audio Plugin SDK. From this interface, French company 3D Sound Labs created a VR Audio Kit providing an High Order Ambisonic 3D [7] spatializer for Unity. The spatializer can handle simultaneously 32 audio sources in a perceptive 3D scene, with a spatial resolution going from order 1 to 6, in the horizontal and vertical plane, delivering a full spherical Auditory 3D perception. The output is binauralized with a default Hrtf characteristic [8].

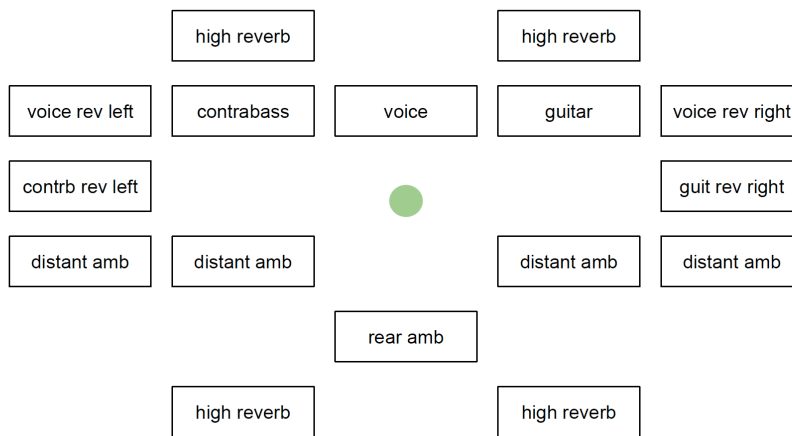
The Alpha version using Unity's native 3D panner runned from end 2015 until june 2016. We developed the Beta version with 3D Sound Labs VR Audio Kit for Oculus Rift since September 2016 and achieved a full Ambisonic 3D version of Empty Room in August 2017. The HTC Vive version started in October 2017.

BiPan-Transpan audio scene transposition

The Beta journey started with a primary test we made with Jean-Christophe Messonnier, sound engineer at the CNSMDP [8] in Paris. Messonnier works on recording and spatialization of complete audio scenes with the Binaural/Transaural system developed by Jean-Marc Lyzwa CNSMDP and Alexis Baskind [9]. We transposed an audio scene previously recorded and spatialized by Messonnier with the Transaural Bi-pan panner, into our VR Auditory Space using 3D Sound Labs VR Audio Kit. The scene consisted of 16 monophonics stems recorded from a Jazz trio, including the direct sounds, the early reflections, the later reflections and the high reverberations.

The placement of the audio stems on the spatialization map Picture 4, creates a very realistic, transparent and homogeneous 3D sound scene trough witch the virtual camera linked to the binaural audio output can move freely in all dimensions, 6 degrees of freedom, between each recorded layer, without audio clicks or directivity loss. Most important of all, in this situation and from the listener's perspective, we entirely get rid off the initial *sweet spot* constraint. In addition to this primary and successful test we decided to add a second recorded sound scene made by Messonnier in our Unity 3D scene. This time with got two pianos, a cello and one harp. We included this scene alongside to the first one. Thus we got two complete different

sound scenes, each one with its own personal reverb and reflection specificities we could position in the 3D space, walking in subjective view from one to another, freely. With 16 stems per audio scene we got a total of 32 stems playing at the same time in the 3D space. These primary tests authorize us to go further and to adapt this technique to achieve Empty Room's Seq 3.



Picture 4 – The Jazz Trio BiPan-TranPan spatial configuration

Working together with Messonnier on Empty Room's Seq 3 we had to adapt his recording and spatializing process to Empty Rooms dramaturgy. Originally Seq 3 takes place in a sort of containment chamber or panic room. Because we need a little bit of reflections to work with the Transaural Bi-pan panner we choose to record a session doing close-up miking on 5 Tibetan percussions instruments, a gong and 4 bowls. For each instrument recording we made a close-up mike in order to take the direct sound and all around we positioned 6 omnidirectional mikes to obtain a 360° primary reflections sound field.

From this session Messonnier build a new sound scene with the Transaural Bi-pan panner placing the 5 elements together with their primary reflections, adding digital reverb to glue the whole scene together and to work on the height reverb perception at +45° elevation. On Picture 3 at the far right we can see the spatial positioning of Seq 3 : the direct sounds are symbolized in red, the primary reflection in grey and the reverb in elevation in blue.

Adapting the spatialization map from 2D to 3D

After Seq 3's finalization we adapt in the same way Empty Room's Seq 1 and 2. Going from a 2D pan positioning 3D sounds in a virtual scene to an Ambisonic 3D soundfield changed several aspects of the sound perception in our VR Auditory Space : now some elements of the spatialization map will be sensed at different heights. Picture 5. This gives more space between them, more clarity and unfolds the space in all 6 dof directions. As a result the amplitude of the sounds are perceived at a lower gain rate. So we had to adjust completely the distance between each element to obtain a well balanced and natural impression of spatial presence for each AudioSource. 3D Sound labs VR Audio Kit force us to bypass Unity's mixer, because all sound sources are mixed through the ambisonic interface and could not be treated individually outside from it.

The VR Auditory Space Perception issues

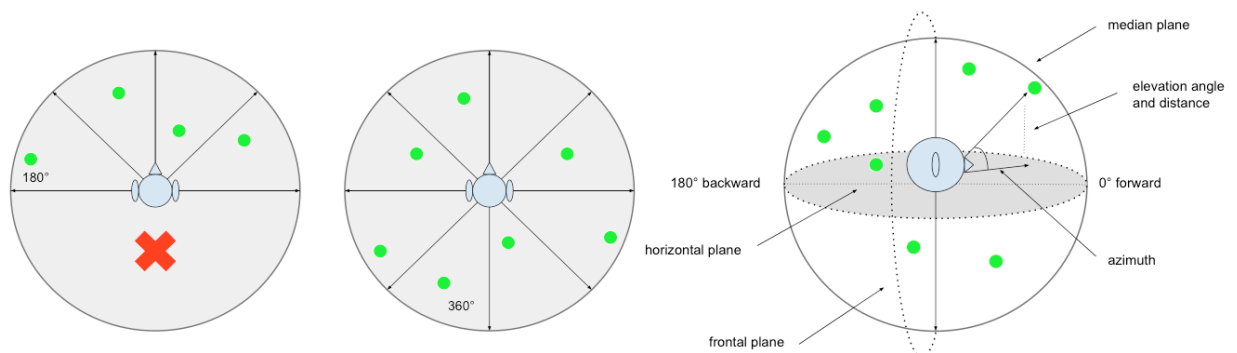
Once we got a stable Beta version for OSX and Oculus Rift DK2 we adapt it to be used with the HTC Vive. Upgrading to HTC Vive gives us a better stereoscopic visual resolution and the possibility to walk very naturally in standing position through our virtual platform in HTC's room scale mode. Displacement could be made physically or with the hand controllers. From this point we have to work now on three important aspects regarding the human immersion in our VR Auditory Scene :

- We have to extend the interaction field from a simple spatial interaction sensed through the listener's displacement to direct touch interactions with the virtual sound field. The question for us is : what do we touch, how do we touch and what for ? Considering that these gestures have to stay in harmony with the compositional purpose.
- Because the body is now full engaged in the virtual scene we have to think about how we want to represent this body through the representation of an avatar. Do we render a full body shape ? Part of the body, or only the hand ? Can the virtual hand become a multi functional tool ?
- How can we organize the VR Auditory Ecosystem as a dynamic 3D workspace for the composer ? Therefore we need to develop a dedicated immersive 3D user interface displaying the virtual spatialization map and all useful elements the composer needs to work out in space.

All these questions are still in processing and will be worked out in the near future.

RESULTS

Empty Room's Alpha and Beta experimentations shows that we can completely integrate the classical principles of the Virtual Auditory Space displayed through loudspeakers configurations and transpose them into a VR3D scene in projection perspective. These two layers of virtuality merge together and form a new environment we called a VR Auditory Space. The VRAS is accessible for the composer and the audience in complete VR immersion with a Stereoscopic Hmd and Binaural audio displayed through headphones. VRAS, Listener and Controllers form our VR Auditory Ecosystem, the basis of our research field.



Picture 5 – Sound positioning in our VR Auditory Space : Left, Unity's 2D distance pan in the horizontal plane without rear perception or elevation. Middle, Ambisonic 2D perception at 360° without elevation. Right, Ambisonic 3D perception at 360° in the horizontal, frontal and median plane.

From our spatialization maps, Picture 3, we show how we can stack and handle for the same musical composition multiples layer of virtual transducers configurations and raise them from 2D to 3D without affecting the general purpose of the sound work. And how we could introduce a non linear auditory experience of the composition, dissolve the sweet spot constraints and mold our spatialization close to any spatial dramaturgy. Following the same principle we show how the listener can sense the VR Auditory Space gradually either with a perception in the horizontal plane in 2D at 180° or 360° or from within a full Ambisonic 3D sphere. Picture 5.

Our experiments state that Ambisonic 3D with a high spatial resolution is the best candidate for our compositional experiments event if we are limited to 32 AudioSources. We also observed that during the Ambisonic 3D tests and demos we used the same default Hrtfs for the binaural output without any major hearing problems reported.

Concerning the use of reverberations, we bypassed Unity's default reverb zone during the Alpha version and designed our sounds with more or less reverb mixed directly on our audio files. The spectral content and the position in the spatialization map accentuate or not the illusion of distance in the field depth. At first this was motivated to preserve the esthetical approach generated by our modular synthesis sound system, at second doing so we spared a lot of precious CPU for the project.

In contrast Messonnier's workflow with the BiPan-TranPan audio scene shows another acoustical approach, how we can spatialize the reverb like an autonomous sound object in the 3D scene using natural recorded reverb and artificial ones in the same scene in conjunction with the direct sounds.

During our experimentations the Ambisonic 3D sound field displaying a high spatial resolution operating in a VR context freed from external loudspeaker constraints acts like a virtual universal host. We managed mono, stereo, quadriphonic, octophonic, and object oriented audio in the same project.

We scratched the surface of our VR Auditory Ecosystem and the potential is promising, there is lot more to investigate for us in the future. We have to optimize the immersion in VR, to define specific 3D tools for the composer and want to involve more dynamically the listener in the compositional process.

CONCLUSIONS

Our VR Auditory Ecosystem is a dynamic environment running on the basis of a projection perspective plane in Unity 3D, translating and rotating with six degrees of freedom (6dof), displaying a VR Auditory Space in High Order Ambisonic 3D with Binaural headphone output. This 3 dimensional sensory scene is accessible in VR immersion with a stereoscopic Hmd, a headphone and tactile controllers.

The Ambisonic3D, Binaural VR Auditory Space represents a new dedicated area for the composer from where he can operate innovative compositional gestures :

The composer can integrate, handle and display numerous audio dissemination formats, such as mono, stereo, multichannel stems, multiphonic, quadriphonic or octophonic within the same VR scene.

The transposition of complex audio scenes originally linked to a sweet spot turn to an open spatial structure and can be perceived from the composer or the audience from any spatial locations in the VR scene.

The manipulation of virtual transducers in the VR scene transforms the spatialization process. It becomes more adaptative and flexible and allows the consideration of a renewed spatial esthetical expression taking full advantage of the topological characteristics from a VR scene rather than from the classical loudspeaker configurations.

All these aspects opens the way to a new harmonization in our VR Auditory Space being at the same time a creation/diffusion platform, suited for complex sound spatialization restitution or a custom made auditory space freed from classical post-production techniques.

ACKNOWLEDGEMENTS

EnsadLab Spatial Media, Ircam, SCAM Brouillon d'un rêve Pierre Shaeffer, La Francogrid, LE CUBE Centre d'art numérique d'Issy-les-Moulineaux, 3D Sound Labs, CNSMDP Conservatoire National de Musique et de Danse de Paris, Institut Français.

REFERENCES

- [1] Christine Webster, François Garnier, Anne Sedes : “Empty Room, how to compose and spatialize electroacoustic music in VR in ambisonic and binaural” (online) : <https://hal.archives-ouvertes.fr/>
- [2] CICM HOA Library – High Order Ambisonics Library (online) : <http://hoalibrary.mshparisnord.fr>
- [3] Spatial Media EnsadLab (online) : <http://spatialmedia.ensadlab.fr>
- [4] Second Life (online) : <https://secondlife.com>
- [5] Unity 3D (online) : <https://unity3d.com/fr>
- [6] Anne Sedes, Pierre Guillot, Elliot Paris : “The HOA Library, review and prospects” Proceedings ICMC|SMC|2014 14-20 September 2014, Athens, Greece
- [7] 3D Sound Labs VR Audio Kit (online) : <http://www.3dsoundlabs.com/category/developers/>
- [8] CNSMDP Conservatoire National Supérieur de Musique et de Danse de Paris (online) : <http://www.conservatoiredeparis.fr/accueil/>
- [9] Alexis Baskind, Thibaut Carpentier, Markus Noisternig, Olivier Warusfel, Jean-Marc Lyzwa : “Binaural and transaural spatialization techniques in multichannel 5.1 production” HAL Archives : <https://hal.archives-ouvertes.fr/hal-01247648/document>