

VR AUDITORY SPACE, SPATIALISATION DU SON EN IMMERSION VR

Christine Webster
Doctorante
Université Paris 8
MUSIDANSE/CICM
EA1572
webster@sfr.fr

Eliott Paris
CICM
Docteur de l'Université Paris 8
eliottparis@gmail.com

David Fierro
Doctorant
Université Paris 8
MUSIDANSE/CICM
EA1572
davidfierro@gmail.com

Sofia Kourkoulakou
Doctorante
INREV, labo AIAC, EDESTA
Université Paris8
EnsadLab Spatial Media
sofia.kourkoulakou@ensad.fr

RÉSUMÉ

Nous présentons dans cet article le projet Eur-ArTeC VRAS ou VR Auditory Space¹, spatialisation du son en immersion VR. Nous décrivons dans cet article le concept du VRAS ainsi que les différentes phases de son développement. Le VRAS est conçu comme un environnement de spatialisation immersif 3D dédié aux compositeur-e-s de musique électronique.

Le VRAS associe le moteur de jeu Unity 3D à HOA Library², la bibliothèque de spatialisation immersive du son réalisée au Labex Arts-H2H entre 2012 et 2015 comme base d'un environnement *in-game* en s'appuyant sur les techniques de spatialisation ambisonique 3D et binaurale.

Notre projet de recherche se place à l'intersection des techniques du jeu-vidéo et de la VR quand elles interrogent la possibilité d'un acte de composition et de spatialisation audionumérique, envisagé dans un contexte d'immersion VR 3D complet, (*full body immersive*). Le projet VRAS forme le 4ème volet expérimental de la thèse création-recherche de Christine Webster³ il est conçu sous la forme d'un partenariat entre les laboratoires CICM/MUSIDANSE Paris 8 et EnsadLabs Spatial Media.

1. INTRODUCTION

La relation entre le son et la Réalité Virtuelle s'est posée dès les années 1990, notamment grâce aux travaux initiés par la NASA. Depuis cette époque les recherches

se sont principalement orientées vers des systèmes de simulations de perception d'un champ acoustique virtuel restitué en écoute binaurale. On citera le Convolver de Crystal River eng. (1992)⁴ spécifiquement dédié à la Réalité Virtuelle. Et plus récemment le projet Evertims⁵ du LIMSI/CNRS portant sur la modélisation du champ acoustique des espaces réverbérés pour la réalité virtuelle.

En 2013, l'apparition des casques stéréoscopiques Oculus Rift et HTC Vive ouvre une nouvelle page de l'écosystème de la Réalité Virtuelle. Grâce à des moteurs de jeu vidéo tel que Unity 3D ou Blender il est possible d'envisager l'immersion du compositeur ou du designer sonore dans une scène de réalité virtuelle en 3D. Se pose alors la question du contrôle de l'espace 3D par le compositeur et de la collaboration en immersion. Des solutions commencent à être proposées dans le domaine de la post-production audio pour le mixage, comme l'application DearVR⁶. Parallèlement des recherches sont menées en réalité virtuelle et réalité augmentée pour explorer une manipulation plus intuitive des sources dans l'espace, comme le projet *Composing space in space* de Giovanni Santini⁷.

Aussi bien au CICM qu'à l'EnsadLab Spatial Media, ces approches sont traitées sur le terrain expérimental de la création artistique. Le travail de Christine Webster, codirigé entre le CICM et l'EnsadLab Spatial Media, est emblématique, basé sur le développement d'Empty Room, œuvre artistique en progrès et dont le présent projet constitue en quelque sorte une forme de généralisation technique se situant au cœur d'un

¹ Voir hoalibrary.mshparisnord.fr/en/vr-auditory-space/.

² Voir hoalibrary.mshparisnord.fr/en/.

³ Composition électro-acoustique expérimentée dans l'espace numérique 3D. Exploration et étude sur les nouvelles techniques de composition émergentes dans un espace numérique 3D. Voir www.theses.fr/s161566.

⁴ Voir www.ece.ucdavis.edu/cipic/spatial-sound/tutorial/hrtfsys/#convtrn.

⁵ Projet Evertims : <http://evertims.github.io/>

⁶ Voir www.dearvr.com/.

⁷ Voir smc2019.uma.es/articles/P2/P2_03_SMC2019_paper.pdf.

processus interdisciplinaire qui fusionne technologies et médiations humaines.

Dans la première partie de notre travail de recherche de thèse, qui s'appuie sur le projet de composition et de spatialisation du son en VR, *Empty Room*, nous avons pu mettre en évidence la solidité et les qualités plastiques d'un dispositif de spatialisation Ambisonique 3D en immersion VR quand il est associé au moteur de jeu Unity 3D. Les techniques d'immersion proposées par Oculus et HTC Vive couplés à un spatialisateur HOA intégré dans Unity 3D, ont contribué à valider l'existence d'un *VR Auditory Space*, un espace auditif virtuel perceptible en immersion VR. Cet espace prolonge la notion de VAS, espace auditif virtuel⁸ tel qu'il a été défini par Begault 2000 [2] et Nicol 2010 [8]. C'est l'espace de la scène sonore virtuelle qui existe dans les dispositifs de projection du son par haut-parleur. En transposant ces dispositifs vers la réalité virtuelle ils héritent d'un certain nombre de qualités qui sont propres au moteur de jeu vidéo et d'une scène immersive, transformant leur potentiel et leur maniabilité.

Pour obtenir ces résultats nous sommes passés par trois paliers d'expérimentation :

Expérience 1 - *Empty Room* version Alpha. Écoute stéréophonique, au casque. Nous avons exploré la spatialisation du son 3D de Unity, qui s'appuie sur un panoramique 2d en intensité et position. Les sons placés dans l'espace 3D sont localisables par l'auditeur uniquement dans l'axe frontal et latéral, sans perception arrière ni en élévation. Les sons placés à l'arrière de l'auditeur sont rabattus dans le champ stéréophonique.

Expérience 2 - *Empty Room* version Beta. Écoute binaurale, au casque. Pour cette version nous avons utilisé le spatialisateur HOA de 3DSoundLabs⁹, le VRAudioKit. Nous avons déployé notre mapping de spatialisation initial dans les 3 dimensions avec une perception du champ avant et arrière et la perception en élévation jusqu'à + 40°/45°. Nous avons également procédé à la transposition d'une scène BipanTranspan [1] en audio orienté objet, vers une scène VR, en collaboration avec Jean-Christophe Meissonnier du CNSMDP. Les bons résultats obtenus en termes de qualité de perception de la position des sources et de la réverbération traitée en tant qu'objet discret nous ont permis de créer la partie 3 d'*Empty Room* en utilisant la même technique.

Expérience 3 - Un acousmonium en immersion VR¹⁰. Cette expérience a pris la forme d'un workshop qui a réuni le CICM Paris 8, le laboratoire IReMUS, la

compagnie MOTUS et le LAM autour d'un projet d'écoutes comparées entre le champ auditif perçu à partir d'un acousmonium et sa reproduction virtualisée en immersion VR.

Nous avons enregistré sur la suite logicielle MotusLab¹¹, une série de tests auditifs et d'interprétations d'œuvres du répertoire électroacoustique réalisées sur l'Acousmonium MOTUS¹². Nous les avons ensuite intégrées dans la modélisation 3D de l'acousmonium réalisé dans Unity par Christine Webster en utilisant les mêmes techniques de spatialisation ambisoniques et binaurales que pour l'expérience 2.

Là encore nous avons pu valider, en fin de session, qu'un certain nombre d'indices de perception binauraux, sont reproduits par le VRAS (externalisation du son, trajectoires, perception avant, arrière, centre, élévation).

Bien que perfectible cette expérience ouvre des perspectives performatives et pédagogiques inédites pour l'interprétation acousmatique.

Les expériences 2 et 3 nous ont montrées qu'il est possible de restituer dans l'espace d'une simulation VR, la perception d'un champ auditif projeté à l'origine par haut-parleur physiques ou virtuels de manière suffisamment convaincante pour approcher ce que Blauert 1983 [3] définissait comme un « *authentic auditory reproduction achievement* ». Tout en gardant à l'esprit que la perception de cette reproduction est relative car elle est dépendante de l'auditeur et des techniques employées.

Arrivés à ce stade, nos investigations nous permettent d'envisager la création d'un environnement immersif spécifique adapté pour le compositeur désirant travailler la spatialisation en immersion VR.

L'ensemble des notions énumérées en introduction ainsi qu'une description détaillée du projet *Empty Room*, sont accessibles via les articles Webster 2017 [9] et Webster 2019 [10].

2. LE PROJET VRAS

Avec le projet VR Auditory Space nous voulons poser les jalons de cet environnement. Nous voulons démontrer comment le compositeur de musique électronique peut manipuler en état d'immersion VR un VRAS - un Virtual Reality Auditory Space, un espace auditif synthétisé numériquement en simulation 3D. Nous voulons aborder les problématiques liées à la manipulation en temps réel de cet espace de spatialisation quand le compositeur est en immersion et ne passe plus par un menu contextuel classique.

En nous appuyant sur les techniques de spatialisation ambisonique 3D et binaurale, augmentées par les caractéristiques propres à la VR, le potentiel d'une telle immersion permettra au compositeur :

⁸ Un espace auditif virtuel (par référence à l'acronyme anglais VAS pour Virtual Auditory Space) est une scène sonore virtuelle constituée d'un ensemble de sources sonores qui n'existent que dans l'espace perceptif de l'auditeur. Cet espace est généré au moyen des technologies de spatialisation sonore (telles que : stéréophonie, technologie binaurale, Wave Field Synthesis ou Higher Order Ambisonics) qui reposent sur des modèles de représentation de la scène sonore. Rozen Nicol 2010.

⁹ 3DSoundLabs partenaire du projet *Empty Room* a été racheté en 2019 par la société Allemande Mimi Hearings. Voir www.mimi.io/.

¹⁰ Vers un Acousmonium en Immersion VR en Ambisonie 3D et Binaural, C.Webster, N. Raboisson, O. Lamarche, P. Couprie, Hugues Genevois. JIM 2020

¹¹ La suite logicielle Motus Lab a été développée par Pierre Couprie, Sorbonne Université Institut de Recherche en Musicologie (CNRS UMR 8223)

¹² Voir motus.fr/.

- d'avoir un accès direct, en immersion et vision stéréoscopique et écoute binaurale au système de spatialisation virtuel.
- de travailler en mode édition/lecture simultanée. Les opérations sont réalisées en temps réel de façon dynamique et rendues visibles dans l'espace 3D.



Figure 1. L'espace d'accueil du VRAS.

2.1. Développement général et positionnement

Le VRAS sera simulé à partir du moteur de jeu Unity3D en mode *runtime*, l'équivalent d'une application autonome. L'utilisateur n'aura pas à se préoccuper de connaître ou non Unity 3D. Il sera directement immergé dans l'environnement de spatialisation avec des fonctions spécifiquement conçues.

Nous développerons à partir du projet HOA du CICM notre propre système de spatialisation en ambisonie 3D immersive sous la forme d'un plugin compatible avec Unity 3D, avec en sortie audio, un décodage du son en binaural pour une écoute restituée au casque.

L'immersion stéréoscopique, le tracking de position et les interactions *in-game* se font avec le HTC Vive et ses manettes. Nous intégrerons également dans le dispositif de spatialisation la problématique de la représentation du sujet par un avatar. Nous étudierons comment l'avatar contribuera à une meilleure appréhension de l'espace VR et comment il pourra servir de contrôleur et de surface de contrôle.

Comme nous voulons privilégier l'écoute et la manipulation des sources virtuelles *in-game*, le VRAS ne fait pas d'interconnexions audio externes en Osc avec des applications telles que Max ou Pure Data dont on peut trouver de nombreux travaux dans la littérature. Nous ne passons pas non plus par des middlewares¹³ tels que Fmod ou Wise spécialement conçus pour la post production de jeu vidéo. Ce positionnement en interne, non dépendant, nous permet de développer une approche singulière de ce que peut être un environnement de composition et de spatialisation en immersion VR. Avec un décodage en binaural de la scène auditive 3D nous pouvons contourner les limites existant au niveau du moteur audio de Unity qui ne gère en sortie physique que le multicanal type *home cinema*, avec un maximum de 8

canaux, compatible avec quelques cartes audio dédiées¹⁴.

2.2. Les étapes de réalisation du VRAS

Pour réaliser le VRAS nous avons réparti le travail sur deux années de développement : pour 2019 nous avons démarré avec une première tranche en février-mars 2019 consacrée au développement du *plugin* HOA Unity, qui fut suivie par une deuxième tranche, entre mars et juillet 2019, consacrée au développement de l'interface 3D immersive utilisateur, la scénographie 3D de l'environnement de spatialisation, un début de prospective concernant l'intégration d'un avatar et le suivi global du projet. L'année 2020 sera consacrée d'une part au développement de la sauvegarde et du chargement des scènes spatialisées ainsi que de l'export et d'autre part à la représentation par avatar du compositeur en immersion, en mono ou double présence (2 avatars en immersion dans la même scène 3D).

3. L'ENVIRONNEMENT DU VRAS

Le VRAS se décompose en trois parties qui sont le plugin HOA Unity, la modélisation d'un monde immersif et son design 3D qui servira de support visuel et spatial à l'interface de spatialisation 3D.

Pour réaliser notre preuve de concept nous avons élaboré au préalable un scénario d'usage simple qui suit un modèle de développement classique avec l'entrée des données, suivi du processus interne et la sortie des données :

Niveau 1 - Importation de fichiers audio multicanaux et leur lecture dans le VRAS. Le compositeur doit être en capacité d'importer dans le VRAS une composition réalisée en amont sur la station audionumérique de son choix créée pour des dispositifs de spatialisation quadraphonique, octophonique ou pour le demi-dôme 16 voies du CICM.

Niveau 2 - Édition dynamique des sources dans le VRAS. Une fois importés dans le VRAS, les fichiers audio seront assignables à une source qui sera rendue visible dans l'espace immersif 3D. Chaque source sera paramétrable en mode édition/lecture simultanée. L'accès aux sources se fait via un menu général immersif 3D en choisissant l'option *preset*, *ajouter source* ou *trajectoire source*. Ces trois modes d'édition de spatialisation pourront cohabiter dans une même scène immersive.

Niveau 3 - Sauvegarde et export de la scène VRAS. Une fois le mapping de spatialisation réalisé, le compositeur doit pouvoir sauver son travail et l'exporter pour qu'il puisse être relu dans le VRAS ultérieurement. L'export vers des stations audionumériques externes sera également étudié.

Ce scénario est envisagé dans un premier temps pour une mono immersion puis sera testé en double immersion à la fin du projet.

¹³ Middlewares Audio comparison charts.

Voir annesoaudio.com/2016/06/12/audio-middleware-comparison-wisefmodfabric/.

¹⁴ Comme carte audio compatible on peut citer la Creative Sound Blaster Z1500.

Le choix des systèmes de spatialisation proposés en preset, qui seront détaillés plus avant dans la section 3.4 sont motivés par l'usage qui en est fait majoritairement dans les conservatoires, universités et live sessions de musique électronique. Nous avons volontairement écarté la stéréo et le 5+1 qui sont plus en correspondance avec les techniques de post-production pour la télévision et le cinéma. L'acousmonium en VR est pour l'instant traité à part mais il pourrait par la suite être intégré dans le VRAS.

3.1. Le plugin HOA Unity

Nous décrivons dans cette section les parties du développement sur lesquels nous avons porté le plus notre attention. Ils s'appuient sur les modèles ambisoniques décrits par Gerzon [6], Blauert [3] et Daniel [5] mais également pour la partie HOA Library, sur les travaux développés au CICM par Guillot, Paris et Colafrancesco [4].

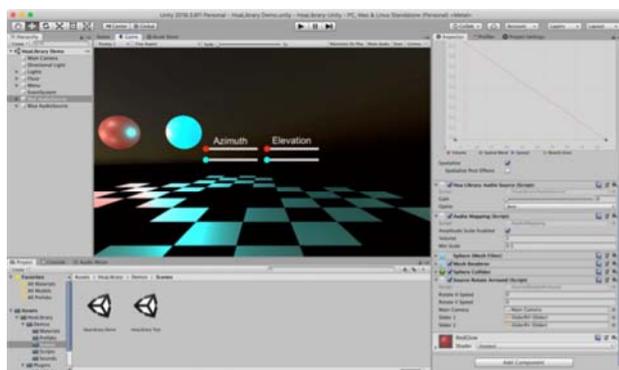


Figure 2. Le plugin HOA Unity

3.1.1. Architecture du plugin Spatialisateur

Idéalement notre plugin HOA Unity doit pouvoir spatialiser des sources sonores ponctuelles, appliquer des transformations et des effets dans le domaine des harmoniques sphériques, être un outil open source.

Nous aurions alors un spatialisateur HOA à trois étages, comprenant l'encodage des sources, le traitement sur les harmoniques et le décodage en binaural. Ce qui différencierait le spatialisateur une fois intégré au VRAS du modèle encodeur/décodeur habituellement en usage pour les moteurs de jeu vidéo.

Pour des questions de contraintes de temps de développement et pour pouvoir jeter les bases d'une preuve de concept VRAS opérationnelle en fin de projet ArTeC, nous reportons le développement de l'accès au contenu des harmoniques sphériques ultérieurement.

Le plugin HOA Unity sera développée à l'ordre 3 pour la version 1. Bien que la résolution spatiale et angulaire soit plus élevée à mesure que l'on monte dans les ordres en HOA, le choix de l'ordre 3 a été motivé pour des raisons de faisabilité dans le temps et d'optimisation des calculs notamment en ce qui concerne le calcul des matrices et de la réponse binaurale, qui sont abordés en fin de section. C'est aussi

l'ordre qui permet le plus facilement de travailler en référence avec un dispositif de haut-parleurs hardware externe.

3.1.2. Rotation de la scène sonore ambisonique

Unity nous permet de récupérer automatiquement toutes les informations relatives à la position d'une source audio via l'objet Unity *AudioSources* et celles relatives à la position de l'auditeur virtuel, via l'objet Unity *Listener*.

C'est à dire l'ensemble des informations relatives à la *matrice de rotation* de la source et de l'auditeur auxquelles il faut ajouter les fonctions de panning 3D, la gestion de la courbe d'atténuation du signal et la gestion des fonctions *minDistance* et *maxDistance* de projection du son de la source.

La rotation de notre scène sonore ambisonique sera gérée par une matrice 4x4 fournie par le SDK¹⁵ de Unity. Cette matrice est déjà transformée ce qui nous évite de la calculer. Les deux matrices travaillent conjointement dans Unity en valeurs relatives, celle de l'*AudioListener* attachée à la caméra virtuelle (elle-même reliée au casque HMD) et celle de l'*AudioSource* qui joue les fichiers sons importés.¹⁶

C'est une pensée audio orientée objets. Nous partons donc sur le développement d'un plugin codé en C++ avec un wrapper en C# utilisant le SDK natif de Unity.

3.1.3. Optimisation

Nous proposons trois processus d'optimisations pour notre HOA Unity, qui ont été abondamment décrits par Daniel, 2001 [5] Colafrancesco 2013 [4] et Guillot 2017 [7] : Basic, maxRe et inPhase. Ces optimisations consistent à diminuer les contributions des lobes des harmoniques opposés à la source sonore virtuelle pour favoriser des conditions d'écoutes variées.

Dans une configuration immersive VR l'optimisation *inPhase* serait la plus pertinente car la contribution des lobes des harmoniques se fait essentiellement en direction de la position de la source, ce qui est idéal pour travailler avec des sources ponctuelles. Mais cette optimisation crée aussi un artefact : à l'ordre 3 la diminution drastique de la contribution des lobes arrières et latéraux déforme le lobe frontal de façon significative, jusqu'à ressembler à un lobe d'ordre 2, agissant à la fois sur la directivité et l'énergie émise par la source.

In fine nous avons décidé de garder la possibilité de choisir entre les trois processus d'optimisations, pour pouvoir rejouer des oeuvres spatialisées sur HOA ayant utilisé l'un de ces processus et ensuite parce que l'optimisation Basic pourra contribuer, en terme d'investigations futures à la manipulation d'un champ diffus en immersion VR.

¹⁵ Software Development Kit

¹⁶ Unity 3D Audio Spatializer SDK - Matrix conventions <https://docs.unity3d.com/Manual/AudioSpatializerSDK.html>

3.1.4. Radius d'intensité

L'atténuation de l'intensité du signal audio en fonction de la distance est gérée par le moteur audio de Unity, (avec les ellipsoïdes 3D minimum/maximum distance). Ce réglage peut être configuré à travers l'interface graphique du logiciel. En revanche, lorsque la distance est inférieure à 1 mètre (fixé de manière arbitraire), on note un effet interne de diminution progressive de la résolution angulaire jusqu'à atteindre une source mono si la distance est nulle. D'autre part pour remédier à la perte significative du niveau des sources quand elles sont spatialisées en ambisonie 3D, nous rajouterons au plugin HOA Unity une fonction de gain se paramétrant directement à partir de l'AudioSource et de l'AudioListener.

3.1.5. Consommation du CPU et Réponse binaurale

Notre décodeur binaural est directement inspiré par celui du projet de la librairie HOA que nous avons adapté pour pouvoir opérer dans un contexte VR avec Unity 3D. C'est un décodage direct dans le domaine des harmoniques sphériques. Les limites spatiales et fréquentielles de cette technique sont connues et ont été identifiées et discutées. Notre objectif n'est pas de trouver une solution qui valide des principes psychoacoustiques pré-établis, mais plutôt d'explorer de nouvelles propositions sur le plan artistique exploitable pour l'utilisateur en temps réel sur son ordinateur. C'est dans le même esprit de faisabilité que nous avons choisi notre réponse binaurale.

En début de projet nous avons rencontré de gros problèmes de consommation du CPU audio. Entre 50% à 36% ce qui nous rendait notre plugin inutilisable. Pour avoir une idée plus juste de l'objectif à atteindre en termes de réduction du CPU audio nous avons procédé à des tests comparatifs entre le pan 2D/3D Unity, notre plugin HOA, le VRAudioKit de 3DSoundLabs et celui de Resonance Audio¹⁷ de Google. Nous avons réussi à gagner plus de 8% de CPU audio en mode release en réduisant le temps des réponses binaurales. Nous avons ensuite testé plusieurs valeurs de réduction pour déterminer le taux de perte de qualité auditive induite. Puis nous avons fixé une balance qualité/consommation qui nous paraissait la plus optimale et l'avons intégré par défaut dans notre spatialisateur.

Le choix de la réponse binaurale finale s'est fait après comparaison entre la base de données IRCAM/Listen et du sujet 1005¹⁸ et une investigation de la base de données SADIE¹⁹. Nous notons que les réponses SADIE ont une taille de 256 échantillons (deux fois moins importante que les échantillons utilisés dans la HOALibrary du CICM). Ce pourrait être une manière de diviser la consommation de CPU de moitié sans faire de réduction ultérieure de la réponse.

Nous procédons à un premier test avec les réponses SADIE, pour un sujet virtuel, *Subject 2*, tête Kemar avec un décodage à l'ordre 3, qui sont concluants.

Nous avons une diminution confirmée du CPU ($\pm 17\%$), nous avons une bonne sensation d'élévation, et de manière générale moins d'effets de trous avec les mouvements de rotation de la tête.

Avec le même sujet à l'ordre 5 nous passons à 30% de consommation de CPU. Au niveau de la perception générale nous ne notons pas une grande différence de rendu par rapport à l'ordre 3. Nous restons à l'ordre 3.

3.1.6. Résultat

Nous avons pu développer un plugin de spatialisation Ambisonic 3D pour Unity qui permet de :

- spatialiser un nombre illimité de sources ponctuelles à l'ordre 3 en fonction de la capacité de l'ordinateur hôte.
- nous avons ramené la consommation du CPU vers le seuil de référence des 17% sur Mac et 30% sur PC en tirant parti de la bibliothèque C++ de traitement matriciel Eigen au sein de la bibliothèque HOA.²⁰

En l'état nous disposons d'une première version du plugin HOA Unity destiné à la réalité virtuelle, que nous intégrerons au VRAS, cette version est disponible sur Github depuis Avril 2019²¹.

3.2. Le design 3D immersif

Nous proposons à l'utilisateur trois scénographies immersives ou *worldspaces* disponibles via le menu principal. Par défaut le programme démarre avec le *worldspace beige*, ou *worldspace 1*. Le *worldspace* contient tous les éléments visuels dont le compositeur en immersion aura besoin pour importer, spatialiser et éditer ses sources.

La taille du *worldspace* s'étend sur une superficie de 1 hectare virtuel, il comprend un décor général et une grille de travail centrale ou *grid*, qui accueille les presets de spatialisation. Le *worldspace* autour de la grille centrale fait office de *bac à sable*, c'est-à-dire de zone où le compositeur peut librement agencer ses sources sonores, sans être dépendant d'une modélisation de système par haut-parleur²².

Pour garantir le confort oculaire du compositeur pendant l'immersion, nous sommes partis sur un principe de ton sur ton à dominante beige, bleu ou blanc cassé. Les objets 3D qui peuplent le *worldspace*, solides de Platon, polyèdres divers, l'amphithéâtre, permettent au compositeur d'éprouver une sensation d'espace non confiné, avec une perspective, ce qui contribue à renforcer cognitivement la sensation de présence dans l'espace virtuel.

¹⁷ Voir resonance-audio.github.io/resonance-audio/.

¹⁸ Voir recherche.ircam.fr/equipements/salles/listen/info_display.php?subject=IRC_1005.

¹⁹ Voir www.york.ac.uk/sadie-project/database.html.

²⁰ Voir eigen.tuxfamily.org/?title=Main_Page.

²¹ Voir hoalibrary.mshparisnord.fr/en/downloads/unity.

²² La notion de bac à sable nous a été inspiré par les mondes virtuels tels que *Second Life*, où les usagers peuvent designer leurs objets 3D directement à partir de la scène immersive.

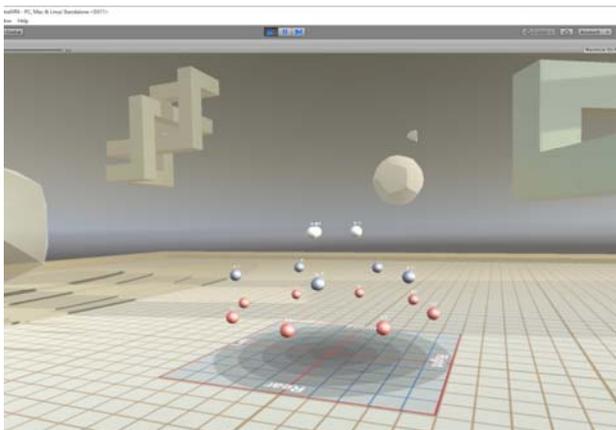


Figure 3. Le workspace VRAS beige avec la grille centrale et le preset demi-dôme 16 voies

Pour la grille de travail centrale, nous nous sommes inspirés de celle que nous avons utilisé tout au long des expériences précédentes 1,2 et 3. Nous avons vérifié et corrigé sa métrique et nous l'avons adapté au contexte du VRAS.

Les icônes des menus sont de couleur vert-d'eau compatibles avec les trois tonalités du workspace. Elles sont toutes translucides pour ne pas boucher la sensation d'espace.

L'apparence des menus suit la même charte couleur que celle des icônes et sont également translucides. Les paramètres des menus se modifient à l'aide du pointeur laser de la manette HTC Vive. L'esthétique émergente est le *low-poly* ; un maillage polygonal en infographie 3D ayant un nombre relativement faible de polygones, ce qui rend l'application en temps réel plus optimisée au niveau de ressources.

3.3. L'interface de spatialisation 3D

Nous décrivons ici les principales fonctions que nous avons implémentées pour l'interface 3D immersive du VRAS.

3.3.1. Le menu in-game

Par défaut le menu apparaît toujours en face de l'utilisateur : si ce dernier bouge, il ramène le menu vers lui à l'aide du pointeur laser par un simple click. L'affichage du menu se trouve à une distance de 20 m et se positionne autour de l'utilisateur à 360° et ce, en tout point où il pourra se trouver dans la scène 3D. Les changements des valeurs de données se font par glissé, au *slider*. Nous avons implémenté 5 menus :

- Menu principal : il permet d'accéder aux presets de spatialisation, aux workspaces, aux fonction lecture et arrêt de lecture de la scène spatialisée, aux paramètres audio source et trajectoires et au menu sauvegarde et chargement de scène ;
- Sous-menu browser : permet d'accéder aux fichiers présents dans l'ordinateur de l'utilisateur, notamment aux fichiers audio pour les assigner à une source dans notre scène immersive. Il est

implémenté dans le menu ajouter source et menu trajectoire ;

- Menu ajouter source : positionne une source dans la scène 3D, gère les paramètres de chargement du son (via le browser) et d'assignation de ce dernier à la source, la fonction de bouclage du son, son volume, son déclenchement dans le temps à xx secondes, les valeurs de minimum et maximum distance et les courbes de décroissances d'intensité du son (linéaire ou logarithmique) ;
- Menu ajouter trajectoire : reprend les mêmes paramètres que le menu précédent mais avec une fenêtre supplémentaire qui gère le sens des trajectoires, leur vitesse et les fonctions de bouclages ;
- Les Menu sauvegarde et chargement de la scène spatialisée seront développés en 2020.

3.3.2. Manette, sélection, principes d'indexation

L'interaction avec les objets 3D et les menus du VRAS se fait avec les manettes du HTC Vive. Pour les 2 manettes de contrôle nous avons codé un système de détection de type FIFO, la première manette prise en main devient automatiquement la manette principale. Pour l'instant nous avons implémentés les contrôles des fonctions du VRAS sur une seule manette.

La sélection d'un objet *AudioSource* et de tout autre objet 3D se fait avec le pointeur laser (*raycasting pointer*) de la manette. Nous utilisons la librairie VRTK²³ pour coder tous les contrôles manettes et les interactions boutons. Tous les objets 3D possédant des boutons peuvent interagir avec notre pointeur laser. Via l'interface usager le pointeur laser pourra contrôler toutes les sources ponctuelles et les changements de valeurs des paramètres du son.

Pour un contrôle plus efficace des sources en mode *edit* nous leur avons attribuées un identifiant unique : S1, S2, etc. pour chaque source et T1, T2, etc. pour chaque trajectoire. La visualisation des identifiants fait toujours face à la camera virtuelle. Chaque fois qu'une source est ajoutée, elle est automatiquement indexée.

Pour les sources faisant parti d'un *preset* nous avons mis en place un code couleur qui indique si un fichier audio est assigné ou non à la source : un affichage blanc par défaut pour non-chargé, un affichage bleu pour chargé.

Par ailleurs notre programme récupère toutes les informations des fichiers audio qui se trouvent sur le disque dur usager. La recherche des fichiers se fait en dehors du temps de chargement des buffers du programme.

Au lancement du programme VRAS nous avons une fonction *late init* qui charge tous les objets présents dans la scène 3D et en fait l'indexation complète en 200 milli-sec. Ensuite le programme passe en mode *play*.

²³ Voir [vrtoolkit.readme.io/](https://github.com/VRToolkit/vrtoolkit.readme.io/).

3.3.3. Codage des fonctions de trajectoires

Pour le VRAS nous avons codé une fonction de trajectoire de source (menu ajouter trajectoire en 3.3.1). À l'aide de la manette HTC et de la fonction *scroll* couplé au pointeur laser. L'utilisateur dessine, pour chaque source créée en mode trajectoire, un parcours visible dans l'espace du worldspace. Comme pour le menu *ajouter source* la liberté de positionnement de la source est totale en plus d'être mobile. L'utilisateur peut assigner autant de sources trajectoires qu'il le désire.

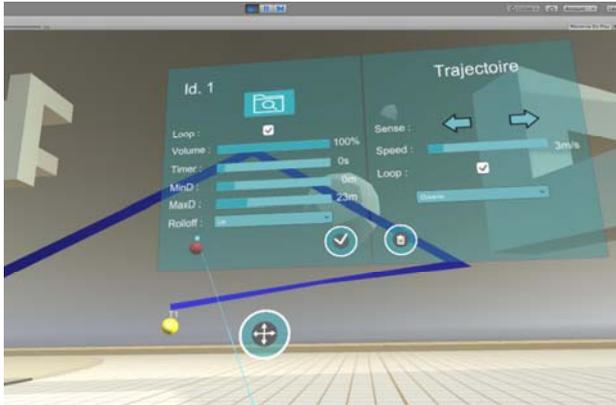


Figure 4. VRAS UI paramétrage de l'AudioSource (sphère jaune) et de sa trajectoire (ligne bleue).

Pour cette fonction nous avons développé un code spécifique qui modifie la précision du mouvement et de son échantillonnage (valeur du point) en fonction de la distance du point par rapport à la position de l'utilisateur. Plus le point est éloigné et plus l'échantillonnage est faible, plus il est proche et plus sa valeur augmente. C'est une sorte d'optimisation de la manipulation de la trajectoire,

Une fois la trajectoire dessinée dans l'espace, nous proposons deux sens de lecture, avant ou arrière. En mode bouclage deux types de boucles sont possibles :

- En mode boucle fermée, l'AudioSource revient à la fin de son parcours en ligne droite à son point de départ ;
- En mode ouvert, l'AudioSource repart directement à son point de départ sans passer par une ligne.

Pour chaque trajectoire créée nous pouvons également régler la vitesse de déplacement de la source.

3.3.4. Limitation de la taille des fichiers audio

Nous avons pu contourner une limitation connue dans Unity concernant la taille maximale des fichiers audio pouvant être joués en mode *play* dans une scène. En effet Unity ne prend pas en compte les fichiers audio qui font plus de 3 minutes, quel que soit leur format ou leur fréquence d'échantillonnage.

Après de nombreux tests pour discriminer d'où le problème pouvait venir, nous avons au final simplement séparé les tâches de chargement du fichier audio du reste du *process* pour les traiter en parallèle tout en restant sur le même processeur. De ce fait, la tâche de

chargement du fichier audio est simplement déviée et peut alors prendre le temps qu'il faut en travaillant en parallèle. Selon la taille du fichier, ce chargement prendra plus ou moins de temps.

C'est une excellente nouvelle pour le projet VRAS, nous pouvons travailler avec des fichiers audio de grande taille qui sont plus en correspondance avec les usages des compositeurs de musique électronique.

3.3.5. Modes de déplacements dans le VRAS

Nous proposons trois modes de déplacement dans le VRAS :

- Mode normal : le déplacement se fait dans toute l'étendue de la scène immersive grâce au système de *tracking* de position du HTC Vive, notamment en utilisant la configuration *Room Scale*. Cette configuration qui privilégie la station debout autorise une déambulation naturelle dépendant uniquement de la longueur du câble HTC qui raccorde le HMD au PC ;
- Mode Joystick humain (expérimental) : dans cette configuration le corps humain se transforme en joystick (il remplace les touches WASD du clavier). Avec un simple mouvement des épaules vers l'avant / arrière / gauche ou droite, bas et haut, l'utilisateur va glisser dans l'espace de la scène 3D, selon les trois axes x,y,z . Ce mode de déplacement permet d'expérimenter des points d'ouïe jamais utilisés pour parcourir une œuvre spatialisée libre ;
- Mode Teleport : l'utilisateur est téléporté vers une position cible dans la scène 3D grâce à la manette de contrôle du HTC Vive.

3.3.6. Résultats

En ce qui concerne la tranche de développement 2019 du VRAS, notre preuve de concept est conforme avec les niveaux 1 et 2 du scénario d'usage proposé. Le niveau 3 sera traité en 2020.

3.4. Les systèmes de spatialisation pour haut-parleurs modélisés pour le VRAS

Nous avons modélisé 4 presets pour dispositifs de spatialisation par haut-parleurs qui sont le plus couramment utilisé en musique expérimentale : quadriphonique, double quadriphonique, octophonique et le demi-dôme 16 voies du CICM. Nous avons réalisé cette modélisation à l'échelle humaine grâce au système métrique embarqué dans Unity : 1m Unity = 1m in *real world*. Ces 4 presets sont accessibles par le menu principal du VRAS.

Le preset vient se poser sur un *mapping* au sol gradué de 10 m x 10 m. À partir du centre, qui est aussi le point d'origine de toute la simulation VRAS nous avons procédé à une graduation circulaire de par 1m.

Unity assigne chaque objet 3D à un système de transformation de coordonnées (cartésiennes, polaires) via un

menu Transform²⁴. Ainsi nous pouvons disposer de façon très précise nos sources sonores dans l'espace de la scène virtuelle en paramétrant la position, la rotation et l'échelle d'un objet.

Nous avons modélisé les trois premiers presets à partir des coordonnées angulaires suivantes :

Quadriphonie		Double Quadri		Octophonie	
Angle	Elévation	Angle	Elévation	Angle	Elévation
45°	0°	45°	0°	0°	0°
135°	0°	135°	0°	45°	0°
225°	0°	225°	0°	90°	0°
315°	0°	315°	0°	135°	0°
		45°	40°	180°	0°
		135°	40°	225°	0°
		225°	40°	270°	0°
		315°	40°	315°	0°

Figure 5. Coordonnées angulaires des trois premiers presets, quadriphonique, double quadriphonique, octophonique.

Nous avons réalisé la modélisation du demi-dôme 16 voies à partir des données qui nous ont été fournies par le CICM. Ce demi-dôme est installé au studio S17 à la MSH Paris Nord. Il a été monté dans le cadre du projet HOA Library²⁵, la bibliothèque de spatialisation immersive du son réalisée au LabEx Arts-H2H conduit par Pierre Guillot, Elliot Paris et Julien Colafrancesco.

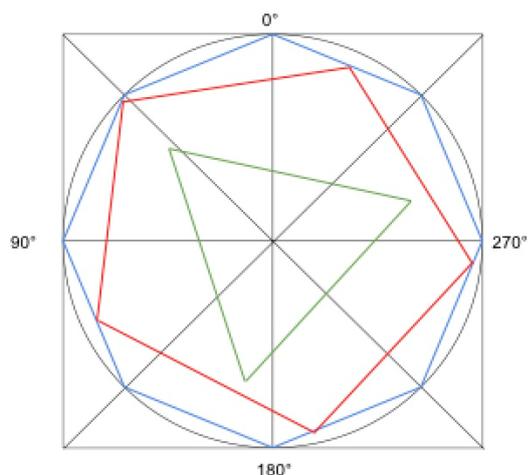


Figure 6. Plan 2D des trois couronnes du demi-dôme 16 voies du CICM.

Il est composé de trois couronnes de haut-parleurs qui répartissent les 16 voies selon une discrétisation irrégulière de la demi-sphère :

- La première couronne forme un Octogone en partant de 0° ;
- La deuxième couronne forme un Pentagone en partant de 335° ;
- La dernière couronne forme un Trigone en partant de 287°.

²⁴ Voir docs.unity3d.com/ScriptReference/Transform.html.

²⁵ Voir hoalibrary.mshparisnord.fr/en/.

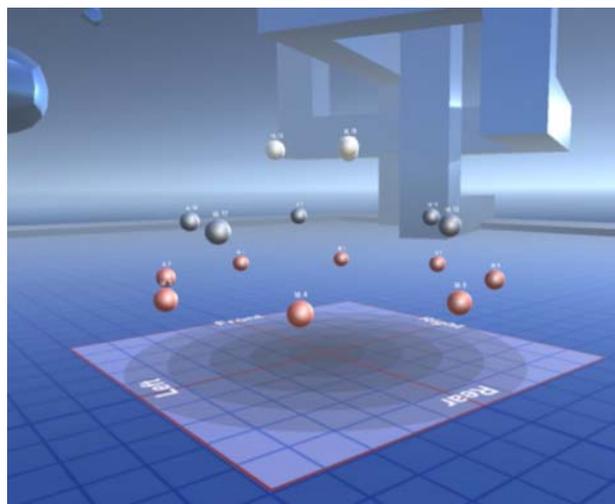


Figure 7. Modélisation 3D du demi dôme 16 voies du CICM dans Unity pour le VRAS.

3.5. Tests d'écoute

Nous décrivons ici les tests de contrôles auditifs que nous avons réalisés tout au long du processus de développement du VRAS. En l'état ces tests ne prétendent pas donner de conclusions mais plutôt des directions dans lesquelles nous nous engageons. Ils sont le fait de l'oreille du musicien et du compositeur nous ne prétendons pas agir ici en tant qu'acousticien.

En ce qui concerne les tests d'investigation du plugin HOA Unity du VRAS, nous avons dans un premier temps utilisé des impulsions de bruits blancs pour effectuer des tests de directivité classiques, que nous avons complété avec un test d'écoute comparée à partir d'une scène Trio Jazz (Larry Grenadier) que nous avons déjà utilisé au cours de l'expérience 2 avec Jean-Christophe Messonnier.

À la fin du développement de notre plugin HOA Unity, nous avons réalisé un deuxième test comparatif avec *Resonance Audio*, en utilisant uniquement les trois directs de la scène Trio Jazz Grenadier, sans les réverbérations.

Notre objectif : nous voulons tester la qualité de la perception de la scène sonore en immersion à partir d'une position de référence ou *sweet spot*, à l'ordre 3, en utilisant le spatialisateur ambisonique de Résonance Audio et notre HOA Unity. Dans la scène immersive nous dirigeons d'abord notre écoute pour évaluer la bonne position des trois sources : la contrebasse doit se trouver à gauche de l'auditeur, la voix au centre et un peu devant lui et la guitare à sa droite. Ensuite nous effectuons une rotation complète de l'auditeur à gauche puis à droite, devant chaque instrument diffusé en solo, puis une dernière rotation dans les deux sens avec tous les instruments. Pour le dernier test l'auditeur s'éloigne du trio en effectuant une marche arrière puis revient vers le trio en marche avant.

Résultats : De manière générale nous constatons que le HOA Unity du VRAS restitue un peu moins de brillance dans les aigus que Resonance Audio, mais sans

que cela soit critique. La localisation latérale est cohérente, il n'y a pas de sensation de trou au centre et la perception arrière est correcte. La scène reste homogène à partir du sweet spot comme en phase d'éloignement. Globalement le HOA Unity est, en termes de perception et de localisation des sources, assez proche de Resonance Audio.

D'autre part nous avons fait un premier test d'écoute avec le Dôme 16 voies du CICM et sa modélisation dans le VRAS à partir d'un extrait de 2' de *Dans la Nef de nos songes* d'Alain Bonardi. Nous voulons tester la sensation d'enveloppement général qui est propre à l'œuvre d'Alain Bonardi, la bonne définition des aigus, les mouvements et les trajectoire fines des sons, la cohérence du placement des sons dans la masse du *nuage* reparti entre les 16 voies, notamment en élévation ainsi que les déplacements de l'auditeur sous le dôme virtuel, avec rotation du corps et petits mouvements de la tête. Dans l'ensemble le retour d'écoute global est plutôt convaincant, latéralement et en élévation moyenne vers 40°, malgré des artefacts perceptibles en élévation haute, vers 90°, en fonction des mouvements de la tête de l'auditeur immergé.

Nous voudrions compléter ce premier test en 2020 avec un protocole d'écoutes comparées beaucoup plus précis et mettre plusieurs œuvres emblématiques du répertoire électroacoustique à l'épreuve en utilisant les quatre presets.

4. CONCLUSIONS ET PERSPECTIVES

En l'état, notre preuve de concept réunit fin 2019 les conditions de base pour donner accès, en mode démonstration, au premier environnement immersif de composition/spatialisation VR, spécialement conçu pour les compositeurs de musique électronique. Nous avons présenté le projet VRAS en novembre 2019 dans le cadre des Rencontres ArTeC au CDA d'Enghien les Bains et pour le Symposium Créativité virtuelle, réalités collectives de l'Ensad. Notre approche, qui propose de considérer l'espace et le potentiel de la réalité virtuelle immersive partagée comme nouveau terrain d'écoute et de création-recherche a suscité un vif intérêt auprès du public comme des chercheurs présents.

La tranche développement VRAS en 2020 introduira la sauvegarde et le partage d'une scène spatialisée, ce qui permettra aux compositeurs de s'en emparer et de tester également l'immersion avec un retour visuel de l'avatar/compositeur en mono ou double immersion. Nous aurons alors un environnement de base complet sur lequel nous pourrions nous appuyer pour prétendre à tout développement futur.

Émergents de l'ensemble de nos expériences et pouvant être raccordés au VRAS et au-delà, nous continuons d'affiner nos recherches en nous focalisant sur les points suivants :

— Perception du son en immersion VR. Le HOA Unity mériterait à terme de bénéficier d'un développement plus appuyé notamment en introduisant des fonctions de directivité des sources et la possibilité de

travailler à des ordres ambisoniques supérieurs à l'ordre 3, surtout si nous voulons procéder à des opérations dans le champ des harmoniques sphériques ambisoniques ;

- L'ambisonie 3D immersive, nous ouvre de nouvelles perspectives de manipulation d'objets sonores. Comment rendre visibles les opérations et les transformations appliquées aux harmoniques en temps réel ? Comment manipuler ces objets dans l'espace 3D ? Les opérations dans le champ des harmoniques sphériques comme le champ diffus par exemple pourraient être un bon départ d'investigation ;
- L'avatar comme contrôleur et agent interactif. Grâce à la nouvelle génération de casques stéréoscopiques, tels qu'Oculus, HTC Vive ou Hololens, nous avons aujourd'hui une qualité de perception du champ 3D en immersion qui permet d'envisager de nouvelles modalités d'interactions entre l'avatar et le monde 3D. En ce qui concerne le VRAS, l'introduction de l'avatar nous amène à penser le compositeur/auditeur en immersion, comme un être écoutant, présent, bilocalisé faisant corps avec l'espace 3D mais également capable d'actions directes sur son environnement de composition. L'avatar, qui se présente comme la seconde peau du compositeur en immersion, ne manifeste pas seulement sa présence mais devient une surface de contrôle à part entière. En tirant parti du potentiel physique, biologique (gestes, regard, parties du corps) et celui de l'enveloppe numérique, nous pouvons développer un compositeur augmenté faisant partie intégrante de l'espace composable numérique. Et par extension en multi-immersion ce sont toutes les modalités de la collaboration dans l'espace qui sont à explorer et définir ;
- Pour une écriture spatiale topologique. La source sonore en immersion VR, l'AudioSource, est la fois un transducteur virtuel qui fixe les *bords* d'un espace auditif virtuel s'il est calqué sur les dispositifs par haut-parleurs hardware mais il est aussi source en-soi, émancipé des formats de spatialisation, capable de s'agencer avec d'autres sources selon des configurations qui permettent de travailler points, lignes, surfaces, perspectives, amas, s'accordant au topos, à la géométrie et géographie de la scène 3D. Pour manipuler ces sources topologiques il faut développer des outils de spatialisation spécifiques. La fonction trajectoire que nous avons créée pour le VRAS est un premier pas dans ce sens.

5. L'EQUIPE DU VRAS

Porteur et chef de projet : Christine Webster, Université de Paris 8, laboratoire EDESTA.

Développement librairie HOA Unity. Elliott Paris, CICM, Université de Paris 8.

Développement interface immersive 3D. David Fierro, CICM, Université de Paris 8.

Création design immersif 3D et avatars. Sofia Kourkoulakou, INREV, Université de Paris 8, EnsadLab Spatial Media.

Encadrement scientifique :

Alain Bonardi, maître de conférences HDR, Université de Paris 8, CICM Musidanse.

Anne Sedes professeur, CICM Musidanse, Université de Paris 8.

François Garnier responsable groupe de recherche Spatial Media EnsadLab.

[10] Webster, C. et Sedes, A. « 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 Darmstadt-Dieburg, Germany 2019. [Voir hal.archives-ouvertes.fr/hal-02276901.]

Texte édité par Tom Mays

6. REFERENCES

- [1] Baskind, A., Carpentier, T., Noisternig, M., Warusfel, O., Lyzwa, J.-M. « Binaural and transaural spatialization techniques in multichannel 5.1 production », 27th Tonmeistertagung, VDT International Convention, Köln, Germany, 2012. [Voir hal.archives-ouvertes.fr/hal-01247648/document]
- [2] Begault, D. R. *3-D Sound for Virtual Reality and Multimedia*, Ames Research Center, Moffett Field (CA), 2000.
- [3] Blauert, J. *Spatial Hearing, The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge (MA), 1997.
- [4] Colafrancesco, J., Guillot, P., Paris, E., Sedes, A., Bonardi, A. « La bibliothèque HOA, bilan et perspectives » JIM 2013, (prix du jeune chercheur 2013), Saint-Denis, France, 2013.
- [5] Daniel, J. « Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia », Thèse de doctorat de l'Université Paris 6, Paris 2001.
- [6] Gerzon, M. « Ambisonics. Part two : Studio techniques », *Studio Sound* 17/8 (1975), p. 24-26, p. 28, p. 30 ; Correction : *Studio Sound* 17/10 (1975), p. 60.
- [7] Guillot, P. « La représentation intermédiaire et abstraite de l'espace comme outil de spatialisation du son », Thèse de doctorat, sous la dir. de A. Sedes et A. Bonardi, université Paris 8 Vincennes Saint-Denis, Paris 2017. [Voir tel.archives-ouvertes.fr/tel-01903017.]
- [8] Nicol, R. « Représentation et perception des espaces auditifs virtuels », Thèse de doctorat, Université du Maine, 2010. [Voir tel.archives-ouvertes.fr/tel-01066312.]
- [9] Webster, C., Garnier, F., Sedes, A. « Empty Room, how to compose and spatialize electroacoustic music in VR in ambisonic and binaural », *The International Journal of Virtual Reality* 17/1 (2017), p. 32-53. [Voir hal.archives-ouvertes.fr/hal-01634943.]