



**Université Lumière Lyon 2**

**Faculté LESLA**

# **Les solutions de spatialisation sonore contemporaines en audionumérique : les cas particuliers du cinéma, du jeu vidéo et de la réalité virtuelle**

Mémoire de Master 2 - Musiques Appliquées Aux Arts Visuels

Soutenu le 9 septembre 2021 par

**Valentin PETITEAU**

Sous la direction de

Martin BARNIER (Université Lumière Lyon 2) et

Jean-Marc SERRE (Université Lumière Lyon 2)

*Devant un jury composé par :*

Gilles Alonzo, compositeur, enseignant au CNSMD de Lyon

Martin Barnier, enseignant-chercheur Université Lyon 2

Marco Busetta, compositeur ATS Studios - Lyon

Emmanuel Deletang, éditeur 22D Music

Jean-Pierre Gaspar, vidéaste-compositeur, enseignant au Master MAAAV Lyon 2

Laetitia Pansanel-Garric, compositrice, enseignante au Master MAAAV Lyon 2

Jean-Marc Serre, compositeur, directeur du Master MAAAV Lyon 2



<b>Introduction</b>	<b>3</b>
<b>I - Naissance et évolution de la spatialisation au cinéma</b>	<b>5</b>
1.1 Les origines de la spatialisation	5
1.2 L'évolution de la monophonie vers la stéréophonie	6
1.3 L'élargissement du champ sonore avec les systèmes de diffusion surround	7
<b>II - Le jeu vidéo : Quelles solutions de spatialisation pour un média interactif ?</b>	<b>13</b>
2.1 Les atouts des intergiciels	13
2.2 La solution binaurale : simuler une perception spatiale optimale	18
2.3 L'ambisonie, un système de spatialisation 3D bénéficiant du son binaural	23
2.4 Utilisation du système ambisonique et du son binaural : retour sur mes travaux personnels	27
<b>III - La spatialisation du son en réalité virtuelle : une immersion nécessaire</b>	<b>31</b>
3.1 Les enjeux de la création sonore pour la réalité virtuelle	31
3.2 Une nouvelle manière de spatialiser le son pour un film : l'exemple de BattleScar	33
3.3 : Une spatialisation exigeante pour le jeu vidéo : l'exemple de Paper Beast	41
<b>Conclusion</b>	<b>47</b>
<b>Remerciements</b>	<b>49</b>
<b>Bibliographie</b>	<b>50</b>
Cinéma	50
Jeu vidéo	50
Réalité virtuelle	52
<b>Annexes</b>	<b>54</b>
<b>Liens vers trois travaux</b>	<b>55</b>
Voix Libre	55
Inoubliable	55
Qu'est ce que la co-gestion? *Projet COGITO*	56

# Introduction

La création sonore dans l'audiovisuel est un domaine riche et diversifié. Bien qu'elle puisse être réduite sous une seule appellation : « le son de ce film était génial », beaucoup de métiers différents sont mis en jeu, incluant des professions techniques et artistiques. Leur but à toutes et tous est de créer une cohérence sonore globale en harmonie avec le visuel, pour que l'auditeur puisse se dire que « le son est bien ». Le son comporte beaucoup de sous-domaines, allant de l'enregistrement audio en passant par le traitement et la manipulation du son puis sa diffusion. Je choisis de m'attarder sur un paramètre qui, bien qu'il puisse entrer en jeu à tous ces niveaux, a une plus grande application sur la diffusion : la spatialisation du son.

Depuis les années 1990 jusqu'à aujourd'hui, la majeure partie de la création, de la manipulation et de la diffusion du son se fait en audionumérique<sup>1</sup>. La plupart des corps de métiers de la production et de la post-production audio travaillent donc en numérique. La composition, le sound design, le mixage et le mastering peuvent être créés au travers de logiciels qu'on appelle des « Stations de Travail Audio-Numériques » (STAN) ou DAW (*Digital Audio Workstation*), représentés par les grandes marques de l'audio comme Steinberg (*Cubase, Nuendo, Dorico*), Avid (*Pro Tools*), Ableton, Reaper et d'autres. Le passage au numérique et l'utilisation de DAW ont donc permis un développement de la spatialisation car ils simplifient la manipulation de flux audio par rapport à l'analogique qui nécessitait un équipement plus imposant. Aujourd'hui des progrès continuent d'être faits, à la fois au cinéma et au travers du jeu vidéo qui pousse à développer un *son 3D* de plus en plus immersif tout en offrant des solutions créatives toujours plus avancées. La récente démocratisation de la réalité virtuelle est un terreau pour le développement des techniques de spatialisation sonore<sup>2</sup>, c'est pourquoi je m'y intéresse dans ce mémoire.

Les méthodes de spatialisation et de diffusion sonore contemporaines sont multiples et s'adaptent aux supports de diffusion : les sources monophoniques et stéréophoniques (toujours d'actualité pour l'enregistrement, la manipulation et la diffusion), le son surround (pour le cinéma ou les systèmes grand public spécialisés) ainsi que le son binaural et ambisonique qu'il faut distinguer<sup>3</sup>. Ces derniers trouvant leur utilité dans le jeu vidéo et la réalité virtuelle entre autres.

La spatialisation audio sera abordée du point de vue des professions de la production (composition musicale, création du sound design) et de la post-production audio (mixage, mastering, intégration), avec un regard porté sur des œuvres cinématographiques

---

<sup>1</sup> C. Roads, *L'audionumérique - Musique et informatique 3e ed.*, Dunod, 2016, p7

<sup>2</sup> P. Fuchs, *Théorie de la réalité virtuelle : les véritables usages*, Paris : Mines Paristech, 2018, p1

<sup>3</sup> P. Fuchs, G. Moreau, JM. Burkhardt. *Le traité de la réalité virtuelle Volume 2*. Presse des Mines, 2006, p 424-429

et vidéoludiques. Un focus sera fait sur la réalité virtuelle, qui peut englober ces deux domaines.

La réalité virtuelle promet un degré d'immersion supérieur par rapport aux supports conventionnels. Par support conventionnel, je fais la distinction entre un jeu vidéo dit *flat screen* (c'est-à-dire qui utilise un écran plat comme périphérique) d'un jeu vidéo en réalité virtuelle qui se joue dans un visiocasque, donc une manière non-conventionnelle -pour l'instant- d'expérimenter ce média. On entend souvent parler en français de « la VR » pour *Virtual Reality*, en revanche on lit plutôt « la RV » pour Réalité Virtuelle dans les écrits scientifiques. J'utiliserai donc plutôt cette dernière abréviation dans mon mémoire.

Je me concentrerai sur la création audiovisuelle contemporaine en étant attentif à la fois à la musique et au sound design en général, incluant les bruitages, les ambiances et les dialogues. Je m'intéresserai à l'importance de la spatialisation audio dans les œuvres audiovisuelles *flat screen* et dans les œuvres en réalité virtuelle en utilisant mon travail pour faire des liens avec ma recherche. Les questions techniques seront donc abordées pour présenter concrètement des solutions à la spatialisation sonore d'œuvres audiovisuelles.

Ces questionnements peuvent être synthétisés en une problématique globale : Comment réaliser une spatialisation sonore idéale dans la création audiovisuelle contemporaine ? Les cas particuliers du cinéma, du jeu vidéo et de la réalité virtuelle.

Le domaine de la spatialisation audio est conséquent, je me focaliserai sur une petite sélection de productions, chacune avec leurs particularités de création et de diffusion, afin d'en déduire les avantages et inconvénients.

Les expériences de spatialisation audio conventionnelle feront l'objet de la première et la deuxième partie. Nous verrons d'abord les solutions de spatialisation pour le son au cinéma, ou sur petit écran, c'est-à-dire pour un point d'écoute fixe. Ensuite nous aborderons le jeu vidéo (en parallèle avec mes travaux personnels de l'année), toujours en temps qu'expérience *flat screen* mais avec la particularité que le point de vue n'est généralement pas fixe, nécessitant des systèmes qui s'adaptent au point de vue des usagers.

La troisième partie abordera les solutions de spatialisation du son pour les expériences en réalité virtuelle. Après une courte présentation de ce média, nous étudierons un film et un jeu vidéo réalisés spécifiquement pour la réalité virtuelle. Le film *BattleScar* et le jeu *Paper Beast*. L'enjeu sera alors d'identifier quelles sont les similitudes et les spécificités par rapport aux œuvres *flat screen*.

# I - Naissance et évolution de la spatialisation au cinéma

## 1.1 Les origines de la spatialisation

La spatialisation du son peut être considérée comme le placement de sources sonores l'espace. D'abord un espace virtuel au sein duquel elles peuvent se mouvoir lors de la création et manipulation du son, puis un espace physique lors de sa restitution. Curtis Road écrit dans son chapitre sur la spatialisation dans *L'Audionumérique* :

« L'art de la spatialisation du son occupe aujourd'hui une position similaire à celle que l'art de l'orchestration occupait au dix-neuvième siècle. Déployer l'espace revient à chorégraphier le son : positionner les sources sonores et animer le son.<sup>4</sup>»

En faisant ce parallèle, il nous fait prendre conscience de la valeur esthétique que peut avoir la spatialisation sonore. L'ampleur de l'espace imaginaire dans lequel peut se déplacer le son dépend des moyens de création et de diffusion de celui-ci, allant d'un point ponctuel pour un son monophonique diffusé sur une seule enceinte à un espace totalement tridimensionnel qui n'a presque aucune limite pour un son créé et diffusé en ambisonique.

L'attention à la spatialisation du son n'est pas nouvelle. Déjà dans les sites grecs où officiaient des Pythies, le son donnait un écho étrange à la voix. Dans les églises romanes du Moyen Âge, l'acoustique était déjà pensée pour une propagation optimale du son. Au XVI<sup>e</sup> siècle, les chœurs de la Basilique Saint-Marc à Venise se répondaient d'un côté à l'autre de la salle dans les compositions d'Adrian Willaert ou d'Andrea Gabrielli. Cet attrait pour la spatialisation de la diffusion sonore a perduré au travers les âges et l'arrivée du haut-parleur au XIX<sup>e</sup> siècle a offert de nouvelles possibilités aux créateurs.<sup>5</sup> Le son spatial a continué de se développer dans des laboratoires de recherche comme le GRM (*Groupe de Recherches Musicales*) et le cinéma a été dans les premiers à bénéficier des avancées dans ce domaine. Aujourd'hui la plupart des salles de cinéma sont équipées au minimum de systèmes dit *surround*, permettant une spatialisation du son à 360° sur un plan horizontal à minima. Le grand public peut aussi y avoir accès à domicile grâce à la démocratisation des systèmes type Home-Cinéma de plus en plus abordables.

Dans l'audiovisuel, la spatialisation du son peut concerner tout ce qui est diffusé par haut-parleurs, oreillettes, casque ou autre dispositif de diffusion sonore. Cela comprend le sound design (bruitage, ambiance), les dialogues et la musique.

---

<sup>4</sup> C. Roads, *op. cit.* p121

<sup>5</sup> *ibid.* p122

## 1.2 L'évolution de la monophonie vers la stéréophonie

Dès l'Exposition Internationale d'Électricité à Paris en 1881, le Théâtrophone de Clément Ader proposait la restitution d'un concert classique à distance par la biais d'un casque audio pour un public ayant payé sa place :

« Ici, « chaque personne reçoit l'impression des deux transmetteurs distincts par l'une et l'autre oreilles » plus ou moins fort selon les déplacements sur la scène, de la même manière que la voix d'un acteur, se trouvant à droite du souffleur, actionne le microphone de droite plus énergiquement que celui de gauche »<sup>6</sup>

Nous constatons que même s'il s'agissait d'un son monophonique qui se déplaçait plus ou moins sur les côtés des oreillettes du casque, il s'agissait déjà des prémices d'une diffusion stéréophonique pour spatialiser le son et donner aux auditeurs la sensation qu'ils étaient dans la salle de concert. Grâce à cet exemple, nous constatons que la spatialisation du son était donc déjà d'actualité un peu avant que le cinéma n'apparaisse.

La diffusion monophonique est à l'origine ponctuelle. Il est donc difficile de recréer un espace sonore car on ne dispose que d'un point fixe dans l'espace incarné par un haut-parleur ou un ensemble de haut-parleurs qui diffusent le même son à la même intensité. Au mixage, on peut donc seulement éloigner ou rapprocher les sources dans l'espace en jouant sur leur gain mais il est impossible de simuler un déplacement horizontal. La seule façon de spatialiser une diffusion monophonique est donc de déplacer physiquement le haut-parleur dans l'espace d'une salle de concert ou de cinéma.

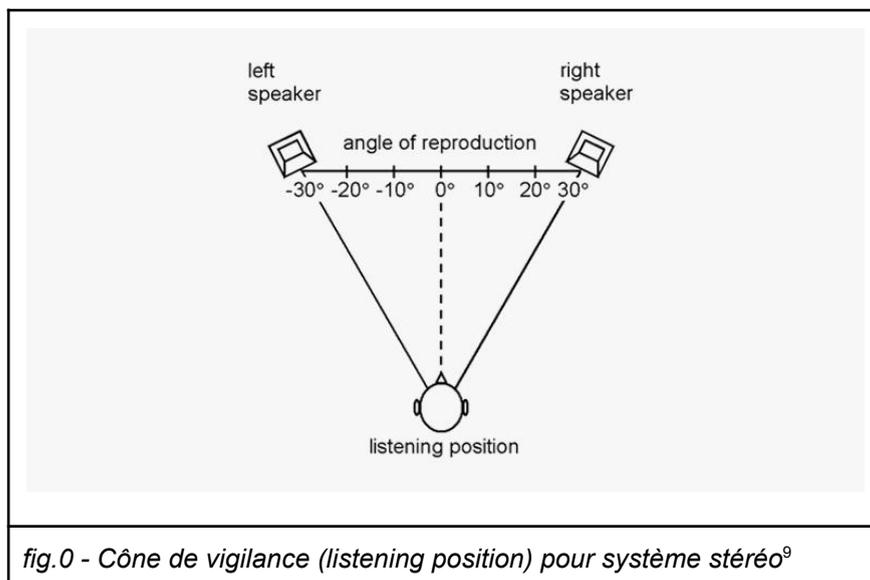
La spatialisation audio concevable dans un studio commence donc vraiment à partir d'une diffusion en stéréo. La stéréophonie a été brevetée par Alan Blumlein en 1931<sup>7</sup>. En stéréo, les objets sonores d'un mix peuvent se déplacer sur un plan horizontal à 180° dès l'étape de la création en studio. Ensuite ce plan horizontal peut être restitué devant l'auditeur dans un espace physique par deux haut-parleurs au minimum. La perception de l'azimut (angle horizontal) est fortement liée au dispositif d'écoute, par exemple, un mix stéréo diffusé sur un seul haut-parleur annihile l'effet de spatialisation verticale et revient à écouter une diffusion monophonique. Au contraire, l'écoute d'un mix stéréo dans un casque audio peut renforcer la perception spatiale, jusqu'à trop l'accentuer si des éléments sont placés totalement à gauche ou à droite. La façon idéale d'entendre un son stéréo est de se situer dans le « cône de vigilance », c'est à dire « Le point où le système de diffusion est calibré de

---

<sup>6</sup> K. Kitsopanidou et G. Pisano. *Les salles de cinéma. Enjeux, défis et perspectives*. Laurent Creton éd., Armand Colin, 2013, p147-178.

<sup>7</sup> C. Roads, *op. cit.* p139

façon optimale et permet la meilleure reconstitution de l'espace sonore »<sup>8</sup>. Pour la stéréo par exemple, le cône de vigilance se situe au sommet d'un triangle équilatéral formé par les deux enceintes et la tête de l'auditeur. [fig.0]



Une fois dans ce cône de vigilance, des objets peuvent apparaître et se déplacer horizontalement à n'importe quel point entre les enceintes, c'est ce que Nicolas Tsingos et Olivier Warusfel appellent des « sources fantômes ». Elles émergent grâce à l'application de retards ou de modifications de gain sur les signaux sonores, des calculs complexes et techniques sur lesquels nous ne nous attarderons pas. La plupart des systèmes multicanaux grand public (stéréophonie, quadraphonie, 5.1, 7.1) sont basés sur cette illusion de sources fantômes<sup>10</sup>

### 1.3 L'élargissement du champ sonore avec les systèmes de diffusion surround

Au cinéma, ces sources fantômes qui émergent d'un système audio de deux haut-parleurs ou plus permettent de créer une image sonore en corrélation avec ce qui est montré à l'image. Ça n'est pas un hasard si en 1931, l'idée de la stéréophonie est venue à Alan Blumlein au cinéma :

« Blumlein avait eu l'idée d'inventer un système binaural, car il lui semblait que lors de la projection d'un film, il était préférable par soucis de réalisme que le son

<sup>8</sup> *ibid.* p147

<sup>9</sup> DPA Microphone. <https://www.dpamicrophones.com/mic-university/stereo-recording-techniques-and-setups>  
Publié le 11/11/2019, consulté le 29/08/2021.

<sup>10</sup> P. Fuchs. *Le traité de la réalité virtuelle Volume 2*. Presse des Mines, 2006, p422

« suive » l'image, et qu'un acteur situé d'un côté de l'écran soit entendu sur le même côté. »<sup>11</sup>

Dans cette citation le terme « binaural » se réfère à l'ancien nom donné à la stéréophonie et non à l'usage qu'on en fait aujourd'hui que nous aborderons plus loin dans ce mémoire. Bien qu'il ait fallu attendre 1935 pour entendre la première application de ce système au cinéma pour du bruitage<sup>12</sup>, donc après le premier disque musical enregistré en stéréo, nous pouvons tout de même affirmer que la spatialisation audio est née pour obtenir un rendu plus réaliste du son au cinéma, dans un but d'immersion du spectateur.

Étonnement, une pratique de spatialisation audio admise dans le cinéma d'aujourd'hui va à l'encontre de l'idée de Blumlein. Ce dernier souhaitait qu'on puisse suivre la voix d'un acteur sur les côtés de l'écran, or aujourd'hui une bonne pratique est de toujours mettre les voix des dialogues dans le haut-parleur central quelque soit la position des acteurs par rapport au cadre.<sup>13</sup> Ceci malgré l'élargissement du champ sonore avec les systèmes de diffusion surround. Une des raisons de cette pratique est qu'elle permet une meilleure distinction des dialogues. Ensuite notre cerveau place lui-même la source sonore dans l'espace où se situe la personne qui parle, c'est ce qu'on appelle « l'effet ventriloque »<sup>14</sup>. Dans mon travail, je respecte la plupart du temps cette règle mais je ne me prive pas de spatialiser les voix différemment dans certains cas. Par exemple dans le film *Voix Libre* de l'ECV Bordeau pour lequel j'ai fait le sound design, j'ai mis les voix au centre pour les dialogues, mais je me suis permis de déroger à cette règle quand le personnage qui parle est hors-champ. Cela peut créer un effet de surprise qui donne de l'importance à cette voix et qui souligne la direction vers laquelle le personnage qui entend cette voix tourne la tête [fig.1].



<sup>11</sup> C. Roads, *op. cit.* p139

<sup>12</sup> A. Blumlein, *Walking and Talking*, Juillet 1935, EMI Archive

<sup>13</sup> C. Tijou, *Cohérence audiovisuelle spatiale au cinéma : Influence de la vitesse et de la nature des objets sonores*. 2016, Mémoire de fin d'étude de l'ENS Louis Lumière. p28

<sup>14</sup> C. Tijou, *op. cit.* p37

Si on place plusieurs voix qui parlent simultanément à différentes positions dans l'espace, cela peut créer un effet de foule qui brouille la compréhension de paroles précises. En y ajoutant des effets comme de la réverbération et de l'écho, les voix de cette foule peuvent prendre de l'ampleur jusqu'à devenir fantastiques ou cauchemardesques, mais l'intelligibilité est encore plus diminuée. J'ai par exemple utilisé cet effet pour créer des sons internes-subjectifs -des sons entendus imaginaires par un personnage et par lui seul<sup>15</sup>- dans les courts métrages *Voix Libre* et *La Récolte* de l'ECV Bordeaux [fig.2].



fig.2 : À gauche *Voix Libre* à 04:40'. À droite *La Récolte* 04:48'

Dans ces deux cas, les personnages entendent des voix qui résonnent dans leur tête. J'ai donc spatialisé les voix en les envoyant autour du spectateur et j'ai utilisé différentes transformations du son (écho, réverbération, pitch-shifting) pour soutenir cet effet.

Ces sons internes-subjectifs qui tournent autour du spectateur sont de bons exemples de l'utilité d'un système dit « surround ». Les systèmes de diffusion surround sont une grande famille dans laquelle nous pouvons classer « tout système émettant le son par plus de deux haut-parleurs »<sup>16</sup>. Par rapport au système stéréo qui demande d'avoir deux haut-parleurs devant l'auditeur, les systèmes surround comme le 5.1 ou le 7.1 utilisent des sources supplémentaires placées derrière l'auditeur. Ils permettent donc de faire entendre du son à 360° autour du spectateur dans une salle de cinéma. En revanche ces deux systèmes (5.1, 7.1) restituent toujours un champ sonore sur un plan horizontal seulement. Si je devais mixer les sons internes-subjectifs évoqués précédemment en surround 5.1, l'effet serait grandement renforcé car je ferais tourner les voix à 360° dans la salle pour immerger le spectateur dans le son et leur faire ressentir ces voix réminiscentes.

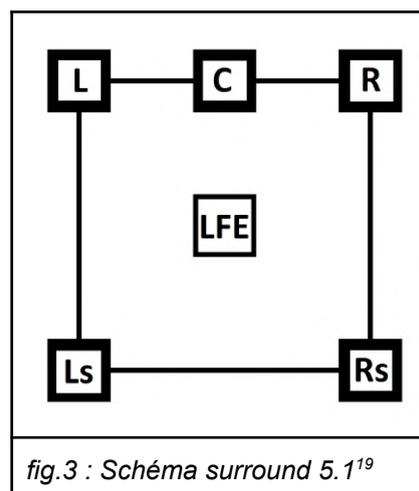
Les premières expérimentations de diffusion du son sur des systèmes multicanaux au cinéma ont été réalisées en 1940 avec un format appelé le *Fantasound* créé spécialement pour le film de Walt Disney : *Fantasia* (1940). Comme le son stéréophonique,

<sup>15</sup> M. Chion, *L'audio-vision - Son et image au cinéma*, 4e ed., Armand Colin, 2017, p67

<sup>16</sup> C. Roads, *op. cit.* p139

le son surround est donc né pour le cinéma et comme pour la stéréo, les premiers enregistrements ont d'abord été musicaux. Ce qui intéressait Leopold Stokowski -chef d'orchestre à l'origine de l'enregistrement de la musique de *Fantasia*- était la reproduction fidèle de l'orchestre, l'immersion dans un concert était présente grâce à l'enregistrement de l'orchestre sur huit pistes différentes (violons/altos, violoncelles, contrebasses, bois, cuivres, percussions et deux prises d'ensemble). Ces huit pistes pouvaient ensuite être diffusées selon la spatialisation choisie en studio sur des ensembles de haut-parleurs pouvant aller de trente à quatre-vingt.<sup>17</sup> Les techniques de diffusion surround plus spectaculaires que la stéréo ont continué de se développer au cours du XX<sup>e</sup> siècle, entre autres pour concurrencer à l'essor de la télévision et continuer d'attirer du public au cinéma. L'entreprise Dolby a joué un rôle majeur dans la diffusion des systèmes audio surround pour le grand public. Elle est à l'origine de multiples formats de stockage et de diffusion, d'abord analogiques puis numériques. Nous ne les citerons pas tous, mais nous pouvons mentionner le Dolby Digital, apparu en 1992, qui a démocratisé le son surround sur support numérique<sup>18</sup>. Le son surround 5.1 Dolby Digital est stocké sur six canaux numériques, il est donc accessible à la fois sur des films optiques pour le cinéma et sur des supports numériques grand public comme le DVD pour les particuliers. Le surround 5.1 est le format surround le plus répandu et accessible mais il existe beaucoup d'autres formats dérivés du même principe comme le 7.1, le 9.1, etc. plus le nombre est élevé, plus il y a d'enceintes et plus la restitution spatiale du son est précise.

Les six canaux du surround 5.1 sont : Gauche (*Left*), Centre (*Center*), Droite (*Right*), Arrière-Gauche (*Left surround*), Arrière-Droite (*Right surround*) et LFE (*Low Frequency Effect*). [fig.3]



<sup>17</sup> *ibid.* p140

<sup>18</sup> *ibid.* p145

<sup>19</sup> Crédit : jynus. URL :

[https://fr.wikipedia.org/wiki/5.1#/media/Fichier:5\\_1\\_channels\\_\(surround\\_sound\)\\_label.svg](https://fr.wikipedia.org/wiki/5.1#/media/Fichier:5_1_channels_(surround_sound)_label.svg) Mise en ligne le 26/05/2006, consulté le 30/08/2021.

L'ordre des canaux peut changer selon une diffusion en cinéma ou en broadcast audiovisuel (Télévision, Internet). En effet, toutes les personnes qui ont la télévision ne sont pas forcément équipées en système de son en 5.1, il est donc nécessaire d'avoir une diffusion compatible avec une conversion en stéréo, ce qui change l'ordre des canaux.

« LFE » est l'abréviation de *Low Frequency Effect*, il s'agit du « .1 » dans 5.1. Ce canal est réservé pour diffuser des sons graves aux fréquences généralement comprises entre 10 Hz et 120 Hz, il a au moins un haut-parleur dédié (parfois plusieurs dans les grandes salles de cinéma ou les systèmes surround avancés, on parle alors 5.2 ou 7.2, etc.). C'est le LFE qui fait trembler le sol et vibrer nos corps lors d'explosions ou d'autres sons graves au cinéma. Ce canal est très intéressant pour créer des sensations fortes mais un usage excessif peut créer une fatigue auditive et physique importante. On n'y met généralement pas de musique ou de sons à hauteur déterminée mais plutôt des nappes de textures sonores ou des coups ponctuels pour renforcer d'autres sons. Il n'est en revanche pas intéressant en termes de spatialisation car nos oreilles ne parviennent pas à percevoir précisément la direction des fréquences graves, la plupart des informations de localisation se situent dans les hautes fréquences aux alentours de 1 kHz et 4 kHz où nos oreilles sont plus sensibles.<sup>20</sup>

Comme j'ai pu le constater lors d'un stage au studio de postproduction audio Piste Rouge, la majorité des programmes audiovisuels destinés au cinéma sont créés en surround 5.1 au minimum. Les ingénieurs du son qui mixent dans ce format doivent contrôler très précisément ce qu'ils veulent faire passer dans chaque canal car des problèmes de phase peuvent vite arriver entre les haut-parleurs avants et arrières. Aujourd'hui, Dolby est toujours à la pointe de la spatialisation audio au cinéma avec par exemple, le format de création et de diffusion Dolby Atmos inauguré en 2012. Les salles de cinéma sont de plus en plus équipées avec ce format qui permet une spatialisation verticale, mais l'expérience n'est pas forcément accessible au grand public car les séances dans les salles utilisant cette technologie sont généralement plus coûteuses que les salles équipées en surround horizontal traditionnel<sup>21</sup>. Ce format propose un son immersif avec jusqu'à 64 enceintes de diffusion placée sur les côtés et au-dessus des spectateurs, ce qui ouvre la dimension verticale pour la spatialisation sonore au cinéma. Les créateurs qui travaillent avec ce format font maintenant ce qu'on appelle un « mix orienté objet », c'est à dire qu'ils n'ont plus à penser précisément au canal de diffusion du son. Ils placent leurs objets sonores dans un espace virtuel directement dans le logiciel Atmos qui offre jusqu'à 128

---

<sup>20</sup> P. Fuchs. *Le traité de la réalité virtuelle Volume 2*. Presse des Mines, 2006, p423

<sup>21</sup> Au cinéma Pathé-Gaumont Lyon Vaise, une séance en salle équipée *Dolby Cinema* coûte 18.40€, contre 13.40€ pour une salle non équipée. Comparaison réalisée sur les séances pour le film *Reminiscence* du 30/08/2021.

pistes audio indépendantes<sup>22</sup>. Ce même logiciel décode en temps réel la position du son au moment de la diffusion pour le jouer sur les haut-parleurs appropriés. Avec cette nouvelle façon de travailler, la spatialisation du son au cinéma se rapproche donc beaucoup de celle du jeu vidéo qui calcule en temps réel la position des objets sonores. La spatialisation Dolby Atmos est aussi disponible pour les particuliers mais elle exige un matériel de pointe pour décoder le son et beaucoup de hauts parleurs pour le diffuser. On parle de systèmes 5.1.4 ou 7.1.4 où le « .4 » correspond à quatre haut-parleurs au plafond (pouvant être réduit à deux).

---

<sup>22</sup> J. Hidalgo. *Dolby's Atmos technology gives new meaning to surround sound, death from above*. Engadget. [en ligne] URL : <https://www.engadget.com/2012-04-26-dolby-atmos-technology-gives-new-meaning-to-surround-sound.html>, publié le 26/04/2012, consulté le 02/08/2021.

## II - Le jeu vidéo : Quelles solutions de spatialisation pour un média interactif ?

### 2.1 Les atouts des intergiciels

Au cinéma, le point d'écoute est fixé par les images. Lorsque les images sont tournées et le montage effectué, le sound designer peut contrôler très précisément ce qu'il veut faire entendre aux spectateurs par rapport aux images. S'il le souhaite, il peut passer du temps à parfaire la spatialisation sonore d'une mouche qui vole dans le cadre, hors du cadre ou qui oscille entre les deux. L'effet sera saisissant au casque et avec un système surround. En revanche, dans un jeu vidéo, le joueur qui se trouve devant une mouche n'a qu'à déplacer son personnage pour qu'elle se retrouve à sa droite ou à sa gauche, devant ou derrière, plus bas ou plus haut, plus loin ou plus près, etc. Tous ces mouvements impossibles à prévoir pour le créateur sonore sont dus à l'interactivité inhérente au support vidéoludique, il a donc fallu trouver de nouveaux moyens pour spatialiser le son par rapport aux images. Des systèmes comme Wwise ou Fmod ont donc émergé pour éviter les dissonances son/image et pour permettre aux créateurs sonores d'aborder la partie audio des jeux vidéo sereinement, dans le but d'apporter un son de qualité à cette œuvre globale qu'est un jeu vidéo.

Comme je l'évoquais dans l'introduction, les sound designers, compositeurs et ingénieurs du son travaillent avec une station de travail audionumérique aussi appelée un DAW pour *Digital Audio Workstation* (l'usage du masculin pour DAW vient probablement du fait qu'un DAW soit aujourd'hui considéré davantage comme un logiciel plutôt qu'une station de travail physique). Bien que tous les DAW aient à peu près les mêmes fonctions de base, chacun est généralement plus adapté à un média et une profession. Pour ma part, j'utilise Cubase de Steinberg, qui est orienté vers la création musicale mais qui permet aussi de faire du sound design et de la post-production audio. Le DAW Pro Tools d'Avid est majoritairement utilisé dans la post-production audio cinématographique et en broadcast audiovisuel (TV, internet). Il est très performant pour la post-production mais est moins accessible pour la composition musicale, même si cela reste possible. Son équivalent chez Steinberg est Nuendo, il est aussi orienté vers la post-production mais il est aussi très utilisé dans la production et la post-production audio vidéoludique.

Nuendo, permet depuis 2015 de communiquer directement avec un programme appelé Wwise (prononcé [wɪz]). Ce programme développé par Audiokinetic est un *middleware* (ou intergiciel pour logiciel intermédiaire en français), il s'agit donc d'un programme qui permet à différents logiciels de communiquer entre eux. Dans un jeu vidéo,

l'intergiciel audio est une couche de logiciel intermédiaire qui s'occupe exclusivement du son, il est lié au moteur de jeu qui effectue les calculs informatiques nécessaires à son fonctionnement.<sup>23</sup> En l'occurrence Wwise est un intermédiaire entre une station de travail audio pour les compositeurs/sound designers et le moteur de jeu vidéo pour les développeurs qui veulent intégrer du son dans leur jeu. Comme pour les DAW, Wwise a un concurrent qui fait à peu près la même chose mais différemment appelé Fmod, développé par Firelight Technologies. À eux deux, Wwise et Fmod sont les leaders actuels du marché. La différence majeure entre ces deux intergiciels est leur modèle économique : Wwise est plus avancé mais moins accessible car il faut faire une demande à l'entreprise pour commercialiser les jeux faits avec celui-ci. Fmod a moins de fonctions avancées mais est plus accessible car il est totalement gratuit tant que les revenus de vente du jeu vidéo ne dépassent pas un certain plafond.

L'apparition de ces intergiciels, Fmod en 1995<sup>24</sup> et Wwise en 2006<sup>25</sup>, a permis une plus grande accessibilité aux sound designers et compositeurs au monde du jeu vidéo. Dans les années 1980, les compositeurs qui travaillaient pour le jeu vidéo devaient à la fois être bons musiciens et bons programmeurs car ils devaient pouvoir gérer les puces sonores des différentes consoles de jeu<sup>26</sup>. Bien qu'il faille toujours quelques notions de programmation pour comprendre comment fonctionne le son dans un jeu vidéo, la transition entre la création sonore cinématographique et vidéoludique pour un compositeur s'est beaucoup simplifiée grâce aux intergiciels. En effet, leur processus de travail se rapproche beaucoup de celui des DAW avec lesquelles ils ont l'habitude de travailler. Cette accessibilité a donc permis aux créateurs sonores, spécialistes dans leur art, d'arriver dans le monde du jeu vidéo avec leurs idées artistiques sans être trop contraints par la technique, cela résulte en une production sonore de plus en plus intéressante pour le jeu vidéo.

Un compositeur comme Olivier Derivière met un point d'honneur à créer des musiques de jeu vidéo qui tirent parti au maximum de la composante interactive de ce média, à l'opposé de la musique de cinéma qu'il considère comme étant « illustrative ».<sup>27</sup> Dans les différents genres de jeux pour lesquels il a travaillé, comme *Remember Me (Action-TPS)*, *Street of Rage (Beat 'em up)* ou *Get Even (Action-FPS)*, il a toujours pensé l'interactivité de la musique avec une progression verticale ou horizontale de celle-ci. Après

---

<sup>23</sup> G. De Prato, *Une technologie de base : l'intergiciel (middleware)*, Hermès, La Revue, 2012/1 (n° 62), p. 101-107. [en ligne] URL : <https://www.cairn.info/revue-hermes-la-revue-2012-1-page-101.htm> Consulté le 20/08/2021.

<sup>24</sup> G. Somberg, *Game Audio Programming - Principles and Practices*, 2016, CRC Press. p103

<sup>25</sup> Site officiel, *Audiokinetic Wwise*, [en ligne] URL : <https://www.audiokinetic.com/fr/about/> Consulté le 23/08/2021.

<sup>26</sup> L. Tobisch, *Comment la musique est devenue un ingrédient essentiel des jeux vidéo*. [en ligne] URL : <https://www.francemusique.fr/culture-musicale/la-musique-des-jeux-vidéo-un-monde-a-decouvrir-66460> Publié le 26/10/2018, consulté le 08/08/2021.

<sup>27</sup> O. Derivière, *Interview d'Olivier Derivière pour L'Écran*. [en ligne] URL : <https://asso-lecran.fr/interview-de-olivier-deriviere> Publié le 16/02/2020, consulté le 23/08/2021.

avoir composé et produit des éléments musicaux dans son DAW, il utilise Wwise pour intégrer la musique dans les jeux vidéo en prenant en compte au maximum le *gameplay* propre à chacun d'eux, c'est-à-dire que la manière d'y jouer doit influencer le son.

En plus de simplement intégrer des sons qui se déclenchent à des moments précis dans le jeu, l'avantage des intergiciels est la possibilité qu'ils offrent de contrôler les paramètres du son en temps réel. Contrairement au cinéma où le sound designer a souvent accès à la prise de son direct captée lors du tournage, le son d'un jeu vidéo est une recreation totale, sans référence. Il faut donc choisir ce qu'on veut faire entendre et comment on veut le faire entendre, le réalisme n'étant pas toujours la bonne option. Par exemple, de la même manière qu'au cinéma, les dialogues importants doivent toujours être audibles peu importe la quantité de sons environnants. Cependant, l'effet ventriloque n'est pas vraiment efficace dans le jeu vidéo à cause de l'immersion du joueur dans l'expérience. Prenons par exemple une séquence de dialogue du jeu *Red Dead Redemption 2* de Rockstar Games, le personnage qu'incarne le joueur parle avec un autre personnage pendant qu'ils chevauchent dans la montagne en pleine tempête de neige.[fig.4]



fig.4 - *Red Dead Redemption 2*, séquence de dialogue dans un blizzard.<sup>28</sup>

Ici, on pourrait choisir de ne pas spatialiser les dialogues dans l'univers du jeu et de les mettre au centre comme au cinéma, mais cela briserait totalement l'immersion du joueur dans cette chevauchée, il aurait l'impression d'entendre des voix-off. Il est malgré tout difficile de prendre le parti d'un rendu totalement réaliste pour cette séquence, car en réalité le son du blizzard couvrirait totalement les voix. Dans ce cas, des contrôles de paramètres

<sup>28</sup> GameArmy, *Red Dead Redemption 2 - Full Game Walkthrough (No Commentary Longplay)*, [en ligne] URL : <https://youtu.be/jZilbLz0m8I?t=378> de 06:18' à 14:45', mis en ligne le 13/11/2018, consulté le 30/08/2021.

du son en temps réel peuvent entrer en jeu. Plutôt que de baisser un à un les sons qui pourraient parasiter la compréhension des dialogues à cet instant, il est plus commode de créer un effet *sidechain* dans l'intergiciel comme on pourrait le faire dans un DAW classique. Il faudra simplement créer un paramètre qui abaissera le volume des autres bus (canaux audio) quand le bus de dialogue sera actif. De cette manière les dialogues restent intelligibles tout en étant spatialisés dans le monde virtuel du jeu.

Avec cet exemple simple de contrôle de niveaux sonores en temps réel, nous pouvons imaginer la puissance des intergiciels pour créer des systèmes audio dynamiques. Les paramètres peuvent contrôler n'importe quel aspect du son (le niveau de réverbération, le niveau d'un filtre passe-bas ou passe-haut, la spatialisation du son qui peut avoir un comportement avancé, etc.).

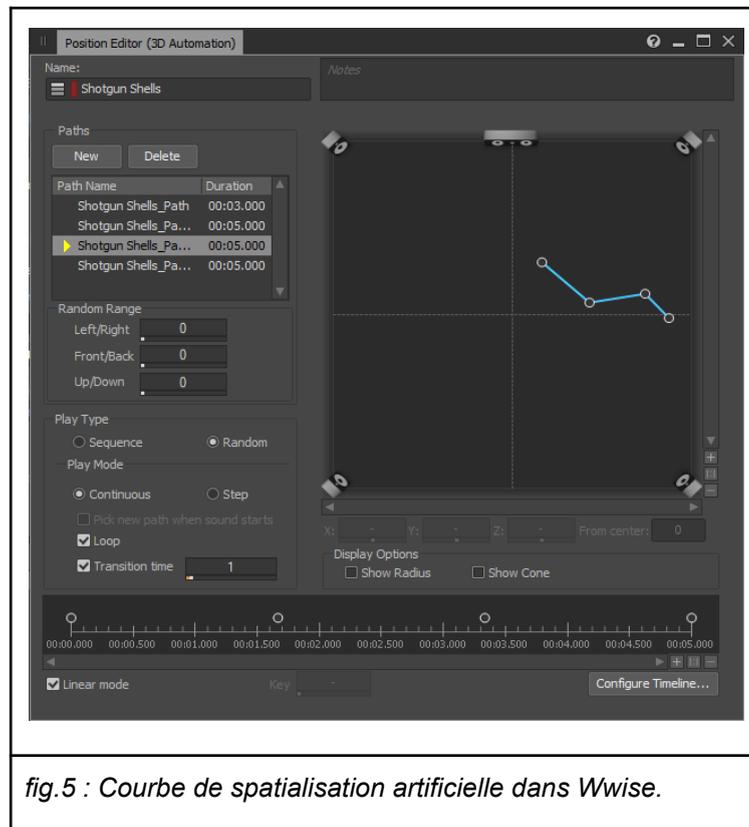
La spatialisation est un aspect central des intergiciels Wwise et Fmod. En effet, une fois que le son est passé du DAW à l'intergiciel, il y a des allers-retours entre l'intergiciel et le moteur de jeu. Ce dernier envoie les coordonnées de position 3D des objets à l'intergiciel qui se charge de jouer les sons dans le champ sonore en conséquence. Les coordonnées sont les éléments basiques que le moteur de jeu renvoie à l'intergiciel pour spatialiser un son, mais il peut aussi envoyer une multitude d'autres données qui peuvent servir à contrôler des paramètres choisis par le sound designer. Par exemple, s'il décide de faire passer une valeur de vitesse de déplacement de l'objet à l'intergiciel, celui-ci pourra simuler un effet Doppler qui pourra être plus ou moins prononcé selon sa configuration.

Dans Wwise et Fmod, le sound designer peut choisir de spatialiser un son ou non, à la manière d'un son « In » diégétique et visible ou d'un son « Off » non-diégétique et acousmatique<sup>29</sup>. S'il décide de spatialiser le son, il a plusieurs solutions. Considérons un FPS (*First Person Shooter* pour jeu de tir à la première personne) dans lequel le joueur tire avec une arme libérant une douille qui rebondira sur le sol. Si le sound designer souhaite spatialiser le rebond de la douille en temps réel sur le sol, le moteur de jeu devra envoyer les coordonnées des rebonds de la douille à chaque tir et pour chaque douille s'il y en a plusieurs. Le sound designer devra ensuite créer une « courbe d'atténuation » qui définira à quelle distance le son est audible (dans le cas de la douille, le son ne serait pas audible au-delà d'1,50 mètre environ). Mais cette méthode de spatialisation ne serait pas pertinente dans le cas d'un FPS, car dans la création vidéoludique, un des objectifs principaux est d'économiser un maximum de ressources du processeur qui fait fonctionner le jeu. Dans ce cas précis, le moteur de jeu devrait envoyer les coordonnées pour chaque douille qui tombe au sol. S'il y a plusieurs personnages qui tirent en même temps, cela ferait une multitude de coordonnées à traiter en même temps pour un son finalement assez insignifiant. Une autre option de spatialisation serait donc de ne pas spatialiser cet élément en temps réel et de créer une « courbe de spatialisation ». C'est-à-dire que le moteur du jeu n'aurait pas à

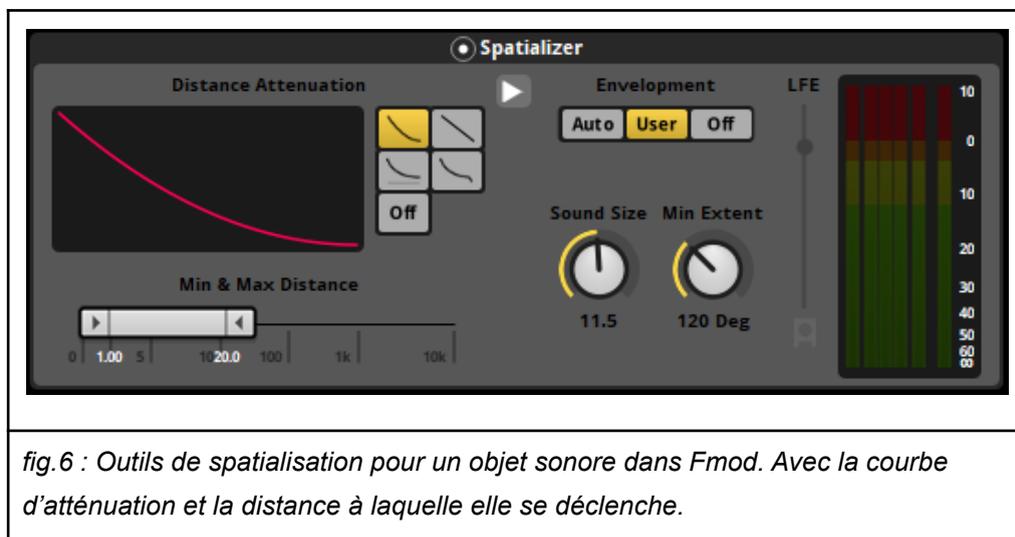
---

<sup>29</sup> M. Chion, *Un art sonore : le cinéma*, Cahiers du cinéma. 2003. p223-234

envoyer les coordonnées des rebonds de la douille pour calculer leur spatialisation en temps réel. Il suffit de tracer des courbes de spatialisation artificielles dans l'intergiciel qui seront appliquées aléatoirement sur le son des douilles pour simuler une spatialisation complexe. [fig.5]



Fmod est l'intergiciel que j'ai utilisé pour l'intégration du son dans le jeu *Lune Noire* de l'école Brassart Lyon pendant mon stage de Master 2. Pour définir le niveau sonore d'un objet et l'espace, Fmod utilise le même système de courbes d'atténuation que Wwise. [fig.6]



J'ai réalisé un exemple simple d'un son spatialisé dans l'espace 3D du moteur de jeu Unity (voir Annexe *Fmod 1.1* à *1.3*). Dans cet exemple, le son est diffusé en stéréo mais les intergiciels prennent aussi en charge la diffusion sur d'autres systèmes de diffusion plus développés. Puisque la position des objets sonores est déterminée par des coordonnées 3D, l'intergiciel peut envoyer le son dans n'importe quel type de système défini en amont par le sound designer. Comme pour le cinéma, du son peut être envoyé dans le canal LFE des systèmes surround.

Les intergiciels audio se perfectionnent constamment proposant des solutions de spatialisation toujours plus immersives. Chez Wwise, cela se matérialise par une nouvelle gamme de fonctionnalités appelée « Wwise Spatial Audio »<sup>30</sup>. Quant à Fmod, il profite des fonctionnalités spatiales avancées du plugin « Steam Audio » de Valve dont la v4.0.0. est paru le 22 juillet 2021. [fig.7]



fig.7 : Le plugin de spatialisation Steam Audio développé par Valve, proposant beaucoup plus de paramètres que le plugin natif de Fmod.

Le plugin de spatialisation Steam Audio prend en compte des facteurs du son très avancés tels que l'influence de l'occlusion, de la réflexion ou de la propagation selon différentes surfaces. Son atout majeur est qu'il prend en compte les HRTF (*Head-Related Transfer Functions*), essentielles pour créer un son binaural immersif, notamment dans les expériences en réalité virtuelle.<sup>31</sup>

## 2.2 La solution binaurale : simuler une perception spatiale optimale

Quand Alan Blumlein a déposé le brevet de la stéréophonie en 1931, il l'a d'abord appelé « son binaural », dans le sens où le son était diffusé sur deux canaux différents. Le terme de stéréophonie est arrivé plus tard et « binaural » a pris un autre sens dans les

<sup>30</sup> Audiokinetic, *Wwise Spatial Audio*. [en ligne] URL : <https://www.audiokinetic.com/fr/products/wwise-spatial-audio/>. Consulté le 23/08/2021.

<sup>31</sup> Steam Audio, *Immersive Audio Solutions*, [en ligne] URL : <https://valvesoftware.github.io/steam-audio/#learn-more>. Consulté le 23/08/2021.

usages sonores contemporains.<sup>32</sup> Dans cette partie nous allons explorer la puissance de la restitution spatiale du son binaural et les usages possibles, en particulier pour le jeu vidéo et la réalité virtuelle. Nous nous attarderons aussi sur le format ambisonique, qu'il ne faut pas confondre avec la diffusion binaurale mais qui va souvent de pair avec elle.

Le binaural, après avoir eu donc pour sens la stéréophonie, est devenu le nom d'un domaine de recherche acoustique au sein duquel on a étudié l'audition humaine dans une chambre anéchoïque<sup>33</sup>. Aujourd'hui on utilise le terme binaural pour nommer une technique de reproduction du son qui a pour objectif d'être la plus fidèle possible à l'audition humaine naturelle, notamment dans sa perception de la spatialisation. Actuellement, l'usage du son binaural est uniquement possible au travers un casque audio ou un autre dispositif avec deux oreillettes directement placées sur les oreilles. Nous verrons par la suite ce qui justifie ce dispositif. Voici d'abord une définition du son binaural par l'Institut de Recherche et Coopération Acoustique/Musique (IRCAM) :

« La restitution réaliste de sources ponctuelles sur casque d'écoute implique l'usage d'une spatialisation binaurale. Celle-ci est basée sur le filtrage dynamique de la source sonore par les fonctions de transfert (*Head-Related Transfer Function*, HRTF) préalablement mesurées sur la tête d'un auditeur ou d'un mannequin. »<sup>34</sup>

Le son binaural est donc avant tout un moyen d'obtenir une restitution spatiale optimale du son. S'il a d'abord été réservé aux milieux de la recherche scientifique, son accessibilité actuelle permet aux créateurs de l'audiovisuel d'envisager son intégration dans leurs œuvres pour atteindre un degré d'immersion convaincant. Avant de parler des HRTF mentionnés dans la citation, principe au centre de la reproduction binaurale du son, je vais m'intéresser aux principes physiques de la perception spatiale du son.

Pour s'extraire de la diffusion spatiale limitée à un champ sonore horizontal – que nous avons vu jusqu'ici avec la stéréo à 180° et le surround à 360° – il faut s'intéresser à la façon dont les êtres humains perçoivent la localisation du son. Imaginons un espace 3D défini par les repères (x, y, z) où sur un plan horizontal nous avons (x) l'abscisse et (y) l'ordonnée, et sur un plan vertical nous avons (z) la cote ou l'altitude.

Le placement du son sur l'axe des ordonnées (y) correspond à l'**azimut** (ou angle horizontal). C'est ce placement qui définit si le son vient plutôt de la droite ou de la gauche de l'auditeur. Sa perception est majoritairement indiquée par la différence du temps d'arrivée du son aux deux oreilles de l'auditeur, s'il atteint d'abord l'oreille droite il sera perçu de ce

---

<sup>32</sup> C. Roads, *op. cit.* p139

<sup>33</sup> *ibid.*

<sup>34</sup> IRCAM, *Technologies de reproduction binaurale*. [en ligne] URL : <https://www.ircam.fr/projects/pages/technologies-de-reproduction-binaurale/> Consulté le 24/08/2021.

côté et inversement. Plus le son mettra du temps à arriver à l'oreille gauche, plus le son sera perçu à droite et inversement. C'est ce qu'on appelle la « différence de temps interaurale » (ou ITD pour *Interaural time difference*). Cette perception est complétée par la différence d'intensité du son qui atteint chaque oreille la « différence de niveau interaurale » (ILD), on l'appelle aussi « l'effet d'ombre de la tête » car si un son vient de l'extrême droite de l'auditeur, la tête absorbera une partie des fréquences avant qu'il n'atteigne l'oreille gauche. L'intensité du son sera donc perçue différemment des deux oreilles ce qui donne une indication à notre cerveau de sa direction.<sup>35</sup>

Le placement du son sur l'axe des abscisses (x) correspond à sa **distance**. C'est-à-dire l'éloignement de la source sonore par rapport à l'auditeur. Les indications de perception de distance sont multiples mais la plus importante est l'amplitude du son par rapport à sa distance. Un son plus éloigné sera tout simplement entendu moins fort qu'un son proche. L'absorption atmosphérique est aussi à prendre en compte, notamment pour les sons entendus en extérieur. Elle aura tendance à atténuer les hautes fréquences pour un son lointain et lui faire perdre ses détails (on peut utiliser un simple filtre passe-bas pour simuler cet effet). Dans le cas d'un son entendu dans un espace réverbérant (comme une grotte par exemple), le rapport entre l'arrivée du signal direct et l'arrivée du signal réverbéré aux oreilles est une indication déterminante de la distance du son (c'est le « pre-delay » dans les plugins de réverbération). Accessoirement, si le son est en mouvement, l'intensité de l'effet doppler peut aussi donner une indication sur sa localisation.<sup>36</sup>

Enfin, le placement du son sur l'axe (z) correspond à l'angle vertical, le **zénith**. C'est-à-dire l'altitude à laquelle est émis le son par rapport à l'auditeur. Chez les humains, l'altitude du son est surtout perçue grâce à ses réflexions sur les pavillons des oreilles et sur les épaules. Suite à la réflexion du son, de courts retards temporels sont créés influant certaines fréquences du spectre quand le son arrive dans nos tympans (particulièrement les hautes fréquences). Notre cerveau détermine ensuite l'altitude de l'objet sonore à partir de ces informations complexes<sup>37</sup>. [fig.7]

---

<sup>35</sup> P. Fuchs, *Le traité de la réalité virtuelle Volume 1*, Presse des Mines, 2006, p73

<sup>36</sup> C. Roads, *op. cit.* 126 à 136

<sup>37</sup> *ibid.* p135

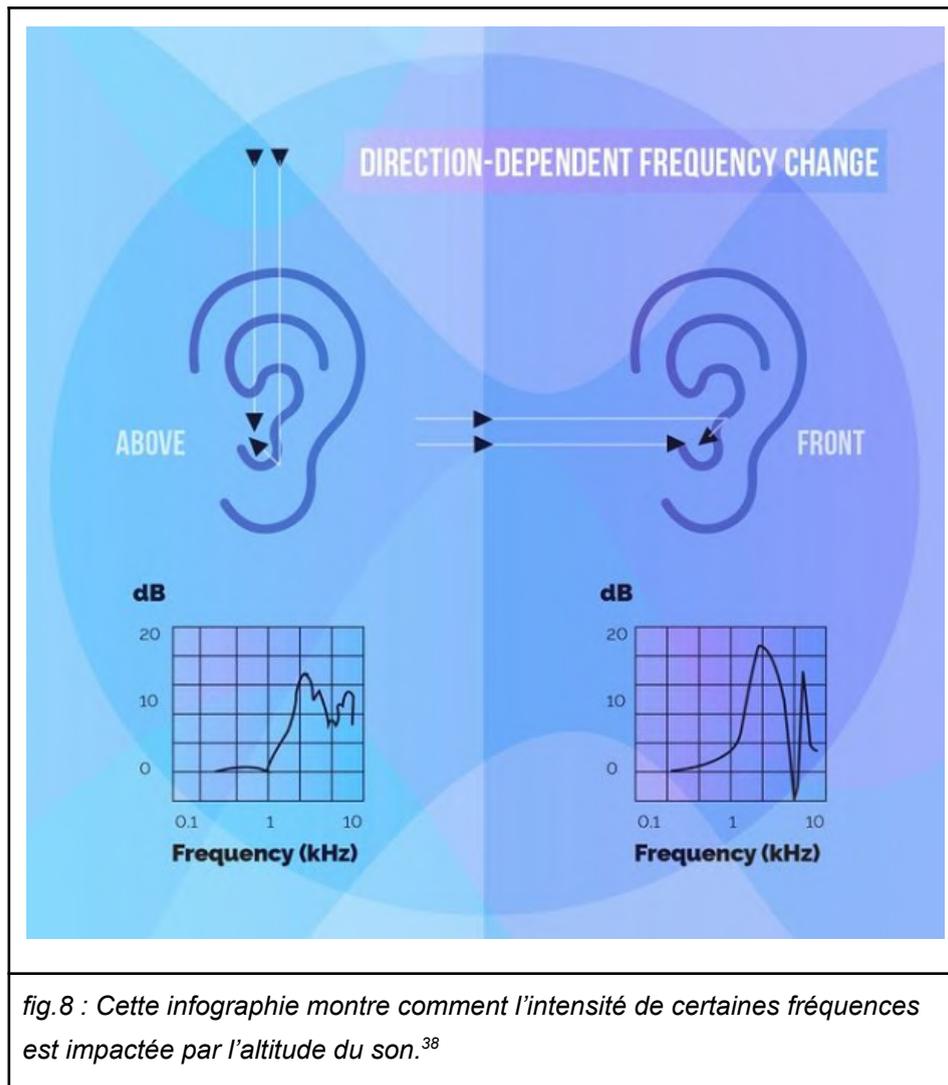


fig.8 : Cette infographie montre comment l'intensité de certaines fréquences est impactée par l'altitude du son.<sup>38</sup>

Avec cette dimension verticale, la restitution binaurale trouve une de ses utilités. Elle permet de simuler les réflexions sur la tête et les épaules de l'auditeur pour créer des indications zénithales. En appliquant des filtres en temps réel, elle rend aussi la spatialisation globale de l'azimut et de la distance plus précises. Pour régler ces filtres au travers desquels le binaural altèrera le son, il faut des données appelées HRTF pour *Head-Related Transfer Functions*.

Les HRTF se traduisent par « fonctions de transfert relatives à la tête ». Elles sont nécessaires au fonctionnement du son binaural car celui-ci vise à simuler une écoute humaine en prenant en compte les déformations du son causées par les caractéristiques physiques humaines (forme de la tête, pavillons des oreilles, réflexion des épaules). C'est pour cette raison qu'on ne peut faire l'expérience du son binaural qu'avec un casque ou des

<sup>38</sup> Source de l'image : Waves, *3D Audio on Headphones: How Does It Work?*. [en ligne] URL : [https://www.waves.com/3d-audio-on-headphones-how-does-it-work?utm\\_source=instagram&utm\\_medium=linkin.bio&utm\\_campaign=88-sale-2021-social&utm\\_term=10082021&utm\\_content=later-19630119](https://www.waves.com/3d-audio-on-headphones-how-does-it-work?utm_source=instagram&utm_medium=linkin.bio&utm_campaign=88-sale-2021-social&utm_term=10082021&utm_content=later-19630119) Publié le 23/06/2016, consulté le 11/08/2021.

écouters. Il faut que le son résultant du processus de filtrage arrive directement dans les tympans, sans être re-déformé par les caractéristiques physiques humaines comme l'est un son diffusé dans l'espace par des haut-parleurs. Les HRTF sont donc des données qui permettent de configurer les filtres qui seront appliqués sur le son pour une restitution binaurale. Le revers de la médaille pour une obtenir une restitution spatiale optimale en binaural, est qu'il faudrait (dans l'idéal) que chaque personne crée son propre profil de HRTF qui corresponde à son anatomie, car il n'y a pas deux oreilles formées pareilles ni de forme de tête globale. [fig.8]

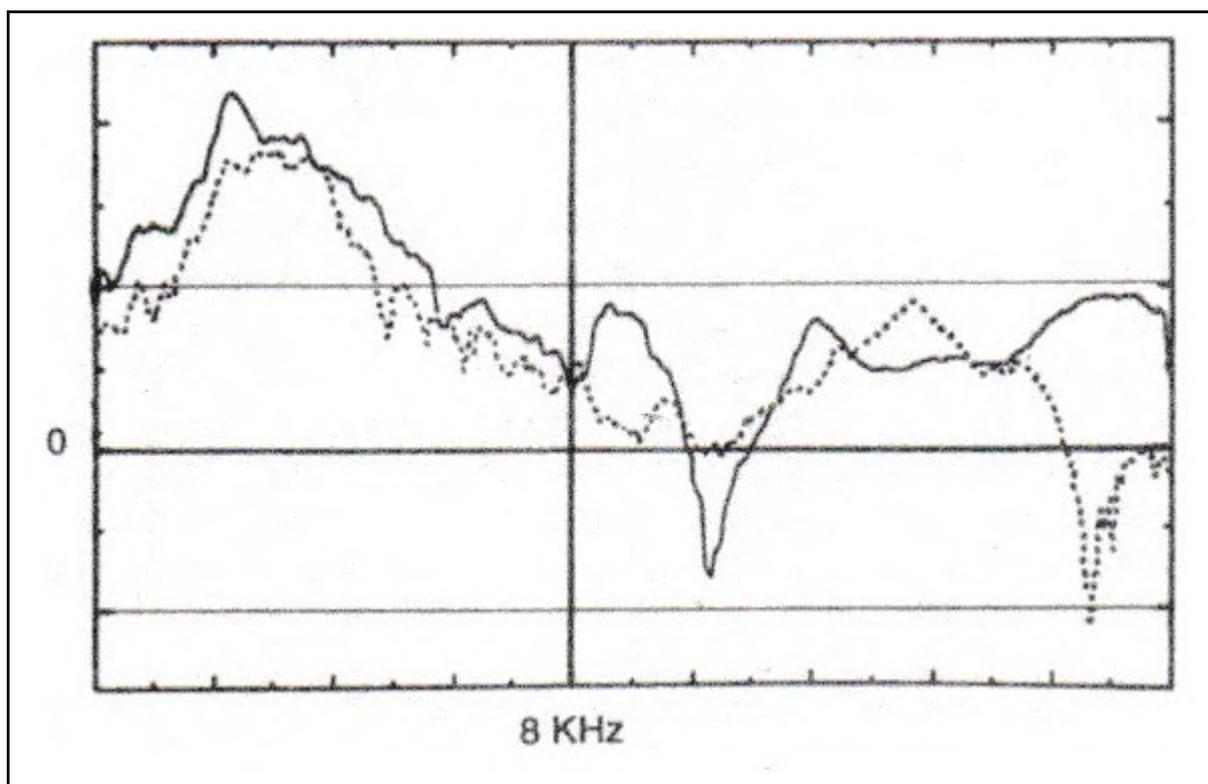


fig.9 : Spectre HRTF du même son pour deux personnes différentes. (Sujet 1 : ligne pleine, Sujet 2 : ligne pointillées). Les disparités dans les hautes fréquences à partir de 8 kHz sont dûes aux anatomies différentes des participants.<sup>39</sup>

Il est possible de créer son propre profil de HRTF mais c'est difficilement accessible car les HRTF doivent être mesurées en laboratoire dans une chambre anéchoïque. Si les HRTF utilisées par le système binaural sont trop loin des nôtres, le son en résultant sera brouillé, avec des erreurs de localisation.

Pour résoudre cette problématique de HRTF adaptées à chacun, des recherches sont actuellement en cours. Le département d'ingénierie électronique de l'Université de York en Angleterre a développé le projet SADIE pour *Spatial Audio for Domestic Interactive*

<sup>39</sup> Source de l'image : C. Road, *op. cit.* p136.

*Entertainment*. Ils ont réalisé une base de données répertoriant les HRTF de différents types de têtes, leurs résultats sont disponibles en libre accès sur leur site donc si la forme de tête d'un des sujets se rapproche de notre forme de tête, nous pouvons télécharger le fichier .SOFA contenant les HRTF correspondantes et l'appliquer à notre décodeur binaural.<sup>40</sup> Le fichier .SOFA pour *Spatially Oriented Format for Acoustics* est un format qui stocke toutes les données nécessaires à l'utilisation optimale d'un système binaural, comprenant des HRTF, des réponses impulsionnelles de salle et d'autres données spécifiques.<sup>41</sup> Au cours du projet SADIE, les chercheurs ont aussi tenté de trouver le meilleur compromis aux HRTF personnalisées. Ils ont déterminé, grâce à des tests sur des participants, que les fonctions de transfert de la tête Neumann modèle KU100 (un système d'enregistrement binaural reproduisant une forme de tête humaine) correspondaient en moyenne à la majorité des HRTF des participants. Google s'est intéressé au projet SADIE et a décidé d'adopter le profil du KU100 comme profil HRTF standard pour son contenu immersif et en réalité virtuelle. Donc lorsque nous faisons l'expérience de contenu en réalité virtuelle ou d'autres vidéos immersives à 360° sur Youtube, si le son est encodé en ambisonique de 1er ordre, il sera filtré selon les fonctions de transfert de la tête KU100.<sup>42</sup>

### 2.3 L'ambisonie, un système de spatialisation 3D bénéficiant du son binaural

Maintenant que le principe du son binaural est expliqué, j'aborde le système qui en tire le plus parti : l'ambisonie. Le système ambisonique a été théorisé par Michel Gerzon et Peter Fellgett en 1975. Voici la définition de la technologie ambisonique donnée par Peter Fellgett dans l'introduction de son article du magazine *Studio Sound* d'août 1975 :

« L'ambisonie est une technologie pour le son surround qui vise spécifiquement à *ne pas* rendre quatre (ou tout autre nombre de) haut-parleurs audibles en tant que sources sonores distinctes. Elle est conçue à l'aide de méthodes d'ingénierie appropriées et de la théorie psychoacoustique qui a montré une bonne valeur prédictive pour utiliser au mieux les canaux de communication disponibles (deux ou plus), et de diffusion sur haut-parleurs (une limitation souvent

---

<sup>40</sup> SADIE II, *Database*. [en ligne] URL : <https://www.york.ac.uk/sadie-project/database.html>. Publié le 13/05/2020, consulté le 10/08/2020.

<sup>41</sup> Audio Engineering Society, *Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions*. [en ligne] URL : [https://www.researchgate.net/publication/236634182\\_Spatially\\_Oriented\\_Format\\_for\\_Acoustics\\_A\\_Data\\_Exchange\\_Format\\_Representing\\_Head-Related\\_Transfer\\_Functions](https://www.researchgate.net/publication/236634182_Spatially_Oriented_Format_for_Acoustics_A_Data_Exchange_Format_Representing_Head-Related_Transfer_Functions). Publié le 7/05/2013, consulté le 21/07/2021.

<sup>42</sup> SADIE, *Spatial Audio filters for the Google YouTube 360/VR pipeline*. [en ligne] URL : <https://www.york.ac.uk/sadie-project/GoogleVRSADIE.html>. Publié en 2017, consulté le 10/08/2021.

oubliée), afin de créer des images acoustiques stables et non colorées dans n'importe quelle position, tout en en gardant les moyens physiques de reproduction du son les moins contraignants possible. Elle revendique une grande liberté quant aux méthodes d'enregistrement et au matériau source, ainsi que la protection du matériau enregistré contre l'obsolescence. »<sup>43</sup>

Ce système avait donc pour but d'améliorer les systèmes de diffusion existants (stéréo, 5.1, etc.) qui n'étaient capables de créer correctement une image fantôme qu'à la position du cône de vigilance. Ils ont élaboré le système « Ambisonics » à partir de calculs complexes d'harmoniques sphériques qui permettent de rendre un son immersif total à partir d'un nombre de canaux de diffusion réduit (à partir de quatre haut-parleurs seulement pour avoir les trois dimensions de l'espace).<sup>44</sup>

L'ambisonie constituée de 4 canaux est dite de « premier ordre », bien qu'elle soit omnidirectionnelle, elle offre une spatialisation moyennement précise. Plus il y a de canaux, plus la localisation des objets sonores sera précise : il y a donc l'ambisonie de « deuxième ordre » avec 9 canaux, de « troisième ordre » avec 16 canaux, etc. Les créateurs sonores qui utilisent ce système travaillent généralement avec un ordre d'ambisonie élevé pour avoir une haute définition de spatialisation dans l'espace virtuel du studio, au moment de la conception. L'ordre final dans lequel sera exporté le son dépendra du support de destination. Par exemple, le premier ordre avec 4 canaux est le standard des vidéos 360° sur Youtube.

Aujourd'hui, la plupart des navigateurs internet proposent la lecture de vidéos semi-interactives à 180° ou 360°. Ces vidéos qui se veulent immersives ont la particularité d'être diffusées en streaming, avec de l'audio et de la vidéo encodés. Contrairement au jeu vidéo, il n'y a donc pas de moteur de jeu qui simule un espace 3D dans lequel le son est spatialisé en temps réel. Il faut alors un moyen de simuler la localisation d'objets sonores dès la post-production audio de ces vidéos immersives, c'est dans cette situation que le sound designer peut choisir d'utiliser le format ambisonique. Le mixage ambisonique consiste en un mixage d'objets sonores dans un espace virtuel. Tous les DAW sont maintenant équipés de plugins qui permettent de spatialiser le son pour l'ambisonique.  
[fig.10]

---

<sup>43</sup> P. Fellgett, *Ambisonics. Part one: general system description*, Studio Sound, Vol. 17 no. 8, p20-22, Août 1975. [en ligne] URL : <https://worldradiohistory.com/Archive-All-Audio/Archive-Studio-Sound/70s/Studio-Sound-1975-08.pdf> consulté le 23/08/2021.

<sup>44</sup> C. Roads, *op. cit.* p147



*fig.10 : VST Multipanner, c'est l'outil de spatialisation de Cubase pour travailler en ambisonique. On choisit directement où placer le son dans une sphère virtuelle de taille variable avec l'auditeur au centre.*

Il est donc possible de faire passer le son au-dessus ou en-dessous de l'auditeur, devant ou derrière sa tête. La spatialisation réalisée dans la sphère autour de l'auditeur peut ensuite être restituée par des haut-parleurs, dans l'idéal à partir d'un ensemble de haut-parleurs qui entourent l'auditeur comme on peut en trouver à l'IRCAM dans l'espace de projection « Espro ». Des conversions peuvent aussi être effectuées pour diffuser le mix ambisonique sur un système surround classique type 5.1 ou même un système stéréo (cependant, l'utilité de mixer en ambisonique pour diffuser sur un système stéréo est discutable). Autrement, le rendu d'un mix ambisonique le plus abordable pour le grand public reste l'écoute dans un casque audio. Dans ce cas, l'utilisation du binaural avec des HRTF bien configurées prend tout son sens car les trois dimensions de l'espace sonore peuvent être reproduites fidèlement dans un casque avec deux oreillettes seulement.

La distinction principale entre l'ambisonie et le binaural est donc que la première est un système qui propose une image sonore du son pouvant être diffusée sur n'importe quel système audio, alors que le deuxième est l'un des modes de restitution du premier, qui vise à reproduire l'audition humaine et qui requiert obligatoirement un casque.

Avec l'accessibilité des technologies ambisoniques et binaurales dans tous les DAW, le rendu spatial haute fidélité commence à se démocratiser. Les vidéos d'ASMR avec du son binaural ou les musiques dites « 8D » qui apparaissent sur Youtube en sont la preuve. Il s'agit généralement de musiques stéréo traditionnelles, qui sont passées dans un plugin de spatialisation ambisonique puis réexportées en stéréo avec l'application de filtres binauraux.<sup>45</sup> L'effet est spectaculaire mais reste selon moi anecdotique car aucune interactivité n'est induite. D'autres expériences plus abouties sont disponibles en ligne, par exemple Dolby Atmos a adapté son système de spatialisation en un format binaural pour les casques audio, une démonstration est disponible sur leur site.<sup>46</sup> Des plateformes de streaming musical (Deezer, Tidal, Amazon Prime Music) commencent aussi à proposer des musiques spatialisées qui peuvent être écoutées au casque en binaural ou avec un système de diffusion multicanal. Les principaux formats utilisés pour la musique spatialisée en streaming sont le Dolby Atmos ou le Sony 360, ils utilisent les mêmes principes que le système ambisonique.<sup>47</sup> Aujourd'hui, le son spatialisé haute fidélité est même accessible directement dans les PC Windows [fig.11], ce qui améliore la spatialisation des applications qui supportent ce traitement.

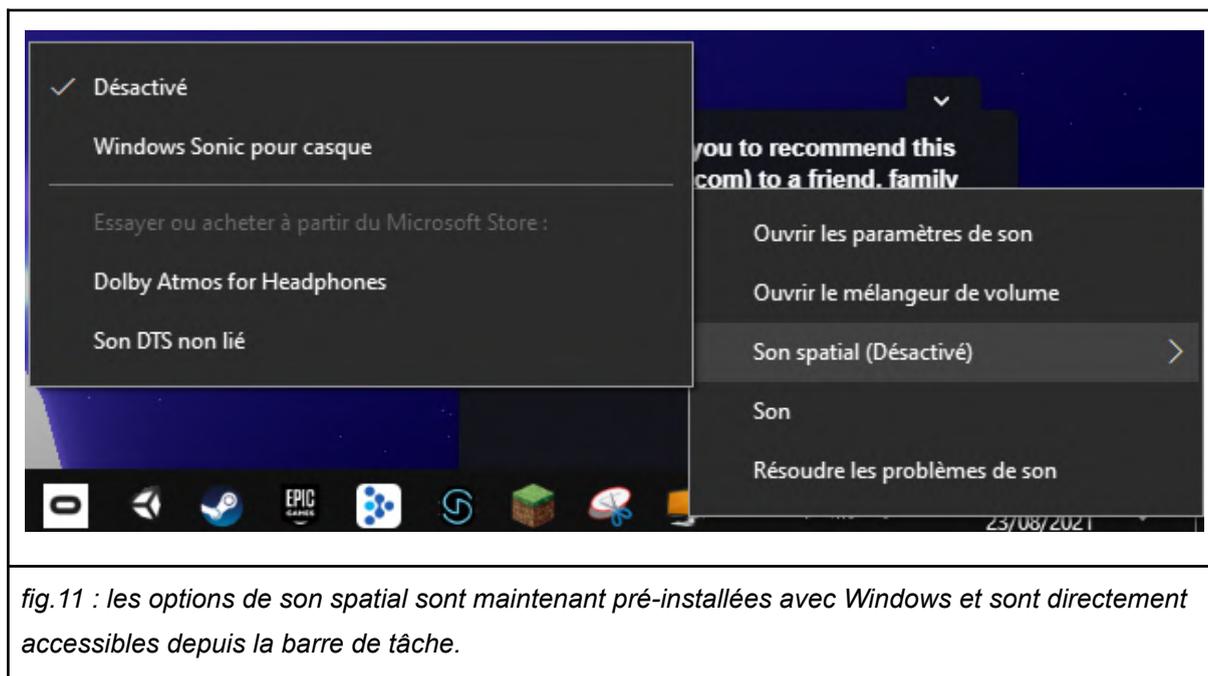


fig.11 : les options de son spatial sont maintenant pré-installées avec Windows et sont directement accessibles depuis la barre de tâche.

Certains jeux vidéo tirent aussi parti de ces technologies pour proposer des esthétiques fortes et uniques. C'est par exemple le cas du jeu paru en 2017 *Hellblade - Senua's Sacrifice* développé par Ninja Theory. On y incarne Senua, souffrant de

<sup>45</sup> Chaîne Youtube 8D TUNES : <https://www.youtube.com/channel/UCrRpYEytIHGyDgNWO6VbHIQ>

<sup>46</sup> Expérience Dolby Atmos pour casque audio : <https://www.dolby.com/atmos-visualizer-music/>

<sup>47</sup> Tall Guy Film, *What is Virtual Surround? - Spatial audio explained*. [en ligne] URL : <https://www.youtube.com/watch?v=E2JneaC6lbl>. Publié le 23/06/2020, consulté le 22/07/2021.

schizophrénie, allant sauver son amant dans les enfers de la mythologie nordique. Le son est rendu de manière binaurale avec comme but de brouiller la limite entre la réalité et la fiction. En nous faisant ressentir ses troubles psychotiques au travers du son binaural, le jeu nous met dans la peau de Senua. Les voix qui tournent constamment dans sa tête et qui parfois sont tellement proches qu'elles chuchotent à son oreille, sont restituées avec précision dans le casque audio des joueurs. Les sons internes subjectifs du personnage deviennent donc les sons internes-subjectifs des joueurs.<sup>48</sup> Une utilisation du dispositif binaural peut donc apporter grandement à la cohérence d'une œuvre, en offrant un degré de précision dans la spatialisation qui peut être utilisé pour soutenir un propos artistique.

## ***2.4 Utilisation du système ambisonique et du son binaural : retour sur mes travaux personnels***

Au cours de mon stage de Master 2, j'ai eu l'occasion d'expérimenter le mixage en ambisonique avec rendu binaural. J'ai fait certaines erreurs qui m'ont aidé à mieux comprendre cette manière de créer du son. Par exemple, je me suis lancé dans la création d'une ambiance ambisonique avec rendu binaural pour le jeu vidéo *Behave* d'E-Art Sup Lyon, or je n'avais pas accès à l'intergiciel pour intégrer le son dans le jeu (c'est un développeur de l'équipe qui s'en occupait). Je l'ai donc faite directement dans mon DAW à la manière d'un son pour un film, avec une pensée à partir d'un point fixe. Le résultat n'était pas convaincant car mon erreur est d'avoir omis la composante interactive du jeu vidéo : les déplacements du personnage dans l'espace. En créant l'ambiance dans mon DAW puis en l'exportant dans un fichier stéréo classique, je l'ai figée. Une fois le son intégré dans le jeu, le rendu binaural n'était donc pas satisfaisant sauf à une seule position à l'origine du son. Dès que le personnage se déplaçait, la localisation des éléments de mon ambiance était brouillée. Avec les éléments cités en amont dans ce mémoire on peut comprendre mon erreur. La meilleure solution pour créer une ambiance binaurale aurait été que j'exporte chaque élément de mon ambiance (goutte d'eau qui tombe sur le sol, grille de ventilation, etc.) pour que le sound designer les intègre dans l'espace 3D du jeu au travers de l'intergiciel. De cette manière la localisation des sons aurait été faite en temps réel selon les déplacements du personnage. Il aurait ensuite fallu appliquer les filtres binauraux aux canaux de diffusion de l'intergiciel avec plugins comme Resonance Audio de Google (utilisant les HRTF de *SADIE*) ou Steam Audio de Valve évoqué plus haut, qui laisse à l'utilisateur le choix des HRTF. J'aurai aussi pu livrer l'ambiance directement avec le format

---

<sup>48</sup> ExServ, Hugo, Maxime. Fin Du Game, *Episode 45 - Hellblade: Senua's Sacrifice*. [en ligne] URL : <https://fin-du-game.lepodcast.fr/episode-45-hellblade-senuas-sacrifice>, Publié le 20/11/2020, consulté le 04/08/2021.

ambisonique sans l'encoder en stéréo car les logiciels sont maintenant capables de lire ce type de fichier multicanaux grâce aux plugins précédemment cités.

J'ai aussi mixé une musique en ambisonique pour le court métrage *Kayak* de l'ENSI car je souhaitais avoir un mix avec un rendu réaliste. Il n'y avait donc pas d'enjeu interactif puisque la destination était un film d'animation traditionnel pour le cinéma. Grâce à l'ambisonie, j'ai pu placer mes instruments dans un espace 3D virtuel restituant une spatialisation réaliste, avec des instruments plus éloignés que d'autres par exemple. Le rendu était convaincant car le style musical était folk et acoustique, mais ce réalisme spatial n'est pas souhaitable pour tous les genres musicaux car nous sommes habitués à entendre des sons transformés par des effets qui augmentent les basses ou d'autres fréquences qui sont moins présentes en situation réelle.

J'utilise aussi parfois l'ambisonie pour spatialiser des éléments précis dans des courts métrages. Par exemple, dans le film *La Récolte* de l'ECV Bordeaux, j'ai utilisé la spatialisation ambisonique pour placer une voix au loin. Celle-ci n'était pas très bien enregistrée (l'acteur était trop proche du micro) et il n'avait pas la possibilité de réenregistrer le son, la spatialisation ambisonique a un peu gommé cet effet.

Pour les deux exemples précédents, je n'ai pas utilisé de filtres binauraux car le son des films est destiné à être diffusé par des haut-parleurs. Comme expliqué en amont, l'ambisonie calcule la localisation des sons dans une sphère virtuelle, le rendu binaural n'est pas obligatoire. Dans Cubase, il est possible de choisir comment est diffusé le son ambisonique : soit en binaural sur un casque (avec choix des HRTF : Standard, Facebook, Youtube ou HRTF personnalisés), soit sur des enceintes stéréo ou surround classiques avec le mode « speaker » [fig.12].

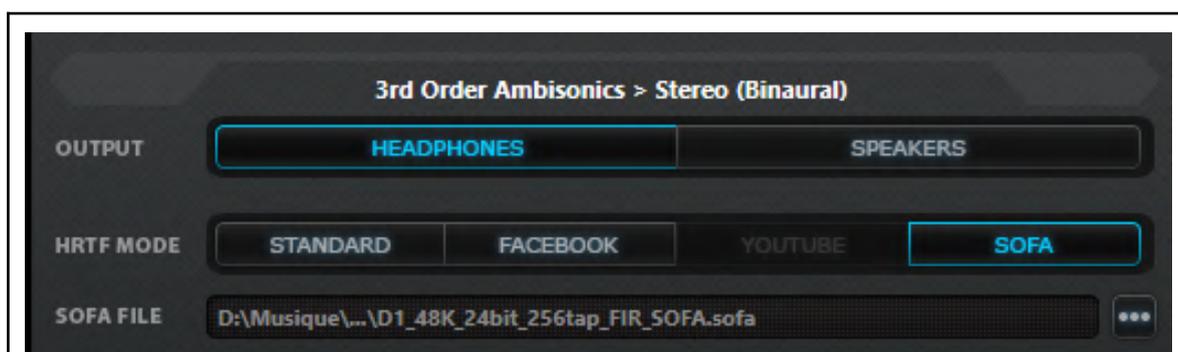


fig.12 : Les différents modes de diffusion quand on travaille en ambisonie dans Cubase. Ici, les HRTF Youtube ne sont pas disponibles car nous sommes en ambisonie de 3e ordre.

Pour achever notre exploration de l'ambisonie et du binaural pour ce mémoire, j'ai créé une ambiance ambisonique avec restitution binaurale disponible sur ma

chaîne Youtube.<sup>49</sup> Pour en faire l'expérience, il faut se munir d'un casque audio ou d'écouteurs. Il s'agit d'une ambisonie de premier ordre, il devrait être possible de repérer des éléments plus ou moins éloignés dans le champ sonore : un chien au loin devant à gauche, des mouches qui virevoltent devant à droite, le cri d'un oiseau derrière à gauche et des bruits d'eau contre le bateau et le ponton tout proche en dessous (le tout par rapport à la direction de référence du début de la vidéo). Une fois ces éléments entendus, il est possible de changer de point de vue (et donc de point d'écoute) en cliquant directement sur la vidéo pour faire défiler l'image à 360°. C'est une des fonctionnalités du lecteur de vidéo 360° Youtube. Si la vidéo est regardée avec un dispositif équipé d'un accéléromètre comme un smartphone, l'orientation du point de vue pourra être contrôlée par l'orientation du smartphone. Pour aller plus loin, j'ai aussi fait une vidéo qui passe en revue les éléments constitutifs de cette ambiance et la manière dont ils sont spatialisés.<sup>50</sup> À partir de 02:07' dans cette deuxième vidéo, je montre l'utilisation d'un casque de réalité virtuelle pour contrôler le point d'écoute. C'est ce qu'on appelle le « Head Tracking » pour « suivi de la tête » en français.

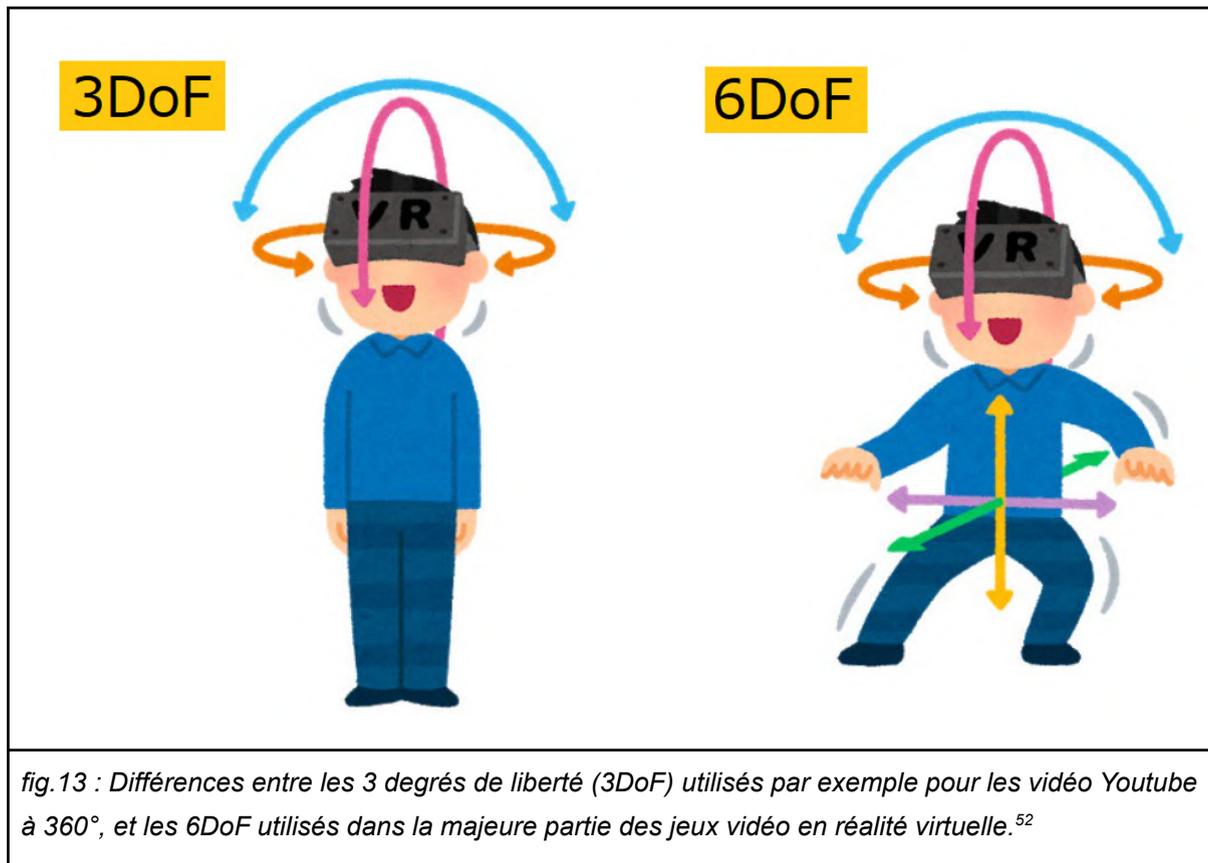
Nous avons donc vu que le format ambisonique est intéressant dans les cas où le point d'écoute de l'auditeur est connu à l'avance. Mais le fait qu'il n'y ait pas de moteur de jeu qui calcule la position des éléments sonores en temps réel bloque toujours l'auditeur à un point fixe dans lequel il ne peut que tourner la tête, c'est ce qu'on appelle le 3DoF (3 *Degrees of Freedom* pour 3 degrés de liberté). On ne considère dans ce dernier que la rotation horizontale, verticale et le roulis de la tête. La réalité virtuelle élargit les possibilités de déplacements des utilisateurs en proposant généralement 6 degrés de libertés (6DoF), c'est-à-dire qu'en plus des déplacements de la tête, la translation du corps de l'utilisateur dans un espace en trois dimensions est prise en compte [fig.13]. Plus il y a de degrés de libertés dans une expérience interactive, plus l'immersion de l'utilisateur est avancée.<sup>51</sup>

---

<sup>49</sup> V. Petiteau, *Ambiance binaurale avec Cubase - 360°* <https://www.youtube.com/watch?v=JP1m25kb2LE>

<sup>50</sup> V. Petiteau, *Ambiance binaurale avec Cubase*, <https://www.youtube.com/watch?v=kBAQIa9fwtg>

<sup>51</sup> P. Fuchs, *Le traité de la réalité virtuelle Volume 1*. Presse des Mines, 2006, p137



<sup>52</sup>Source de l'image : Blog de Pupuru. *Qu'est-ce que le DoF (degré de liberté) de la RV ? Explication de 3DoF et 6DoF.* [en ligne] URL : <https://pupuru-blog.com/wp-content/uploads/2018/12/d.png>. Publié le 26/09/2019, consulté le 26/08/2021.

## III - La spatialisation du son en réalité virtuelle : une immersion nécessaire

### 3.1 Les enjeux de la création sonore pour la réalité virtuelle

Puisque nous avons exposé les principes de l'ambisonie et du son binaural, nous allons maintenant présenter un média qui a beaucoup à gagner en utilisant ces technologies : la réalité virtuelle.

Ayant toujours été intéressé par les nouvelles technologies et les œuvres audiovisuelles, ma première expérience d'une application en RV est un souvenir marquant. Le dispositif que j'avais utilisé était assez rudimentaire comparé à ceux disponibles aujourd'hui, il s'agissait simplement d'un téléphone placé dans un support avec deux lentilles qui permettent d'obtenir une image stéréoscopique (avec de la profondeur). [fig.13]



fig.13 : Casque de RV basique utilisant un téléphone comme écran.<sup>53</sup>

La résolution de l'écran du téléphone était très basse, l'accéléromètre qui contrôle les mouvements de la tête était approximatif et offrait seulement 3DoF. L'application *Google Daydream* n'était pas optimale, mais l'expérience était tout de même sensationnelle. C'est entre autres grâce à l'immersion bluffante offerte d'emblée par les images en relief que cette technologie a un haut potentiel d'émerveillement.

<sup>53</sup> Source de l'image : Alfonso De Tomas, [en ligne] URL : [https://fr.123rf.com/photo\\_53175942\\_front-view-of-an-open-virtual-reality-headset-with-a-smart-phon-e-ready-for-vr-experience-lenses-are.html](https://fr.123rf.com/photo_53175942_front-view-of-an-open-virtual-reality-headset-with-a-smart-phon-e-ready-for-vr-experience-lenses-are.html). Consulté le 25/08/2021.

La réalité virtuelle a bénéficié de l'explosion du développement des smartphones depuis les années 2010, notamment grâce aux écrans haute définition de moins en moins cher à produire. Ce développement a permis aux constructeurs de proposer des casques de RV aux résolutions acceptables à des prix décents, ouvrant l'accès à cette technologie à un plus grand public.<sup>54</sup>

L'objectif principal de la RV est d'immerger les utilisateurs dans ses applications. Philippe Fuchs, professeur aux Mines - Paristech et spécialiste de la RV en France, propose de définir trois niveaux d'immersion et d'interaction en RV : les sensori-motrices, les cognitives et les fonctionnelles.<sup>55</sup> La prise en compte de chacun de ces niveaux donne des clés intéressantes pour les créateurs qui veulent proposer des expériences en réalité virtuelle, mais ils peuvent paraître assez hermétiques pour ceux qui ne voudraient pas s'attarder sur des considérations scientifiques. Pourtant il me semble que c'est nécessaire pour comprendre les enjeux du média. Les chercheurs français font d'ailleurs tout pour mettre à disposition au plus grand nombre le fruit de leurs recherches pour améliorer le domaine. C'est la démarche des membres de l'AFXR (*Acteurs Français de la XR*, « XR » dans le sens de « Extended Reality » ou « Réalité Étendue », qui englobe la RV et d'autres systèmes immersifs). On peut ainsi trouver sur le site de l'AFXR leur manifeste ainsi que des textes enrichissant sur la réalité virtuelle comme *Le Traité de la Réalité Virtuelle* en 5 volumes, auquel je me réfère beaucoup dans ce mémoire.<sup>56</sup> Je suis convaincu que la définition de l'immersion en RV profiterait d'un autre niveau qui désignerait le « degré d'émerveillement ». C'est le sujet de la thèse de Léa Dedola, doctorante CIFRE à l'Université Lumière Lyon II, elle explore le niveau d'immersion émotionnel, qui pourrait peut être toucher davantage les créateurs d'expériences en RV.<sup>57</sup>

Les applications de la RV sont multiples, pour le grand public elle se cantonne majoritairement au jeu vidéo pour l'instant, mais elle est utilisée depuis longtemps dans les domaines de la recherche, de la formation et de la médecine. Le cinéma commence lui aussi à s'approprier le média, avec de plus en plus d'œuvres en RV qui sont remarquées et qui gagnent des prix au sein de festivals prestigieux (Cannes, Annecy, *Sundance*, *Tribeca*, etc.). Avec l'arrivée sur le marché de casque de RV vendus au prix d'une console de jeu classique, les particuliers sont de plus en plus équipés pour faire l'expérience d'applications et de films dédiés à ce support. Des salles d'arcade spécialisées en réalité virtuelle apparaissent aussi un peu partout en France. Cet engouement encourage donc les grands

---

<sup>54</sup> P. Fuchs, *Théorie de la réalité virtuelle : les véritables usages*, Paris : Mines Paristech, 2018, p18-19

<sup>55</sup> P. Fuchs, *Le traité de la réalité virtuelle Volume 2*. Presse des Mines, 2006, p32

<sup>56</sup> AFXR, *Publications Clés*, <https://www.afxr.org/page/863801-publications-cles>

<sup>57</sup> L. Dedola, *Tête-à-tête avec Léa Dedola*, site de l'Université Lyon 2, <https://www.univ-lyon2.fr/recherche/magazine/numero-1/tete-a-tete-avec-lea-dedola>, 29/01/2021.

studios de création d'œuvres vidéoludiques et cinématographiques à s'intéresser sérieusement à ce média car le domaine devient viable.<sup>58</sup>

Comme je l'expliquais plus haut, une des raisons d'utiliser la RV est la qualité d'immersion qu'elle offre. Le son est au centre de cet enjeu immersif, au même niveau que la fidélité visuelle. La spatialisation, qui nous intéresse dans ce mémoire, fait partie des caractéristiques du son décisives pour immerger l'utilisateur dans une expérience en RV. Pour Philippe Fuchs :

« [La spatialisation du son] C'est l'un des composants les plus importants dans la production de son pour la réalité virtuelle. On peut la comparer à la production d'images en relief pour le canal visuel. [...] La référence reste la stéréo binaurale qui permet de reproduire à l'aide d'un simple casque stéréo et d'un filtrage spécifique des signaux, la direction de provenance d'un son synthétisé ou enregistré. Ce système permet de reproduire aussi bien des sons provenant de l'avant que de l'arrière, du dessus ou du dessous de l'auditeur. Néanmoins, la qualité du résultat est très dépendante de l'auditeur puisque les filtres à appliquer sont fortement dépendants de la morphologie de la personne. »<sup>59</sup>

P. Fuchs affirme donc que le son stéréo binaural que nous avons évoqué en amont est la solution idéale pour réaliser une spatialisation du son en RV, c'est aussi la solution la plus abordable car elle ne demande qu'un casque stéréo. Pour créer une spatialisation idéale sans l'aide d'un casque audio, il faudrait faire l'expérience d'applications en RV au milieu d'une sphère de haut-parleurs qui restituerait le son dans l'espace pour simuler l'audition humaine omnidirectionnelle. Cette solution est bien sûr presque impossible sans se trouver dans un laboratoire de recherche acoustique.

### ***3.2 Une nouvelle manière de spatialiser le son pour un film : l'exemple de BattleScar***

Le cinéma en RV a un statut ambigu car il se trouve entre le jeu vidéo et le cinéma traditionnel. En effet, dans le cinéma traditionnel le point de vue est toujours fixé par les images, mais en RV la composante interactive doit être présente. On ne peut pas fixer les images sans prendre en compte les mouvements du spectateur car la diffusion sur ce support n'aurait presque plus aucun intérêt par rapport à un support conventionnel.

---

<sup>58</sup> Serge R., *Tendances VR/AR 2021 : À quoi faut-il s'attendre cette année ?*. [en ligne] URL : <https://www.realite-virtuelle.com/tendances-vr-ar-2021/>, Publié en janvier 2021, consulté le 25/08/2021

<sup>59</sup> P. Fuchs, *Le traité de la réalité virtuelle - Volume 3*, Presse des Mines, 2006, p94-95

Il faut donc que les réalisateurs arrivent à tirer parti de la singularité du visionnage d'un film en réalité virtuelle. Selon P. Fuchs, les personnes qui réalisent des films doivent se détacher du réflexe de vouloir simplement « raconter une histoire », il faut plutôt qu'elles se dirigent vers l'idée de « faire vivre une histoire ». Un film en RV réussi doit être une véritable expérience dans un environnement virtuel.<sup>60</sup>

Pour ce mémoire j'ai décidé d'étudier un film en réalité virtuelle sélectionné et primé en festivals. Je m'intéresserai surtout au son et sa spatialisation dans le but d'identifier quelques bonnes pratiques dans ce domaine.

## *BattleScar*

*BattleScar - Punk was invented by girls* est un film d'animation en RV réalisé par Nico Casavecchia et Martin Allais, avec Mercedes Arturo à la direction artistique. Il a été achevé en 2018 puis a fait partie de beaucoup de sélections officielles de festivals comme la Mostra de Venise, Sundance, Tribeca, Annecy, etc. Il a remporté quelques prix et est disponible pour le public sur différentes plateformes depuis le 14 janvier 2021. *BattleScar* est produit par le studio Atlas V et a été en partie développé à Lyon, coproduit par Arte France entre autres, il a bénéficié des financements du CNC et du fonds régional Auvergne-Rhône-Alpes Cinéma.

Voici le synopsis du film disponible dans le kit de presse officiel :

« *BattleScar* suit une année de la vie de Lupe, une Portoricaine-Américaine vivant à la fin des années 1970 à New York. Le film s'appuie sur le journal intime manuscrit de Lupe pour nous guider dans ses expériences au travers l'année 1978, alors que [dans sa fugue] elle rencontre Debbie, une autre jeune fugueuse vivant dans la ville. Debbie fera découvrir à Lupe la scène punk de la *Bowery* et l'exposera aux mondes secrets qui coexistent dans le *Lower East Side* à la fin des années 70's. *BattleScar* est un drame sur le passage à l'âge adulte qui explore le thème de l'identité au travers l'utilisation de l'animation et des environnements immersifs en réalité virtuelle. »<sup>61</sup>

Les mondes secrets évoqués dans le synopsis correspondent à des lieux de la culture punk underground de New York. Le film est découpé en trois chapitres pour une durée totale de

---

<sup>60</sup> P. Fuchs, *Réalité virtuelle : une révolution scientifique*, La Méthode scientifique, France Culture. De 30:53' à 31:05'. [en ligne] URL : <https://www.franceculture.fr/emissions/la-methode-scientifique/realite-virtuelle-une-revolution-scientifique>. Publié le 19/10/2016, consulté le 28/04/2021.

<sup>61</sup> *BattleScar* press kit, *Extended synopsis*. [en ligne] URL : [https://static1.squarespace.com/static/5a2680f5914e6b6824a7254f/t/5a2aa73624a694e5fd262125/1512744789541/BATTLESCAR\\_Sundance\\_V8-1.pdf](https://static1.squarespace.com/static/5a2680f5914e6b6824a7254f/t/5a2aa73624a694e5fd262125/1512744789541/BATTLESCAR_Sundance_V8-1.pdf), consulté le 04/08/2021.

32:36'. La musique originale et le sound design ont été créés par le studio *Antfood* basé à New-York. C'est un studio déjà expérimenté dans le son pour la réalité virtuelle car ils ont fait le son de l'application *Google Earth VR* avec des ambiances sonores et des musiques spatialisées en ambisonie de 3<sup>e</sup> ordre.<sup>62</sup>

Le court métrage est disponible dans deux versions différentes : une version en streaming sur Youtube (lue avec le lecteur 360°) et une version « standalone » (une application qui se suffit à elle-même pour être lue, à la manière d'un jeu vidéo). Malgré le fait que la version streaming soit accessible gratuitement sur la chaîne Youtube d'ARTE<sup>63</sup>, je ne la recommande pas car il y a des différences audio et vidéo importantes par rapport à la version standalone. Cette dernière est disponible sur différentes plateformes comme l'Oculus Store<sup>64</sup> ou Steam pour 5€.

La version streaming est beaucoup moins immersive car la vidéo n'est pas stéréoscopique, c'est-à-dire que même dans un casque de RV l'image reste plate. De plus, le *head tracking* (suivi des mouvements de la tête avec un casque de RV) s'effectue seulement sur 3DoF contre 6DoF pour la version standalone. Et enfin, le son n'est pas adaptatif du tout en streaming, il est mis à plat en stéréo classique et entendu en mode *headlocked*, c'est-à-dire que les mouvements de la tête ne sont pas pris en compte pour l'audio, seul le point de vue change mais pas le point d'écoute. L'interactivité sonore est donc inexistante et cela brise l'immersion, ce qui aurait pu être corrigé car, comme nous l'avons constaté avec mon exemple d'ambiance binaurale, Youtube peut prendre en charge les fichiers ambisoniques de 1<sup>er</sup> ordre. Nous pouvons donc nous demander comment le studio *Antfood* a conçu la spatialisation sonore : ont-ils travaillé en ambisonique à la manière de mon exemple d'ambiance binaurale ou ont-ils utilisé un intergiciel avec une spatialisation en temps réel à la manière d'un jeu vidéo ? J'ai contacté le studio *Antfood* mais je n'ai pas eu de réponse à ce jour. Je ne peux donc que faire des suppositions mais le fait que la version streaming n'ait pas de son interactif me laisse penser qu'un intergiciel a été utilisé.

Je me concentrerai donc sur la version standalone, bien plus immersive sur les plans sonores et visuels. À la présentation du film à la Mostra de Venise, une installation avait été mise en place pour mettre les spectateurs dans l'ambiance du film [fig.14].

---

<sup>62</sup> Antfood, *Google Earth VR*. [en ligne] URL : <https://www.antfood.com/work/emerging-tech/google-earth-vr>. Consulté le 20/08/2021.

<sup>63</sup> ARTE, *Punk was invented by girls | BattleScar | Vidéo 360°*. [en ligne] URL : <https://www.youtube.com/watch?v=KcGIT9wTbbs>. Publié le 14/01/2021, consulté le 04/08/2021.

<sup>64</sup> Oculus Store, *BattleScar* [https://www.oculus.com/experiences/quest/2985592331490872/?locale=fr\\_FR](https://www.oculus.com/experiences/quest/2985592331490872/?locale=fr_FR)

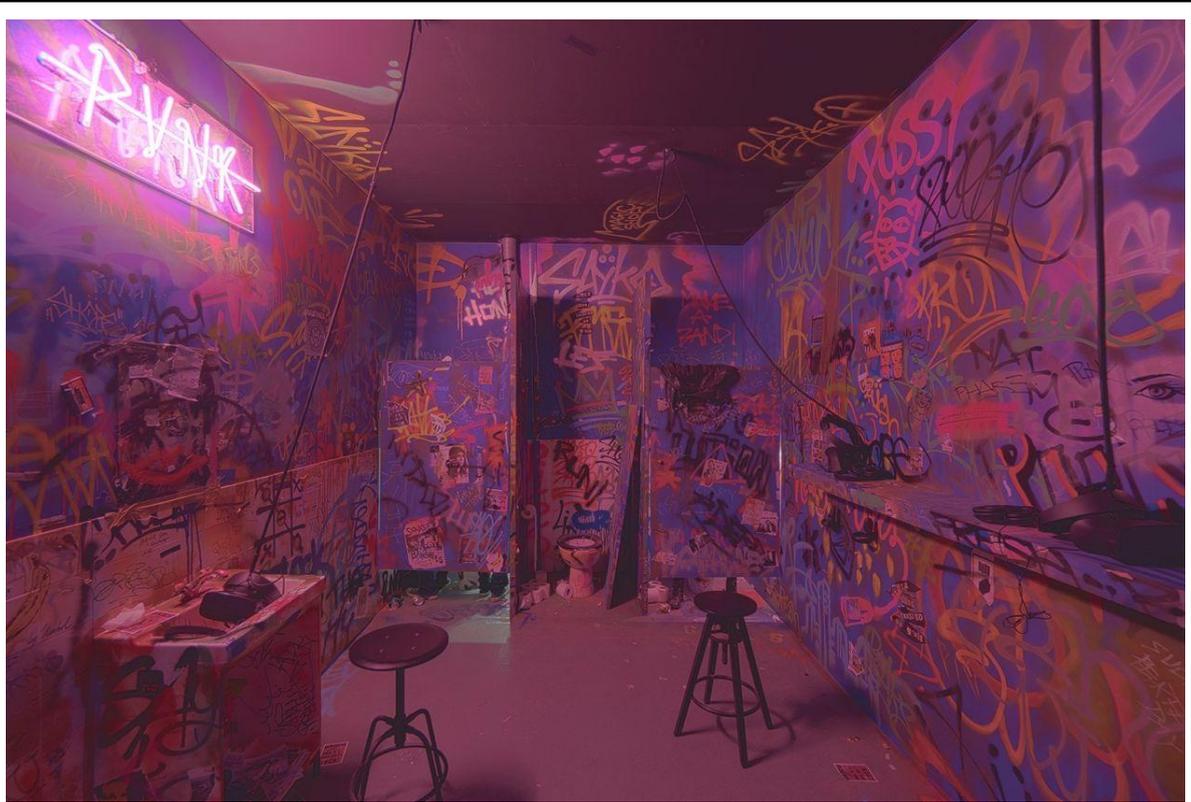


fig.14 : L'installation BattleScar à la Mostra de Venise par le Studio PHI qui a recréé les WC du club punk new-yorkais CBGB.<sup>65</sup> Des casques de RV pour regarder le film sont disposés sur la tablette à droite et dans le lavabo à gauche

Cette excellente idée permettait de lier le monde réel au monde virtuel dans lequel les spectateurs étaient projetés une fois le casque de RV sur la tête.

Dans le film, la plupart des éléments sonores sont spatialisés, mais la qualité de la spatialisation est assez inégale. Commençons par l'aspect le plus réussi du son spatialisé : les ambiances. Leur précision est un atout pour le film car elles permettent au spectateur de s'immerger dans l'espace où est vécue l'histoire. Chaque espace (bar, extérieur, vieil appartement de squat, prison, etc.) est très bien caractérisé. Le passage dans le squat à partir de 3:55' par exemple est très efficace [fig.15], il y a un *roomtone* constant qui caractérise la pièce, auquel s'ajoutent des éléments ponctuels spatialisés comme des bruits de moteurs et de sirènes de police à l'extérieur qui ancrent l'appartement dans un espace urbain malfamé. Puisque le film est en RV, le spectateur peut tourner la tête et diriger son écoute vers ces sons, ce qui améliore beaucoup l'immersion globale (et ce qu'il manque dans la version streaming).

<sup>65</sup> Source de l'image : Frédéric Segard. [en ligne] URL : <https://phi.ca/fr/studio/projets/phi-a-la-mostra-de-venise/>. Consulté le 04/08/2021..

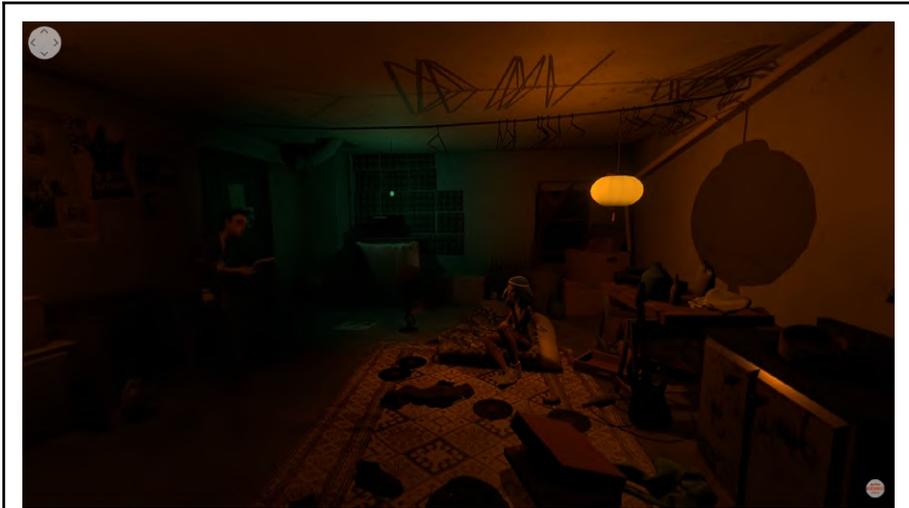


fig.15 : l'appartement squat à l'ambiance sonore réussie. 4:20'

À partir de 3:30' un autre plan utilise intelligemment la spatialisation d'un bruit d'ambiance [fig.16]. Dans cette scène, un son d'ambiance urbaine est aussi entendu mais il n'est pas diffusé comme celui de la scène mentionnée précédemment. Au lieu d'être un son global qui entoure le spectateur, le son d'ambiance est émis depuis un point fixe : la carte postale « New-York » collée dans le journal intime de Lupe. Le son de la ville sort donc du carnet et si le spectateur se rapproche de la carte postale, il entend le son de plus en plus fort. Cette idée est très intéressante car elle oriente le regard du spectateur. De plus, elle permet d'économiser beaucoup de moyens car une simple image sonore en cohésion avec l'image fixe de la carte postale permet au spectateur de se plonger dans l'ambiance du passage raconté.

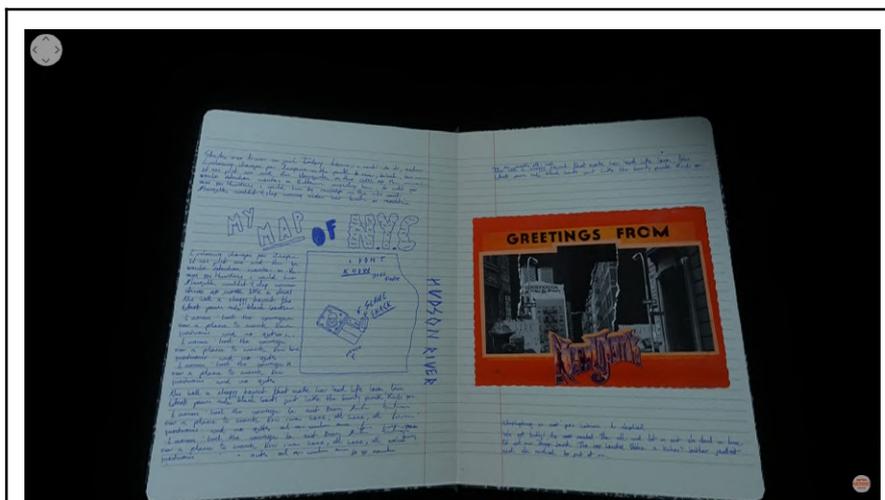
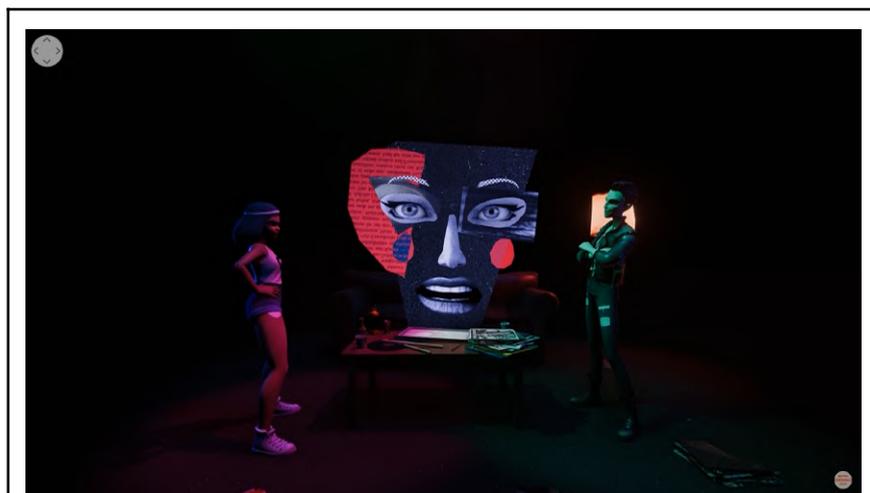


fig.16 : Dans ce plan à 3:30', l'ambiance de la ville vient de la carte postale de New-York sur la page de droite.

Intéressons nous maintenant au traitement de la voix. Une des originalités du film est que tout est raconté par une seule actrice. Elle prend en charge la voix off de la narratrice et les voix des différents personnages, c'est Rosario Dawson pour la version anglaise et Jehnny Beth pour la version française. Pour moi, cela créer un « un effet contes », à la manière d'un conte du soir raconté par une seule personne. Les images matérialisent ce qu'on pourrait s'imaginer en entendant l'histoire, avec des cadrages et des jeux d'échelles surréalistes. En termes de spatialisation de la voix, on retrouve une pratique courante du cinéma qui consiste à mettre la narratrice au centre, sans effet supplémentaire (pas de réverbération par rapport à l'espace montré par les images). En revanche, certains dialogues sont spatialisés et traités avec de la réverbération pour les ancrer dans des espaces acoustiques. Mais le rendu est inégal, la plupart du temps les voix sont entendues beaucoup trop sur les côtés par rapport à ce qu'on voit à l'image. Pire, il arrive que la spatialisation soit tellement confuse qu'on a du mal à identifier le personnage qui parle. Le choix de prendre une seule narratrice pour incarner plusieurs personnages aurait dû être un prétexte pour exploiter au maximum les possibilités de spatialisation sonore, en caractérisant les différents personnages selon leurs emplacements par exemple. Mais la spatialisation est défectueuse car certaines voix sont localisées au même endroit sans explication rationnelle. Par exemple à 16:55' [fig.17], la voix de Lupe est spatialisée (trop) à gauche et la voix de Debbie (trop) à droite, ce qui reste quand même logique par rapport au placement des personnages dans le cadre quand on regarde le dialogue. Mais quand Elda Stilleto prend la parole elle émet du même endroit que Debbie à droite alors qu'elle est représentée au centre du cadre. Ce genre d'erreurs de spatialisation brise l'immersion mise en place par le reste de la direction artistique. Dans la version en streaming du film, la spatialisation des voix est inexistante car toutes les voix sont ramenées au centre.



*fig.17 : Lupe à gauche, Elda Stilleto au milieu et Debbie à droite. 16:55'*

Concernant la spatialisation des bruitages, il y a les mêmes problèmes que pour la voix. La spatialisation est présente mais imprécise, ce qui brise l'immersion à certains moments. La plupart du temps, les sources sonores sont trop espacées par rapport aux objets visualisés à l'écran. Les plans à 02:55' [fig.18] ou à 04:35' [fig.19] font entendre des exemples évidents de spatialisation approximative. Ces scènes se déroulent en face du spectateur mais les sons sont entendus comme s'ils étaient espacés à 180° de chaque côté. Le problème est audible dans la version streaming et la version en RV, où il faut tourner la tête vers un point dans le vide pour être en face de la source du son.



*fig.18 : Dans le métro à 02:55'. Lupe se cogne dans le portique sur la gauche, le son est spatialisé totalement à gauche au lieu d'être devant à gauche.*



*fig.19 : 04:38'. Au moment où chaque club se retourne, les sons sont spatialisés complètement à droite et à gauche, ce qui est incohérent par rapport à l'image qui les montre en face.*

La musique aussi est concernée par la spatialisation sonore. Il y a deux catégories de musiques dans le court métrage : la musique *headlocked* et la musique spatialisée. La musique *headlocked* ne réagit pas aux mouvements du spectateur, elle a ses propres critères de spatialisation figés lors de la production en studio. Elle correspond à la musique de fosse du cinéma classique. Le passage du concert à 26:54' par exemple utilise de la musique de fosse, ce qui donne un effet clip à la scène. Les éléments musicaux spatialisés, quant-à eux, correspondent à la musique d'écran. Ils sont très présents dans le film car la musique punk est au centre de l'histoire, mais ils souffrent des mêmes problèmes de spatialisation imprécise que les bruitages et la voix.

Mon premier contact avec le film a été positif, le style est travaillé avec de beaux graphismes. La lumière, la composition et la narration sont inspirées, il y a beaucoup d'idées originales de mises en scène avec des jeux très intéressants sur les échelles au sein des plans. Cette réussite est due à l'implication des créateurs pour faire un film qui tire réellement parti des possibilités de la RV :

« Ne venez pas à la RV avec des idées préconçues sur les autres médias. Vous devez vous familiariser avec les possibilités de la RV - alors jouez à des jeux vidéo, allez voir des pièces de théâtre... et essayez de penser dans l'espace. »<sup>66</sup>

Avec ces paroles de Nico Casavecchia, on comprend l'état d'esprit vers lequel il faut se tourner pour créer une œuvre pour ce média. Malgré ce premier contact positif, l'expérience du film est un peu ternie par une sonorisation qui ne paraît pas aboutie. Surtout à cause d'une spatialisation du son approximative qui crée des dissonances audio/visuelles et nous sort de l'immersion offerte par le reste de la direction artistique. Il serait intéressant d'avoir un retour des sound designers qui ont travaillé sur ce projet pour savoir s'ils ont eu des difficultés lors de la production. Qu'est-ce qui expliquerait ces problèmes de point d'écoute imprécis ? Est-ce que le montage a changé après que le son ait été créé ?

Il y a donc encore une marge de progression pour perfectionner ce média du cinéma en réalité virtuelle qui, rappelons-le, est encore jeune.

---

<sup>66</sup> N. Casavecchia, *BattleScar - Made with Unity*. [en ligne] URL : <https://unity.com/fr/madewith/battlescar>. Consulté le 26/08/2021.

### 3.3 : Une spatialisation exigeante pour le jeu vidéo : l'exemple de Paper Beast

Le jeu vidéo lui en revanche n'est pas nouveau dans le paysage de la création audiovisuelle. Il s'est beaucoup développé depuis sa popularisation dans les années 1980 et l'immersion a toujours été un jalon important de son développement. Même si la majorité du jeu vidéo a évolué en *flat screen* (terme utilisé dans le langage courant pour définir un jeu joué sur écran plat, en opposition à la RV), l'immersion s'est développée au travers de graphismes aux détails toujours plus poussés et un son de plus en plus précis dans sa restitution. L'idée de la réalité virtuelle appliquée au jeu vidéo découle de cette volonté d'immersion toujours recherchée dans le jeu vidéo. Elle est en germe depuis longtemps, on trouve un premier essai de commercialisation de casque de RV destiné au jeu vidéo dès l'année 1995 avec le *Virtual Boy* de Nintendo [fig.20], malgré le fait qu'il ait été un échec commercial, cela montre l'intérêt constant du média pour cette technologie.



La spatialisation audio est un atout déterminant pour l'immersion, l'utilisation du son 3D dans un jeu vidéo paraît donc être une application évidente. Le son binaural y est de plus en plus utilisé, même dans les jeux qui ne sont pas en réalité virtuelle. Avec la réalité virtuelle, on ne

<sup>67</sup> Crédit photo : Evan-Amos. [en ligne] URL : <https://upload.wikimedia.org/wikipedia/commons/thumb/4/44/Virtual-Boy-Set.jpg/1024px-Virtual-Boy-Set.jpg?uselang=fr>. Publié le 22/09/2011, consulté le 26/08/2021.

fait qu'ajouter un casque offrant un degré d'immersion supplémentaire via une vision 3D stéréoscopique et la prise en compte des mouvements de l'utilisateur.

Les jeux vidéo en RV bénéficient donc des innovations sur l'audio immersif tout en étant des agents de son développement. Contrairement au cinéma en RV qui est encore en recherche d'un équilibre entre le cinéma et le jeu vidéo, le jeu vidéo a déjà ses outils de création dédiés avec les moteurs de développement de jeu vidéo et les intergiciels qui prennent en charge certains aspects spécifiques du développement (comme le son avec Wwise ou Fmod).

L'immersion visuelle et sonore de certains jeux vidéo *flat screen* est déjà tellement poussée qu'il peuvent supporter un portage en réalité virtuelle sans avoir tout à remanier. C'est le cas par exemple du jeu *Hellblade - Senua's Sacrifice* mentionné en amont. Bien qu'il soit un jeu *flat screen* à l'origine, une version en RV a aussi été mise sur le marché par la suite. Elle se contente uniquement de nous projeter dans l'univers visuel du jeu mais le gameplay et le son restent les mêmes. L'expérience est donc encore plus saisissante car l'immersion visuelle vient s'ajouter à l'immersion sonore déjà présente grâce au son binaural.

### *Paper Beast*

Avec le développement de la RV et l'apparition d'un marché viable, des expériences de qualité dédiées à la RV émergent. Contrairement à un jeu comme *Hellblade* pour lequel la réalité virtuelle peut paraître anecdotique (car elle n'avait pas été pensée à l'origine), les jeux vidéo développés spécifiquement pour la RV sont beaucoup plus intéressants car ils prennent en compte tous les aspects immersifs de ce média.

Pour aborder un exemple de jeu vidéo en réalité virtuelle, j'ai choisi de parler du jeu *Paper Beast*. Il a été développé par le studio montpelliérain Pixel Reef, fondé et dirigé par Eric Chahi, un créateur pionnier dans le domaine vidéoludique depuis les années 1990. Le jeu est sorti au cours de l'année 2020, d'abord sur PS4 puis sur PC. Une version *flat screen* du jeu, jouable sans casque de RV, a aussi été rendue disponible par la suite. *Paper Beast* a donc fait le chemin inverse de *Hellblade*, d'abord en réalité virtuelle puis sur écran plat traditionnel. Il a été sélectionné dans des festivals prestigieux comme les DICE Awards et a gagné des prix, entre autres un Pégase pour « Meilleur univers de jeu vidéo » par l'Académie des arts et techniques du jeu vidéo en 2021. Voici le pitch du jeu :

« [*Paper Beast* est] Un jeu d'aventure et d'exploration sur la vie sauvage dans un univers onirique. Quelque part dans les profondeurs abyssales d'Internet, la vie a

fleuri. Soyez le premier à explorer cette terre vierge où la nature occupe une place centrale. Une puissance inconnue semble bouleverser son équilibre. Joignez vos forces à celles de créatures intrigantes et attachantes pour résoudre les énigmes de *Paper Beast*. »<sup>68</sup>

Le but du jeu est donc d'explorer un univers fictif émergeant des Big Data, pour en découvrir ses enjeux et interagir avec les créatures qui le peuplent.

Le concept fort du jeu est qu'il se déroule dans un univers entièrement simulé en temps réel, cela va des comportements des créatures aux conditions météorologiques qui influent sur le terrain. À l'inverse du cinéma, presque rien n'est défini précisément à l'avance quand le joueur parcourt le jeu. Son univers est régi par un ensemble de systèmes qui interagissent entre eux et deux joueurs différents ne vivront pas exactement la même expérience. Le but du jeu est de comprendre comment manipuler tous ces éléments de façon directe ou indirecte, avec des outils donnés au fur et à mesure de la progression.

L'équipe de développement était composée de 15 personnes dont 4 affectées à la partie sonore, ce qui est beaucoup pour une production indépendante de cette envergure. On retrouve Roly Porter à l'écriture et la composition musicale ainsi que le groupe japonais TsuShiMaMiRe pour la musique punk rock qui apparaît occasionnellement. La créatrice sonore Floriane Pochon était à la composition sonore et musicale en collaboration avec R.Porter, et Clément Duquesne s'est occupé de la partie conception sonore et programmation audio. Il a donc créé des systèmes audio et a intégré le son dans le jeu avec l'intergiciel Wwise. Sa conférence « *Sonoriser l'écosystème VR de Paper Beast* » donnée au festival Addon 2021 sera ma source principale pour cette partie.

« Le point de départ pour le design des systèmes audio a été l'immersion en réalité virtuelle. La synchronisation, la spatialisation et la propagation étaient des outils pour que les joueurs et les joueuses puissent accepter cette nouvelle réalité. »<sup>69</sup>

Encore une fois, la spatialisation audio fait partie des critères essentiels à l'immersion dans la RV. Nous allons voir de quelle manière Clément Duquesne et son équipe ont relevé ce défi de spatialiser un son immersif qui s'adapte en temps réel à une expérience qui se veut imprévisible.

---

<sup>68</sup> Pixel Reef, *Paper Beast*, [en ligne] <https://www.pixelreef.fr/paper-beast/> consulté le 27/08/2021.

<sup>69</sup> C. Duquesne, [#ADDONLIVE] *Sonoriser l'écosystème VR de Paper Beast*, de 07:50' à 08:31' URL : <https://www.youtube.com/watch?v=yU2472sVBTg>, mis en ligne le 19/05/2020, consulté le 04/08/2021.

Puisque l'univers procédural du jeu provoque des situations imprévisibles, il fallait que l'audio soit dynamique et adaptatif aussi : Clément a résolu cette question en créant ce qu'il a appelé des *soundsources*, ce qui pourrait bien être une solution idéale pour spatialiser le son en RV. Les *soundsources* sont des émetteurs de sons adaptatifs, leur position est calculée en temps réel à partir des coordonnées des objets envoyées par le moteur de jeu. Jusqu'ici, c'est le même principe que la spatialisation normale d'éléments sonores dans un jeu vidéo, mais Clément a eu la bonne idée de définir le centre de gravité des ces *soundsources* grâce à des paramètres nourris en temps réel par le moteur physique.

Le centre de gravité est une notion essentielle pour l'adaptativité du système de spatialisation audio de *Paper Beast*, il permet d'avoir un rendu audio extrêmement précis. Par exemple, les sons des mouvements des créatures sont directement liés à ce système de *soundsources*. Ce système permet d'avoir des sons très précis pour chaque articulation quand la créature en papier ou en métal se déplace. Cette précision de localisation audio rend les créatures très crédibles et participe à l'immersion totale du joueur. [fig.21]

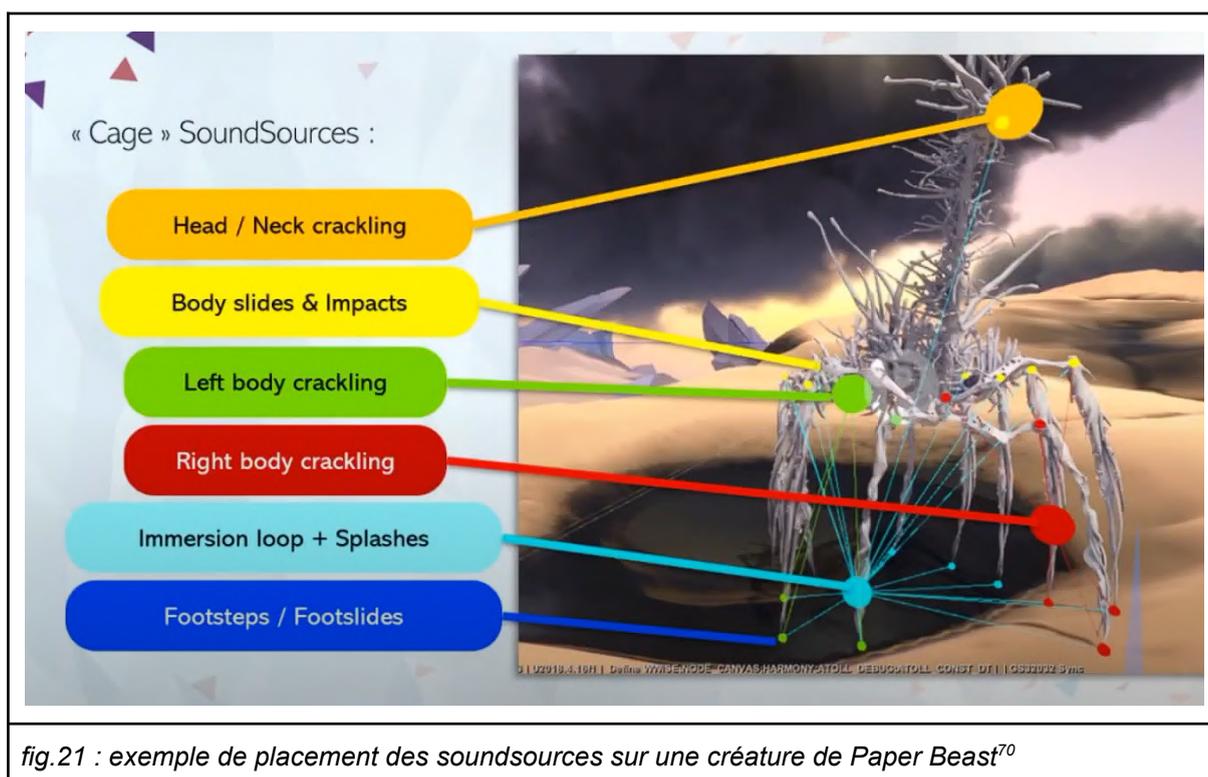


fig.21 : exemple de placement des soundsources sur une créature de *Paper Beast*<sup>70</sup>

La méthode généralement utilisée pour les jeux *flat screen* consistant à émettre les bruits d'une créature depuis un centre de gravité unique n'est pas convaincante une fois dans la réalité virtuelle. Il est donc crucial en RV d'être le plus spécifique possible dans le placement de ces sons.

<sup>70</sup> C. Duquesne, *op. cit.*, 23:36'

En jeu, la spatialisation des éléments extrêmement précise et réaliste est aussi due à l'utilisation de techniques binaurales. Pour illustrer cette influence des filtres binauraux j'ai réalisé une analyse du spectre d'un son en jeu [fig.22]. Sur les images avec l'analyse du son, nous constatons que les hautes fréquences sont atténuées quand l'objet se trouve derrière la tête du personnage.

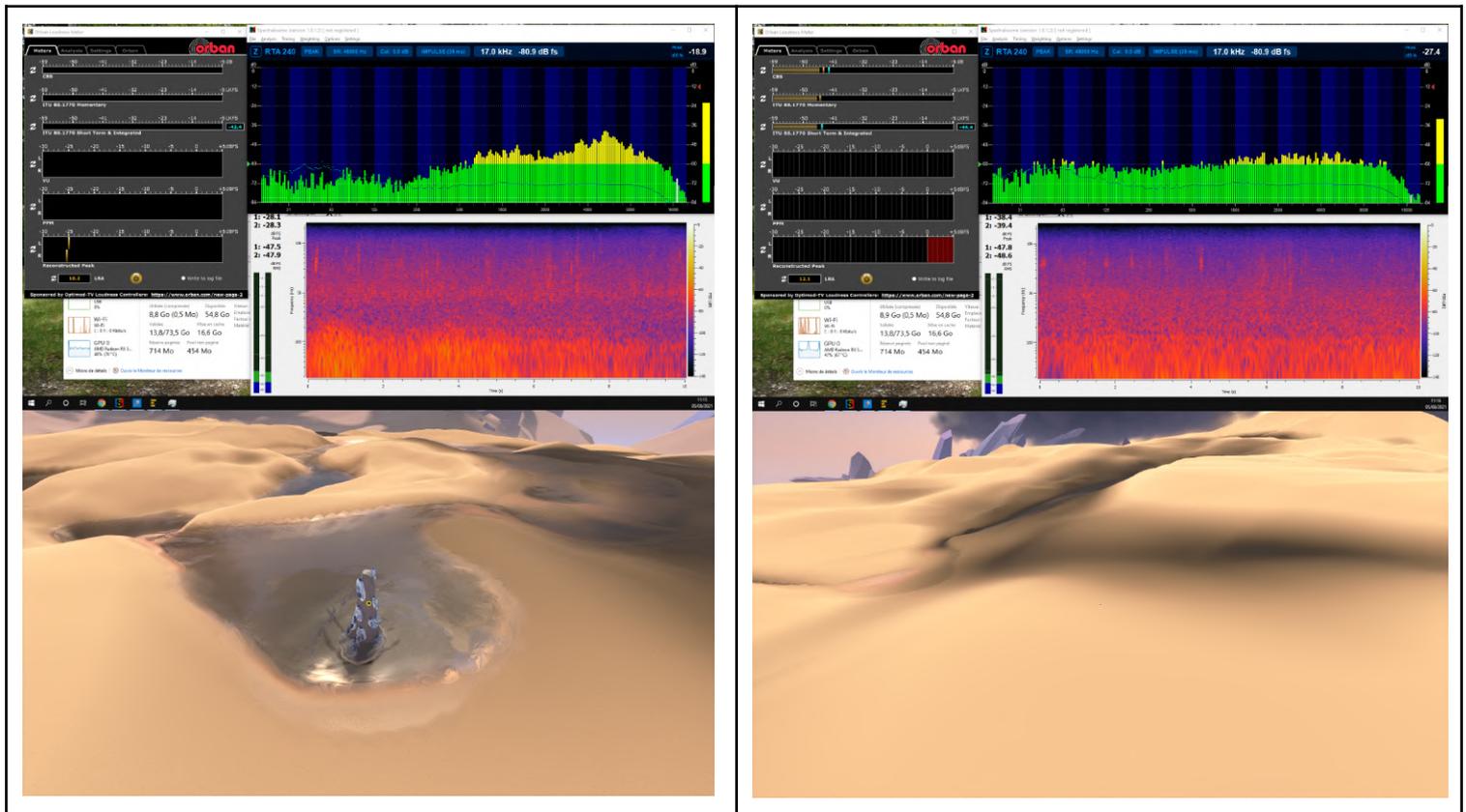


fig.22 : À gauche, l'élément émetteur de son est devant le joueur. À droite le joueur s'est retourné à 180°, l'émetteur est donc derrière. On distingue que le pic de fréquences autour de 6000 Hz est atténué quand l'élément émet de derrière le joueur.

Pour chercher de bonnes pratiques de spatialisation du son en RV, nous pouvons aussi nous intéresser aux sons d'ambiances de *Paper Beast*. Une des difficultés dans la gestion de l'audio en temps réel pour les jeux vidéo est qu'on ne peut pas jouer un nombre de sons illimité simultanément car le rendu audio est limité par la puissance de calcul du support de destination (comme nous l'avons vu avec l'exemple des douilles dans le chapitre sur les intergiciels). Il faut partager cette puissance de calcul avec le rendu visuel entre autres. Pour le cas particulier des ambiances sonores extérieures, contrairement à un espace intérieur où on a un nombre limité de sources sonores ponctuelles (appareils électriques, animaux, etc.), les ambiances extérieures peuvent en contenir un nombre incalculable. Il faut donc trouver des solutions ingénieuses pour rendre ces ambiances convaincantes à l'oreille. Il est bien sûr possible d'enregistrer directement une ambiance réelle avec un micro ambisonique pour l'intégrer tel quel dans le jeu, mais le résultat ne sera

pas adaptatif et fonctionnera seulement si l'univers du jeu est réaliste. Dans *Paper Beast*, la solution a plutôt été de brouiller la perception spatiale des ambiances extérieures. Des expériences récentes ont montré que les performances de localisation de notre système auditif décroissent lorsque le nombre de sources sonores simultanées augmente<sup>71</sup>. Pour que le joueur ne puisse pas identifier où un son commence et s'arrête, l'équipe de sound design a donc décidé de superposer plusieurs couches de sons avec différentes techniques de spatialisation. De cette façon, quand le joueur bouge la tête, il n'entend pas un seul fichier audio qui s'adapte au mouvement, mais un enchevêtrement d'éléments qui réagissent tous de façons différentes. Ces ambiances sont constituées de 3 couches différentes :

- Des sons binauraux non spatialisés (*headlocked*). Ils ne s'adaptent pas aux changements de point d'écoute du joueur et sont utilisés essentiellement pour les textures homogènes comme le vent ou le sable. Il ne faut pas qu'ils aient des détails saillants qui seraient reconnaissables et vers lesquels le joueur voudrait se tourner.
- Des *quads* spatialisés, ce sont des fichiers qui diffusent un son d'ambiance sur 4 canaux. Les 4 canaux émetteurs sont placés en hauteur aux 4 points cardinaux de la scène à sonoriser. Ils sont ensuite passés dans un plugin binaural ce qui donne de la profondeur à la scène.
- Des bruitages 3D spatialisés aléatoires et ponctuels qui apportent une grande crédibilité aux ambiances sonores. Bien que ces petits éléments soient excellents pour donner la sensation de présence dans l'espace, ils sont à utiliser avec parcimonie car si le joueur cherche la source du son et qu'il ne la trouve pas, cela aurait l'effet inverse en le sortant de l'immersion.

En jeu, l'effet est saisissant et l'immersion est véritablement présente grâce à ces paysages sonores complexes créés en utilisant différentes techniques de spatialisation.

En nous intéressant à ces quelques solutions de spatialisation sonore dans *Paper Beast*, nous réalisons qu'il faut trouver des solutions différentes de celles traditionnellement utilisées pour les expériences non RV. Il faut à tout prix pousser la qualité de la spatialisation pour éviter que le joueur sente qu'il écoute un simple fichier audio. C'est ce qui a pêché selon moi dans le film RV *BattleScar*, avec un son qui ne sonne pas assez abouti à cause de la spatialisation et qui brise l'immersion à certains moments. L'immersion impose de mettre de l'importance sur tous les sons, les ambiances par exemple n'ont pas le rôle d'arrière-plan comme elles peuvent l'avoir sur d'autres supports. En RV, elles sont vraiment tout autour de nous et il faut réussir à donner l'illusion de présence.

---

<sup>71</sup> P. Fuchs, *Le traité de la réalité virtuelle - Volume 1*, Presse des Mines, 2006, p73

# Conclusion

Au cours de ce mémoire, nous avons fait un rapide tour d'horizon des solutions de spatialisation sonore contemporaines en audionumérique. Nous nous sommes concentrés sur les solutions adaptées aux œuvres cinématographiques et vidéoludiques en lien avec l'expérience que j'ai pu acquérir en tant qu'étudiant au MAAAV. Nous avons aussi fait un focus sur la réalité virtuelle, un support qui peut englober les domaines cinématographiques et vidéoludiques. Il y a bien sûr d'autres solutions de spatialisation que nous n'avons pas étudiées et qui correspondent à d'autres domaines spécifiques comme la production musicale, radiophonique, théâtrale, ou tout autre domaine qui diffuse du son.

Le cinéma a joué un rôle majeur dans le développement du son spatial et sa diffusion au grand public. Les techniques utilisées comme la stéréophonie et le surround sont devenues les techniques conventionnelles utilisées dans tous les domaines sonores. Le cinéma continue d'être un moteur du développement du son spatial, notamment avec la démocratisation des systèmes immersifs comme le Dolby Atmos, qui supprime les limitations du plan de diffusion sonore horizontal. Nous pouvons espérer que la diffusion du son sur un ensemble de haut-parleurs qui entourent complètement les spectateurs sera bientôt la norme accessible dans toutes les salles de cinéma.

Pour spatialiser du son dans les jeux vidéo où, contrairement au cinéma, le point de vue n'est pas fixé, il faut utiliser des techniques de spatialisation en temps réel. La solution idéale est l'utilisation d'intergiciels qui permettent aux créateurs sonores d'aborder ce média sereinement. Bien sûr, l'utilisation d'un intergiciel n'est pas obligatoire pour réaliser le son d'un jeu vidéo, mais elle fait gagner du temps à toutes les professions et offre la possibilité de repousser les limites de la qualité sonore. L'accessibilité actuelle des techniques binaurales, notamment grâce aux plug-ins libres d'accès des intergiciels, permet aux créateurs d'envisager leur intégration dans leurs œuvres pour atteindre une qualité de spatialisation et donc d'immersion toujours plus convaincante. Ce mode de restitution sonore trouve une réelle utilité dans le jeu vidéo, ce qui promet un bel avenir à la qualité du son dans les productions actuelles et futures. Nous avons pu voir avec des exemples concrets tirés de ma pratique que l'ambisonie et le binaural pouvaient aussi être utilisés dans d'autres domaines que le jeu vidéo.

La réalité virtuelle bénéficie totalement des avancées citées précédemment car elle tire parti des intergiciels et du mode de restitution binaural en stéréo. Nous avons pu voir avec le film *BattleScar* qu'elle était adaptée au cinéma, mais qu'il fallait tout de même repenser la manière de travailler par rapport au cinéma traditionnel. La spatialisation du son est un point à affiner en particulier car elle a un impact important sur l'expérience d'un

contenu en réalité virtuelle. Avec le jeu vidéo *Paper Beast*, nous avons vu l'exemple d'une spatialisation audio qui aboutit en réalité virtuelle. Nous avons repéré quelques points clés sur lesquels il faut se concentrer pour optimiser le son d'un jeu en RV, particulièrement au travers avec la création de systèmes audio cohérents par rapport au moteur de jeu et la création d'ambiances sonores soignées.

Pour conclure nous pouvons affirmer qu'une bonne spatialisation audio est bénéfique à l'expérience globale d'une œuvre qu'elle soit cinématographique ou vidéoludique. Les innovations comme le son binaural et la flexibilité des intergiciels permettent de concevoir un son spatial idéal. En réalité virtuelle, qui peut réunir les deux domaines cités, un son bien spatialisé participe au potentiel d'émerveillement du média. Pour délivrer une spatialisation audio optimale, il faut jouer sur des paramètres précis qui peuvent paraître imperceptibles sans outils d'analyse (comme les filtres binauraux), mais ce travail est gratifiant car elle pousse l'auditeur à se dire que « le son est bien ».

# Remerciements

Je souhaite remercier toutes les personnes qui m'ont accompagné pendant ces deux années au master MAAAV. Particulièrement mes camarades de la promo 2019-2021 avec qui nous avons passé de très bons moments malgré le contexte sanitaire. Merci à Emma Marichal pour ses relectures et ses conseils dans l'écriture ainsi que la mise en page de ce mémoire. Merci à mes parents pour me soutenir dans ce que j'entreprends.

Je remercie aussi tous les étudiants et étudiantes d'écoles différentes, qui m'ont fait confiance pour la création sonore sur leurs projets. Ces périodes de stage ont été enrichissantes humainement et professionnellement.

Un grand merci à l'ensemble de l'équipe pédagogique du master MAAAV, en particulier Jean-Marc Serre, Laetitia Pansanel Garric, Jean-Pierre Gaspar, Martin Barnier, et Frédéric Finand qui m'a fait découvrir l'existence du son binaural, faisant germer l'idée de ce mémoire.

# Bibliographie

## Cinéma

CHION Michel. *L'audio-vision - Son et image au cinéma, 4e ed.*, Armand Colin, 2017

CHION Michel. *Un art sonore : le cinéma*, Cahiers du cinéma. 2003. p223-234

HIDALGO Jason. *Dolby's Atmos technology gives new meaning to surround sound, death from above*. Engadget. [en ligne] URL :

<https://www.engadget.com/2012-04-26-dolby-atmos-technology-gives-new-meaning-to-surround-sound.html>. Publié le 26/04/2012, consulté le 02/08/2021.

KITSOPANIDOU Kira, PISANO Giusy. 3 - *L'émergence du hors-film sur grand écran ou la « nouvelle » polyvalence des salles de cinéma*, dans : Laurent Creton éd., *Les salles de cinéma. Enjeux, défis et perspectives*. Paris, Armand Colin, « Recherches », 2013, p.

147-178. [en ligne] URL :

<https://www-cairn-info.bibelec.univ-lyon2.fr/les-salles-de-cinema--9782200287580-page-147.htm>. Consulté le 08/08/2021.

TALL GUY FILM. *What is Virtual Surround? - Spatial audio explained*. [en ligne] URL :

<https://www.youtube.com/watch?v=E2JneaC6lbl>. Publié le 23/06/2020, consulté le 22/07/2021.

TIJOU Clément. *Cohérence audiovisuelle spatiale au cinéma : Influence de la vitesse et de la nature des objets sonores*. Mémoire de fin d'étude de l'École Nationale Supérieure Louis Lumière. 2016. [en ligne] URL :

[https://www.ens-louis-lumiere.fr/sites/default/files/2017-08/Tijou\\_Son\\_2016.pdf](https://www.ens-louis-lumiere.fr/sites/default/files/2017-08/Tijou_Son_2016.pdf). Publié en 2016, consulté le 26/07/2021.

## Jeu vidéo

BUFFONI Louis-Xavier. *Ambisonics as an Intermediate Spatial Representation (for VR)*. [en ligne] URL :

<https://blog.audiokinetic.com/fr/ambisonics-as-an-intermediate-spatial-representation-for-vr/>. Publié le 02/08/2016, consulté le 23/28/2021.

DE PRATO Giuditta, *Une technologie de base : l'intergiciel (middleware)*, Hermès, La Revue, 2012/1 (n° 62), p. 101-107. [en ligne] URL : <https://www.cairn.info/revue-hermes-la-revue-2012-1-page-101.htm>. Consulté le 20/08/2021.

DERIVIERE Oliver, *Interview pour L'Écran*. [en ligne] URL : <https://asso-lecran.fr/interview-de-olivier-deriviere>. Publié le 16/02/2020, consulté le 23/08/2021.

FELLGETT Peter. *Ambisonics. Part one: general system description*, Studio Sound, Vol. 17 no. 8, pp. 20-22, Août 1975. [en ligne] URL : <https://www.michaelgerzonphotos.org.uk/articles/Ambisonics%201.pdf>. Publié en 2009, consulté le 23/08/2021.

FIN DU GAME, (ExServ, Hugo, Maxime). *Épisode 45 - Hellblade: Senua's Sacrifice*. [en ligne] URL : <https://fin-du-game.lepodcast.fr/episode-45-hellblade-senuas-sacrifice>, Publié le 20/11/2020, consulté le 04/08/2021.

IRCAM, *Technologies de reproduction binaurale*. [en ligne] <https://www.ircam.fr/projects/pages/technologies-de-reproduction-binaurale/>. Consulté le 24/08/2021.

SADIE, *Spatial Audio filters for the Google YouTube 360/VR pipeline*. [en ligne] URL : <https://www.york.ac.uk/sadie-project/GoogleVRSADIE.html>. Publié en 2017, consulté le 10/08/2021.

STEINBERG Cubase. *3D Mixes for Ambisonics*. [en ligne] URL : [https://steinberg.help/cubase\\_pro\\_artist/v10/en/cubase\\_nuendo/topics/surround\\_sound/surround\\_sound\\_ambisonics\\_c.html](https://steinberg.help/cubase_pro_artist/v10/en/cubase_nuendo/topics/surround_sound/surround_sound_ambisonics_c.html). Consulté le 21/07/2021.

TOBISCH Léopold. *Comment la musique est devenue un ingrédient essentiel des jeux vidéo*. France Musique. [en ligne] URL : <https://www.francemusique.fr/culture-musicale/la-musique-des-jeux-video-un-monde-a-decouvrir-66460>. Publié le 26/10/2018, consulté le 08/08/2021.

WWISE - site officiel, *Audiokinetic Wwise*, [en ligne] <https://www.audiokinetic.com/fr/about/> Consulté le 23/08/2021.

## Réalité virtuelle

ALLAIS Martin, CASAVECCHIA Nico. *BattleScar press kit - Extended synopsis*. [en ligne]

URL :

[https://static1.squarespace.com/static/5a2680f5914e6b6824a7254f/t/5a2aa73624a694e5fd262125/1512744789541/BATTLESCAR\\_Sundance\\_V8-1.pdf](https://static1.squarespace.com/static/5a2680f5914e6b6824a7254f/t/5a2aa73624a694e5fd262125/1512744789541/BATTLESCAR_Sundance_V8-1.pdf), consulté le 04/08/2021.

ARTE. *Punk was invented by girls | BattleScar | Vidéo 360°*. [en ligne] URL :

<https://www.youtube.com/watch?v=KcGIT9wTbbs>. Publié le 14/01/2021, consulté le 04/08/2021.

AUDIO ENGINEERING SOCIETY, *Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions*. [en ligne] URL :

[https://www.researchgate.net/publication/236634182\\_Spatially\\_Oriented\\_Format\\_for\\_Acoustics\\_A\\_Data\\_Exchange\\_Format\\_Representing\\_Head-Related\\_Transfer\\_Functions](https://www.researchgate.net/publication/236634182_Spatially_Oriented_Format_for_Acoustics_A_Data_Exchange_Format_Representing_Head-Related_Transfer_Functions), publié le 7/05/2013, consulté le 21/07/2021.

CASAVECCHIA Nico. *BattleScar - Made with Unity*. [en ligne] URL :

<https://unity.com/fr/madewith/battlescar>. Consulté le 26/08/2021.

CNC. « *BattleScar* » : *un court d'animation en VR à découvrir en ligne*. [en ligne]

[https://www.cnc.fr/creation-numerique/actualites/battlescar---un-court-danimation-en-vr-a-decouvrir-en-ligne\\_1477797](https://www.cnc.fr/creation-numerique/actualites/battlescar---un-court-danimation-en-vr-a-decouvrir-en-ligne_1477797). Publié le 10/06/2021, consulté 04/08/2020.

DEDOLA Léa. *Tête-à-tête avec Léa Dedola*. [en ligne] URL :

<https://www.univ-lyon2.fr/recherche/magazine/numero-1/tete-a-tete-avec-lea-dedola>. Publié le 29/01/2021, consulté le 28/07/2021.

FUCHS Philippe. *Réalité virtuelle : une révolution scientifique*, La Méthode scientifique, France Culture. [en ligne] URL :

<https://www.franceculture.fr/emissions/la-methode-scientifique/realite-virtuelle-une-revolution-scientifique>. Publié le 19/10/2016, consulté le 28/04/2021.

FUCHS Philippe, MOREAU Guillaume, BERTHOZ Alain. *Le traité de la réalité virtuelle Volume 1 : L'Homme et l'environnement virtuel*. Presse des Mines, 2006, Mathématique et informatique.

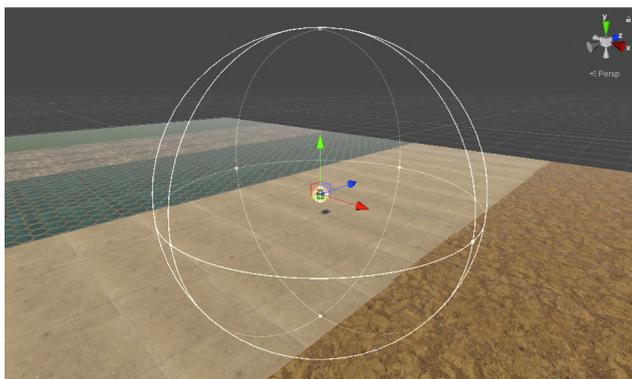
FUCHS Philippe, MOREAU Guillaume, BURKHARDT Jean-Marie. *Le traité de la réalité virtuelle Volume 2 - L'interfaçage : l'immersion et l'interaction en environnement virtuel*. Presse des Mines, 2006, Mathématique et informatique.

DUQUESNE Clément. [#ADDONLIVE] *Sonoriser l'écosystème VR de Paper Beast*. [en ligne] URL : <https://www.youtube.com/watch?v=yU2472sVBTg>. Publié le 19/05/2020, consulté le 04/08/2021.

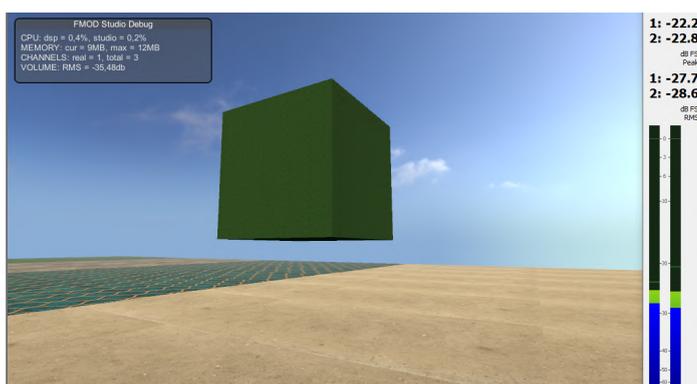
PIXEL REEF, *Paper Beast*, [en ligne] <https://www.pixelreef.fr/paper-beast/> consulté le 27/08/2021.

R.Serge, *Tendances VR/AR 2021 : À quoi faut-il s'attendre cette année ?* [en ligne] URL : <https://www.realite-virtuelle.com/tendances-vr-ar-2021/>. Publié en janvier 2021, consulté le 25/08/2021.

# Annexes



1.1 *Fmod* : Dans le moteur de jeu Unity. La sphère blanche autour du cube vert est la représentation de l'ampleur du son dans l'espace virtuel. Elle correspond aux valeurs de la courbe d'atténuation définie dans l'intergiciel Fmod



1.2 *Fmod* : Dans le jeu avec une vue à la première personne. Le cube vert émet du son juste devant nous. Comme le confirme le VU-mètre, les canaux gauche-droite sont à peu près égaux.



1.3 *Fmod* : Le personnage s'est tourné de 90° vers la droite. Le cube émetteur de son est donc hors champ sur la gauche, il émet depuis cette nouvelle position. Ce changement est visible sur le VU-mètre avec le canal de gauche plus élevé.