

Mémoire de fin d'études
École Nationale Supérieure Louis Lumière
**Adaptation rapide à une spatialisation
binaurale non-individualisée : cas de la
réalité virtuelle appliquée au jeu vidéo**

Guillaume ANDRE
Directeur interne : Alan BLUM
Directeur externe : Renaud SÉGUIER
Rapporteurs : Laurent MILLOT et Éric URBAIN

promotion SON 2017

Remerciements

Je voudrais remercier Alan Blum pour son accompagnement tout au long de ce mémoire, ainsi que Renaud Séguier pour avoir accepté de m'encadrer et de mettre à ma disposition les moyens de 3D Sound Labs. Merci à Corentin Guézénoc et Thomas Braure pour leur aide et leurs compétences techniques qui m'ont été très précieuses. Merci à Vincent Zanetto pour son incroyable expertise et à Papa, Nono, Pierre et Hubert pour les cours de maths plus que nécessaires, ainsi qu'à Juju, Boz et Maman. Merci à Louis Martin et à l'équipe de DontNod pour leur éclaircissement professionnel sur le secteur vidéoludique. Merci enfin à tout ceux qui ont pris le temps de venir passer le test qui constitue la partie pratique de ce mémoire !

Résumé

La réalité virtuelle est une technologie en plein essor. Les premiers modèles à destinations du grand public viennent tout juste de sortir, et l'industrie vidéoludique promet des expériences toujours plus réalistes qui poussent de plus en plus loin l'exploit technologique. Parmi ces promesses, se trouve celle de l'utilisation de la spatialisation binaurale, qui permettrait une localisation sonore beaucoup plus naturelle et précise en simulant l'impact de la morphologie d'un individu sur le son qui parvient à ses oreilles. Cette spatialisation serait en outre renforcée par l'utilisation du suivi des mouvements de la tête. Cependant, il est actuellement trop compliqué de proposer à chaque utilisateur une réponse binaurale adaptée à sa morphologie, condition pourtant nécessaire pour que l'apport de la spatialisation binaurale soit pertinent, ou ne risque pas au contraire de dégrader la perception du son dans l'espace. Les constructeurs estiment que ce problème est contournable en supposant qu'un joueur serait capable de s'adapter très rapidement à une réponse générique différente de la sienne. C'est dans ce contexte que ce mémoire propose une réflexion sur la place du binaural dans le jeu vidéo et la réalité virtuelle, pour ensuite tester l'hypothèse d'une capacité d'adaptation rapide grâce à des tests perceptifs.

Mots-clés : binaural, spatialisation, jeux vidéo, réalité virtuelle, adaptation, personnalisation

Abstract

Virtual reality is rising. Displays aiming for a large audience are barely starting to appear, and the video game industry is promising more and more realistic experiences, pushing further the technical feat. Among these promises is binaural sound, which could allow a much more natural and accurate sound localization. Furthermore, this spatialization would be reinforced by the use of headtracking. However, it is too complicated at the moment to offer a personal binaural response suitable one's morphology, which is an important condition to insure binaural spatialization effectiveness. Makers are saying that this issue is easily avoided by estimating that a player could adapt quickly to a generic response that doesn't match his. In this environment, this master thesis aim to question the purpose and interest of binaural use in video game and virtual reality, then test this hypothesis about rapid adaptation capacity using perception tests.

Keywords : binaural, spatialization, video game, virtual reality, adaptation, personalization

Sommaire

Introduction	7
1 Spatialisation binaurale : le choix de l'approche adaptative	10
1.1 Principe de la spatialisation binaurale	10
1.1.1 Indices de localisation du son	12
1.1.2 Binaural statique et dynamique	15
1.1.3 Quel format de fichier pour le binaural?	17
1.1.4 L'approche adaptative contre l'individualisation	18
1.2 Méthode de l'individualisation	18
1.2.1 Lourdeur du dispositif	18
1.2.2 Transportabilité des mesures	19
1.2.3 Pistes à explorer	20
1.3 Approche adaptative	21
1.3.1 Temps d'adaptation	22
1.3.2 Apport de la vision	22
1.3.3 Apport du binaural dynamique	24
1.3.4 Selection d'HRTF	24
1.4 Conclusion	25
2 Binaural, jeux vidéo et réalité virtuelle	26
2.1 Du rôle du contenu audio dans le jeu vidéo	27
2.2 Utilisation du binaural dans le jeu vidéo	31

2.2.1	Apport sur la spatialisation	32
2.2.2	Le binaural comme parti pris esthétique	33
2.2.3	Quand le son prend le dessus : les jeux audio	36
2.3	Petite histoire de la réalité virtuelle	38
2.4	Repenser le son pour la réalité virtuelle	42
2.5	Réalité virtuelle et spatialisation binaurale : un rapproche- ment instinctif	45
2.6	Conclusion	47
3	<i>Cubes</i> : tests perceptifs sur la spatialisation binaurale pour un jeu en réalité virtuelle	48
3.1	Mise en contexte	49
3.2	Contenu du jeu	51
3.3	Description de l'expérience et résultats attendus	52
3.4	Protocole de mesure	54
3.4.1	Outils utilisés	54
3.4.2	Phase de sélection d'HRIR	55
3.4.3	Phases de jeu	59
3.5	Analyse des résultats	66
3.5.1	Sélection d'HRIR	66
3.5.2	Phases de jeu	68
3.5.3	Conclusion	76
	Conclusion générale	77
	Annexes	80
	ANNEXE 1 : Sélection d'HRIR : mise en place	80
	ANNEXE 2 : Sélection d'HRIR : résultats complets	87

ANNEXE 3 : Projet Unity de <i>Cubes</i>	89
ANNEXE 4 : Phases de jeu : résultats complets	99
Temps de jeu global	99
Différence d'angle observée	106

Bibliographie	109
----------------------	------------

Introduction

Les dispositifs de réalité virtuelle à destination du grand public, tels que l'Oculus Rift d'Oculus VR, le HTC de Valve et à des tarifs beaucoup plus abordables les dispositifs pour smartphones sont en plein essor. Preuve de cet engouement, la toute récente conférence AES à Los Angeles en septembre 2016 qui a fait de l'audio pour la réalité virtuelle et augmentée son thème principal¹

Cette nouvelle technologie promet de remettre en cause la façon de penser le jeu vidéo, en proposant un nouveau rapport à l'image et au son, mais aussi de nouveaux mécanismes d'interaction entre le joueur et l'objet interactif, vidéoludique ou non.

Les constructeurs semblent s'accorder cependant sur un aspect : l'expérience sonore passera par l'utilisation des techniques de spatialisation binaurale qui permettent, en tenant compte de l'incidence de la tête d'un individu sur le son, de filtrer et recréer artificiellement au casque des sons provenant de n'importe quelle direction (c'est la spatialisation binaurale), ce qui est impossible avec une stéréo classique.

1. *Audio for virtual and augmented reality, 2016 AES Conference*, programme officiel de la conférence, 2016 URL : http://www.aes.org/conferences/2016/avar/downloads/Full_AVAR_Conference_Program_V3.pdf

Les fabricants promettent une immersion et une expérience renforcée grâce à l'utilisation de ce procédé, déjà parfois employé dans des contextes radiophonique ou accompagnant de la vidéo interactive ou non. De plus l'emploi de HRIR de façon quasi-systématique dans le cadre de la VR est envisageable car la plupart des dispositifs sont basés sur un système d'écoute au casque, condition essentielle pour son application.

Un problème se pose cependant : les HRIR (pour Head Related Transfer Function, la fonction de transfert de la tête) sont, par essence, personnelles, car elles dépendent de la forme de la tête, qui est propre à chacun. L'établissement d'une HRIR élaborée pour chaque utilisateur de la VR ne semble pas envisageable, car elle requiert des dispositifs lourds (chambre anéchoïque, micros de mesure spéciaux qui peuvent être insérés dans les conduits auditifs) qui sont démesurés par rapport à l'équipement possédé par le grand public.

Les compagnies de VR proposent une solution : le joueur peut se passer d'une HRIR personnalisée car il serait possible de s'habituer à celle d'un autre individu.² Ce phénomène d'accoutumance est démontré³ synthèse binaurale, mais prend normalement un temps d'adaptation non-négligeable. Le support vidéo interactif serait alors un moyen d'accélérer l'adaptation, grâce au headtracking, qui suit les mouvements de la tête, ainsi qu'à l'effet ventriloque, qui stipule que notre cerveau fait naturellement corrélérer ce qu'il voit et ce qu'il entend dans l'espace.

C'est cette hypothèse que ce mémoire cherche à vérifier : peut-on réellement s'habituer à l'HRIR d'un autre, et cette prise en main est-elle effectivement suffisamment rapide (de l'ordre de quelques minutes) pour ne pas

2. *Oculus Audio SDK Documentation*, manuel à destination des développeurs pour la plateforme Oculus, v1.1.4, p12, 2017

3. Blum A., Katz B., Warufsel O., "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training presence", CFA-DAGA, 2004

gêner l'utilisateur ? La performance d'un joueur à un jeu sera-t-elle impactée si il joue avec une HRIR qui ne lui correspond pas ? L'utilisation des HRIR elle-même apportent-elles réellement un plus dans l'expérience de jeu ?

Pour répondre à ces problématiques, on se penchera dans un premier temps plus en détail sur la technologie binaurale et sur son approche adaptative, tout en expliquant pourquoi on se détourne de la personnalisation d'HRIR. Se posera ensuite la question de la place du son binaural dans le jeu vidéo et dans la réalité virtuelle. Enfin, des tests perceptifs basés sur une expérience vidéoludique en réalité virtuelle (*Cubes*) seront réalisés afin de déterminer si les joueurs parviennent, à leur insu, à s'adapter rapidement à une HRIR qui n'est pas la leur.

Chapitre 1

Spatialisation binaurale : le choix de l'approche adaptative

Dans cette partie, on commencera par rappeler brièvement le fonctionnement général de la spatialisation binaurale, puis l'approche de l'individualisation et celle de l'adaptation seront présentées et critiquées afin de justifier le choix de cette dernière comme base de l'expérience du mémoire.

1.1 Principe de la spatialisation binaurale

La spatialisation binaurale est un procédé technique qui permet en théorie de recréer un espace tridimensionnel lors d'une écoute avec un casque audio. Là où une écoute stéréo classique ne permet d'entendre un déplacement du son que sur le plan horizontal, une écoute binaurale permettra de recréer un déplacement sur le plan vertical (zénithal) et sur l'ensemble du plan horizontal (avant comme arrière). Cette forme de spatialisation nécessite obligatoirement un casque audio, car elle permet de s'affranchir de l'impact qu'a la pièce sur le son dans le cas d'une écoute avec des enceintes. Des solutions d'écoute dites transaurales sont proposées dans le cadre d'une

diffusion avec des enceintes, mais elles ne seront pas développées dans ce mémoire, car elles n'en font pas l'objet.

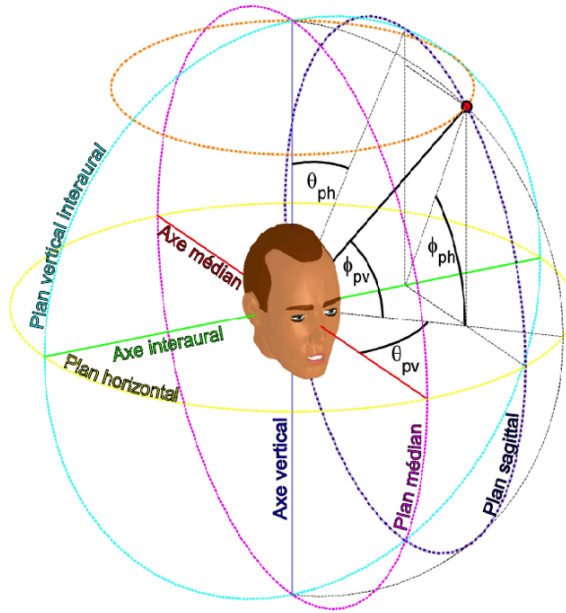


fig1. Les différents axes de repère, extrait de la thèse de P. Guillon ¹

Ce procédé est rendu possible par l'utilisation d'une HRTF (Head Related Transfer Function) ou d'une HRIR (Head Related Impulse Response), qui correspondent respectivement à la fonction de transfert et à la réponse impulsionnelle créée par l'interaction entre le son et le corps d'un individu, plus particulièrement sa tête. En parvenant à capturer cette réponse pour chaque oreille et pour chaque position possible autour d'un individu, on peut

1. Guillon P. *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*. Thèse sous la direction de Simon L., Université du Maine, département acoustique, Novembre 2009, p.5

donc recréer artificiellement les indices temporels et spectraux propres à un individu qui lui permettent de localiser une source sonore dans l'espace.

Pour obtenir cette information, le protocole de mesure le plus répandu consiste à placer des micros dans les conduits auditifs d'un sujet (ou d'une tête artificielle) et d'émettre des stimuli sonore de toutes les directions jugées nécessaires en azimuth et en élévation pour couvrir assez précisément le champ d'écoute. On récupère alors une réponse impulsionnelle par oreille et par position pour un sujet donné.²

Car il s'agit bien en effets d'indices propres à un individu, c'est à dire qu'une HRTF³ pourra grandement changer selon la morphologie du sujet, et que si la spatialisation qui en découle peut paraître adéquate pour une personne sur laquelle a été effectuée la mesure, ce constat n'est pas forcément valable pour une autre. C'est là une des grandes difficultés qui s'oppose à une utilisation systématique de ce procédé dans les contenus diffusés au casque, et une des problématiques soulevées dans ce mémoire et auxquelles l'on va tâcher de répondre.

1.1.1 Indices de localisation du son

Pour localiser un son, nous faisons appels à des indices acoustiques basés sur la différence de perception entre les deux oreilles dans le domaine temporel, de niveau et fréquentiel⁴ :

2. Andreopoulou A., Begault D., Katz B., "Inter-Laboratory Round Robin HRTFF Measurement Comparison", *IEEE J Selected Topics in Signal Processing*, vol. 9, no. SI : *Spatial Audio*, 2015

3. En réalité, il ne s'agit pas d'une HRTF ou d'une HRIR, mais d'un ensemble d'HRTF ou d'HRIR, chacune correspondant à une mesure dans une direction précise qui ensemble forment la réponse binaurale d'un individu. Par souci d'alléger la formulation, c'est à cet ensemble que l'on se réfère quand on parle d'une HRTF ou d'une HRIR dans ce mémoire

4. Rugeles Ospina F, *Individualisation de l'écoute binaurale : création et transformation des indices spectraux et des morphologies des individus*, Thèse sous la direction de

- L'ITD (Interaural Time Delay), qui permet une localisation latérale en comparant le temps que met un son à parvenir à chaque oreille. Cet indice est précis dans les fréquences basses (jusqu'à 1,5KHz environ)

- L'ILD (Interaural Level Difference), qui permet de prendre le relais dans les fréquences plus aiguës, en analysant la différence de niveau perçues entre les deux oreilles.



fig2. ITD

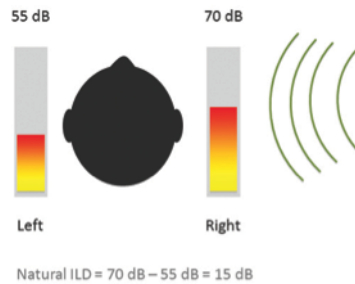


fig3. ILD

Katz B., Université Pierre et Marie Curie, département sciences mécaniques, acoustique, électronique et robotique, soutenu en 2016, pp11-20

- La localisation du son est également renforcée par des petits mouvements de tête inconscients, qui permettent de déterminer de façon dynamique la position d'une source sonore en comparant la façon dont elle est entendue de plusieurs angles d'écoute.

- Les indices spectraux servent quant à eux principalement pour la localisation en élévation et avant/arrière en prenant en compte l'impact de notre corps sur la réponse fréquentielle du son. C'est un paramètre qui nécessite un apprentissage et qui est propre à chacun, selon sa morphologie, là où les deux indices précédents sont quantifiables de façon plus générale, même si ils restent en partie dépendants de la morphologie (la taille du crâne et donc la distance entre les deux oreilles peut influencer sur l'ITD et l'ILD).⁵

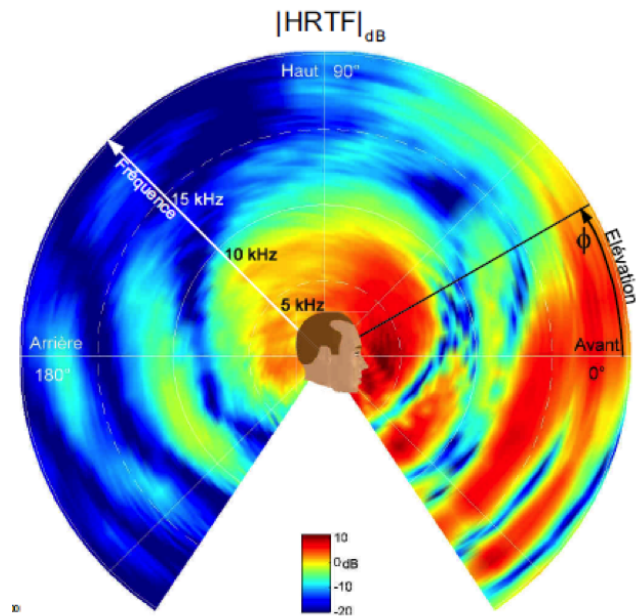


fig4. Indices spectraux

5. Blauert J., *Spatial hearing : The Psychophysics of Human Sound Localization*, édition revisitée, 1997

1.1.2 Binaural statique et dynamique

D'un côté, il y a le binaural statique, c'est à dire qui correspond à une scène sonore dont le point d'écoute est fixe. Pour cela, il est possible de recourir à un enregistrement en binaural natif, fait à l'aide de micros placés dans les conduits auditifs. À l'écoute, l'auditeur entend la scène avec la même spatialisation qu'au moment de l'enregistrement, il est impossible de replacer dans l'espace n'importe quel élément sonore puisqu'il s'agit d'une prise de son globale (comme par exemple lorsque l'on place un couple stéréophonique dans un prise de son musicale ou d'ambiance). Dans ce cadre là, la question de la mesure des HRIR n'est plus pertinente, puisque de toutes façons il est impossible de les exploiter pour retravailler la spatialisation du son. Il est également possible de faire du binaural statique de synthèse, en respatialisant le son de façon artificielle en se basant sur la réponse binaurale d'un individu précis.

C'est un méthode qui peut être utilisée dans le cadre de la fiction ou du documentaire radio, ou encore parfois pour accompagner de la vidéo, mais qui présente des limites dès qu'il est possible pour le spectateur d'évoluer plus précisément dans le média, dans le cadre par exemple de la vidéo 360° ou du jeu vidéo dans lequel le point de vue est tributaire de l'utilisateur. Un point de vue sonore fixe avec une image changeante est en effet assez perturbant.

De l'autre côté, le binaural dynamique permet de se déplacer dans une scène sonore et de réactualiser en permanence la positions d'objets sonore par rapport à la position de l'auditeur. Chaque élément sonore est enregistré séparément puis filtré selon les caractéristiques d'une HRIR précise pour une position donnée entre l'utilisateur et l'objet. C'est la méthode qui est utilisée dans le cadre du jeu vidéo par exemple, en faisant évoluer la spatialisation du son en fonction de ce qui est affiché à l'écran. Cette méthode est aussi

implantée dans le cadre de la réalité virtuelle, en se basant sur des techniques de suivi des mouvements de la tête de l'auditeur, ou "headtracking", terme employé dans la littérature sur le sujet et qu'on utilisera par la suite dans ce mémoire. Ainsi, la spatialisation du son dépend directement de la direction dans laquelle regarde l'utilisateur. C'est logiquement cette méthode qui retiendra donc notre attention pour la suite de ce mémoire.⁶

Le problème de cette méthode réside dans le fait que la mesure d'HRIR est faite sur un nombre fini de points autour du sujet, il faut donc résoudre le problème du passage d'un point à un autre lors d'un déplacement dans l'espace sonore, et ce sans que la latence de calcul entre le mouvement effectué et l'adaptation du son qui en résulte soit suffisamment importante pour être entendue, ou brise la cohérence entre le message sonore et l'interface visuelle. L'ordre de grandeur maximal que peut prendre cette latence a été chiffrée aux alentours de 30ms.⁷

Pour recréer artificiellement les positions intermédiaires entre les points mesurés, il existe plusieurs solutions : une HRIR peut être décomposée sous la forme d'un retard en série avec un filtre à phase minimale pour chaque position dans l'espace et chaque oreille. On peut faire des interpolations entre deux points de l'espace en temps réel à l'aide de filtres à réponse finie (FIR) ou infinie (IIR). En mettant à jour les coefficients de ces filtres, on peut changer la position du son dans l'espace. La mise à jour des filtres est un facteur qui risque de faire apparaître des artefacts (clics). Des solutions à ce problème ont été proposées dans la thèse de Véronique Larcher sur la spatialisation des sons pour la réalité virtuelle⁸

6. Honda A., Shibata H, Hidaka S., Gyoba J., Iwaya Y et Suzuki Y., "Effects of head movement and proprioceptive feedback in training of sound localization", *iperception*, Juin 2013

7. Larcher V. *Techniques de spatialisation des sons pour la réalité virtuelle*, thèse sous la direction de Polack J.-D., Université Paris VI, soutenu en mai 2001, p73

8. Larcher V, *op. cit.*, pp37-76

On peut également utiliser un système HOA (pour High Order Ambisonic), basé sur le principe de l'ambisonic dans lequel est simulé un champ sonore en 3 dimensions sous la forme de quatre canaux qui contiennent sa description mathématique, c'est le B-Format. En augmentant l'ordre de l'ambisonic, on augmente le nombre de canaux, et on crée un maillage spatial plus précis.⁹

1.1.3 Quel format de fichier pour le binaural ?

Afin de rendre transportable et exploitable universellement les données mesurées pour l'HRTF d'un individu, le regroupement BiLi, un groupe recherche sur le binaural composé notamment de France Télévisions, Radio France, Orange Labs, l'IRCAM et le CNSM a proposé un format, appelé SOFA (Spatially Oriented Format for Acoustics).^{10 11}

Il présente une grande flexibilité, car il encapsule dans ses métadonnées les conditions de mesures de la réponse binaurale, et non pas forcément une grille normalisée de points dans l'espace, ce qui veut dire qu'on peut lui donner une réponse mesurée dans n'importe quel laboratoire, pour peu qu'on connaisse précisément les conditions de mesure. Le format est basé sur les concepts de *Listeners* (l'auditeur, par exemple la tête), de *Receivers* (récepteurs, les oreilles du sujet ou les micros de mesure), d'*Emitters* (émetteurs, haut parleur), *Source* (l'ensemble des émetteurs), *Room* (la pièce). L'ensemble de ces facteurs permet d'avoir un fichier contenant des données suffisamment précises sur la constitution de la réponse binaurale utilisée. C'est donc vers ce format que l'on se penchera pour intégrer du binaural

9. Hollerweger F., "An introduction to Higher Order Ambisonic", article en ligne, 2008, URL : <http://flo.mur.at/writings/HOA-intro.pdf>

10. <http://www.bili-project.org>, site internet du projet Bili

11. Majdak P., Noisternig M., "SOFA, Spatially Oriented Format for Acoustics", version 0.6, 2015, URL : <https://www.sofaconventions.org/mediawiki/index.php>

dans la partie pratique du mémoire.

1.1.4 L'approche adaptative contre l'individualisation

Maintenant que la spatialisation binaurale est définie et que le problème de son intégration dans un projet avec de la spatialisation en temps réel est résolu, il convient de se poser la question suivante : comment choisir les HRIR que l'on va intégrer dans l'expérience du mémoire ? Deux approches sont proposées : celle de la personnalisation, qui consiste à créer une HRTF propre à chaque individu désireux d'écouter du contenu spatialisé en binaural, et celle de l'adaptation, dans laquelle on considère qu'il est possible de s'habituer à une HRTF non-individualisée. Ces deux solutions seront détaillées par la suite.

1.2 Méthode de l'individualisation

De façon utopique, si il existait une manière accessible à tous qui permette de mesurer facilement une réponse binaurale personnalisée, ce mémoire n'aurait pas lieu d'être. Cependant, les méthodes actuelles sont encore loin de permettre cela, et les obstacles principaux au développement du binaural personnalisé sont décrits ci-après.

1.2.1 Lourdeur du dispositif

La mesure d'une HRTF nécessite une installation qui ne peut être mise en place qu'en laboratoire, dans une chambre anéchoïque équipée d'une structure métallique sphérique sur laquelle sont disposées une ou plusieurs enceintes mobiles ou fixes (on compte selon les installations jusqu'à trente cinq enceintes).

Le sujet (ou la tête artificielle dans certains cas) est placé au centre de la sphère avec des micros de mesure dans les conduits auditifs et doit rester parfaitement immobiles pendant que des sons sont émis de toutes les incidences qui sont jugés nécessaires pour constituer une réponse binaurale personnalisée suffisamment précise. Cela va de quelques centaines à plus de deux mille mesures.

Cette procédure élimine l'individualisation comme une méthode utilisable pour le plus grand nombre. Il leur est en effet impossible d'accéder au laboratoire de mesure, ou d'accéder à l'individualisation depuis leur salon.

1.2.2 Transportabilité des mesures

La mise en place du format SOFA (2013) étant récente comparée au début des études et des mesures sur les HRTF, il existe autant de protocoles de mesures et de formats que de laboratoires, ce qui peut conduire à un problème lors de l'intégration d'une réponse binaurale dans un projet extérieur à son laboratoire d'origine, toutes les bases de données n'ayant pas encore été converties.

Plus gênant encore, on constate que pour des mesures effectuées avec la même tête artificielle dans des centres de recherche différents, on obtient des différences significatives de la réponse finale obtenue. De même, des mesures effectuées par une même équipe à plusieurs années d'intervalles présenteront des différences non négligeables.¹²

Ces écarts peuvent s'expliquer par le nombre trop important de paramètres pouvant changer d'un protocole de mesure à l'autre (placement des micros dans les oreilles, fréquence d'échantillonnage, nombre de haut-

12. Andreopoulou A., Begault D., Katz B., "Inter-Laboratory Round Robin HRTFF Measurement Comparison", *IEEE J Selected Topics in Signal Processing*, vol. 9, no. SI : *Spatial Audio*, pp. 895-906, 2015

parleurs utilisés, nombres de positions mesurés, distance entre la tête et les enceintes, positions retenues pour les mesures en azimuth et en élévation, nature des stimuli utilisés...)

Autre problème : celui du casque utilisé. En effet, la réponse en fréquence d'un casque audio change considérablement entre des écouteurs bon marché et un casque hifi ou utilisé dans un cadre professionnel. Or, pour bien restituer une réponse binaurale personnalisée, il est important de mettre en place une égalisation qui permette de compenser la réponse en fréquence du casque. Il n'y a pas encore, à l'heure actuelle de base de données suffisamment exhaustive ou de programme informatique qui propose une égalisation automatique pour les casques utilisés dans un cadre de loisirs.¹³

1.2.3 Pistes à explorer

Dans son mémoire *Création d'une interface interactive à usage domestique dans le but de créer un modèle HRTF personnalisé*¹⁴, Vincent Zanetto propose une approche hybride, dans laquelle il serait possible de créer une réponse binaurale personnalisable en combinant les éléments de différentes HRIR déjà existantes.

De même, dans sa thèse *Individualisation de l'écoute binaurale : création et transformation des indices spectraux et des morphologies des individus*¹⁵, Felipe Rugeles Ospina essaye de proposer une solution permettant d'aboutir à une HRIR individualisée en se basant uniquement sur la connaissance morphologique d'un sujet (en faisant une modélisation 3D). Cette thèse part du

13. Rugeles Ospina F, *Individualisation de l'écoute binaurale : création et transformation des indices spectraux et des morphologies des individus*, Thèse sous la direction de Katz B., Université Pierre et Marie Curie, département sciences mécaniques, acoustique, électronique et robotique, soutenu en 2016, p34

14. Zanetto V., *Création d'une interface interactive à usage domestique dans le but de créer un modèle HRTF personnalisé*, mémoire effectué sous la direction de Blum A., École Nationale Supérieure Louis Lumière, soutenu en juin 2017

15. Rugeles Ospina F, *op. cit.*

constat que la réponse binaurale peut être caractérisée et catégorisée uniquement grâce à la morphologie d'un individu. En se basant sur des HRIR déjà mesurées et dont la morphologie du sujet est connue, il serait possible d'obtenir une nouvelle HRIR en se basant sur les différences de morphologie. Cela permet de se passer des structures de mesures très lourdes, mais demande tout de même une puissance et une durée de calcul en amont de l'utilisation trop importante pour la plupart des ordinateurs à usage domestique. Il est cependant imaginable qu'à terme, grâce à l'évolution de la technologie ce genre de solution puisse être employée par le plus grand nombre.

Il s'agit là finalement de solutions que l'on pourrait qualifier d'hybrides plutôt que personnalisées, car elles se basent sur des réponses binaurales déjà existantes, qui sont ensuite modifiées pour créer une nouvelle réponse non pas totalement individuelle, car non-mesurée, mais en tout cas améliorée.

1.3 Approche adaptative

Si on compare les performances immédiate de localisation sonore d'un individu avec une HRIR personnalisée et une qui ne l'est pas dans le cadre d'une écoute statique, on s'aperçoit effectivement qu'il est bien plus efficace dans le premier cas, et ce d'autant plus si une égalisation de correction a été appliquée au casque sur lequel le test est effectué.

Cependant, l'approche adaptative part du principe qu'un individu soumis à une HRIR qui n'est pas la sienne finira par s'y habituer et présentera alors des performances de localisations similaires qu'avec sa propre réponse binaurale, ou en tout cas suffisamment satisfaisante pour ne pas altérer son expérience. Pour palier aux défauts de cette réponse, qui à l'avantage d'être bien plus simple à mettre en pratique, et pour essayer d'accélérer ce phénomène, on aura recours à plusieurs éléments présentés dans cette partie.

1.3.1 Temps d'adaptation

La plasticité du système auditif a été démontrée à plusieurs reprises chez l'être humain, et les expériences généralement proposées, telles que l'inversion du signal reçu par l'oreille droite et l'oreille gauche, ou la mise en place de moules qui influencent l'incidence du son sur les oreilles et donc la perception de la localisation, sont concluantes, et montrent même que l'apprentissage observé est pérenne, c'est à dire qu'en renouvelant les tests après un arrêt de l'expérience, les performances observées sont similaires. Cependant, ces tests se déroulent sur des périodes de plusieurs jours voir plusieurs semaines.¹⁶

C'est le problème majeur de l'approche adaptative. En effet, l'adaptation à une HRIR non-individualisée n'est pas un processus instantané, ce qui peut poser problème dans le cadre d'une activité de type loisir, où il n'est pas désirable de passer plus de temps à s'adapter à une réponse binaurale qu'à se consacrer à son divertissement. Cependant, on note qu'il est possible de faciliter cette adaptation en s'appuyant sur d'autres éléments que l'audition pure.¹⁷

1.3.2 Apport de la vision

La vision permet un gain considérable dans la précision de la localisation d'un son, et ce de façon instantanée. Cela viendrait de l'effet ventriloque, qui suppose que notre cerveau fait automatiquement corrélérer dans l'espace un son et l'image qui s'y rapporte (comme par exemple l'utilisation du canal centre pour les dialogues dont on voit les interlocuteurs apparaître à l'image

16. Parseihian G., *Sonification binaurale pour l'aide à la navigation*, Thèse sous la direction D'Alessandro C., spécialité acoustique, Université Pierre et Marie Curie, octobre 2012, pp47-49

17. Blum A., *Étude de la plasticité du système auditif en localisation sonore. Application au problème de l'individualisation en synthèse binaurale*, rapport de DEA, Université d'Aix-Marseille 2, 2003

au cinéma et qui nous paraîtront cohérents même si les personnages sont décentrés). Il a été étudié et démontré que la précision dans le repérage de la spatialisation d'un son est supérieure quand une image lui est attachée. De même, lorsque l'on décorrèle une image et le son qui lui est associé, il y a une tendance naturelle à les réassocier mentalement.¹⁸

Plus intéressant encore, la perception de la spatialisation du son est dépendante de la vision. Certaines études avancent cette hypothèse, par exemple en comparant la perception de la spatialisation entre des personnes aveugles de naissance et des personnes voyantes placés dans le noir. Il en résulte que la perception est la même sur le plan azimuthal, mais que dans le cas de l'élévation les personnes voyantes ont un avantage, ce qui pourrait montrer un apprentissage de ce paramètre plus précis lié à la vision.¹⁹

Dans une autre expérience, des sujets qui portent des lunettes compressant leur champ de vision voient leur capacité à localiser des sons dans le plan azimuthal diminuer de façon comparable à leur vue. Cependant, cette expérience avait lieu après que les sujets aient porté les lunettes plusieurs jours.²⁰

Tous ces articles tendent vers la même conclusion sur la prédominance de la vision sur la localisation dans l'espace sonore. Cet avantage pourra donc largement être mis à profit dans le cadre du jeu vidéo et de la réalité virtuelle, tout du moins dans les cas où les sources sonores sont rattachés à un objet physique, et qui est visible à l'écran. Un problème récurrent de la spatialisation binaurale est celui de la confusion avant/arrière dans le cadre

18. Nguyen K., Suied C., Viaud-Delmon I., Warusfel O., "Spatial audition in a static virtual environment : the role of auditory-visual interaction.", *JVRB - Journal of Virtual Reality and Broadcasting* n° 5, 2009

19. Zwiers M. P., Van Opstal A. J., Cruysberg J. R. MR., "A spatial Hearing Deficit in Early-Blind Humans", *The Journal of Neuroscience*, Vol. 21 RC142, 2001

20. Zwiers M. P., Van Opstal A. J., Paige G. D., "Plasticity in human sound localization induced by compressed spatial vision", *Nature Neuroscience* 6, pp175 - 181, 2003

d'une écoute à l'aveugle, mais il pourrait donc possiblement être en partie résolu dans notre cas.

1.3.3 Apport du binaural dynamique

Avec le binaural dynamique, on introduit la notion d'interactivité, et donc de pouvoir se positionner par rapport à un son pour le localiser avec plus de précision (par exemple avec du headtracking). Cela permet, par rapport au binaural statique de retrouver un élément de localisation du son en plus, à savoir les mouvements de tête, qui permettent de se repérer de façon relative, et non plus seulement absolue.

Les études mises en places proposant des tests de localisation avec des HRIR non-individualisées tout en utilisant du headtracking sont plutôt encourageantes, car elles montrent des signes d'adaptation assez rapides, avec des sessions de test pouvant être diminuées jusqu'à l'ordre de la dizaine de minutes.²¹

1.3.4 Selection d'HRTF

Une fois la décision prise de se passer d'individualisation pour se reposer sur des réponses déjà existantes, il faut déterminer comment les choisir. Plusieurs études ont été effectuées pour rendre compte de la compatibilité d'une HRTF pour une population donnée. C'est notamment le cas d'une étude qui fait appel à 46 sujets issus de la base de données LISTEN de l'IRCAM, qui regroupe une soixantaine de réponse binaurales correspondant chacune à un individu dont l'HRTF a été mesurée dans les mêmes conditions à l'IRCAM. Ces réponses binaurales on l'avantage de toutes être disponibles

21. Parseihian G., *Sonification binaurale pour l'aide à la navigation*, Thèse sous la direction D'Alessandro C., spécialité acoustique, Université Pierre et Marie Curie, octobre 2012, pp51-53, p76

au format SOFA.

Dans cette expérience, chacun des sujets effectue des tests perceptifs avec sa propre HRTF ainsi que celle des autres puis les juge subjectivement. De cette étude découle le constat suivant : chaque sujet réagit effectivement bien à sa propre réponse, mais également que certaines réponses sont adaptées à un nombre important de personnes. De ces réponses compatibles avec beaucoup de sujets découlent les "golden ears" : un regroupement de sept HRTF qui couvrent un panel de morphologies suffisamment large pour que au moins l'une d'entre elles soit considérée comme bonne par n'importe quel sujet. C'est elles qui seront utilisées dans le cadre de ce mémoire comme références.²²

1.4 Conclusion

Après avoir passé en revue les approches de la personnalisation et de l'adaptation des HRIR, nous avons pu constater l'impossibilité pour le plus grand nombre d'accéder à une spatialisation binaurale individualisée facilement. En faisant le choix de se baser sur une méthode adaptative, on constate malheureusement que c'est un choix qui prend un temps non négligeable, mais qui peut être amélioré soit par l'apport de la vision, soit par celui du binaural dynamique. Peut être que la combinaison de ces deux facteurs permettrait un apprentissage encore plus rapide, c'est donc ce que nous allons essayer de mettre en place dans la partie pratique de ce mémoire. Pour cela, on proposera une série de tests perceptifs sur la spatialisation binaurale à l'aide d'un dispositif de réalité virtuelle, ce qui induit un support visuel et du suivi des mouvements de la tête.

22. Schönstein D., Katz B., "Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale.", Société Française d'Acoustique, 10ème Congrès Français d'Acoustique, Lyon, France. 2010

Chapitre 2

Binaural, jeux vidéo et réalité virtuelle

Définir la façon d'intégrer de la spatialisation binaurale dans des médias interactifs (en l'occurrence le jeu vidéo, et plus précisément sur dispositif de réalité virtuelle) est un premier pas, mais encore faut-il déterminer l'intérêt qu'apporte l'utilisation de cette technologie. C'est ce qu'on se propose de rechercher dans cette deuxième partie, en commençant par se questionner sur la place du son dans une expérience vidéoludique, puis en analysant les propositions déjà faites d'intégration de spatialisation binaurale dans des jeux vidéo pour en mesurer l'apport spécifique. Enfin, on se demandera dans quelle mesure ce raisonnement est extensible au support de la réalité virtuelle.

2.1 Du rôle du contenu audio dans le jeu vidéo

La place de l'audio dans le jeu vidéo est un sujet très vaste, qui mériterait de faire l'objet d'un mémoire à part entière. C'est d'ailleurs le cas de plusieurs travaux effectués au sein de l'école Louis Lumière, qui se sont posés la question de l'écriture du son dans le jeu vidéo de façon générale¹, ou qui proposent une approche recentrée sur certaines composantes essentielles de la bande sonore d'un jeu, comme la musique², le rôle de la voix³, ou bien encore le lien entre le son et l'apparition d'une émotion dans un jeu⁴.

Avant de poser la question du son, il convient de se demander dans un premier temps ce qui définit un jeu vidéo. Tous ces mémoires s'accordent pour dire que ce qui le différencie des autres médias est l'interactivité, c'est à dire le fait que le joueur puisse agir sur ce qu'il voit et entend, là où il ne sera que spectateur dans le cas du cinéma. Cependant, ils relèvent également que cette définition n'est pas suffisamment exhaustive, car il existe des jeux qui ne contiennent pas d'audio, ou pas de graphismes. Ce serait alors le système de règles établit par le jeu, couplé à l'interaction qui le rendrait distinct des autres médias. Cependant, cette définition correspond à celle d'un jeu de société. Dans son traité sur la philosophie des jeux vidéo, Mathieu Tricot propose la définition suivante, sous forme d'un faux dialogue entre Socrate et Mario :

1. Berger C., *L'écriture sonore dans les jeux vidéo*, mémoire effectué sous la direction de Halbwachs J.-P., École Nationale Supérieure Louis Lumière, soutenu en juin 2010

2. Soulier A., *Musique et jeux vidéo, de l'interactivité d'un média singulier*, mémoire effectué sous la direction de Rouchouse J., École Nationale Supérieure Louis Lumière, soutenu en juin 2016

3. Meyer C., *Du mutisme au dialogue : les interactions vocales dans le jeu vidéo*, mémoire effectué sous la direction de Coduys T., École Nationale Supérieure Louis Lumière, soutenu en juin 2015

4. Palacin B., *Le rôle du son dans l'apparition de la peur dans le jeu vidéo*, mémoire effectué sous la direction de Gazeau C., École Nationale Supérieure Louis Lumière, soutenu en juin 2013

"- S : Le jeu vidéo, il me semble que nous le reconnaissons d'abord comme une certaine forme d'expérience, une «expérience instrumentée», bien sûr, qui a besoin de l'écran et de la machine de calcul pour se produire. Et, avec cela, nous nous mettons dans un certain état, un «état ludique» qui ne ressemble à aucun autre. Au fond, ce n'est pas très différent de ce que produisent la lecture ou le cinéma. Ce sont aussi des technologies culturelles.

- M : Comment cela ?

- S : Elles exploitent un dispositif technique particulier, le livre, le film, la salle de projection, l'écran, pour produire une certaine forme d'expérience. Disons un état livresque ou un état filmique. Et ces états instrumentés sont des états de retrait vis-à-vis du cours ordinaire de la vie sociale ; ce qui n'a pas été moins reproché à la lecture et au cinéma qu'aux jeux vidéo.

- M : Bien sûr !

- S : Et, pourtant, ces états ne sont pas les mêmes. Ils ne se ressemblent pas. Donc, du côté du jeu vidéo, je ne sais pas comment on pourrait le définir sans prendre en compte ces formes d'expérience, qui sont ressenties par le joueur. On doit se dire que le jeu vidéo propose une expérience spécifique, un jeu avec des univers simulés, engendrés par le calcul, où il y a de l'habileté souvent, et puis une certaine profondeur dans la simulation. Sinon, ce sont juste des jeux sur ordinateur."⁵

Cette définition est assez complexe, mais elle présente l'intérêt de mettre en avant, par ce qu'elle appelle "l'état ludique", l'importance de la relation entre le joueur et le jeu, tout comme elle met en avant cette relation dans le cas du cinéma ou de la lecture, créant ainsi un parallèle entre tous ces arts. Plus que le réalisme d'un jeu vidéo, ou la qualité technique de sa programmation, de ses graphismes et de sa réalisation sonore ce qui compte est de

5. Triclot M., *Philosophie des jeux vidéo*, Paris, Zones, p6, 2011

créer un univers cohérent afin d'instaurer un pacte entre le joueur et le jeu.⁶

C'est en effet cela qui va permettre une expérience immersive et plaisante, car le joueur accepte les incohérences entre le jeu et le monde réel, et accepte, le temps que dure sa partie, d'intégrer l'univers qui lui est présenté à sa propre réalité.

C'est vrai pour le jeu, comme ça l'est pour le cinéma ou la lecture, et c'est ce qui explique qu'un joueur puisse faire un retour aussi positif d'une partie de quelques minutes de *Tétris* (développé par Alekseï Pajitnov en 1984) où il se contentera d'aligner des cubes en écoutant une musique 8bit qui s'efforce tant bien que mal de reproduire Korobeiniki de Nikolaï Alekseïevitch Nekrassov, que d'une session de plusieurs heures de jeu sur un jeu AAA (équivalent d'un blockbuster au cinéma) comme *Dark Souls III* (développé par From Software en 2016), qui présente des graphismes ultra-réalistes, une musique interprétée par un orchestre symphonique, des sons spatialisés en 7.1, et une histoire d'une riche complexité qui s'étend sur plus de 100 heures de jeu.

Bien sûr, entre ces deux extrêmes, il y a une quantité quasi-infinie de jeux, dont le pacte est unique selon le genre, la date de sa parution, le support sur lequel il est supposé être joué entre autres. C'est donc au développeur du jeu de définir et de mettre en place ce pacte qui permettra d'instaurer "l'état ludique" proposé par M. Triclot, via des lignes de codes informatiques qui définiront par calcul un environnement et les règles qui le régiront. Transgresser ce pacte, du côté du joueur comme celui du développeur revient à briser l'expérience et à faire prendre conscience au joueur de sa non-appartenance à l'expérience qui lui est proposée, à l'instar de l'ac-

6. Triclot M., *Philosophie des jeux vidéo*, Paris, Zones, 2011, p91

teur qui brise le quatrième mur au cinéma ou au théâtre. Cela peut d'ailleurs parfois être fait exprès, afin de faire prendre une nouvelle dimension à l'interaction créée par le jeu. Charles Meyer propose une analyse de deux de ces jeux dans lequel le pacte est brisé par l'utilisation du son dans son mémoire, *The Stanley Parable* (Galactic Cafe, 2013) et *Metal Gear Solid 2 : Sons of Liberty* (Konami, 2001).⁷ Dans ces jeux, le voix du narrateur s'adresse directement au joueur plutôt qu'à son avatar, ce qui le désarçonne dans un premier temps, puis le conduit à analyser son expérience d'interaction entre le personnage qu'il dirige et le monde du jeu auquel il joue autrement qu'au premier degré.

Si cette réflexion est valable pour n'importe quel jeu dans sa globalité, c'est donc aussi à elle qu'il faudra se rattacher quand vient le moment de penser sa réalisation sonore. En tant que Lead Sound (responsable de la partie sonore d'un jeu vidéo), c'est en ayant ce pacte en tête qu'il faudra créer et penser l'intégration de chaque son afin de respecter l'univers proposé par le directeur créatif. Ce n'est pas parce que l'on dispose d'outils technologiques permettant d'intégrer des fonctionnalités très complexes qu'il faudra systématiquement en faire usage, ou parce que l'on peut implanter une quantité de son de quasi-infinie qu'il faut engorger tout l'espace disponibles avec des sons qui ne sont pas forcément judicieux dans le contexte de l'expérience vidéoludique. Avoir une richesse de son démesurée par rapport au reste du jeu peut nuire à l'expérience, et inversement.

Les sons utilisés dans le cadre d'un jeu peuvent être classifiés en plusieurs catégories, selon ce à quoi on les rattache. Ainsi, on fera un distinguo entre les **sons entités**, liés à l'action d'un personnage où à un objet faisant partie intégrante du jeu, les **sons décors**, qui créent l'ambiance du décor composant le jeu et les **sons narratifs**, qui sont déclenchés quand une action

7. Meyer C., *Du mutisme au dialogue : les interactions vocales dans le jeu vidéo*, pp7-8

précise est effectuée.

Cependant, ce classement n'est pas suffisant, car il faut également prendre en compte le rôle auquel est destiné le son, et ce qu'il soit entité, décor ou narratif. Chaque son sert à renforcer l'**immersion** du joueur, mais ne remplit pas toujours uniquement cette fonction. Il peut également être **illustratif** et correspondre à un objet ou une situation liés aux graphismes, ou bien à une commande entrée par le joueur. Enfin, les sons **informatifs** permettent au joueur de mieux comprendre l'environnement dans lequel il se trouve, avec par exemple un tutoriel dans lequel une voix expliquera au joueur comment contrôler son avatar.

Ce double classement, issu du mémoire de Cédric Berger⁸, quand on l'associe à la notion de pacte ludique, permet de définir et de penser de façon judicieuse le rôle et l'intégration de chaque élément audio d'un jeu vidéo.

2.2 Utilisation du binaural dans le jeu vidéo

En ayant en tête la définition mise en place précédemment de l'implémentation du son dans un jeu, on se penche maintenant sur l'utilisation et l'intégration d'un procédé technique de spatialisation pour le son. Quelle est la plus-value apportée par cette technique, quand est-il judicieux de s'en servir et quels sont les exemples que l'on peut trouver de son utilisation ?

8. Berger C., *L'écriture sonore dans les jeux vidéo*, mémoire effectué sous la direction de Halbachs J.-P., École Nationale Supérieure Louis Lumière, soutenu en juin 2010, pp28-31

2.2.1 Apport sur la spatialisation

C'est en effet l'amélioration la plus logique. Dans tous les cas, l'apport du binaural passe forcément par la spatialisation, mais on parle ici de l'intérêt purement technique qu'apporte la technologie. Le binaural permet de downmixer un mixage multicanal (qui est la norme pour la plupart des jeux 3D aujourd'hui, alors que la majorité des joueurs ne disposent que d'un système stéréo chez eux) vers une écoute stéréo dans le cadre d'une écoute au casque. Le binaural de synthèse permet alors de transformer en temps réel le rendu multicanal tel qu'il serait entendu en stéréo passé à travers un filtre binaural.

Quelques jeux proposent cette fonctionnalité, mais le terme spatialisation binaurale n'étant pas très répandu dans le cadre du grand public, elle est la plupart du temps indiquée comme simple "écoute au casque" dans les menus de réglages audio à la disposition de l'utilisateur. Il n'est pas question ici de choisir le jeu d'HRIR qui va être utilisé, et qui est donc arbitrairement choisi par le développeur. C'est une des raisons principales pour laquelle on préfère dans ce mémoire s'intéresser à la question de l'adaptation plutôt que celle de la personnalisation. Ci-dessous, quelques exemples de la fonctionnalité telle qu'elle est proposée dans les options :

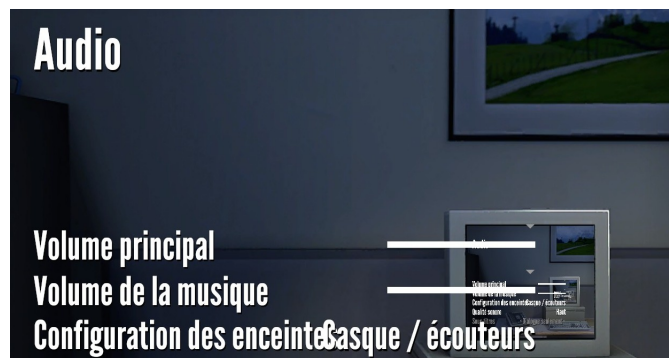


fig5. Menu audio du jeu *Stanley Parable* (Galactic Cafe, 2013)



fig6. Menu audio du jeu *The Beginner's Guide* (Everything Unlimited, 2015)

À noter que cette fonctionnalité est loin d'être systématique, parce qu'elle est compliquée à mettre en place et nécessite des outils techniques qui ne sont pas accessibles à tous les développeurs d'une part, et parce qu'elle n'est pas toujours nécessaire d'autre part. En effet, faire le choix d'une spatialisation binaurale ou même faire un mixage multicanal pour accompagner un jeu dont les graphismes sont en 2D ne présente que peu d'intérêt, au même titre qu'une ambiance réaliste ou une musique symphonique comme bande son d'un jeu en pixel art pourra paraître étrange au joueur. On en revient à la question du pacte ludique, le son doit être cohérent avec le reste du jeu (à moins bien sûr de vouloir créer un effet particulier, mais celui doit être réfléchi et assumé).

2.2.2 Le binaural comme parti pris esthétique

Dans cette partie, il est important de faire la distinction entre les enregistrements binauraux natifs et l'utilisation du binaural de synthèse. Le premier implique en effet un point d'écoute fixe là où l'autre permet de faire évoluer la spatialisation en temps réel. C'est deux technologies peuvent être utilisées de concert, mais répondent à des enjeux différents.

Dans le cas des enregistrements natifs, on trouve par exemple le jeu *Hellblade, Senua's Sacrifice* (Ninja Theory, 2017), ancré dans la mythologie nordique et dans lequel le personnage principal que l'on incarne est une jeune femme atteinte de psychose. Elle va alors peu à peu plonger dans un enfer peuplé de ses propres hallucinations. Pour faire état de ses problèmes mentaux, l'équipe son a décidé d'enregistrer les voix qu'elle entend dans sa tête en binaural. En effet, en rencontrant des personnes atteintes de problèmes mentaux, celles-ci leur ont révélé que les voix qu'elles peuvent entendre viennent souvent d'un point fixe, et non pas situé à l'intérieur de la tête. En essayant via l'enregistrement binaural natif de reproduire cette sensation de voix étranges qui viennent d'un point précis dans l'espace, les designers sonores du jeu cherchent à reproduire avec fidélité l'aspect dérangeant et inquiétant de ces voix qui viendront perturber l'aventure du joueur.

À l'opposé total de cette esthétique, on trouve le contre-exemple parfait au paragraphe précédent disant que binaural et 2D ne font pas bon ménage. Le jeu *Inside* (Playdead, 2017) est une succession de tableau horizontaux dans lesquels le joueur incarne un enfant qui s'enfuit d'un camp pour atteindre un but inconnu. Le jeu ne comporte en effet pas de paroles, le joueur est libre d'interpréter lui-même la façon dont il comprend cette aventure. Les graphismes très sombres et étouffants, composés essentiellement d'espaces clos, contrastent avec la largeur des ambiances sonores, souvent constitués de réverbérations très larges et d'ambiances enregistrées en binaural au milieu desquelles résonnent les pas du petit garçon seul. Ce contraste permet de marquer le monde angoissant et incompréhensible dans lequel évolue ce petit garçon, les graphismes montrant l'étau qui se resserre sur lui et la largeur de l'espace sonore dans laquelle on n'entend quasiment que lui accentue sa fragilité et l'impression qu'il peut être entendu et attrapé à tout moment.



fig7. Capture d'écran d'un tableau angoissant du jeu *Inside*
(Playdead, 2017)

Le binaural de synthèse est plus largement répandu comme réducteur stéréo comme cela a été dit ci-dessus, mais il est également parfois utilisé comme élément de spatialisation pour seulement certains éléments précis de jeu. Cette pratique est notamment trouvable dans les jeux d'horreur-épouvante. Dans le jeu *Alien Isolation* (Creative Assembly, 2014), basé sur la série de films *Alien* réalisés successivement par Ridley Scott, James Cameron, David Fincher et Jean-Pierre Jeunet, le joueur incarne la fille d'Ellen Ripley, personnage principal du premier film, et se rend sur un vaisseau spatial où elle se retrouve rapidement piégée et où un Alien, extra-terrestre monstrueux, rôde. Sans arme efficace contre la bête, elle doit alors se fier à son audition pour entendre arriver son ennemi, dont les déplacements sont d'autant plus terrifiants qu'ils sont spatialisés en binaural, et que le joueur peut parfois entendre l'Alien se glisser juste derrière lui ou faire résonner une bouche d'aération proche.

Dans un genre totalement différent, de nombreux jeux de voitures proposent une spatialisation binaurale, afin d'essayer de représenter le plus fidèlement possible l'impression de vitesse et le son des moteurs, les fans du

genre cherchant à reproduire via le jeu la sensation qu'ils auraient au volant d'une voiture hors de prix. Ici, le binaural est souvent accompagné de volants de courses, et mêmes parfois de gilets qui contiennent des haut-parleurs pour reproduire les vibrations et la composante grave du son de voiture. C'est le cas de la série de jeu *DIRT* (Codemasters, 2009-2012) et des jeux de simulation *F1* (Codemasters, 2011-2016).

2.2.3 Quand le son prend le dessus : les jeux audio

Un aspect souvent oublié du jeu vidéo est celui des jeux audio, qui ont été dans un premier temps introduits comme jeux pour les aveugles, mais dont on trouve de nos jours quelques exemples qui sont destinés à un public large. Ces jeux ont des difficultés à se vendre, car les joueurs ont souvent l'impression qu'il s'agit de jeu "tronqués", dans lesquels les développeurs n'auraient pas pris le temps d'ajouter des graphismes, alors qu'il s'agit en fait d'un support extrêmement intéressant notamment pour le genre du jeu d'horreur-épouvante, dont on a déjà vu plus haut qu'il était idéal pour la mise en place d'une spatialisation binaurale.

Deux de ces jeux sont d'ailleurs cités dans le mémoire de Baptiste Palacin sur la peur dans le jeu vidéo : *Papa Sangre* (Somethin' Else, 2013) et *BlindSide* (Epicycle, 2012)⁹.

L'absence d'image dans ces jeux permet d'exploiter le son et la spatialisation binaurale pour décupler l'imaginaire du joueur afin que celui-ci se crée mentalement ses propres démons. Dans *Papa Sangre*, le joueur incarne un défunt qui traverse l'au delà pour sauver l'âme de la personne qui compte le plus pour lui dans une atmosphère inquiétante, remplie de monstres. Le

9. Palacin B., *Le rôle du son dans l'apparition de la peur dans le jeu vidéo*, mémoire effectué sous la direction de Gazeau C., École Nationale Supérieure Louis Lumière, soutenu en juin 2013, p145

joueur a également la possibilité de sauver d'autres âmes, au risque de réveiller les monstres se trouvant à proximité. Le dilemme auquel il fait face est purement sonore, puisque la seule raison qui l'empêcherait de faire ce geste altruiste est la peur d'entendre un son de monstre. Le joueur se déplace à l'aide d'une interface qui lui permet de tourner sur lui-même et d'avancer, faisant évoluer la spatialisation des sons qui se trouvent autour de lui, et qui seront donc son point de repère pour déterminer dans quelle direction aller.



fig8. Interface de *Papa Sangre*, le seul élément visuel du jeu (Somethin' Else, 2013)

BlindSide pousse le concept encore plus loin en ne proposant aucune interface graphique. Dans ce jeu sur mobile ou un couple se réveille complètement aveugle et cherche à comprendre la raison de leur cécité (et celle pour laquelle des monstres assoiffés de sang sont là, ce qui justifie le genre horrifique), la seule façon de se déplacer est de faire pivoter son téléphone, ce qui permettra de faire pivoter son avatar d'autant grâce aux gyroscopes intégrés.

Plus récemment, *A Blind Legend* (Dowino, 2015), essaie de se démarquer du genre horrifique en glissant le joueur dans la peau d'Edward Blake, un chevalier aveugle, ayant pour seul guide sa fillette, qui sert à la fois à indiquer la direction à prendre et à décrire le paysage environnant, permettant de

s’imaginer plus facilement les scènes. Ce jeu est constitué uniquement de spatialisation en synthèse binaurale. Cependant, si les scènes sonores sont très bien restituées, la spatialisation devient bien plus grossière dans les phases de combats (un ennemi vient de la gauche, de la droite, ou de devant, mais il n’y a pas d’entre-deux) et de déplacements (en quart de tour). Il aurait été intéressant d’essayer de pousser le concept au maximum dans les moments d’interaction, et de faire plus confiance à la capacité de localisation sonore du joueur.

On a pu voir que les exemples d’utilisation de spatialisation binaurale dans le jeu vidéo ne sont pas très nombreux, car elle n’est pas toujours déclarée comme telle par les développeurs, qui cherchent à mettre en retrait l’aspect trop technique dans la communication auprès des joueurs. De plus, il s’agit d’une technologie encore naissante dans ce milieu, puisque les exemples les plus anciens qui ont été montrés datent de 2009, ce qui peut expliquer que les entreprises soient encore discrètes sur son utilisation et son implémentation. Enfin, cette timidité peut aussi s’expliquer par la nécessité d’une écoute au casque pour apprécier pleinement l’expérience, qui bien qu’étant une pratique répandue reste loin d’être systématique.

2.3 Petite histoire de la réalité virtuelle

Tout comme dans la partie précédente, il semble important de commencer par prendre du recul pour définir la notion de réalité virtuelle et savoir d’où elle provient. Il ne faut pas oublier que le terme réalité virtuelle tel qu’il est prononcé dans cet écrit vient de l’anglais *virtual reality* (terme proposé par l’essayiste et chercheur en informatique Jaron Lanier en 1986), et que si *reality* se traduit bien par réalité, *virtual* ne se traduit pas par virtuel mais par quasi. Le terme choisit par les anglais ne fait donc pas état d’une fausse

réalité, une réalité recréée, mais plutôt d'une quasi-réalité. La traduction française ne tient donc pas compte de ce sens dans sa dénomination.

Le concept quant à lui date de bien avant la création de l'informatique, Platon est le premier à décrire de façon implicite la réalité virtuelle. Dans l'*allégorie de la caverne*, des personnes attachées ne peuvent voir que leur projection ainsi que celles d'autres sous forme d'ombres projetées sur un mur à l'aide d'un feu. Ils n'entendent que les échos produits par les parois de la caverne, leurs sens visuels et auditifs se retrouvent biaisés par un dispositif immersif et altérant sur le plan sensoriel.¹⁰ Un rapprochement est possible également entre un passage du *Discours de la méthode* dans lequel Descartes questionne la réalité qui pourrait selon lui n'être que des impressions envoyées par un démon (esprit trompeur) et les Principes de la connaissance humaine dans laquelle Berkeley expose sa Théorie de la connaissance et son immatérialisme, résumé par la phrase "Esse est percipi aut percipere", "être, c'est être perçu ou percevoir". Pour Berkeley, nous ne percevons que des idées et nous ne pouvons rien concevoir en dehors d'elles. On ne peut donc pas affirmer l'existence du monde extérieur, car nous ne le percevons qu'à travers l'idée que nous nous en faisons. Ces deux écrits font donc cas d'une réalité "virtuelle", qui n'est perçue que par le biais de nos sens et de l'analyse qui en découle.^{11 12}

Il faudra attendre un peu pour la mise en place de systèmes de réalité virtuelle, qui remonte à 1962 avec l'invention du sensorama par Morton Heilig permettant de diffuser une image stéréoscopique, un son stéréo, mais également des odeurs et du vent durant un film diffusé pour un spectateur seul.

10. Platon, Livre VII de La République, *L'allégorie de la caverne*, Grèce, IV siècle av. J.-C.

11. Descartes R., *Discours de la méthode*, France, 1637

12. Berkeley G., *Principes de la connaissance humaine*, Angleterre, 1710

C'est dans les années 90 que les premiers dispositifs tels qu'on les connaît aujourd'hui sont prototypés, mais la lourdeur des équipements nécessaire est décourageante, et il faudra attendre 2012 pour que Palmer Lucky annonce le développement de l'Oculus Rift.¹³



fig9. Photo du sensorama

Les systèmes de réalité virtuelle actuels sont constitués d'une dalle vidéo reliée à un masque que l'on attache autour du visage. L'écran projette une image stéréoscopique (c'est à dire que l'écran est scindé en son milieu, chacune des images correspond à ce que voit un des deux yeux) déformée afin d'éviter la distorsion optique créée par la proximité entre l'écran et les yeux, puis passée à travers des lentilles qui permettent d'augmenter le champ visuel et qui permettent de faire une mise au point à l'infini. Des gyroscopes disposés sur le casque permettent de récupérer les données des mouvements de têtes effectués par l'utilisateur.

13. Fayet M.-H., de la Rochère B., *Réalité virtuelle, année zéro!* et *Réalité virtuelle, pour y faire quoi?*, vidéo, Videosystem, 1991



fig10. l'Oculus Rift, premier né d'une génération de casque de réalité virtuelle à usage domestique

C'est systèmes dispositifs aujourd'hui de se commercialiser pour le plus grand nombre, malgré des prix rebutants au premier abord (700 euros pour un Oculus Rift, sans compter l'ordinateur qui permet de le faire tourner, et au moins un smartphone d'une valeur de quelques centaines d'euros pour une solution portable), et de plus en plus d'entreprises se lancent dans la course à la réalité virtuelle (Valve avec le HTC, Sony avec le Playstation VR, sans compter les innombrables marques qui proposent des supports type Google Cardboard).



fig11. le Google Cardboard est un simple support en carton qui incorpore deux lentilles sur lequel on vient poser son smartphone. Des versions plus élaborées permettent de régler l'écartement des lentilles et la mise au point, pour corriger certains problèmes de vue

2.4 Repenser le son pour la réalité virtuelle

La question de la place à accorder au son dans ce domaine est une problématique relativement nouvelle, tout du moins en ce qui concerne les dispositifs à disposition du grand public cités en introduction de ce mémoire, et qui ne sont sur le marché que depuis peu (fin mars/début avril 2016 pour l'*Oculus Rift* et le *HTC Valve*, et juin 2014 pour les premières solutions mobiles type *Google Carboard*). Il y a donc assez peu de recul en ce qui concerne la réflexion purement esthétique du son sur ces dispositifs, et l'on se contente trop souvent de prendre en compte uniquement la dimension technologique. Cette réflexion est valable pour les jeux vidéo comme pour les vidéos interactives à 360°. Il n'est en effet pas rare de trouver des vidéos dont l'image est interactive mais dont la bande-son est une stéréo statique, c'est à dire que les éléments qui la composent ne se déplaceront pas avec l'image, ce qui est entraîné un effet de gêne pour le spectateur.

Un exemple parmi d'autres serait la vidéo *Invasion!*¹⁴, dans laquelle un petit lapin se retrouve confronté à l'arrivée d'une soucoupe volante puis des aliens, qu'il parvient finalement à mettre en déroute. Pour accompagner cette vidéo se trouve une bande-son stéréo, ce qui fait que le spectateur entendra toujours le vaisseau à sa droite et ce quel que soit ce qu'il regarde. Ce choix arbitraire limite l'expérience du spectateur et lui en fait prendre conscience, ce qui aura tendance à l'éloigner du pacte mentionné dans la partie précédente.

Car c'est ici un des changements majeurs, et qui conditionne la façon dont il faut orienter la réflexion sur l'écriture pour la réalité virtuelle : la prise en compte des mouvements de la tête de l'utilisateur par le biais de capteur gyroscopiques (le headtracking). Il va donc être nécessaire de créer des scènes

14. Darnell E., *Invasion!*, 2016, États-Unis

sonores qui seront étroitement liées à ces déplacements, tout en déterminant quels sont les éléments qui devront en être esclaves ou indépendant. Cette réflexion reste finalement assez similaire à celle qu'on peut mener dans un jeu vidéo 3D classique, où certains composants sont dépendants de la position de la caméra qui délivre un angle de vue pour le joueur. Il s'agit la plupart du temps des sons illustratifs et des sons entités, plus rarement les sons décors, et très rarement les sons informatifs ou narratifs. Si cette fonctionnalité ne pose pas de problème dans le cadre du jeu, ou du film d'animation, dans lequel tout l'audio est créé en post-production, elle en devient problématique dans le cadre des vidéos à 360° basées sur des prises de vue réelles. Quel dispositif de prise de son utiliser pour pouvoir ensuite faire varier en même temps que le point de vue, le centre de l'attention sonore ?

Un début de réponse a été apporté par la société Arkamys, qui dans le cadre du regroupement Bili (pour Binaural Listening, un projet de recherche collaboratif sur l'écoute binaurale), qui propose d'effectuer l'enregistrement sonore à l'aide d'un dispositif ambisonic, basé sur un ensemble de trois microphones bidirectionnels et un microphone omnidirectionnel. Il est alors possible en post-production, en encodant et décodant les signaux de créer des microphones virtuels. Arkamys propose une solution logicielle qui associe la direction de ce microphone virtuelle à la position de la caméra. Cette solution, vidéo à l'appui¹⁵, fonctionne mais n'est pour l'instant utilisable que sur un système utilisant un téléphone Android. Cette solution est intéressante technologiquement, mais elle implique que le système de prise de son soit en permanence accolé au système de prise de vue. En voulant tout capturer absolument, le résultat est globalement médiocre. La meilleure méthode pour effectuer de la prise de son en parallèle d'une capture de vidéo à 360° n'a

15. Jouinot O., *Vidéo et son 360°, Headtracking*, vidéo youtube à regarder avec un téléphone Android, juillet 2016

pas encore de solution idéale (avec une solution de prise de son classique, où cacher le preneur de son et les microphones?). Une solution pourrait être de prendre uniquement du son témoin et de recréer en post-production une scène virtuelle.

En plus du headtracking, c'est le couplage avec un casque vidéo qui instaure une grande différence avec un jeu vidéo ou une vidéo interactive classique. La proximité du dispositif crée une intimité et place l'utilisateur au coeur de l'expérience, avec un rapport beaucoup plus égo-centré à l'objet qui lui est proposé et qu'il faut prendre en compte.

Quand le jeu vidéo "classique" propose à part égale des jeux 2D et des jeux 3D, eux-mêmes séparés entre jeux à la 1ère personne (jeu vu à travers la vision de l'avatar) et à la 3ème personne (vision englobant l'avatar), la réalité virtuelle semble se concentrer uniquement sur cette dernière catégorie. Tout est fait pour pousser le joueur à se dire qu'il incarne le personnage plutôt que de simplement prendre son contrôle. Il en va de même pour les vidéos 360° dans lesquelles le spectateur peut avoir le sentiment d'être présent sur place alors qu'il se sentira seulement témoin face à une vidéo vue sur un écran. C'est dans ce sens que doivent travailler conjointement le son, l'image, et l'interaction via les mouvements de la tête et l'utilisation de manette, pour garantir à l'utilisateur l'expérience la plus immersive possible.

Afin de parvenir à ces résultats qui cherchent à fausser nos sens, des outils technologiques très pointus sont utilisés, sans pour autant que ceci soient toujours suffisamment perfectionnés. Le risque est alors de se retrouver dans la vallée de l'étrange.¹⁶ Cette expression est issue d'une théorie développée par Masahiro Mori, qui stipule que plus un robot humanoïde sera réaliste, plus

16. Masahiro Mori, "La vallée de l'étrange", *Revue d'anthropologie et d'histoire des arts*, n° 15, pp26-33, 16 mai 2012

ses défauts nous paraîtront monstrueux et disproportionnés. Un robot sera plus facilement accepté comme tel si il est clairement différent d'une représentation humaine ou s'il est suffisamment irréprochable pour faire illusion. Entre ces deux extrêmes se trouve la vallée de l'étrange.

Ce principe est aussi imputable aux expériences en réalité virtuelle : comment en effet réussir à tromper à la perfection la vision, l'audition, l'oreille interne? Cette tromperie se remarque d'autant plus que tous nos sens ne sont pas stimulés par la réalité virtuelle, et notamment le toucher. Tout ça au risque de perdre l'intérêt de l'utilisateur ou même de le rendre malade en perturbant son oreille interne. Les premiers tests en réalité virtuelle montraient des signes de nausées chez le sujet au bout de quelques minutes, à cause d'une latence entre les déplacements de la tête et le suivi des capteurs gyroscopiques qui actualisent l'image, ce qui perturbe la stabilité. Ce phénomène a été amélioré avec l'évolution du matériel mais n'est pas encore totalement endigué. Peut-être faudrait-il alors clairement assumer ce manque technologique et ne pas clamer un rendu plus vrai que nature mais proposer un média qui offre une expérience unique.

2.5 Réalité virtuelle et spatialisation binaurale : un rapprochement instinctif

L'émergence de cette technologie dans l'industrie vidéoludique est finalement assez proche en date de l'apparition des systèmes de réalité virtuelle pour grand public, ce qui peut être en partie expliquer leur rapprochement naturel. Au lieu de parler de binaural, les développeurs font souvent référence au terme de "son 3D"¹⁷, comme pour mettre l'accent sur l'accès

¹⁷. *Oculus Audio SDK Documentation*, manuel à destination des développeurs pour la plateforme Oculus, v1.1.4, p12, 2017

à une nouvelle dimension, qui était pourtant déjà accessible avant l'avènement de la réalité virtuelle.

Pragmatiquement, l'accompagnement quasi-systématique de casque audio ou d'écouteurs avec les expériences en réalité virtuelle en fait un candidat parfait pour une utilisation de la spatialisation binaurale qui sera alors utilisée dans des conditions correctes, justifiant également l'engouement des développeurs pour cette technologie. Enfin, cette mise en parallèle s'explique aussi et surtout par le besoin d'une reproduction du son égocentrée, en rapport avec ce qui a été décrit dans la partie précédente. Le binaural cherchant par essence à reproduire une écoute qui se veut naturelle, son utilisation dans les expériences en réalité virtuelle paraît logique, à condition que son effet soit suffisamment probant, malgré l'impossibilité d'avoir accès à des HRTF personnalisées et à des casques égalisés. Le binaural permettrait aussi de distinguer d'avantage pour le joueur les sons illustratifs et/ou d'entités, des sons narratifs et informatifs qui ne dépendent pas du headtracking, permettant à celui-ci de mieux faire la distinction entre les sons provenant ou non de l'environnement virtuel qui l'entoure.

Il faut cependant faire attention à ne pas se reposer trop sur le binaural en prenant seulement en compte son principe théorique, et en s'attendant à ce que, dans son avancement technologique actuel, il soit suffisant. Dans sa présentation donnée à l'occasion de la rencontre sur la réalité augmentée et la réalité virtuelle *Vision Summit*, Nicolas Tsingos (directeur des technologies audio chez Dolby Laboratories) met en garde sur l'utilisation de solutions techniques trop simplifiées pour implémenter du "son 3D" en réalité virtuelle et sur la dégradation qu'elle risque d'apporter à l'expérience de l'utilisateur.¹⁸

18. Tsingos N., *Audio challenges in virtual and augmented reality*, présentation vidéo, février 2016

2.6 Conclusion

Cette partie a permis de faire d'abord un rappel sur la définition de la notion de jeu vidéo et sur le rôle que le son y tient précisément, afin de mieux déterminer la place d'un outil technologique comme la spatialisation binaurale dans ce média. Ce raisonnement a ensuite été étendu aux encore très récents systèmes de réalité virtuelle, dont la grammaire sonore s'avère, malgré quelques différences, assez similaire à celle employées dans les jeux vidéo en 3D.

La justification du choix de la spatialisation binaurale par adaptation a également été étayée, en montrant l'impossibilité matérielle d'implémenter des solutions personnalisées pour le plus grand nombre. Une mise en garde à été formulée sur l'avancée technologique actuelle, qui n'est peut être pas suffisante pour une implémentation efficace et acceptable en condition de jeu, au risque de proposer une expérience finale insatisfaisante.

Chapitre 3

Cubes : tests perceptifs sur la spatialisation binaurale pour un jeu en réalité virtuelle

Après avoir effectué un tour d’horizon sur la technologie binaurale qui nous a conduit à faire le choix de l’adaptation plutôt que de la personnalisation et avoir questionné la pertinence et la place de l’utilisation d’HRIR dans le jeu vidéo et la réalité virtuelle, les découvertes et hypothèses formulées vont être mises à l’épreuve dans une partie expérimentale. Après avoir expliqué la raison qui a conduit à faire ces test perceptifs, leur mise en place et leur contenu sera détaillé d’un point de vue général et technique en précisant les résultats escomptés, que l’on comparera à l’analyse de ceux obtenus.

3.1 Mise en contexte

"In a perfect world, yes, we'd all have custom HRTFs measured that match our own body and ear geometry precisely, but in reality this isn't practical. While our HRTFs are personal, they are similar enough to each other that a generic reference set is adequate for most situations, especially when combined with head tracking.

*Most HRTF-based spatialization implementations use one of a few publicly available data sets, captured either from a range of human test subjects or from a synthetic head model such as the KEMAR."*¹

Traduction : *"Dans un monde parfait, oui, nous aurions tous des HRTF personnalisées mesurées qui correspondraient à la géométrie de notre corps et de nos oreilles précisément, mais cela n'est pas réalisable en pratique. Bien que nos HRTF soient individuelles, elles sont suffisamment similaires entre elles pour qu'un jeu de référence générique soit approprié dans la plupart des situations, et plus particulièrement quand on le combine avec le suivi des mouvements de la tête.*

La plupart des implémentations de spatialisation basées sur des HRTF utilisent une des quelques banques de données disponibles au public, qui s'appuient soit sur des tests effectués sur un panel de sujets ou sur un modèle de tête artificielle comme la KEMAR "

1. *Oculus Audio SDK Documentation*, manuel à destination des développeurs pour la plateforme Oculus, v1.1.4, p12, 2017

C'est ce paragraphe, placé dans l'introduction d'un document à l'attention des programmeurs audio et sound designers par Oculus pour introduire le kit de développement logiciel audio de leur casque de réalité virtuelle qui est à l'origine de cette partie pratique de mémoire. En effet, le constructeur considéré comme pionnier et qui compte parmi les leaders sur le marché de cette nouvelle forme de loisirs, se permet en quelques phrases, et sans justifications précises ou citations de documents scientifiques d'affirmer que le principe d'adaptation à des HRIR non-individualisées est une solution viable et qui n'altère pas l'expérience ni les performances du joueur qui utilise son système, grâce notamment au headtracking qui permettrait de palier aux défauts de cette approche, quel que soit le jeu d'HRIR générique pris pour référence, sans prendre en compte l'égalisation du casque audio utilisé et sans faire mention du temps d'adaptation. Le joueur serait donc capable de s'adapter très rapidement.

Partant de ce constat et à l'issue de la recherche bibliographique effectuée dans les parties précédentes sur la notion d'apprentissage et de plasticité de l'écoute, il me semble intéressant de vérifier ce parti pris en proposant une expérience simple qui permettrait de tester les performances de jeu d'un sujet selon qu'elle lui soit proposée en spatialisant le son avec un jeu d'HRIR qui lui correspond ou pas, tout en reproduisant les circonstances d'un jeu en réalité virtuelle (support visuel et headtracking), c'est à dire celles auxquelles le sujet est ordinairement confronté dans l'utilisation de ce dispositif. La comparaison de ces deux parties permettrait de valider ou non la thèse du constructeur et la capacité d'adaptation rapide à une HRIR non-individualisée.

3.2 Contenu du jeu

Cubes est un jeu vidéo en réalité virtuelle avec une mécanique de jeu très simple : le joueur est dans une pièce, et autour de lui se trouve une cible (en l'occurrence un cube) qui émet un son, spatialisé en binaural. Pour trouver le cube, il doit le localiser auditivement, puis se tourner sur lui-même afin de se mettre face à l'objet. Il faut ensuite le fixer quelques instant pour que celui-ci disparaisse puis réapparaisse ailleurs sur la carte (en dehors de son champ de vision ou non), et ainsi de suite. Le joueur est fixe, c'est à dire qu'il reste en permanence au centre de la pièce, ses mouvements de têtes ne changent que l'orientation de la caméra. Il n'y a pas de manette de jeu.

Ce jeu repose donc essentiellement sur des mécanismes liés à la capacité de localisation auditive quand les cubes apparaissent hors du champs de vision. On considérera que la vision prend le pas sur l'audition lorsque le cube est à portée de vue.

Il s'agit d'un jeu préexistant au mémoire, programmé sur le moteur de jeu Unity dont le code a été gracieusement fourni par 3D Sound Labs, ainsi que leur module d'intégration d'HRIR.

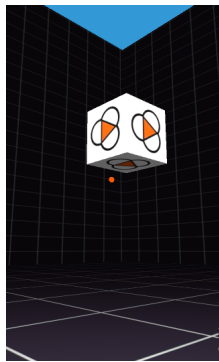


fig12. Capture d'écran de *Cubes*. Le point orange sous le cube représente l'endroit où regarde le joueur

3.3 Description de l'expérience et résultats attendus

Dans cette expérience, on souhaite donc montrer si un individu est capable de s'adapter rapidement à une HRIR qui n'est pas la sienne. Pour cela, l'expérience est divisée en trois parties distinctes, entrecoupées chacune d'une pause de quelques minutes.

Sélection d'HRIR

- Dans un premier temps, le sujet écoute une série de fichiers audio contenant un son décrivant un cercle sur le plan horizontal puis sur le plan médian, spatialisés avec sept jeux d'HRIR différents et qu'il évalue subjectivement. Le jeu lui est ensuite présenté brièvement afin qu'il se familiarise avec son mécanisme avant les sessions de test.

Phase de jeu

- La moitié des sujets effectue une séance de jeu où il doit repérer 30 cibles avec l'HRIR qu'il aura jugée comme étant la meilleure, tandis que l'autre moitié joue avec celle qu'il a considérée comme la pire.

- La dernière séance est la même que la deuxième, mais en donnant à chaque sujet l'HRIR qu'il n'a pas utilisée à l'étape précédente. Cela permet de vérifier que le phénomène que l'on souhaite observer est indépendant de l'ordre dans lequel le sujet passe les tests.

En comparant les résultats obtenus entre les deux sessions, on s'attend à trois issues possibles :

- Il y a une différence claire entre la session de jeu avec la bonne HRIR et celle avec la mauvaise, mais elle se résorbe de façon significative au cours du temps, et la thèse proposée par Oculus sera validée, car au vu de la durée des sessions de jeu (deux minutes par session environ), on pourra considérer que le phénomène d'adaptation est très rapide, suffisamment pour ne pas impacter l'expérience d'un joueur lorsqu'il à une pratique de loisir.

- Une différence est observée, mais elle ne se résorbe pas au cours de l'expérience : dans ce cas, la thèse d'Oculus est réfutée, et la solution de l'adaptation par rapport à celle de la personnalisation dans le cadre du jeu vidéo et de la réalité virtuelle est remis en cause, car elle sera montrée comme insuffisante.

- Enfin, et ce serait l'observation la plus problématique : le niveau de performance est le même, avec une HRIR adaptée ou non. Cela aurait plusieurs causes possibles : soit la méthode d'implémentation de l'HRIR n'est pas bonne et dégrade trop la réponse capturée par rapport à une écoute naturelle, et c'est le principe même d'utilisation de spatialisation binaurale dans le jeu vidéo et la réalité virtuelle qui est remise en cause, soit le problème est au niveau du protocole expérimental décrit ci-après et des données qui sont récoltées, et c'est l'expérience proposée qui est à revoir.

3.4 Protocole de mesure

3.4.1 Outils utilisés

Afin de mener à bien cette expérience, on utilise deux logiciels différents : Unity et Puredata

Unity est le moteur de jeu sur lequel est développé *Cubes*. Un moteur de jeu est un ensemble de composants logiciels qui permettent ensemble de structurer l'architecture d'un logiciel, dans notre cas un jeu vidéo. Ces composants permettent de gérer séparément le moteur de rendu 2D et/ou 3D, le moteur physique, le moteur son, l'animation, l'intelligence artificielle, la gestion du réseau et de la mémoire. Les intérêts des moteurs de jeu sont multiples. Ils permettent, entre autres, de développer simultanément plusieurs versions d'un même jeu pour plusieurs plates-formes différentes, mais aussi de réutiliser des éléments de code dans d'autres jeux. Les moteurs de jeu présentent des interfaces simple et intuitives, dans lesquelles les développeurs peuvent facilement implémenter des objets visuels et/ou sonores. Ils vont ensuite leur donner les comportements qu'ils souhaitent à l'aide de scripts écrits dans des langages de haut niveau, ce qui facilite grandement le travail d'implémentation.

Dans le cas de Unity, ces scripts peuvent être écrits en C# ou en JavaScript, le langage choisi dans notre cas étant le C#, car les parties de codes déjà fournies étaient écrites dans ce langage (le C# ou C sharp est un langage de programmation orienté objet dérivé du C++ proposé par Microsoft). Cette plateforme de développement présente l'avantage d'être gratuite sans limitation logicielle, et surtout de posséder une communauté aussi nombreuse qu'active.

Puredata est un logiciel de programmation graphique très simple d'utilisation, basé sur le langage C, qui est capable de gérer les signaux entrant et sortant d'une machine. À la façon dont Unity permet de rassembler plusieurs objets reliés entre eux pour former un jeu, Puredata permet d'assembler très facilement des fonctionnalités pour créer un programme. Cette application va permettre de créer un lecteur son dans la phase de sélection d'HRIR.

3.4.2 Phase de sélection d'HRIR

Les HRIR sélectionnées pour cette première partie du test sont issues de la base de données LISTEN de l'IRCAM, il s'agit des 7 golden ears, des HRIR qui sont supposées couvrir un panel de morphologie suffisamment large pour qu'on moins une corresponde à n'importe quel sujet. Il s'agit des réponses *irc 1008*, *irc 1013*, *irc 022*, *irc 1031*, *irc 1032*, *irc 1048* et *irc 1053*, qui sont toutes disponibles au format SOFA sur le site internet sofaconventions.org. Le choix de cette banque parmi d'autres se justifie par la présence de l'étude sur les golden ears² d'une part, mais également par la citation de cette banque de données comme une de celles utilisables dans le manuel d'Oculus.³

Au cours de cette phase, le sujet est mis face à un lecteur audio créé sous Puredata. 7 instances de ce lecteur se trouvent devant lui, chacune jouant successivement un burst de bruit rose de 50ms suivi d'un silence de 450ms avec des micro-fades d'entrée et de sortie. Le son décrit un cercle sur le plan horizontal par pas de 30° puis un arc de cercle sur le plan médian, par pas 15°. À l'aide d'une échelle de mesure sous forme d'une échelle graduée, le

2. Schönstein D., Katz B., "Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale.", Société Française d'Acoustique, 10ème Congrès Français d'Acoustique, Lyon, France. 2010

3. *Oculus Audio SDK Documentation*, manuel à destination des développeurs pour la plateforme Oculus, v1.1.4, p12, 2017

sujet estime subjectivement la qualité de reproduction du cercle. Chacune de ces instances correspond à l'enregistrement du son effectué avec une des 7 HRIR introduites au paragraphe précédent, réparties de façon aléatoire et non-identifiée pour le sujet. Une fois qu'il a noté chacune des 7 HRIR, il lui est demandé de refaire la même opération face à une nouvelle série de lecteur, qui contiennent les mêmes HRIR présentées dans un ordre différent. Une moyenne de la note attribuée aux deux instances est effectuée pour chaque jeu d'HRIR, et la mieux notée et la pire sont conservées pour la suite de l'expérience. Cette double notation permet entre autres de s'assurer que le sujet est cohérent dans sa notation. Avant de commencer le test, le sujet lit la feuille suivante :

"Vous allez écouter plusieurs fichiers audio, chacun de ces fichiers contenant une série de sons qui décrit un cercle sur le plan horizontal puis un demi-cercle sur le plan vertical, séparé par une courte pause.

Ces sons sont spatialisés en tenant compte de la forme du crâne et du corps d'un autre sujet, qui a parfois une morphologie différente de la vôtre et qui entend donc différemment.

À l'issue de chaque écoute, il vous est demandé de noter subjectivement la qualité de reproduction des cercles à l'aide de l'échelle de notation présentée à côté de chaque bouton lecteur (Le trajet est-il bien circulaire dans le cas horizontal ? Percevez-vous une sensation d'élévation du son dans le cas vertical ?). La note comporte 10 paliers, la case la plus haute de l'échelle de notation étant la réponse qui pour vous combine le mieux ces deux facteurs.

Chacun de ces fichiers peut être rejoués autant de fois que vous le souhaitez, et la note que vous lui attribuez peut être modifiée à tout moment."

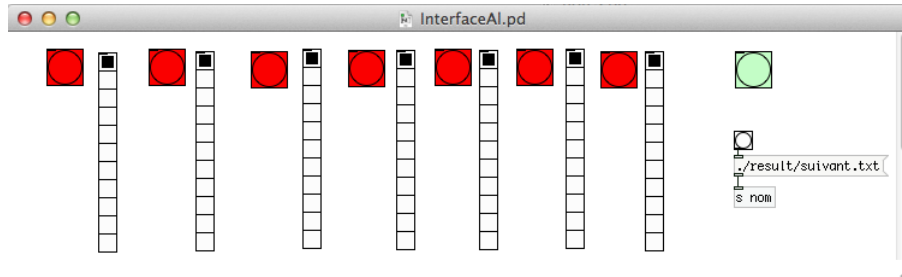


fig13. Interface graphique proposée au sujet. Le fonctionnement en détail est développé en ANNEXE 2

Idéalement, il aurait fallu que les sujets possèdent un jeu d'HRIR personnalisé, pour que la comparaison soit réellement basée sur la différence entre une réponse binaurale très adaptée contre une non-individualisée. Cependant, il m'est impossible dans le cadre de ce mémoire d'accéder à un tel panel, d'où la décision d'effectuer cette comparaison sur deux jeux sélectionnés de façon subjective par le sujet.

Le choix a été fait de se passer de support vidéo et de headtracking pour cette phase du test, qui sont précisément les éléments qu'on cherche à mettre en évidence comme étant la cause d'une adaptation rapide à une réponse binaurale non-individualisée dans les sessions de jeu. Les inclure dans cette phase n'apparaît donc pas pertinent, d'autant plus que s'ils s'avèrent réellement efficaces, ils rendront la notation des différents jeux d'HRIR plus difficile.

Pour cette expérience, on utilise un casque audio Sennheiser HD-25 (impédance 70 ohms) branché sur la prise mini-jack de l'ordinateur de test (carte son intégrée, macbook pro 2011). Le niveau de sortie de l'ordinateur est réglé de façon subjective avec le son de l'expérience. Une fois celui fixé, on utilise un son de référence (un bruit rose à un niveau de -9dBFS crête) et on mesure

un niveau de pression correspondant de 84,5dBA intégration slow. Le sujet n'a pas la possibilité de modifier le niveau sonore durant l'expérience.



fig14. dispositif de calibration

Un problème se pose cependant : les HRIR utilisées dans les sessions ne sont pas au format SOFA mais au format CFGA, qui est le format propriétaire développé par 3D Sound Labs. 3D Sound Labs a pu fournir des versions des réponses binaurales de la base LISTEN converti du format SOFA au format CFGA, mais il est possible que cette conversion entraîne une modification non-négligeable de la façon dont sera perçue la spatialisation. Pour être sûr que les conditions d'écoute soient les mêmes entre cette phase de sélection et les sessions de jeu, les sons diffusés par les lecteurs ont été spatialisés avec des HRIR au format CFGA.

Pour cela, une session de jeu Unity à été créée, dans laquelle une sphère émettant le son du test effectue une rotation autour de la caméra. C'est le son de cette session qui a été exporté, sa sortie audio est enregistrée directement avec le logiciel d'enregistrement Audacity au format .WAV, fréquence d'échantillonnage 48kHz et quantification 24 bits (qui sont également les réglages des paramètres audio de Unity).

3.4.3 Phases de jeu

Chaque sujet va donc passer deux sessions de jeu courtes au cours desquels il devra repérer 30 cubes, une fois avec la réponse binaurale qu'il a le mieux noté, et une fois avec celle qu'il a jugé la moins satisfaisante. L'ordre dans lequel ces deux tests seront passés va varier selon les sujets, pour s'assurer que celui-ci n'a pas d'influence sur les critères mesurés. Avant de passer les tests, le sujet effectue une partie de jeu qui n'est pas évaluée, afin qu'il puisse se familiariser avec les mécaniques de jeu et le système de réalité virtuelle.

Quatre paramètres à observer durant les sessions sont proposés afin d'évaluer de façon significative la capacité d'adaptation à un jeu d'HRIR non-individualisé :

- Le temps global mis pour compléter la partie.
- Le temps mis entre la découverte de chaque cube.
- La différence entre l'angle idéal à faire en déplaçant sa tête pour trouver le prochain cube et l'angle effectif.
- L'évolution de cette différence d'angle au cours du temps.

Pour pouvoir mener à bien cette campagne de test, il a été nécessaire de légèrement modifier le code du jeu *Cubes* tel qu'il était fait par l'équipe de 3D Sound Labs. Le script proposé à l'origine utilise un script qui téléporte au hasard le prochain cube une fois le précédent trouvé. Ce n'est pas très pratique pour comparer deux sessions de jeu. À la place, une liste de 30 positions dans l'espace déterminés par un vecteur composés d'un angle et d'une distance est proposée. Le sujet est confronté à ces 30 positions dans les deux cas, mais dans un ordre différent.

Ce raisonnement part du principe qu'il est impossible pour le sujet de se rendre compte qu'il doit faire les mêmes mouvements mais dans le désordre entre les deux séances. En outre, ce choix permet de comparer la durée et la différence d'angle pour un même mouvement de tête entre deux cubes avec une réponse binaurale différente. Les positions du cubes sont réparties par pas de 30° (de 0° à $\pm 180^\circ$) sur le plan azimutal et par pas de 15° sur le plan de l'élévation (de -15° à $+75^\circ$, en deçà le cube serait sous le sol du plateau de jeu et donc impossible à repérer. La limite maximale de 75° est fixée car elle correspond déjà à un position du cube quasi-verticale pour le joueur).

Pour chaque mouvement absolu sur le plan azimutal, toutes les positions absolues en élévation son testées (par exemple, pour un mouvement de $\pm 120^\circ$ en azimuth, le test comportera des cubes placés à 0° , $\pm 15^\circ$, $\pm 30^\circ$ et $\pm 45^\circ$ en élévation. $\pm 60^\circ$ n'est pas systématiquement testé, vu qu'il correspond en soi à un mouvement quasi-uniquement sur le plan de l'élévation). Les listes de positions des cubes sont données en ANNEXE 3.

Pour simplifier les mesures, l'éloignement des cubes n'est pas pris en compte : chaque cube se trouve sur une sphère d'un diamètre fixe. Cela permet de faire des mesures prenant en compte et permettant de comparer uniquement l'angle azimutal et l'angle d'élévation, quitte à ce que, si les mesures sont concluantes, l'étude de l'éloignement fasse l'objet d'une prochaine expérience.

Pour effectuer les mesures, on ajoute au jeu deux scripts. Le premier crée un chronomètre qui se déclenche à partir de l'instant où le premier cube est repéré, et qui relève la durée entre chaque cube (cette donnée n'apparaît pas à l'écran pendant la phase de jeu) avant de consigner ces données dans un fichier texte. Le deuxième permet de relever les données du headtracker toutes les 5ms à compter du moment où le premier cube est repéré et de reporter ces coordonnées dans un fichier texte. Le choix de cette vitesse de rafraîchissement se justifie ainsi : pour calculer l'angle de rotation de la tête, on a décidé de s'intéresser à la tendance du mouvement effectué au cours des 20 premières millisecondes de déplacement qui semblent plus révélatrices de la localisation à laquelle le sujet pense avoir entendu le cube par rapport au mouvement global, qui est de toutes façons parasité par l'apparition du cube dans le champ de vision, ce qui entraînera une correction immédiate de la part du sujet.

Dix-sept exports différents du jeu sont donc à faire, un par jeu d'HRIR, que l'on double avec le deuxième parcours de cibles possible (quatorze versions). Les trois autres versions sont des exports avec trois des sept jeux d'HRIR et une téléportation du cube aléatoire, pour la brève séance de familiarisation, afin qu'elle ne puisse en aucun cas être celle qui corresponde aux choix du sujet.

Le choix a été fait sur cette trentaine de cibles de faire en sorte que quelques-unes apparaissent dans le champ de vision de l'expérimentateur en début de partie, puisque que cela fait partie de l'hypothèse formulée comme quoi c'est un des facteurs qui l'aideront à s'adapter plus rapidement. Cependant, ces mouvements de tête ne seront pas comptabilisés dans le dépouillement des données, car on considère que dans un tel cas, la vue prédomine sur l'audition, par conséquent le rôle de celle-ci dans le repérage de la cible n'est plus significatif.

Le signal sonore qui a été choisi pour accompagner le cube n'est pas un son de test, il ne s'agit ni d'un bruit blanc, ni d'un bruit rose, d'un bruit gaussien ou d'un burst. Il a été décidé de garder le son qui accompagnait le jeu à la base. Il s'agit d'un son de fiction qui permet d'illustrer le cube et de le rendre plus "vivant". Il est composé d'un bruit de fond assez régulier, pour que le sujet puisse tout de même être capable de localiser la cible, ponctué de son brefs type sound design de droïde (dans l'idée des sons émis par RD2D2 dans la saga *Star Wars*), qui apportent un côté ludique à l'expérience. Ce choix s'explique par le désir de maintenir un aspect ludique relativement proche de ce que pourrait expérimenter un sujet en situation de jeu. En effet, un bruit blanc n'a rien de sympathique, qui plus est sur la durée, et de plus, ce n'est clairement pas le genre de son représentatif de ceux auxquels serait confronté la plupart du temps une personne en situation de loisir. Un fichier .WAV contenant ce son est fourni avec cet écrit.

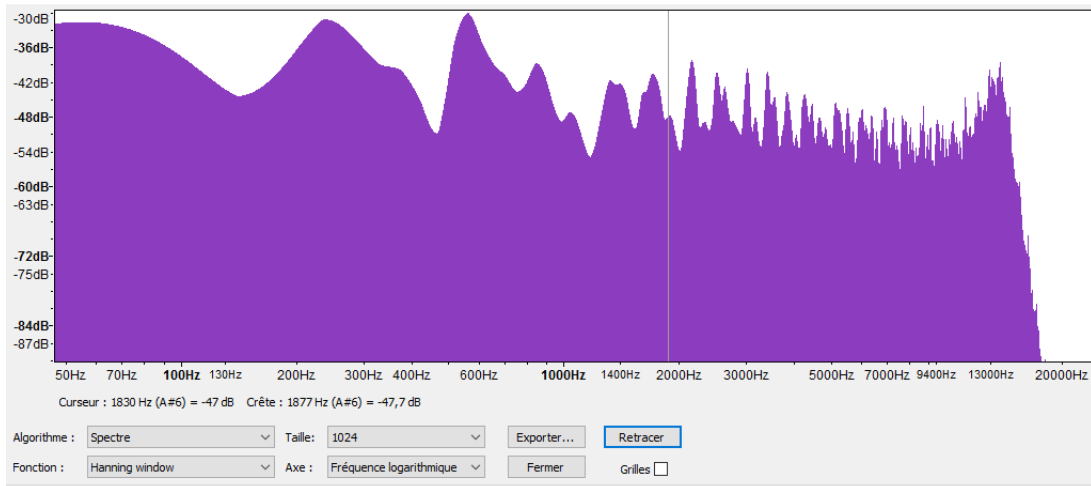


fig15. Spectre du signal sonore utilisé : bruit constant

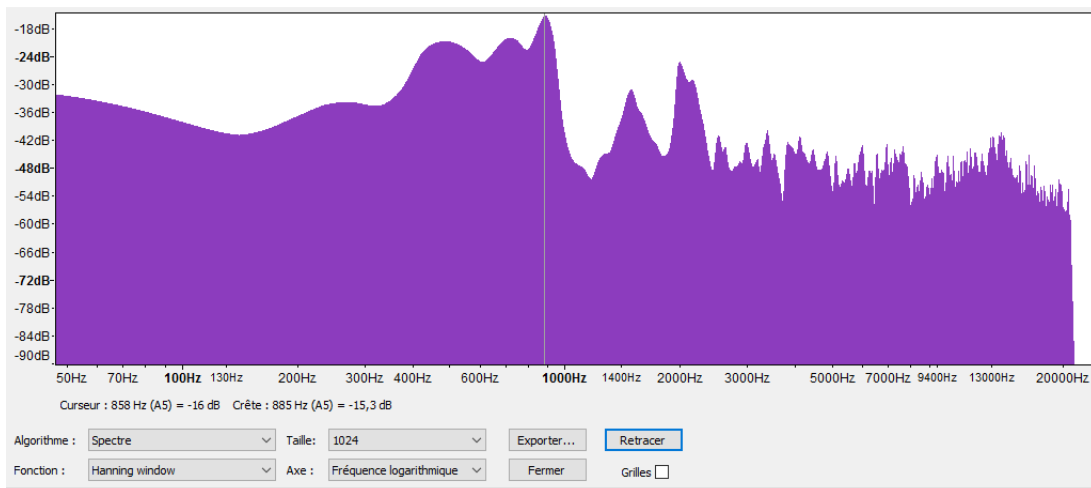


fig16. Spectre du signal sonore utilisé : exemple de bruit ponctue.

Noter la différence d'échelle.

Le spectre du bruit de fond est plus chargé en basses qu'en hautes fré-

quences, mais reste assez large bande. Les sons ponctuels créent régulièrement des informations supplémentaires avec une partie transitoire large spectre et des pics compris entre 1kHz et 3kHz.

Avant de commencer cette phase de l'expérimentation, le texte suivant est présenté au sujet :

"Vous allez enchaîner deux phases de jeu successives au cours desquelles il vous faudra repérer 30 cubes en vous fiant à votre audition.

Pour le bien de l'expérience, il vous est demandé de "jouer le jeu" et d'essayer de faire en sorte de ne déplacer la tête que dans la direction vers laquelle vous entendez les cibles et d'éviter les mouvements de tête superflu.

Le jeu se termine lorsque le point qui sert de viseur se transforme de façon permanente en cercle. Merci!"

Cette expérimentation se fait sur un système de réalité virtuelle sur téléphone portable, constitué d'un casque équipé de lentilles dans lequel on vient glisser le téléphone auquel on a branché un casque audio via la prise mini-jack. Dans notre cas, le casque est un Homido V2, qui permet de régler l'écartement entre les lentilles à l'aide d'une molette, mais également de corriger leur position selon les problèmes de vue (myopie, ou hypermétropie), la première session de jeu non-évaluée permettant de faire ces réglages.⁴

Deux téléphones différents sont utilisés, pour éviter les problèmes de

4. <http://www.homido.com/en/medias/file/homido-user-manual-web-d4a4.pdf>, manuel d'utilisation du casque Homido

surchauffé et les pannes de batterie lors de l'enchaînement des tests. Il s'agit de deux Galaxy S7 de Samsung afin d'être sûr de la constance de la qualité des gyroscopes, de la dalle vidéo, du convertisseur audio et du processeur. Le Galaxy S7 utilise le système d'exploitation Android 6.0 *Marshmallow*, auquel est rajoutée une surcouche propriétaire de Samsung. Un des téléphone a été acheté dans le cadre du mémoire et l'autre a été prêté.



fig17. Casque Homido ouvert avec un Samsung Galaxy S7 monté sur le support prévu à cet effet

Dans les deux cas, le casque audio utilisé est un Sennheiser HD-25 (impédance 70 ohms), le niveau d'écoute est calibré de façon à ce qu'il soit le même sur les deux téléphones, et on fait en sorte que le volume sonore soit verrouillé durant l'expérience à l'aide de l'application *Volume Locker*, de même que la luminosité de l'écran. Le niveau est fixé de façon subjective, par comparaison avec le niveau de la phase de sélection d'HRIR.

3.5 Analyse des résultats

31 sujets ont participé à l'expérience, mais seulement 30 série de résultats sont exploitables, car un des sujets n'a pas pu aller jusqu'au bout de l'expérience car trop dérouté par le système de réalité virtuelle (sujet 17). Parmi les sujets retenus, aucun n'utilise de façon régulière de système de réalité virtuelle (soit aucune expérience, soit essayé à l'occasion de salon ou par des amis). 20 d'entre eux sont des professionnels dans le domaine de l'audio. Aucun sujet n'a dit avoir de problème d'audition connu. La fourchette d'âge des participants est comprise entre 20 et 40 ans. À l'issue de l'expérience et sans avoir été sollicités, 25 sujets ont dit avoir trouvé l'expérience de jeu amusante, on peut donc considérer que l'aspect ludique de cette phase a été respecté.

3.5.1 Sélection d'HRIR

Le tableau présentant les résultats complets de la phase de sélection d'HRIR s'est présenté ci-dessous. En plus de déterminer quels jeux d'HRIR vont être utilisés pour la phase suivante, cette étape a permis de vérifier si les sujets étaient cohérents dans leur notation en comparant les deux notes données pour une même HRIR. On constate qu'en moyenne sur tous les sujets, la réponse varie de 22% pour une même spatialisation binaurale.

	meilleure HRIR	pire HRIR	erreur moyenne (%)
sujet 1	1008	1022	39
sujet 2	1053	1022	16
sujet 3	1013	1048	24
sujet 4	1053	1013	16
sujet 5	1008	1022	33
sujet 6	1048	1008	21
sujet 7	1048	1032	33
sujet 8	1013	1022	35
sujet 9	1022	1032	21
sujet 10	1053	1032	11
sujet 11	1008	1013	21
sujet 12	1013	1053	30
sujet 13	1031	1048	21
sujet 14	1048	1053	20
sujet 15	1031	1022	13
sujet 16	1013	1022	19
sujet 17	1031	1022	
sujet 18	1031	1032	19
sujet 19	1048	1022	14
sujet 20	1031	1008	13
sujet 21	1048	1022	26
sujet 22	1022	1032	19
sujet 23	1013	1008	11
sujet 24	1008	1022	20
sujet 25	1048	1031	10
sujet 26	1008	1032	24
sujet 27	1032	1048	27
sujet 28	1008	1031	16
sujet 29	1032	1008	24
sujet 30	1008	1022	23
sujet 31	1008	1022	26
		MOYENNE TOTALE	22
		MOYENNE RETENUE	19

fig18. Résultats de la phase de sélection d'HRIR

Ce chiffre n'est pas négligeable, mais il témoigne tout de même d'une cohérence dans la notation des sujets. S'il ne jugent pas toujours de façon identique le même son, ils ne sont pas non plus incohérents au point de juger un HRIR comme étant la meilleure sur un série de test, puis la pire sur la suivante, à quelques exceptions près. Afin d'éviter de ramener du bruit

dans les mesures suivantes, on a décidé de retirer de la suite des calculs les quelques sujets les plus incohérents de cette phase. Les sujets 1, 5, 7, 8 et 12, qui présentent tous une erreur moyenne supérieure ou égale à 30% et ont donc donné les notes les plus disparates ne seront pas pris en compte dans les résultats qui vont suivre. Une fois ces sujets retirés, la moyenne d'erreur tombe à 19% pour les 25 sujets restants.

Cette phase a fait l'objet de retour non sollicité de la part de certains sujets : sur les 31 qui ont passé le test, 28 ont affirmé ne percevoir aucune élévation pour aucune spatialisation binaurale ou ressentir une légère variation fréquentielle qu'ils ne parvenaient pas à associer à une sensation d'élévation. Ils percevaient cependant des trajectoires circulaires différentes sur le plan azimutal, ce qui signifie que leur notation s'est principalement basée sur ce critère.

3.5.2 Phases de jeu

On constate des différences entre les sessions de jeu effectuées avec les HRIR qualifiées de meilleures et de pire par les sujets. Les deux tableaux suivants présentent le temps global moyen mis par l'ensemble des sujets pour effectuer chacune des deux séances, mais aussi selon l'HRIR avec laquelle il passe chaque session. Un deuxième tableau donne la même information pour le temps moyen entre chaque cube. Le détail complet est disponible en ANNEXE 4.

	temps 1 (s)	temps 2 (s)	différence (s)	progression (%)
MOYENNE TOTALE	104,69	87,66	-17,10	16,26
MOY. MEILLEURE	90,52	87,33	-3,19	3,52
MOY. PIRE	116,63	88,00	-28,63	24,55
ÉCART TYPE TOTAL	26,16	16,35	20,15	13,94
ÉCART TYPE MEILLEURE	16,09	13,89	19,09	10,00
ÉCART TYPE PIRE	27,92	18,60	8,59	9,21

fig19. temps moyen global et progression entre les deux sessions de jeu

	temps moyen 1 (s)	temps moyen 2 (s)	différence (s)	progression (%)
MOYENNE TOTALE	4,19	3,51	-0,68	16,31
MOY. MEILLEURE	3,79	3,52	-0,27	7,19
MOY. PIRE	4,59	3,49	-1,09	23,84
ÉCART TYPE TOTAL	1,05	0,65	0,81	13,93
ÉCART TYPE MEILLEURE	1,12	0,56	0,76	9,98
ÉCART TYPE PIRE	0,64	0,75	0,34	9,19

fig20. temps moyen entre chaque cube et progression entre les deux sessions de jeu

Tout d'abord, il y a une différence globale assez importante entre les deux sessions de jeu : les sujets mettent en moyenne 17,10s de moins pour compléter la deuxième session de jeu quel que soit le jeu d'HRIR avec lequel ils l'effectuent, soit 0,68s de moins en moyenne entre chaque cube, ou encore une progression de l'ordre de 16% ce qui n'est pas négligeable sur une expérience d'une durée d'environ 1min30. Il y a donc bien un apprentissage qui s'effectue entre les deux sessions de jeu. Il faut à présent déterminer à quoi il est dû, et si ce constat est valide.

Si la différence de cette moyenne de temps est assez importante dans la première session entre les sujets jouant avec la meilleure HRIR et ceux jouant avec la pire (les joueurs commençant avec la réponse qu'ils ont jugée

la meilleure finissent en moyenne 26s plus tôt que les autres, soit 0,8s de moins par cube en moyenne), elle se résorbe grandement et devient quasiment nulle lors de la deuxième session (moins d'1s d'écart global et 0,02s par cube). Cela prouverait la supériorité de la meilleure HRIR dans un premier temps, mais tendrait à démontrer que dans un laps de temps rapide (environ 3min de jeu) cette supériorité se résorbe et que le choix de l'HRIR prend moins d'importance, ce qui irait dans le sens de l'hypothèse défendue par Oculus en introduction de cette partie.

Le choix de faire alterner les sujets entre bonne et mauvaise HRIR est finalement critiquable : il serait en effet intéressant d'observer si cette tendance se retrouve si l'on compare une population confrontée uniquement au jeu d'HRIR le plus pertinent et une autre confrontée uniquement à la pire. En effet, les résultats observés peuvent aussi montrer un bond de la progression des joueurs qui commencent par leur plus mauvaise HRIR avant de passer à la meilleure et un frein de la progression des sujets qui se retrouvent soudainement dans une situation où la spatialisation du son est dégradée pour eux.

Les écarts type constatés sur ces moyennes sont assez importants, de l'ordre de la dizaine de seconde sur les temps globaux relevés, et de la seconde sur les temps moyen entre chaque cube. Cette importance montre une répartition assez disparate de la performance des différents sujets, ce qui rend plus difficile de conclure sur la tendance d'apprentissage observée. On note que ces avec la pire HRIR que l'écart type est le plus resserré, et donc que les joueurs ont la performance la plus constante.

Sur les 25 sujets pris en compte dans les mesures, 5 ont été moins performants lors de la deuxième session par rapport à la première (cf. ANNEXE 4). Il est intéressant de noter que ces cinq sujets ont tous passé la première session de jeu avec leur meilleure HRIR puis la deuxième avec leur pire. Cet échantillon représente donc la moitié des sujets sélectionnés qui ont passé les

session dans cet ordre. On pourrait y voir une déstabilisation liée à la perte de qualité de localisation dans le plan azimutal.

La remarque sur la difficulté à percevoir la sensation d'élévation persiste à la fin de l'expérience pour les sujets ayant évoqué ce problème à la phase précédente. Les HRIR utilisées sont pourtant les mêmes que dans le test de la partie pratique du mémoire de Vincent Zanetto, qui a fait passer des tests similaires à ses sujets, et dans lesquels ceux-ci ont bien perçu une sensation d'élévation différente selon les réponses utilisées.⁵

Après recherche, Il s'avère que ce problème provient d'une erreur lors de l'export de la session de jeu, ce qui entraîne une perte notable dans la spatialisation du son en élévation mais pas sur le plan azimutal. Un élément marqué "binaural direct", et qui n'est pas explicité dans la documentation technique du plug-in de spatialisation binaurale de 3D Sound Labs est à l'origine de ce problème.⁶ Il est donc envisagé de reconduire les tests entre le rendu de l'écrit et la soutenance orale de ce mémoire, auquel on annexera les nouveaux résultats obtenus.

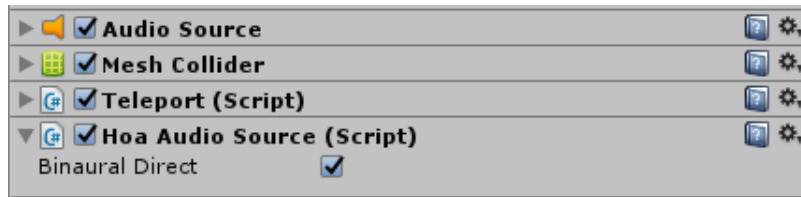


fig20. Le paramètre "binaural direct" associé à chaque source sonore que l'on veut spatialiser en binaural et qui n'est pas coché par défaut

5. Zanetto V., *Création d'une interface interactive à usage domestique dans le but de créer un modèle HRTF personnalisé*, mémoire effectué sous la direction de Blum A., École Nationale Supérieure Louis Lumière, soutenu en juin 2017

6. *VRAudioKit for Unity, revision 1*, manuel d'utilisation du plug-in de spatialisation binaurale de 3D Sound Labs, mars 2016 URL : http://3dsoundlabs.com/wp-content/uploads/vraudiokit/VRAudioKit_Unity_Documentation.pdf

Malgré ce défaut conséquent constaté, la question se pose toujours de la capacité d'adaptation du sujet sur le plan azimutal, mais aussi en élévation grâce aux éléments inhérents à l'interactivité dont il dispose (headtracking et support visuel).

Les sujets qui n'ont pas ressenti d'élévation dans la phase de sélection ont pour la plupart déclaré qu'ils ressentaient un changement de la réponse en fréquence du son quand celui était à une élévation différente de 0° , mais qu'ils ne l'associaient pas avec un changement d'élévation en perception. Pour palier au manque d'informations sur le plan de l'élévation, les sujets ont pour la plupart adopté très rapidement une posture adaptative : ils se déplacent sur le plan azimutal jusqu'à ne plus sentir de déplacement du son dans l'espace puis balaient leur regard sur le plan de l'élévation si le cube est hors de portée du regard.

Cela permet de compenser le manque d'informations de spatialisation, mais va par contre fausser le calcul de la différence entre l'angle idéal à faire en déplaçant sa tête pour trouver le prochain cube et l'angle effectif, qui perd donc son intérêt comme élément de mesure de la progression du joueur. Ce facteur ne sera donc finalement pas calculé pour cette série de tests, mais sera à prendre en compte lors de la nouvelle campagne de mesure effectuée entre le rendu de cet écrit et la soutenance orale.

Cette difficulté de perception dans l'axe d'élévation se retrouve notamment sur un mouvement présent dans les deux sessions de jeu : le cube se retrouve au-dessus du joueur. C'est le mouvement qui a causé le plus de difficulté à l'ensemble des joueurs. Ci-dessous un tableau comparatif entre le temps moyen mis entre chaque mouvement et ce mouvement en particulier par l'ensemble des sujets (détails en ANNEXE 4) :

	temps moyen 1 (s)	temps moyen 2 (s)	mouv vertical 1 (s)	mouv vertical 2 (s)
MOYENNE TOTALE	4,19	3,51	13,17	6,43
MOY. MEILLEURE	3,79	3,52	9,66	7,04
MOY. PIRE	4,59	3,49	16,68	5,82
ÉCART TYPE TOTAL	1,05	0,65	12,09	2,70
ÉCART TYPE MEILLEURE	1,12	0,56	14,85	2,71
ÉCART TYPE PIRE	0,64	0,75	5,22	2,56

fig21. temps moyen entre chaque cube comparé au temps mis pour trouver le cube à la verticale

On s'aperçoit que ce mouvement prend quasi-systématiquement un temps beaucoup plus important que les autres à effectuer (jusqu'à 60 secondes pour un sujet), ce qui prouve le problème de localisation dans le plan vertical avec l'implantation de spatialisation binaurale telle qu'elle a été effectuée dans le cadre de ce test. L'écart type est de l'ordre de grandeur de la moyenne, ce qui montre l'invalidité de cette mesure en particulier. Cela correspond avec le fait que dans le cas de l'élévation pure, les sujets raisonnent de façon logique plutôt qu'en se basant sur ce qu'ils entendent. Au bout d'un certain temps, ils finissent par déduire que s'ils ne voient pas le cube autour d'eux, c'est que celui-ci se trouve au dessus.

Partant de ces constatations, on se propose de refaire les premiers calculs en excluant ce déplacement vertical. On obtient les tableaux comparatifs suivants (données complètes en ANNEXE 4) :

CALCUL DU TEMPS GLOBAL SUR L'ENSEMBLE DES POSITIONS				
	temps 1 (s)	temps 2 (s)	différence (s)	progression (%)
MOYENNE TOTALE	104,69	87,66	-17,10	16,26
MOY. MEILLEURE	90,52	87,33	-3,19	3,52
MOY. PIRE	116,63	88,00	-28,63	24,55
ÉCART TYPE TOTAL	26,16	16,35	20,15	13,94
ÉCART TYPE MEILLEURE	16,09	13,89	19,09	10,00
ÉCART TYPE PIRE	27,92	18,60	8,59	9,21

CALCUL DU TEMPS GLOBAL SANS LA POSITION VERTICALE				
	temps 1 (s)	temps 2 (s)	différence (s)	progression (%)
MOYENNE TOTALE	90,40	81,24	-10,49	10,14
MOY. MEILLEURE	81,36	81,51	0,16	-0,19
MOY. PIRE	98,23	80,96	-17,27	17,58
ÉCART TYPE TOTAL	20,49	15,10	16,91	12,78
ÉCART TYPE MEILLEURE	12,75	12,10	19,05	12,71
ÉCART TYPE PIRE	23,02	17,78	6,91	7,63

fig22. Comparaison du temps global selon qu'on intègre ou non la position verticale

CALCUL DU TEMPS MOYEN SUR L'ENSEMBLE DES POSITIONS				
	temps moyen 1 (s)	temps moyen 2 (s)	différence (s)	progression (%)
MOYENNE TOTALE	4,19	3,51	-0,68	16,31
MOY. MEILLEURE	3,79	3,52	-0,27	7,19
MOY. PIRE	4,59	3,49	-1,09	23,84
ÉCART TYPE TOTAL	1,05	0,65	0,81	13,93
ÉCART TYPE MEILLEURE	1,12	0,56	0,76	9,98
ÉCART TYPE PIRE	0,64	0,75	0,34	9,19

CALCUL DU TEMPS MOYEN SANS LA POSITION VERTICALE				
	temps moyen 1 (s)	temps moyen 2 (s)	différence (s)	progression (%)
MOYENNE TOTALE	3,80	3,43	-0,37	8,11
MOY. MEILLEURE	3,50	3,42	-0,07	2,01
MOY. PIRE	4,10	3,44	-0,66	14,22
ÉCART TYPE TOTAL	0,83	0,64	0,78	15,49
ÉCART TYPE MEILLEURE	0,96	0,52	0,80	12,96
ÉCART TYPE PIRE	0,38	0,75	0,49	12,77

fig22. Comparaison du temps moyen selon qu'on intègre ou non la position verticale

On constate alors que la disparité entre les deux sessions de jeu s'ame-
nuise : la progression globale entre les deux phases de jeu passe de 16% à
10%, et la tendance s'inverse même pour les sujets commençant le jeu par
leur meilleure HRIR puis leur pire : d'une légère progression, on passe à une
très faible régression (qu'on pourrait plutôt considérer comme une stagna-
tion).

Comme précédemment, on observe une disparité sur les résultats de la
première phase de test entre les sujets ayant le meilleur jeu d'HRTF et les
autres (17s d'écart en moyenne globale, et 0,6s de plus en moyenne entre
chaque cube) et la deuxième phase de jeu (même durée moyenne globale et
entre chaque cube). Les conclusion tirées avant le retrait du trajet de tête
à la verticale dans le calcul des données restent donc pertinentes, dans une
moindre mesure et avec les incertitudes qui ont été relevées auparavant.

L'écart type observé reste le même, ce qui veut dire que même en retirant
cette position qui présente pourtant la plus grande disparité en terme de
performance (entre 63,31s et 4,25s pour le repérer durant la première session
et entre 10,59s et 2,64s pour la deuxième), les temps mis pour compléter les
deux sessions de jeu restent donc assez disparates selon les sujets.

3.5.3 Conclusion

Malgré un défaut dans la mise en place de l'expérience, une légère tendance d'adaptation dans un laps de temps très court (entre 2 et 5 min de jeu) semble se dégager si on prend en compte majoritairement les déplacements sur le plan azimutal (les déplacements en élévation étant biaisés), ce qui serait un premier pas dans la direction de l'hypothèse soutenue par Oculus de l'apport de la vision et du headtracking à la rapidité de cette capacité d'adaptation.

Ces résultats sont à prendre avec prudence, car à la place d'une progression globale, ils pourraient aussi montrer un changement de performance avec le changement d'HRIR (perte de progression pour ceux qui passent le test d'abord avec la spatialisation binaurale qu'ils ont jugé la plus pertinente, et progression forte pour ceux qui commencent avec a pire).

Pour confirmer ou infirmer cette tendance et vérifier la nature des chiffres obtenus, on propose de renouveler l'expérience entre le rendu de cet écrit et la soutenance orale en corrigeant les paramètres de spatialisation choisis à l'export et en modifiant le protocole de test : au lieu de faire passer à chaque sujet un test avec sa meilleure puis sa pire HRIR, on crée deux populations. La première passera le jeu uniquement avec sa meilleure HRIR, et la deuxième uniquement avec la pire. Ainsi, il sera possible de comparer la progression des deux populations plus facilement. On essaiera en priorité d'avoir des sujets qui n'ont pas effectué la première expérience.

Conclusion générale

L'adaptation à une spatialisation binaurale non-individualisée trouve dans la réalité virtuelle et le jeu vidéo des candidats idéaux pour son application grâce à leur caractère interactif et à l'association du son à l'image. La première partie de ce mémoire a permis de mettre en avant l'impossibilité d'accéder à une spatialisation binaurale personnalisée pour le grand public dans l'état actuel de la technologie et les raisons qui peuvent pousser à penser que l'adaptation du joueur à un jeu d'HRIR générique est une solution viable dans le cadre d'une activité vidéoludique et/ou en réalité virtuelle.

L'expérience présentée dans ce mémoire a eu pour but de mettre en avant l'accélération de cet apprentissage par le biais de ce support, et les résultats obtenus pourraient abonder dans cette direction, malgré un problème dans la mise en place de l'expérience. Les performances des différents sujets étant assez variées, potentiellement à cause de ce problème, il n'est pas possible d'exploiter les mesures obtenues avec une confiance totale. Afin de s'assurer des tendances observées dans cette première série de tests, une nouvelle expérience va être menée et dépouillée avant la soutenance orale, puis annexée à cet écrit. Cette nouvelle expérience corrigera le biais observé précédemment, et va proposer un protocole légèrement différent pour éviter de créer une confusion lors de l'obtention des résultats.

En dépit des applications et des exemples montrés dans la deuxième partie de ce mémoire, l'utilisation de la spatialisation binaurale dans le cadre du jeu vidéo reste rare. Pour comprendre les raisons de ce manque d'engouement pour une technologie permettant théoriquement d'améliorer la perception de la spatialisation (ce qui serait un apport réel pour l'interactivité entre le joueur et l'environnement du jeu), j'ai eu l'occasion de rencontrer quelques sound designer et lead sound (c'est le responsable de la partie sonore d'un jeu) qui m'ont confirmé leur frilosité quant à son utilisation.

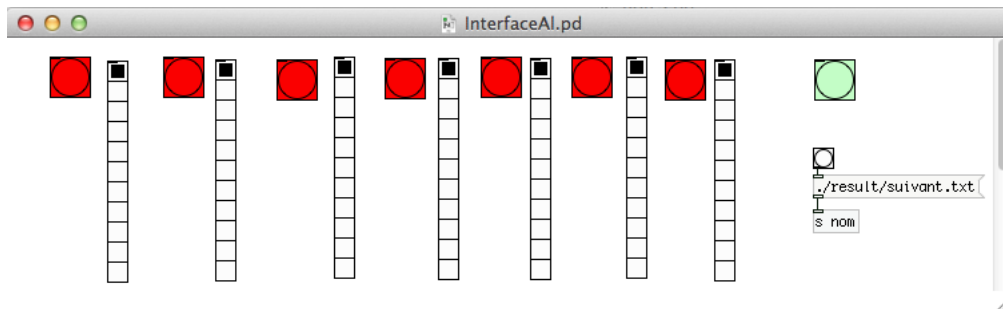
Celle-ci s'explique dans le cadre du jeu vidéo par l'absence de contrôle sur le mode de diffusion du son utilisé par le joueur, alors que la spatialisation binaurale requiert une écoute au casque pour fonctionner. Cependant, dans le cadre de la réalité virtuelle, l'écoute au casque est quasi-systématique, il est donc légitime de penser que ce support puisse intégrer de la spatialisation binaurale. La raison pour laquelle il n'est pas rare de trouver des expériences en réalité virtuelle mixées en stéréo est que les professionnels craignent le changement de timbre entraîné par la spatialisation binaurale, qui peut affecter la qualité du travail de la matière sonore. Le manque de contrôle sur le choix du jeu d'HRIR utilisé dans la plupart des solutions d'intégration de la spatialisation binaurale dans les moteurs de jeu a également été pointé du doigt.

C'est dans ce cadre qu'une solution comme celle proposée par 3D Sound Labs qui laisse ce paramètre accessible aux développeurs et aux sound designer pourrait peut être les faire changer d'avis. Une autre piste à explorer également pourrait être d'essayer d'améliorer la mesure et l'accès à une réponse binaurale personnalisée pour un usage domestique, et faciliter son intégration dans les moteurs de jeu.

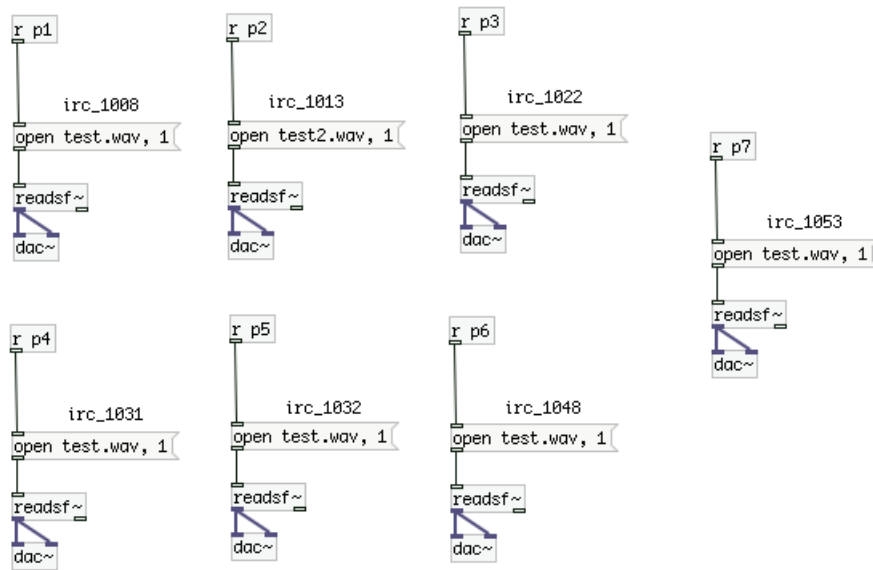
Annexes

ANNEXE 1 : Sélection d'HRIR : mise en place

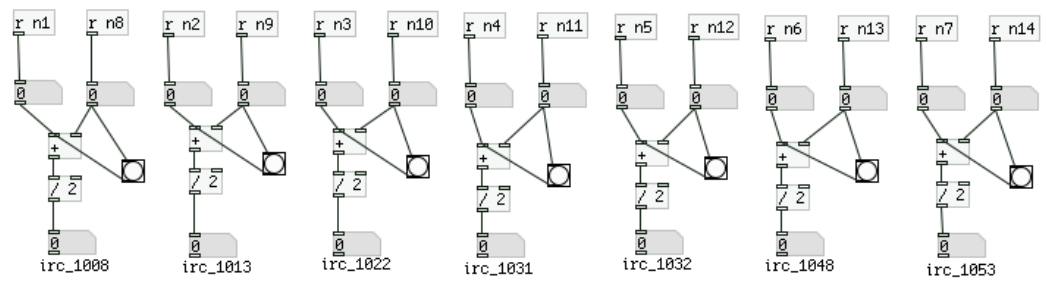
Pour réaliser ce test, on crée un programme sous Puredata constitué de trois patches : l'interface utilisateur, le lecteur, et le relevé des notes attribuées.



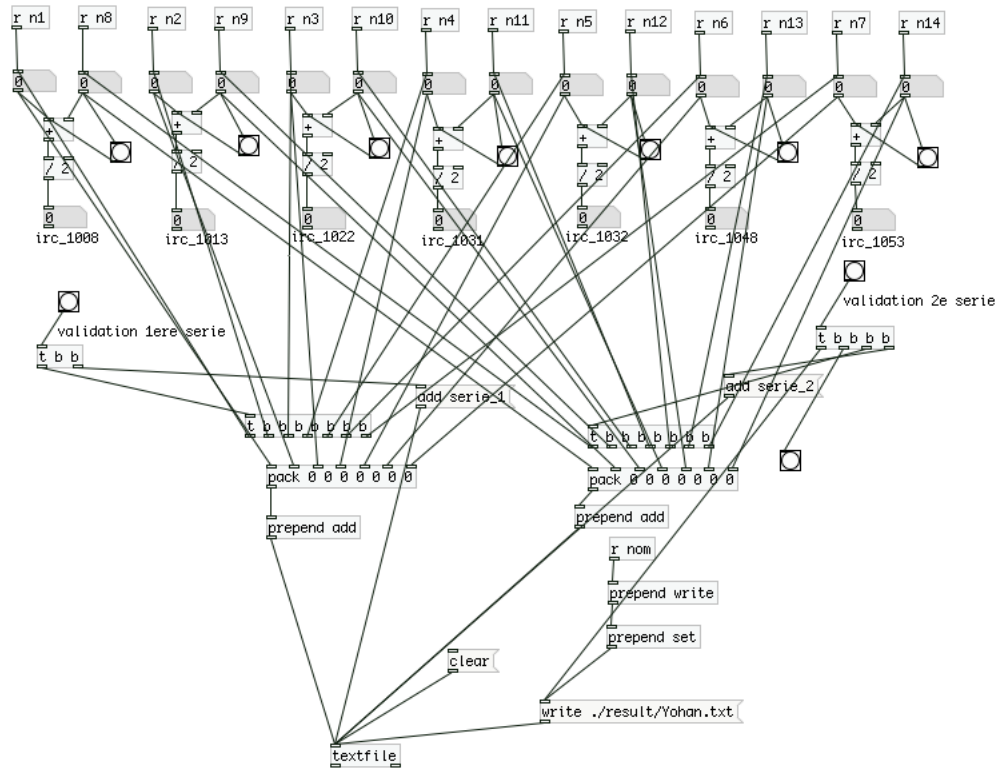
Interface graphique proposée au sujet



Lecteur des fichiers audio



Notes attribuées par le sujet



**Patch complet incluant la partie permettant de récupérer les notes
dans un fichier texte**

Chacun des boutons de lecture de l'interface utilisateur est relié à un objet [send] qui le relie à un des lecteur du patch de lecture. Chacun de ces lecteur lance un fichier audio constitué d'une série de bruits roses de 50ms suivis d'un silence de 450ms qui décrivent un cercle sur le plan horizontal par pas de 30°, puis un arc sur le plan médian par pas de 15°, avec des fades d'entrée et de sortie. Le tout spatialisé à l'aide de l'une des sept HRIR golden ears de la base LISTEN de l'IRCAM, selon le lecteur. Le bouton vert à droite permet de valider les choix du sujet une fois que celui-ci a finalisé sa décision. On lui propose alors une deuxième interface identique avec laquelle il doit juger les mêmes HRIR mais dans un ordre différent.

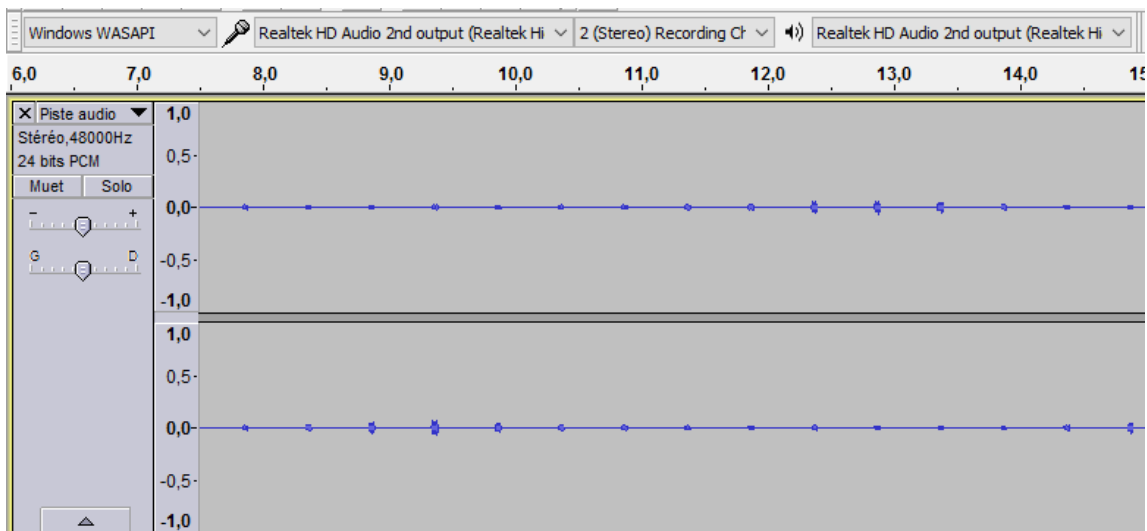
L'échelle graduée accolée à chaque bouton permet de lui attribuer une note entre 0 et 10 selon la qualité de reproduction du cercle, qui est envoyée par un objet [send] au patch qui les relève. Celui effectue la moyenne des deux notes qui correspondent au même fichier audio, puisque le sujet note deux fois chaque jeu d'HRTF. Dans Pure Data, la note maximale se trouve en bas de l'échelle de notation. Or il est plus instinctif pour le sujet de choisir le haut de l'échelle comme meilleure note. **La meilleure note correspond donc à 0 et la pire à 10 dans les données relevées.** Quand les deux séries sont validées, un fichier texte est automatiquement créé avec le nom qui est entré dans la case présente dans le patch de l'interface (dans l'exemple ci-dessus, "suivant.txt").

À partir de ces données, il est facile de savoir la note moyenne entre les deux séries, et de sélectionner la meilleure et la pire spatialisation binaurale pour le sujet.

```
serie_1;  
7 7 8 3 10 9 3;  
serie_2;  
0 9 8 5 5 2 7;
```

Fichier texte obtenu pour le sujet 1, avec les notes données respectivement à la réponse irc 1008, 1013, 1022, 1031, 1032, 1048 et 1053

Pour enregistrer les sons de ce test, on passe par le logiciel Audacity avec lequel on enregistre directement la sortie audio de l'ordinateur sur lequel se trouve le projet Unity



Exemple d'enregistrement sur Audacity. L'enregistrement s'effectue en 48kHz/24bits, et on règle les entrées et sorties du logiciel pour récupérer directement la sortie audio de Windows

Dans notre projet Unity, on rajoute un objet appelé "Sphere" auquel on attache un script qui permet de lui faire effectuer soit un cercle sur le plan horizontal, soit un cercle sur le plan médian, dont on peut gérer l'éloignement et la vitesse de déplacement. On ajuste ces deux paramètres pour faire en sorte que le son de test soit émis aux positions voulues.

```
1  using System.Collections;
2  using System.Collections.Generic;
3  using UnityEngine;
4
5  public class CercleH : MonoBehaviour {
6      float timeCounter = 0;
7      float speed;
8      // Use this for initialization
9      void Start () {
10         speed = 1f;
11     }
12
13     // Update is called once per frame
14     void Update () {
15         timeCounter += Time.deltaTime*speed;
16
17         float x = Mathf.Sin(timeCounter) * 5;
18         float y = 1;
19         float z = Mathf.Cos(timeCounter) * 5;
20
21         transform.position = new Vector3(x, y, z);
22     }
23 }
24
```

Script qui fait décrire une cercle horizontal à la sphère

```

1  using System.Collections;
2  using System.Collections.Generic;
3  using UnityEngine;
4
5  public class CercleVertFront : MonoBehaviour {
6      float timeCounter = 0;
7      float speed;
8      // Use this for initialization
9      void Start()
10     {
11         speed = 0.5f;
12     }
13
14     // Update is called once per frame
15     void Update()
16     {
17         timeCounter += Time.deltaTime*speed;
18
19         float x = 0;
20         float y = Mathf.Sin(timeCounter) * 5;
21         float z = Mathf.Cos(timeCounter) * 5;
22
23         transform.position = new Vector3(x, y, z);
24     }
25 }
26

```

Script qui fait décrire une cercle sur le plan médian à la sphère

Une fois les deux mouvements enregistrés, ils sont montés dans Audacity pour constituer un seul fichier audio que l'on exporte au format .WAV, 48kHz/24bits. L'opération est effectuée sept fois, une fois avec chaque jeu d'HRIR. Pour rappel, il s'agit des *irc 1008*, *irc 1013*, *irc 022*, *irc 1031*, *irc 1032*, *irc 1048* et *irc 1053*.

ANNEXE 2 : Sélection d'HRIR : résultats complets

	sujet 1						
	irc_1008	irc_1013	irc_1022	irc_1031	irc_1032	irc_1048	irc_1053
serie_1	7	7	8	3	10	9	3
serie_2	0	9	8	5	5	2	7
erreur	7	2	0	2	5	7	4
moyenne	3,5	8	8	4	7,5	5,5	5
erreur moyenne	3,9						
erreur moyenne (%)	39						

exemple de résultats obtenus à l'issue de la phase de sélection

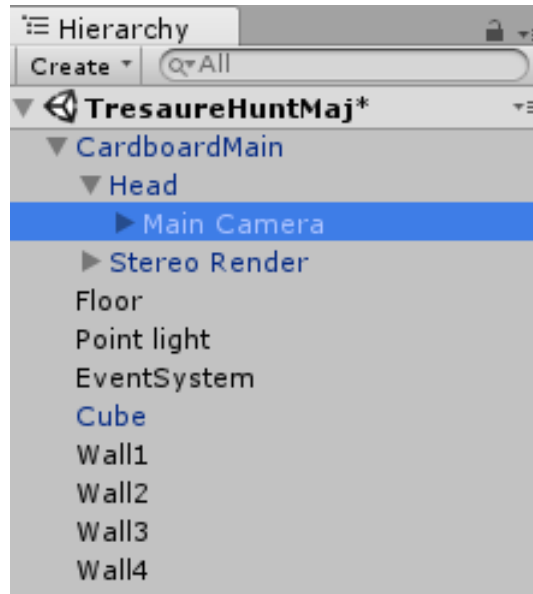
Pour chaque sujet, on obtient un tableau similaire à celui ci-dessus. En vert, le jeu d'HRIR considéré comme le meilleur par le sujet (on rappelle que les notes sont inversées et qu'ici 0 et la note maximale et 10 la plus faible) et en rouge le pire. Quand comme dans ce cas deux réponses sont ex aequo, on prend la réponse avec la plus faible erreur pour effectuer la phase de jeu. Les HRIR sélectionnées et l'erreur moyenne en pourcentage de chaque sujet sont reportées dans le tableau ci-dessous.

	meilleure HRIR	pire HRIR	erreur moyenne (%)
sujet 1	1008	1022	39
sujet 2	1053	1022	16
sujet 3	1013	1048	24
sujet 4	1053	1013	16
sujet 5	1008	1022	33
sujet 6	1048	1008	21
sujet 7	1048	1032	33
sujet 8	1013	1022	35
sujet 9	1022	1032	21
sujet 10	1053	1032	11
sujet 11	1008	1013	21
sujet 12	1013	1053	30
sujet 13	1031	1048	21
sujet 14	1048	1053	20
sujet 15	1031	1022	13
sujet 16	1013	1022	19
sujet 17	1031	1022	
sujet 18	1031	1032	19
sujet 19	1048	1022	14
sujet 20	1031	1008	13
sujet 21	1048	1022	26
sujet 22	1022	1032	19
sujet 23	1013	1008	11
sujet 24	1008	1022	20
sujet 25	1048	1031	10
sujet 26	1008	1032	24
sujet 27	1032	1048	27
sujet 28	1008	1031	16
sujet 29	1032	1008	24
sujet 30	1008	1022	23
sujet 31	1008	1022	26
		MOYENNE TOTALE	22
		MOYENNE RETENUE	19

liste des HRIR sélectionnées par chaque sujet à l'issue des deux phases de sélection. Dans la troisième colonne l'erreur moyenne de notation entre les deux phases. En rouge les sujets qui ne sont pas pris en compte.

ANNEXE 3 : Projet Unity de *Cubes*

Le projet Unity est constitué d'un ensemble de "game objects", que l'on fait interagir entre eux et auxquels on peut incorporer différents composants qui correspondent à des scripts plus ou moins pré-codés. Le projet utilisé dans le contexte de l'expérience contient un objet cube (*Cube*), un correspondant à chacun des murs et au sol qui délimite le champ de vision du joueur (*Floor* et *Wall*), une lumière qui permet d'éclairer la scène (*Point Light*), un objet qui détermine les actions que peut effectuer le sujet (*Event-System*) et un objet permettant de mettre en place les conditions pour que la caméra s'adapte à un système de réalité virtuelle (*CarboardMain*), lui même constitué de deux sous-élément : un qui décrit le point central qui constitue le joueur (*Head*) et un qui permet d'assurer le rendu de la stéréoscopie (*Stereo Render*).



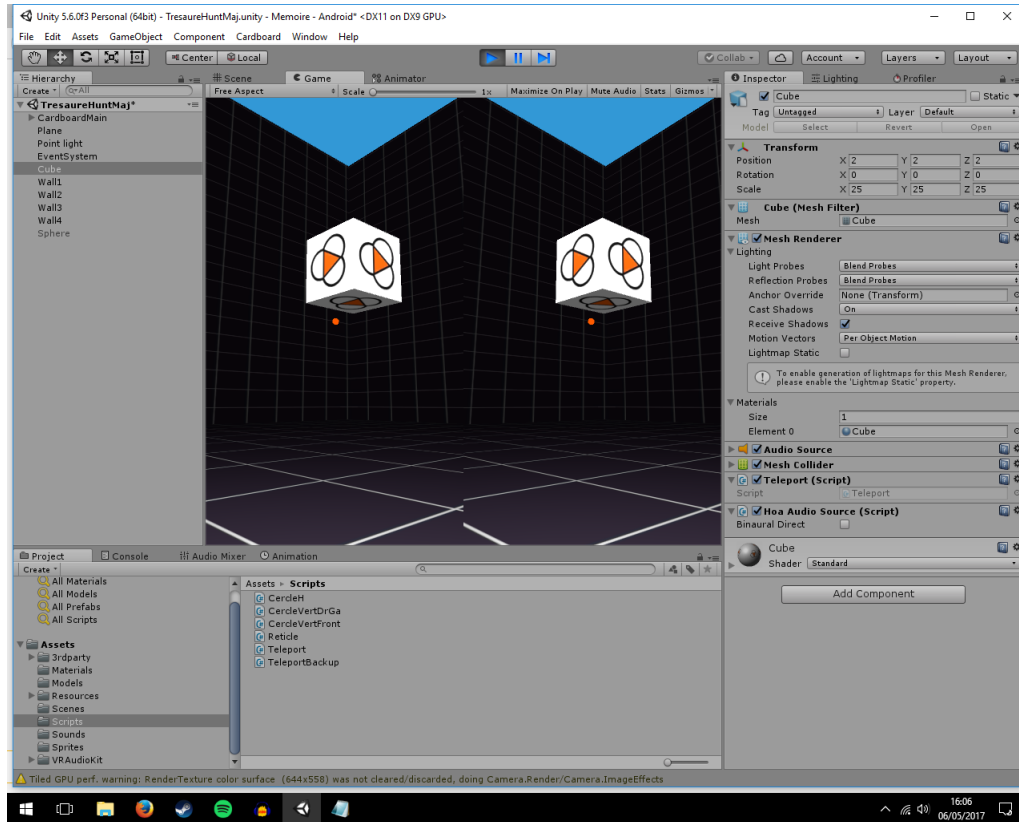
Liste des "game objects"

Dans cette description du projet, on va se concentrer essentiellement sur les objets *Cube* et *Head*, qui sont ceux qui sont intéressants dans le cadre de notre expérience, et aux scripts concernant les tests et l'audio. On fournira en outre les scripts qui ont été écrits ou modifiés dans le cadre du mémoire.

L'objet *Cube* contient trois éléments pertinents : une source audio, pour laquelle indique l'emplacement du son du cube, et deux scripts, l'un créé par 3D Sound Labs pour gérer l'utilisation de l'ambisonic dans la spatialisation du son de la source, et un script qui gère sa téléportation.

L'objet *Head* contient lui-même un autre objet, *Main Camera*, qui se décompose en sous-élément contenant la caméra correspondant à chaque œil, et le point qui indique à l'utilisateur ce qu'il est en train de regarder (*Reticle*). C'est sur *Main Camera* qu'on vient mettre le composant *Audio Listener* qui correspond au point d'écoute durant la session de jeu. C'est à cet *Audio Listener* qu'il faut adjoindre un script créé par 3D Sound Labs pour implanter la spatialisation binaurale du son. Ce script permet, entre autres de déterminer quel jeu d'HRIR est utilisé pendant la session de jeu (il doit être au format CFGA, le format propriétaire de 3D Sound Labs). C'est donc ce paramètre qu'on va changer pour faire les différents exports. En outre, ce script permet aussi de changer l'ordre des HOA (High Order Ambisonics), et par conséquent la finesse du maillage spatial. Ce réglage est donc poussé au maximum. Il faudrait sans doute le descendre dans le cas d'une utilisation dans un jeu plus gourmand en ressource.

C'est l'objet *Reticle* qui va lui actionner la téléportation du cube, en lançant la fonction de téléportation lorsque il y a une collision entre ce point central du regard et un cube.



Vue de la session Unity

Les scripts ci-dessous sont ceux qui ont été ajoutés à la session dans le cadre de ce mémoire. Celui-ci permet de récupérer la position du headtracker toutes les 5ms (Il se déclenche et s'arrête lorsque le premier et le dernier cube sont trouvés) pour pouvoir ensuite calculer les angles correspondant aux mouvements de la tête.

```

1  using System.Collections;
2  using System.Collections.Generic;
3  using UnityEngine;
4  using System.IO;
5
6  public class HeadtrackerData : MonoBehaviour
7  {
8
9      // Use this for initialization
10     Quaternion rotationTete;
11     Vector3 vecteurPosition;
12     float counter = 0f;
13     float x;
14     float y;
15     float z;
16     float pi = Mathf.PI;
17     StreamWriter file;
18     string filePath; // chemin où est sauvé le fichier texte
19     const float RefreshRate = 0.005f; // fréquence de relevé de la position de la tête
20
21     void Start()
22     {
23         InvokeRepeating("SendPosition", 0, RefreshRate);
24         filePath = Application.persistentDataPath + "/HeadTracker.txt";
25         file = new StreamWriter(filePath, true);
26         file.WriteLine("Start\r\n");
27     }
28
29     void SendPosition()
30     {
31         vecteurPosition = GameObject.Find("Head").transform.eulerAngles;
32
33         x = vecteurPosition[0];
34         if (x < 180)
35         {
36             x = -x * (pi / 180);
37         }
38         else
39         {
40             x = (360 - x) * (pi / 180); // azimuth en radian entre 0 et 2pi
41         }
42         y = vecteurPosition[1] * pi / 180; // azimuth en radian entre pi/2 et -pi/2
43
44         counter = Time.time;
45
46         file.WriteLine("{0} {1} {2} \r\n", counter, y, x); // print temps, azimuth, elevation
47     }
48 }
49
50

```

Script permettant de récupérer l'orientation de la tête toutes les 5ms

Ce script correspond à la modification de la fonction de téléportation, pour que le cube suive un parcours déterminé à l'avance, comparé à l'ancien script de téléportation. Il contient également la fonction qui crée et récupère le chronomètre lors de l'épreuve. Pour comparaison, on donne d'abord le script de téléportation aléatoire.

```
15 using UnityEngine;
16 using System.Collections;
17 using UnityEngine.UI;
18 using System.IO;
19
20
21
22 [RequireComponent(typeof(Collider))]
23 public class Teleport : MonoBehaviour
24 {
25     StreamWriter file;
26     string filePath;
27     void Start()
28     {
29
30         filePath = Application.persistentDataPath + "/Chrono.txt";
31     }
32
33
34     void LateUpdate()
35     {
36         Cardboard.SDK.UpdateState();
37         if (Cardboard.SDK.BackButtonPressed)
38         {
39             Application.Quit();
40         }
41     }
42
43     public void TeleportRandomly()
44     {
45         float distance = 1.5f * Random.value + 1.5f;
46         float azimuth = 2.0f * Mathf.PI * Random.value;
47         float elevation = 0.4f * Mathf.PI * Random.value;
48
49         Vector3 direction = new Vector3();
50         direction.x = Mathf.Sin(azimuth) * Mathf.Cos(elevation);
51         direction.y = Mathf.Sin(elevation);
52         direction.z = Mathf.Cos(azimuth) * Mathf.Cos(elevation);
53
54         Vector3 origin = new Vector3(0.0f, 0.25f, 0.0f);
55         transform.localPosition = origin + direction * distance;
56     }
57
58 }
```

Script original de téléportation aléatoire

```

59
60     int i = 0;
61     float counter = 0f;
62     float distance;
63     float azimuth;
64     float elevation;
65     float pi = Mathf.PI;
66     Vector3 CoordoneesSphDp;
67     Vector3 CoordoneesSphAr;
68     Vector3 direction;
69     Vector3 origin;
70     Vector3 sum;
71
72
73
74     public void TeleportNotRandomly()
75     {
76         /*Vector3[] positionArray = new Vector3[32]; // liste de positions 1
77         positionArray[0] = new Vector3(2.5f, 0, 0); // point de départ
78         positionArray[1] = new Vector3(2.5f, 0, pi/12); // 0° azimuth, + 15° elevation
79         positionArray[2] = new Vector3(2.5f, -pi/6, 0); // - 30° azimuth, 0° elevation
80         positionArray[3] = new Vector3(2.5f, -pi/2, pi/6); // - 90° azimuth, + 30° elevation
81         positionArray[4] = new Vector3(2.5f, pi, -pi/4); // + 180° azimuth, - 45° elevation
82         positionArray[5] = new Vector3(2.5f, 2*pi/3, pi/12); // +120° a, +15° e
83         positionArray[6] = new Vector3(2.5f, pi/6, pi/3); // +30° a, +60° e
84         positionArray[7] = new Vector3(2.5f, -pi/6, -pi/6); // -30° a, -30° e
85         positionArray[8] = new Vector3(2.5f, pi/2, -pi/12); // +90° a, -15° e
86         positionArray[9] = new Vector3(2.5f, -pi/3, 0); // -60° a, 0°e
87         positionArray[10] = new Vector3(2.5f, -2*pi/3, -pi/4); // -120° a, -45°e
88         positionArray[11] = new Vector3(2.5f, pi, pi/12); // 180° a, +15° e
89         positionArray[12] = new Vector3(2.5f, pi/3, pi/4); // +60° a, +45° e
90         positionArray[13] = new Vector3(2.5f, -pi/2, -pi/6); // -90° a, -30° e
91         positionArray[14] = new Vector3(2.5f, 5*pi/6, -pi/12); // +150° a, -15° e
92         positionArray[15] = new Vector3(2.5f, pi/6, pi/4); // +30° a, +45° e
93         positionArray[16] = new Vector3(2.5f, 0, -pi/4); // 0° a, -45° e
94         positionArray[17] = new Vector3(2.5f, 2*pi/3, 0); // 120° a, 0° e
95         positionArray[18] = new Vector3(2.5f, -pi/6, pi/3); // -30° a, +60° e
96         positionArray[19] = new Vector3(2.5f, -5*pi/6, 0); // -150° a, 0° e
97         positionArray[20] = new Vector3(2.5f, pi/2, -pi/6); // +90° a, -30° e
98         positionArray[21] = new Vector3(2.5f, 0, -pi/6); // 0° a, -30° e
99         positionArray[22] = new Vector3(2.5f, 2*pi/3, pi/6); // +120° a, +30° a
100        positionArray[23] = new Vector3(2.5f, -pi/3, -pi/6); // -60° a, -30° e
101        positionArray[24] = new Vector3(2.5f, 5*pi/6, pi/4); // +150° a, 45° e
102        positionArray[25] = new Vector3(2.5f, pi/6, -pi/12); // +30° a, -15° e
103        positionArray[26] = new Vector3(2.5f, pi, -pi/6); // 180° a, -30° e
104        positionArray[27] = new Vector3(2.5f, -5*pi/6, pi/6); // -150° a, +30° e
105        positionArray[28] = new Vector3(2.5f, -pi/2, -pi/4); // -90° a, -45° e
106        positionArray[29] = new Vector3(2.5f, pi/3, pi/12); // 60° a, 15° e
107        positionArray[30] = new Vector3(2.5f, pi, 0); // 180° a, 0° e
108        positionArray[31] = new Vector3(2.5f, -pi/2, 0); // -90° a, 0° e*/

```

```

109
110
111 Vector3[] positionArray = new Vector3[32]; // liste de positions 2
112 positionArray[0] = new Vector3(2.5f, 0, 0); // point de départ
113 positionArray[1] = new Vector3(2.5f, 0, pi / 12); // 0° azimuth, + 15° elevation
114 positionArray[2] = new Vector3(2.5f, -pi / 2, 0); // -90° a, 0° e
115 positionArray[3] = new Vector3(2.5f, pi, 0); // 180° a, 0° e
116 positionArray[4] = new Vector3(2.5f, pi / 3, pi / 12); // 60° a, +15° e
117 positionArray[5] = new Vector3(2.5f, 0, pi / 6); // 0° a, +30° e
118 positionArray[6] = new Vector3(2.5f, -pi / 2, -pi / 4); // -90° a, -45° e
119 positionArray[7] = new Vector3(2.5f, -5 * pi / 6, pi / 6); // -150° a, +30° e
120 positionArray[8] = new Vector3(2.5f, pi, -pi / 6); // 180° a, -30° e
121 positionArray[9] = new Vector3(2.5f, pi / 6, -pi / 12); // +30° a, -15° e
122 positionArray[10] = new Vector3(2.5f, 5 * pi / 6, pi / 4); // +150° a, 45° e
123 positionArray[11] = new Vector3(2.5f, -pi / 3, -pi / 6); // -60° a, -30° e
124 positionArray[12] = new Vector3(2.5f, 2 * pi / 3, pi / 6); // +120° a, +30° a
125 positionArray[13] = new Vector3(2.5f, pi / 2, -pi / 6); // +90° a, -30° e
126 positionArray[14] = new Vector3(2.5f, -5 * pi / 6, 0); // -150° a, 0° e
127 positionArray[15] = new Vector3(2.5f, -pi / 6, pi / 6); // -30° a, +30° e
128 positionArray[16] = new Vector3(2.5f, 2 * pi / 3, 0); // 120° a, 0° e
129 positionArray[17] = new Vector3(2.5f, 0, -pi / 4); // 0° a, -45° e
130 positionArray[18] = new Vector3(2.5f, pi / 6, pi / 4); // +30° a, +45° e
131 positionArray[19] = new Vector3(2.5f, 5 * pi / 6, -pi / 12); // +150° a, -15° e
132 positionArray[20] = new Vector3(2.5f, -pi / 2, -pi / 6); // -90° a, -30° e
133 positionArray[21] = new Vector3(2.5f, pi / 3, pi / 4); // +60° a, +45° e
134 positionArray[22] = new Vector3(2.5f, pi, pi / 12); // 180° a, +15° e
135 positionArray[23] = new Vector3(2.5f, -2 * pi / 3, -pi / 4); // -120° a, -45° e
136 positionArray[24] = new Vector3(2.5f, -pi / 3, 0); // -60° a, 0° e
137 positionArray[25] = new Vector3(2.5f, pi / 2, -pi / 12); // +90° a, -15° e
138 positionArray[26] = new Vector3(2.5f, -pi / 6, pi / 12); // -30° a, +15° e
139 positionArray[27] = new Vector3(2.5f, pi / 6, pi / 3); // +30° a, +60° e
140 positionArray[28] = new Vector3(2.5f, 2 * pi / 3, -pi / 12); // +120° a, -15° e
141 positionArray[29] = new Vector3(2.5f, pi, -pi / 4); // + 180° azimuth, - 45° elevation
142 positionArray[30] = new Vector3(2.5f, -pi / 2, pi / 6); // - 90° azimuth, +30° elevation
143 positionArray[31] = new Vector3(2.5f, -pi / 3, 0); // - 60° azimuth, 0° elevation
144

```



```

145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
}
}
}

if (i==31)
{
    transform.localPosition = new Vector3(0f, 0.5f, 1.5f);
    Time.timeScale = 0;
    StreamWriter file = new StreamWriter(filePath, true);

    file.WriteLine("fin de session \r\n");
    file.Close();
    Application.Quit();
}

if (i < positionArray.Length)
{
    CoordoneesSphDp += positionArray[i];
    CoordoneesSphAr = positionArray[i + 1];

    distance = CoordoneesSphAr.x;
    azimut = CoordoneesSphAr.y + CoordoneesSphDp.y;
    elevation = CoordoneesSphAr.z + CoordoneesSphDp.z;

    direction.x = Mathf.Sin(azimut) * Mathf.Cos(elevation);
    direction.y = Mathf.Sin(elevation);
    direction.z = Mathf.Cos(azimut) * Mathf.Cos(elevation);

    Vector3 origin = new Vector3(0.0f, 1f, 0.0f);

    transform.localPosition = origin + direction * distance;
    i++;
    counter = Time.time;
    file = new StreamWriter(filePath, true);
    file.WriteLine("{0} \r\n", counter);
    file.Close(); // inscrit le temps écoulé dans un fichier texte sur une nouvelle ligne
}
}
}

```

Script modifié avec ajout du chronomètre

Enfin, les deux tableaux ci-dessous comportent les positions successives utilisés dans la première et le deuxième session de jeu. Les positions surlignées en vert correspondent à celle qui sont visibles à l'écran dès leur apparition, et qui ne sont donc pas pris en compte dans les calculs de temps, puisque on part du principe que la vision prend le pas sur l'audition dans ce cas. Ces positions sont relatives, c'est à dire qu'elles n'indiquent non pas la position des cubes de façon absolue dans l'espace, mais leur position par rapport au cube précédent, et donc le mouvement à effectuer pour les repérer.

	mouvements 1	
	azimuth (°)	élévation (°)
0	0	0
1	0	15
2	-30	0
3	-90	30
4	180	-45
5	120	15
6	30	60
7	-30	-30
8	90	-15
9	-60	0
10	-120	-45
11	180	15
12	60	45
13	-90	-30
14	150	-15
15	30	45
16	0	-45
17	120	0
18	-30	60
19	-150	0
20	90	-30
21	0	-30
22	120	30
23	-60	-30
24	150	45
25	30	-15
26	180	-30
27	-150	30
28	-90	-45
29	60	15
30	180	0

	mouvements 2	
	azimuth (°)	élévation (°)
0	0	0
1	0	15
2	-90	0
3	180	0
4	60	15
5	0	30
6	-90	-45
7	-150	30
8	180	-30
9	30	-15
10	150	45
11	-60	-30
12	120	30
13	90	-30
14	-150	0
15	-30	30
16	120	0
17	0	-45
18	30	45
19	150	-15
20	-90	-30
21	60	45
22	180	15
23	-120	-45
24	-60	0
25	90	-15
26	-30	15
27	30	60
28	120	-15
29	180	-45
30	90	30

Liste des positions des cubes dans les deux parties. En vert les cubes apparaissant directement à l'écran

ANNEXE 4 : Phases de jeu : résultats complets

Temps de jeu

Les temps de jeu ont été mesurés de façon globale et moyenne (temps mis pour repérer chaque cube). On ne prend pas en compte dans le calcul du temps écoulé les cubes apparaissant directement dans le champ de vision du joueur. Ci-dessous un exemple de données recueillies pour un sujet, puis les données globales (les données complètes pour tous les sujets sont disponibles sur demande) :

SUJET 1 - Série 1 - meilleure HRIR				
position	azimuth (élévation	temps écoulé (s)	temps mouvement (s)
0	0	0	38,25	
1	0	15	40,04	1,79
2	-30	0	42,59	2,55
3	-90	30	47,07	4,48
4	180	-45	52,23	5,16
5	120	15	55,80	3,57
6	30	60	67,06	11,25
7	-30	-30	68,33	1,27
8	90	-15	71,62	3,29
9	-60	0	74,72	3,11
10	-120	-45	80,02	5,29
11	180	15	88,30	8,28
12	60	45	93,07	4,77
13	-90	-30	96,85	3,77
14	150	-15	100,23	3,39
15	30	45	104,60	4,36
16	0	-45	106,02	1,43
17	120	0	108,72	2,69
18	-30	60	112,26	3,55
19	-150	0	117,50	5,24
20	90	-30	120,10	2,60
21	0	-30	122,31	2,21
22	120	30	125,65	3,34
23	-60	-30	128,12	2,46
24	150	45	132,68	4,56
25	30	-15	134,66	1,98
26	180	-30	139,48	4,82
27	-150	30	143,27	3,80
28	-90	-45	145,78	2,51
29	60	15	147,89	2,10
30	180	0	151,88	4,00
TOTAL			103,83	
			TPS MOYEN	4,15

Temps total et moyen relevé pour le sujet 1. En vert, les positions qui apparaissent à l'écran et ne sont pas pris en compte dans le chronomètre.

	ordre de passage	temps 1 (s)	temps 2 (s)	différence (s)	progression (%)
sujet 1	meilleure/pire	103,83	84,51	-19,32	18,61
sujet 2	pire/meilleure	119,51	87,97	-31,54	26,40
sujet 3	meilleure/pire	135,22	136,64	1,43	-1,06
sujet 4	pire/meilleure	106,03	89,41	-16,61	15,67
sujet 5	meilleure/pire	136,05	120,50	-15,55	11,43
sujet 6	pire/meilleure	94,65	83,38	-11,28	11,91
sujet 7	meilleure/pire	81,65	69,00	-12,65	15,49
sujet 8	pire/meilleure	121,65	131,48	9,83	-8,08
sujet 9	meilleure/pire	75,62	64,56	-11,06	14,63
sujet 10	pire/meilleure	72,98	57,05	-15,94	21,84
sujet 11	meilleure/pire	90,55	74,33	-16,22	17,91
sujet 12	pire/meilleure	82,55	72,21	-10,34	12,53
sujet 13	meilleure/pire	73,52	78,37	4,84	-6,59
sujet 14	pire/meilleure	129,36	83,09	-46,27	35,77
sujet 15	meilleure/pire	82,86	77,20	-5,66	6,83
sujet 16	pire/meilleure	78,02	63,49	-14,53	18,63
sujet 17					
sujet 18	meilleure/pire	83,84	77,41	-6,44	7,68
sujet 19	pire/meilleure	106,38	97,35	-9,03	8,49
sujet 20	meilleure/pire	92,61	85,55	-7,06	7,63
sujet 21	pire/meilleure	163,99	87,59	-76,40	46,59
sujet 22	meilleure/pire	97,47	107,56	10,08	-10,35
sujet 23	pire/meilleure	113,14	87,98	-25,16	22,24
sujet 24	meilleure/pire	76,97	77,77	0,81	-1,05
sujet 25	pire/meilleure	125,60	88,89	-36,71	29,23
sujet 26	meilleure/pire	105,34	85,09	-20,25	19,22
sujet 27	pire/meilleure	176,05	116,96	-59,08	33,56
sujet 28	meilleure/pire	97,83	101,21	3,38	-3,46
sujet 29	pire/meilleure	114,89	80,93	-33,96	29,56
sujet 30	meilleure/pire	89,18	80,27	-8,91	10,00
sujet 31	pire/meilleure	115,61	82,16	-33,45	28,93
	MOYENNE TOTALE	104,69	87,66	-17,10	16,26
	MOY. MEILLEURE	90,52	87,33	-3,19	3,52
	MOY. PIRE	116,63	88,00	-28,63	24,55
	ÉCART TYPE TOTAL	26,16	16,35	20,15	13,94
	ÉCART TYPE MEILLEURE	16,09	13,89	19,09	10,00
	ÉCART TYPE PIRE	27,92	18,60	8,59	9,21

Temps total mis part chaque sujet pour compléter chaque session de jeu. En rouge les sujets qui ne sont pas pris en compte dans le calcul et en orange ce qui ont une moins bonne performance sur la deuxième session de jeu.

	ordre de passage	temps moyen 1 (s)	temps moyen 2 (s)	différence (s)	progression (%)
sujet 1	meilleure/pire	4,15	3,38	-0,77	18,55
sujet 2	pire/meilleure	4,78	3,52	-1,26	26,36
sujet 3	meilleure/pire	5,41	5,47	0,06	-1,11
sujet 4	pire/meilleure	4,24	3,58	-0,66	15,57
sujet 5	meilleure/pire	5,44	4,82	-0,62	11,40
sujet 6	pire/meilleure	3,79	3,34	-0,45	11,87
sujet 7	meilleure/pire	3,27	2,76	-0,51	15,60
sujet 8	pire/meilleure	4,87	5,26	0,39	-8,01
sujet 9	meilleure/pire	3,02	2,58	-0,44	14,57
sujet 10	pire/meilleure	2,92	2,28	-0,64	21,92
sujet 11	meilleure/pire	3,62	2,97	-0,65	17,96
sujet 12	pire/meilleure	3,30	2,89	-0,41	12,42
sujet 13	meilleure/pire	2,94	3,13	0,19	-6,46
sujet 14	pire/meilleure	5,17	3,32	-1,85	35,78
sujet 15	meilleure/pire	3,31	3,09	-0,22	6,65
sujet 16	pire/meilleure	3,12	2,54	-0,58	18,59
sujet 17					
sujet 18	meilleure/pire	3,35	3,10	-0,25	7,46
sujet 19	pire/meilleure	4,26	3,89	-0,37	8,69
sujet 20	meilleure/pire	3,70	3,42	-0,28	7,57
sujet 21	pire/meilleure	6,56	3,50	-3,06	46,65
sujet 22	meilleure/pire	3,90	4,30	0,4	-10,26
sujet 23	pire/meilleure	4,53	3,52	-1,01	22,30
sujet 24	meilleure/pire	3,08	3,11	0,03	-0,97
sujet 25	pire/meilleure	5,02	3,56	-1,46	29,08
sujet 26	meilleure/pire	4,21	3,40	-0,81	19,24
sujet 27	pire/meilleure	7,04	4,68	-2,36	33,52
sujet 28	meilleure/pire	3,91	4,05	0,14	-3,58
sujet 29	pire/meilleure	4,60	3,24	-1,36	29,57
sujet 30	meilleure/pire	3,57	3,21	-0,36	10,08
sujet 31	pire/meilleure	4,62	3,29	-1,33	28,79
	MOYENNE TOTALE	4,19	3,51	-0,68	16,31
	MOY. MEILLEURE	3,79	3,52	-0,27	7,19
	MOY. PIRE	4,59	3,49	-1,09	23,84
	ÉCART TYPE TOTAL	1,05	0,65	0,81	13,93
	ÉCART TYPE MEILLE	1,12	0,56	0,76	9,98
	ÉCART TYPE PIRE	0,64	0,75	0,34	9,19

Temps moyen mis part chaque sujet entre chaque cube dans les deux session de jeu. En rouge les sujets qui ne sont pas pris en compte dans le calcul et en orange ceux dont la performance décroît sur la deuxième session de jeu.

	ordre de passage	temps moyen 1 (s)	temps moyen 2 (s)	mouv vertical 1 (s)	mouv vertical 2 (s)
sujet 1	meilleure/pire	4,15	3,38	11,25	4,53
sujet 2	pire/meilleure	4,78	3,52	11,53	6,31
sujet 3	meilleure/pire	5,41	5,47	19,51	7,30
sujet 4	pire/meilleure	4,24	3,58	9,24	5,87
sujet 5	meilleure/pire	5,44	4,82	11,73	19,57
sujet 6	pire/meilleure	3,79	3,34	12,15	3,19
sujet 7	meilleure/pire	3,27	2,76	3,93	4,61
sujet 8	pire/meilleure	4,87	5,26	4,84	15,74
sujet 9	meilleure/pire	3,02	2,58	5,97	2,80
sujet 10	pire/meilleure	2,92	2,28	15,43	3,02
sujet 11	meilleure/pire	3,62	2,97	15,09	3,93
sujet 12	pire/meilleure	3,30	2,89	6,14	3,23
sujet 13	meilleure/pire	2,94	3,13	3,65	5,06
sujet 14	pire/meilleure	5,17	3,32	12,58	4,72
sujet 15	meilleure/pire	3,31	3,09	7,00	6,76
sujet 16	pire/meilleure	3,12	2,54	8,3	4,15
sujet 17					
sujet 18	meilleure/pire	3,35	3,10	4,25	4,14
sujet 19	pire/meilleure	4,26	3,89	8,43	5,21
sujet 20	meilleure/pire	3,70	3,42	6,01	2,64
sujet 21	pire/meilleure	6,56	3,50	8,81	5,50
sujet 22	meilleure/pire	3,90	4,30	6,01	7,12
sujet 23	pire/meilleure	4,53	3,52	26,11	2,68
sujet 24	meilleure/pire	3,08	3,11	8,44	9,85
sujet 25	pire/meilleure	5,02	3,56	30,52	3,99
sujet 26	meilleure/pire	4,21	3,40	17,21	8,32
sujet 27	pire/meilleure	7,04	4,68	63,31	13,61
sujet 28	meilleure/pire	3,91	4,05	9,72	10,59
sujet 29	pire/meilleure	4,60	3,24	24,46	6,57
sujet 30	meilleure/pire	3,57	3,21	15,09	8,32
sujet 31	pire/meilleure	4,62	3,29	8,31	3,45
	MOYENNE TOTALE	4,19	3,51	13,17	6,43
	MOY. MEILLEURE	3,79	3,52	9,66	7,04
	MOY. PIRE	4,59	3,49	16,68	5,82
	ÉCART TYPE TOTAL	1,05	0,65	12,09	2,70
	ÉCART TYPE MEILLEURE	1,12	0,56	14,85	2,71
	ÉCART TYPE PIRE	0,64	0,75	5,22	2,56

temps moyen entre chaque cube comparé au temps mis pour trouver le cube à la verticale. En rouge les sujets qui ne sont pas pris en compte dans le calcul et en orange ceux dont la performance décroît sur la deuxième session de jeu.

CALCUL DU TEMPS GLOBAL SANS LA POSITION VERTICALE					
	ordre de passage	temps 1 (s)	temps 2 (s)	différence (s)	progression
sujet 1	meilleure/pire	92,58	79,98	-12,60	13,61
sujet 2	pire/meilleure	107,98	81,66	-26,32	24,38
sujet 3	meilleure/pire	115,71	129,34	13,64	-11,79
sujet 4	pire/meilleure	96,79	83,54	-13,24	13,68
sujet 5	meilleure/pire	124,32	100,93	-23,39	18,82
sujet 6	pire/meilleure	82,50	80,19	-2,32	2,81
sujet 7	meilleure/pire	81,65	64,39	-17,26	21,14
sujet 8	pire/meilleure	116,81	115,74	-1,07	0,91
sujet 9	meilleure/pire	69,65	61,76	-7,89	11,33
sujet 10	pire/meilleure	57,55	54,03	-3,53	6,13
sujet 11	meilleure/pire	75,46	70,40	-5,06	6,70
sujet 12	pire/meilleure	76,41	68,98	-7,43	9,73
sujet 13	meilleure/pire	69,87	73,31	3,43	-4,92
sujet 14	pire/meilleure	116,78	78,37	-38,41	32,89
sujet 15	meilleure/pire	75,86	70,44	-5,42	7,15
sujet 16	pire/meilleure	69,72	59,34	-10,38	14,89
sujet 17					
sujet 18	meilleure/pire	79,59	73,27	-6,33	7,95
sujet 19	pire/meilleure	97,95	92,14	-5,81	5,93
sujet 20	meilleure/pire	86,60	82,91	-3,69	4,27
sujet 21	pire/meilleure	155,18	82,09	-73,09	47,10
sujet 22	meilleure/pire	91,46	100,44	8,97	-9,81
sujet 23	pire/meilleure	87,03	85,30	-1,73	1,99
sujet 24	meilleure/pire	68,53	67,92	-0,60	0,88
sujet 25	pire/meilleure	95,08	84,90	-10,18	10,70
sujet 26	meilleure/pire	88,13	76,77	-11,36	12,89
sujet 27	pire/meilleure	112,74	103,35	-9,38	8,32
sujet 28	meilleure/pire	88,11	90,62	2,51	-2,85
sujet 29	pire/meilleure	90,43	74,36	-16,07	17,77
sujet 30	meilleure/pire	74,09	71,95	-2,14	2,89
sujet 31	pire/meilleure	107,30	78,71	-28,59	26,64
	MOYENNE TOTALE	90,40	81,24	-10,49	10,14
	MOY. MEILLEURE	81,36	81,51	0,16	-0,19
	MOY. PIRE	98,23	80,96	-17,27	17,58
	ÉCART TYPE TOTAL	20,49	15,10	16,91	12,78
	ÉCART TYPE MEILLEURE	12,75	12,10	19,05	12,71
	ÉCART TYPE PIRE	23,02	17,78	6,91	7,63

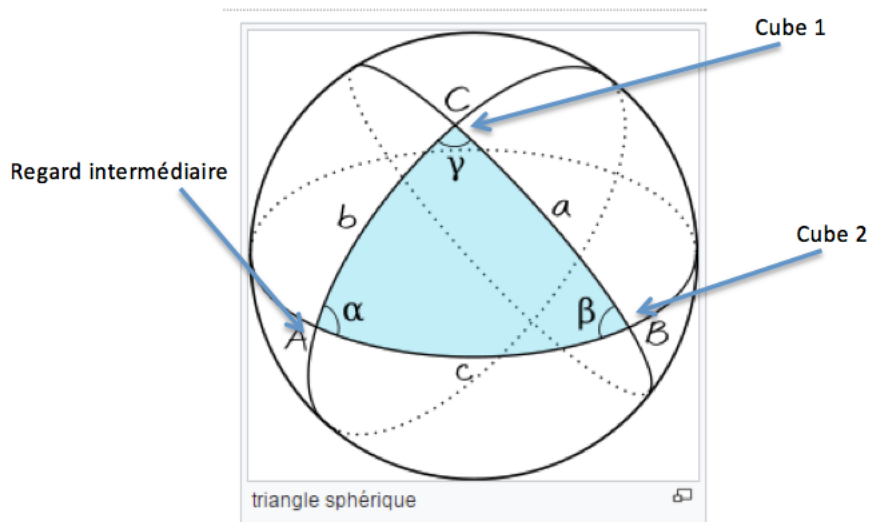
Temps global mis part chaque sujet pour compléter les deux session de jeu sans prendre en compte le déplacement vertical. En rouge les sujets qui ne sont pas pris en compte dans le calcul et en orange ce qui ont une moins bonne performance sur la deuxième session de jeu.

CALCUL DU TEMPS MOYEN SANS LA POSITION VERTICALE					
	ordre de passage	temps moyen 1 (s)	temps moyen 2 (s)	différence (s)	progression (%)
sujet 1	meilleure/pire	3,86	3,33	-0,53	13,73
sujet 2	pire/meilleure	4,50	3,40	-1,10	24,44
sujet 3	meilleure/pire	4,03	5,51	1,48	-36,72
sujet 4	pire/meilleure	4,24	3,54	-0,7	16,51
sujet 5	meilleure/pire	5,18	4,24	-0,94	18,15
sujet 6	pire/meilleure	3,44	3,4	-0,04	1,16
sujet 7	meilleure/pire	3,24	2,70	-0,54	16,67
sujet 8	pire/meilleure	4,87	4,92	0,05	-1,03
sujet 9	meilleure/pire	2,90	2,58	-0,32	11,03
sujet 10	pire/meilleure	2,40	2,27	-0,13	5,42
sujet 11	meilleure/pire	3,14	2,98	-0,16	5,10
sujet 12	pire/meilleure	3,18	2,86	-0,32	10,06
sujet 13	meilleure/pire	2,91	3,06	0,15	-5,15
sujet 14	pire/meilleure	4,87	3,31	-1,56	32,03
sujet 15	meilleure/pire	3,16	2,94	-0,22	6,96
sujet 16	pire/meilleure	2,9	2,49	-0,41	14,14
sujet 17					
sujet 18	meilleure/pire	3,32	3,08	-0,24	7,23
sujet 19	pire/meilleure	4,08	3,91	-0,17	4,17
sujet 20	meilleure/pire	3,61	3,51	-0,10	2,77
sujet 21	pire/meilleure	6,47	3,46	-3,01	46,52
sujet 22	meilleure/pire	3,81	4,21	0,40	-10,50
sujet 23	pire/meilleure	3,63	3,60	-0,03	0,83
sujet 24	meilleure/pire	2,86	3,11	0,25	-8,74
sujet 25	pire/meilleure	3,96	3,61	-0,35	8,84
sujet 26	meilleure/pire	3,67	3,23	-0,44	11,99
sujet 27	pire/meilleure	4,70	4,35	-0,35	7,45
sujet 28	meilleure/pire	3,67	3,83	0,16	-4,36
sujet 29	pire/meilleure	3,77	3,13	-0,64	16,98
sujet 30	meilleure/pire	3,09	3,03	-0,06	1,94
sujet 31	pire/meilleure	4,47	3,32	-1,15	25,73
	MOYENNE TOTALE	3,80	3,43	-0,37	8,11
	MOY. MEILLEURE	3,50	3,42	-0,07	2,01
	MOY. PIRE	4,10	3,44	-0,66	14,22
	ÉCART TYPE TOTAL	0,83	0,64	0,78	15,49
	ÉCART TYPE MEILLEURE	0,96	0,52	0,80	12,96
	ÉCART TYPE PIRE	0,38	0,75	0,49	12,77

Temps moyen entre chaque cube mis part chaque sujet pour compléter les deux session de jeu sans prendre en compte le déplacement vertical. En rouge les sujets qui ne sont pas pris en compte dans le calcul et en orange ce qui ont une moins bonne performance sur la deuxième session de jeu.

Différence d'angle observée

Le principe du calcul est le suivant : le regard du joueur correspond à un point projeté dans un repère sphérique de rayon constant sur lequel sont disposés les cubes. On relève trois jeux de coordonnées dans ce repère sphérique : le point correspondant au cube que vient de repérer le joueur, le point correspondant au nouveau cube qui vient d'apparaître et enfin, on relève le regard du joueur quelques millisecondes après qu'il ait amorcé son déplacement vers le nouveau cube. La situation est décrite dans la figure suivante :



représentation du regard du joueur à trois instants différents

On cherche l'angle gamma qui correspond à l'angle entre la trajectoire de regard idéale à faire d'un cube à l'autre et le trajet amorcé par le sujet. On connaît A (Azim. A ; Haut. A) , B (Azim. B ; Haut. B) , C (Azim. C ; Haut. C), et O (0,0) qui correspondent aux valeurs de la position de la tête

qui sont relevées toutes les 5ms et à l'origine du repère sphérique.

Dans un premier temps on va déterminer les angles a,b,c

$$a = (\text{BOC})$$

$$b = (\text{COA})$$

$$c = (\text{AOC})$$

Ces angles, exprimés en radian, sont aussi les longueurs des cotés du triangle sphérique.

$$\text{Pour l'arc } a : \cos a = \sin(\text{Haut. B}) * \sin(\text{Haut. C}) + \cos(\text{Haut. B}) * \cos(\text{Haut. C}) * \cos(\text{Azim. C} - \text{Azim. B})$$

Il faut faire un test pour choisir la bonne solution :

$$\cos a = X$$

$$\text{Si } a < \text{Pi} \text{ alors } a = \arccos(X)$$

$$\text{Si } a > \text{Pi} \text{ alors } a = 2 * \text{Pi} - \arccos(X)$$

Une fois déterminé les 3 arcs a, b, c on peut calculer les angles grâce aux formules de trigonométrie sphérique

$$\text{Cos}(\text{gama}) = (\cos c - \cos a \cdot \cos b) / (\sin a \cdot \sin b)$$

On ne fait qu'exposer le principe du calcul ici, car il n'a pas lieu d'être réalisé avec les mesures qui ont été effectuées. C'est le calcul qu'il faudra faire dans la nouvelle campagne de mesure entre cet écrit et la soutenance orale.

Bibliographie

Ouvrages

Blauert J., *Spatial hearing : The Psychophysics of Human Sound Localization*, édition revisitée, 1997

Triclot M., *Philosophie des jeux vidéo*, Paris, Zones, 2011

Platon, Livre VII de La République, *L'allégorie de la caverne*, Grèce, IV siècle av. J.-C.

Descartes R., *Discours de la méthode*, France, 1637

Berkeley G., *Principes de la connaissance humaine*, Angleterre, 1710

Articles

Blum A., Katz B., Warusfel O., "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training presence", CFA-DAGA, 2004

Nguyen K., Suied C., Viaud-Delmon I., Warusfel O., "Spatial audition in a static virtual environment : the role of auditory-visual interaction.", *JVRB - Journal of Virtual Reality and Broadcasting* n° 5, 2009

Andreopoulou A., Begault D., Katz B., "Inter-Laboratory Round Robin HRTFF Measurement Comparison", *IEEE J Selected Topics in Signal Processing, vol. 9, no. SI : Spatial Audio*, 2015

Zwiers M. P., Van Opstal A. J., Cruysberg J. R. MR., "A spatial Hearing Deficit in Early-Blind Humans", *The Journal of Neuroscience, Vol. 21 RC142*, 2001

Zwiers M. P., Van Opstal A. J., Paige G. D., "Plasticity in human sound localization induced by compressed spatial vision", *Nature Neuroscience* 6, 2003

chönstein D., Katz B., "Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale.", Société Française d'Acoustique, 10ème Congrès Français d'Acoustique, Lyon, France. 2010

Honda A., Shibata H, Hidaka S., Gyoba J., Iwaya Y et Suzuki Y.,
"Effects of head movement and proprioceptive feedback in training of
sound localization", *iperception*, Juin 2013

Masahiro Mori, "La vallée de l'étrange", *Revue d'anthropologie et
d'histoire des arts*, n° 15, 16 mai 2012

Mémoire ou thèses

Blum A., *Étude de la plasticité du système auditif en localisation sonore. Application au problème de l'individualisation en synthèse binaurale*, rapport de DEA, Université d'Aix-Marseille 2, 2003

Guillon P. *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*. Thèse sous la direction de Simon L., Université du Maine, département acoustique, Novembre 2009

Rugeles Ospina F, *Individualisation de l'écoute binaurale : création et transformation des indices spectraux et des morphologies des individus*, Thèse sous la direction de Katz B., Université Pierre et Marie Curie, département sciences mécaniques, acoustique, électronique et robotique, soutenu en 2016

Larcher V. *Techniques de spatialisation des sons pour la réalité virtuelle*, thèse sous la direction de Polack J.-D., Université Paris VI, soutenu en mai 2001

Zanetto V., *Création d'une interface interactive à usage domestique dans le but de créer un modèle HRTF personnalisé*, mémoire effectué sous la direction de Blum A., École Nationale Supérieure Louis Lumière, soutenu en juin 2017

Parseihian G., *Sonification binaurale pour l'aide à la navigation*, Thèse sous la direction D'Alessandro C., spécialité acoustique, Université Pierre et Marie Curie, octobre 2012

Berger C., *L'écriture sonore dans les jeux vidéo*, mémoire effectué sous la direction de Halbwachs J.-P., École Nationale Supérieure Louis Lumière, soutenu en juin 2010

Soulier A., *Musique et jeux vidéo, de l'interactivité d'un média singulier*, mémoire effectué sous la direction de Rouchouse J., École Nationale Supérieure Louis Lumière, soutenu en juin 2016

Meyer C., *Du mutisme au dialogue : les interactions vocales dans le jeu vidéo*, mémoire effectué sous la direction de Coduys T., École Nationale Supérieure Louis Lumière, soutenu en juin 2015

Palacin B., *Le rôle du son dans l'apparition de la peur dans le jeu vidéo*, mémoire effectué sous la direction de Gazeau C., École Nationale Supérieure Louis Lumière, soutenu en juin 2013

Sources internet

Audio for virtual and augmented reality, 2016 AES Conference,
programme officiel de la conférence, 2016
URL : [http://www.aes.org/conferences/2016/avar/downloads/
Full_AVAR_Conference_Program_V3.pdf](http://www.aes.org/conferences/2016/avar/downloads/Full_AVAR_Conference_Program_V3.pdf)

Hollerweger F., "An introduction to Higher Order Ambisonic", article en
ligne, 2008,
URL : <http://flo.mur.at/writings/HOA-intro.pdf>

Projet Bili
URL : <http://www.bili-project.org>

Majdak P., Noisternig M., "SOFA, Spatially Oriented Format for
Acoustics", version 0.6, 2015,
URL : <https://www.sofaconventions.org/mediawiki/index.php>

Vidéothèque

Fayet M.-H., de la Rochère B., *Réalité virtuelle, année zéro !* et *Réalité virtuelle, pour y faire quoi ?*, vidéo, Videosystem, 1991

Darnell E., *Invasion !*, 2016, États-Unis

Jouinot O., *Vidéo et son 360°, Headtracking*, vidéo youtube à regarder avec un téléphone Android, juillet 2016

Tsingos N., *Audio challenges in virtual and augmented reality*, vidéo youtube, février 2016

Documentation technique

Oculus Audio SDK Documentation, manuel à destination des développeurs pour la plateforme Oculus, v1.1.4, p12, 2017

manuel utilisateur casque Homido,

URL : <http://www.homido.com/en/medias/file/homido-user-manual-web-d4a4.pdf>

VRAudioKit for Unity, revision 1, manuel d'utilisation du plug-in de spatialisation binaurale de 3D Sound Labs, mars 2016

URL : http://3dsoundlabs.com/wp-content/uploads/vraudiokit/VRAudioKit_Unity_Documentation.pdf

Ludothèque

Tétris, développé et édité par Alekseï Pajitnov, 1984

Dark Souls III, développé et édité par From Software, 2016

The Stanley Parable, développé et édité par Galactic Cafe, 2013

Metal Gear Solid 2 : Sons of Liberty, développé et édité par Konami, 2001

The Beginner's Guide développé et édité par Everything Unlimited, 2015

Hellblade, Senua's Sacrifice développé et édité par Ninja Theory, 2017

Inside développé et édité par Playdead, 2017

Alien Isolation développé par Creative Assembly et édité par Sega, 2014)

DIRT développé et édité par Codemasters, 2009-2012

F1 développé et édité par Codemasters, 2011-2016

Papa Sangre développé et édité par Somethin' Else, 2013

BlindSide développé et édité par Epicycle, 2012

A Blind Legend développé et édité par Dowino, 2015