

La vidéo à 360°, réflexions sur la place du son dans ce nouveau média.

Comment la vidéo 360° offre-t-elle de nouvelles contraintes et possibilités par rapport au cinéma pour la mise en scène sonore ?

Mémoire de fin d'études

Corvo LEPESANT-LAMARI Département Son Promotion Agnès VARDA – 2017

Remerciements

Je tiens avant tout à remercier mes directeurs de département et la direction des études de la Fémis pour m'avoir fait confiance quant au choix du sujet de mon mémoire.

Pour leur aide, le temps passé et les discussions fort instructives qui ont nourri ma réflexion: Sylvain Lambinet (École Nationale Supérieure Louis Lumière), Alexandre Dazzoni (Université Paris-Sud, Orsay) et Philippe Fuchs (MINES ParisTech).

Barbara Turquier pour les conseils et contacts précieux.

Laurence Berreur et Delphine Benroubi pour leur suivi.

Geoffrey, Grégoire, Léonard, Louise et Mikhael, mes camarades de promo avec qui nous avons partagé nos doutes, nos erreurs, nos difficultés et notre bonne humeur pendant quatre ans et plus particulièrement cette dernière année.

Enfin je veux remercier tous ceux, personnels de l'école ou amis ou famille, qui m'ont soutenu de près ou de loin pendant les recherches sur ce sujet et la rédaction de mon mémoire.

Préambule

Le choix d'un sujet de mémoire n'est certainement pas une chose aisée, il faut qu'il nous nourrisse dans un intérêt pédagogique, qu'il soit un moyen d'aborder le métier que nous souhaitons faire plus tard, qu'il soit digne d'intérêt pour le lecteur mais aussi qu'il soit source de plaisir et non un fardeau dans son élaboration. J'ai donc cherché un sujet dans l'air du temps, un domaine nouveau, à défricher, mais qui me permette aussi de nourrir une réflexion à la fois technique et artistique.

C'est pour ces raisons que j'ai décidé de travailler sur la vidéo 360°. Nouveau média à la fois futuriste et novateur qui est entré très rapidement dans nos mœurs même s'il n'en est encore qu'à ses balbutiements. Il déchaîne les passions et la curiosité à la fois des spectateurs, des producteurs mais aussi des institutions qui s'en sont emparés très rapidement. De par sa nouveauté et l'intérêt qu'elle suscite, la vidéo 360° est un univers qui évolue très vite. Entre le moment où j'ai commencé les recherches sur le sujet, il y a un an déjà, et le jour où j'écris ces lignes les situations technologiques, économiques et sa diffusion ont bien évoluées.

J'ai donc dans ce mémoire essayé de compiler et de rendre compte des réflexions, des remarques et des découvertes que j'ai pu faire durant cette année à suivre l'émergence de ce nouveau moyen d'expression. La forme de mon plan est certainement peu académique et les sujets abordés hétérogènes mais c'est le meilleur moyen que j'ai trouvé pour vous transmettre mes questionnements et mes apprentissages de l'année passée dans des domaines à la fois techniques, artistiques, scientifiques...

Je souhaite qu'il vous fasse réfléchir et vous questionner tout en - je l'espère - vous instruisant, bonne lecture !

Sommaire

introduction	1
I – Qu'est ce que la vidéo 360° ?	
1 – Réalité virtuelle et vidéo 360°	4
2 – 360°, vraiment ?	7
3 – Utilisations actuelles de la vidéo 360°	10
II – (dé)placement dans l'espace et audition	
1 – Écoute binaurale et champ auditif	12
2 – La réverbération, un marqueur important de l'espace	17
3 – Équilibrioception et proprioception	19
III – Vidéo 360° et mise en scène sonore, quelques pistes	
1 – La question cruciale des points de vue et d'écoute	21
2 – Comment guider le spectateur dans la narration grâce au son	22
3 – Au sujet de la voix	24
IV – Les solutions techniques actuelles pour le son en vidéo 360°	
1 – La synthèse binaurale	24
2 – L'ambisonie	27
3 – Logiciels spécifiques	28
V – Réflexions sur un nouveau système pour le son à 360°	
1 – Max/MSP et le Spat	31
2 – Mon projet	33
Conclusion	35
Bibliographie	37
Filmographie	39
ANNEXE I : Liens vers les vidéos 360°	40
ANNEXE II · Comparatif technique de différents visiocasques	42

Introduction : la question de l'immersion au cinéma

« IMMERSION n.f. : Action d'immerger, de plonger dans un liquide ; résultat de

cette action. »

Paul Robert, Le Petit Robert.

Le concept d'immersion est difficile à définir, le sens compris de tous dans lequel nous allons l'utiliser est celui de transporter le spectateur dans un univers, une vie, un point de vue qui n'est pas le sien, qui ne lui est pas habituel et ce de manière totale comme si l'on plongeait dans un autre corps.

Le cinéma dès ses débuts entretient un lien fort avec l'immersion, les vues Lumière, filmées à hauteur d'homme, cherchent à faire vivre aux spectateurs des moments qu'ils ne connaissent pas forcément, parfois loin d'eux. L'immersion n'est pas encore forcément recherchée particulièrement à l'époque, comme elle peut l'être aujourd'hui, mais le vent de panique qui a soufflé dans le salon indien du Grand Café lors de la projection de *L'Arrivée d'un train à La Ciotat* témoigne d'un effet d'immersion des spectateurs. Ceux-ci se sont vus à la place de la caméra, un train leur fonçant dessus. On imagine mal aujourd'hui une salle hurler lors du visionnage de ce film mais cette idée de transporter le spectateur ailleurs et de lui faire vivre des émotions « comme si vous y étiez » n'a jamais quitté le septième art, cherchant toujours plus d'immersion dans le divertissement au fur et à mesure que les spectateurs s'habituent aux innovations technologiques.

Au départ les films sont projetés dans des foires avec un bonimenteur qui raconte les aventures projetées des héros et héroïnes comme un grand livre

1

d'images commentées. Le lieu n'est pas forcément propice à l'immersion, du bruit, des lumières parasites, un écran petit et une projection de mauvaise qualité sont souvent au rendez-vous. L'apparition de salles de cinéma dédiées permet d'augmenter l'effet d'immersion, il n'y fait pas tout de suite noir mais au moins l'ambiance est sombre, l'écran s'agrandit, l'effet d'échelle par rapport au spectateur apparaît et l'on se met à voir arriver de plus en plus d'effets destinés à intégrer le spectateur dans l'action. Dès les années 10, des longs métrages de fiction en procédé anaglyphe¹ comme *The Power of Love* de Joseph A. Golden voient le jour. Ce procédé a aussi été utilisé pour tourner certaines scènes du Napoléon d'Abel Gance malheureusement coupées au montage, mais Gance avance tout de même dans l'idée d'immersion avec une projection sous forme de triptyque avec trois écrans et trois projecteurs offrant à l'image un ratio de 4:1. L'avènement des formats panoramiques, qui offrent une meilleur immersion car s'approchant plus de la vision humaine que les formats « carrés », arrive dans les années 50 avec notamment le Cinemascope. D'autres formats n'ont pas survécu mais recherchaient aussi le même sentiment d'immersion comme le Cinérama venu des Etats-Unis ou le Kinopanorama de Russie en utilisant plusieurs projecteurs diffusant sur un écran courbe afin d'augmenter le sentiment d'englobement du spectateur. Au milieu des années 80, une nouvelle étape est franchie avec l'apparition de lieux dédiés à ce cinéma de grand spectacle comme le Futuroscope ou la Géode. Mais depuis quelques années, en plus de l'explosion du cinéma stéréoscopique, on commence à voir apparaître des salles reprenant les concepts de ces parcs d'attractions : fauteuils qui bougent, ventilateurs, projections d'eau, diffusion d'odeurs... afin d'immerger encore plus le spectateur dans le film, c'est le cinéma « 4D » auguel une salle est dédiée dans le nouveau multiplex Pathé La Villette.

_

¹ Procédé permettant une vision stéréoscopique grâce à l'utilisation de lunettes équipées de filtres (souvent cyan et rouge) lors de la projection.

Le son aussi joue un rôle très important dans la sensation d'immersion. C'est, comme on le verra après, un sens toujours actif et qui capte à 360°. L'arrivée du son synchrone au milieu des années 20 a permis aux metteurs en scène d'avoir un nouvel outil pour raconter leurs histoires mais aussi d'apporter du réalisme et de la crédibilité aux images. Un personnage qui bouge les lèvres et dont on entend la voix est bien évidemment plus captivant qu'un texte écrit entre guillemets sur un carton. Mais le vrai avantage du son pour l'immersion c'est qu'il peut sortir de l'écran pour, paradoxalement, y faire rentrer le spectateur. C'est dès la fin des années 30 qu'apparaissent les premiers systèmes de diffusion multicanaux pour le cinéma avec notamment le Fantasound développé par les studios Disney pour *Fantasia* et qui leur vaudra même un oscar spécial. Par la suite arriveront les systèmes que l'on connaît aujourd'hui, d'abord quatre, puis six, puis huit, puis *n* canaux avec le Dolby Atmos qui permet une meilleure précision dans le positionnement des sources dans la salle qu'au niveau de l'écran. Le spectateur ainsi entouré par les sons du film se retrouve donc pris dans l'espace qui est celui de l'action et de l'écran.

Pour finir, il ne faut pas restreindre l'immersion a l'identification à un autre être humain mais elle peut aussi chercher à nous emmener dans un univers infiniment petit où l'on va pouvoir s'identifier à une fourmi comme dans *Microcosmos : Le Peuple de l'herbe* de Claude Nuridsany et Marie Pérennou. Ou dans un style totalement différent, à l'intérieur du corps humain lors d'un rapport sexuel avec *Love* de Gaspard Noé.

I – Qu'est ce que la vidéo 360°?

1 – Réalité virtuelle et vidéo 360°

« La réalité virtuelle est un domaine scientifique et technique exploitant l'informatique et des interfaces comportementales en vue de simuler dans un monde virtuel le comportement d'entités 3D, qui sont en interactions en temps réel entre elles et avec une personne en immersion pseudo-naturelle. »

Philippe Fuchs, Les casques de réalité virtuelle et de jeux-vidéo, p.21.

Nous allons nous servir de cette définition, certes, peu digeste et très technique pour nous permettre de définir la réalité virtuelle et par opposition la vidéo 360°. On limite souvent la réalité virtuelle aux visiocasques² alors que son principe est bien plus large. Il s'articule autour d'une boucle entre l'Homme et l'environnement virtuel, le logiciel temps réel va créer des stimulis sensoriels (images, sons, vibrations...) en réaction à des interfaces motrices contrôlées par l'utilisateur (capteurs de mouvements, gyroscopes, joysticks...). En clair, si je lève mon bras, un capteur va le détecter puis faire apparaître l'image d'une main sur l'écran de mon visiocasque, le cercle est sans fin car les actions motrices de l'utilisateur vont donner lieu à des stimulis sensoriels qui vont à leur tour inciter l'Homme à réagir, la boucle est bouclée. Bien évidemment ce concept qui existe depuis une trentaine d'année a d'abord été destinées aux milieux professionnels et militaires à de fins de formation et d'entraînement, on peut notamment

4

² Dispositif englobant l'intégralité du champ de vision et permettant de voir une image stéréoscopique, plus communément appelé « casque de réalité virtuelle ».

penser aux simulateurs de vol, avant d'être destiné à des applications plus « grand public » à commencer par le jeu vidéo.

Mais la réalité virtuelle a fini par dépasser le cadre du seul jeu-vidéo pour donner lieu à des objets audiovisuels d'un nouveau genre, où il y a une narration bien définie mais où « le spectateur face à son écran devient par moments, lorsqu'il est sollicité, une sorte d'acteur qui indique ses choix au déroulement du récit. » de là est né le concept de spect'acteur. Un mot-valise qui désigne un nouveau mode de consommation des contenus audiovisuels comme par exemple le web-documentaire, qui est somme toute une forme très basique de réalité virtuelle, où le spectateur interagit avec la narration.

La vidéo 360°, elle, laisse le spectateur passif dans son visionnage face au déroulement du récit mais lui permet tout de même de regarder où il le souhaite. Il peut agir sur la forme en déplaçant le champ de vision et en faisant tourner la scène sonore mais pas sur le fond, l'histoire. Il est donc un spectateur immergé mais inactif. La vidéo 360°, tout comme la réalité virtuelle d'ailleurs, n'est pas forcément liée à un rendu visuel stéréoscopique ni à un visiocasque même si l'effet d'immersion est bien évidemment moindre, des plateformes comme celles de YouTube ou Facebook permettent de regarder des vidéos à 360° sur un ordinateur en déplaçant le champ visuel grâce à sa souris. On peut aussi parler des « cardboards », dispositifs en carton dans lesquels on insère son smartphone afin de créer un simili visiocasque permettant à tout un chacun d'expérimenter ces nouveaux contenus.

À l'inverse de la réalité virtuelle qui se base souvent sur des environnements générés numériquement en 3D, la vidéo 360° utilise la prise de vue réelle. Pour ce faire l'utilisation de caméras spécifiques est nécessaire, il en existe deux types : les caméras dites « omnidirectionnelles » qui ne possèdent

5

³ Pierre Barboza, *Fiction interactive « métarécit » et unités intégratives*, p.99.

qu'un seul capteur au dessus duquel se trouve un miroir en forme de cône qui permet de projeter une image à 360° dans l'objectif de la caméra, et d'autres systèmes eux constitués de plusieurs caméras placées sur une sphère et dont les angles de prise de vue se recoupent afin de former une image à 360°.



Une caméra omnidirectionnelle et un exemple de rendu d'image. (sources : wikimedia.org, cs.columbia.edu.



La caméra 360° multiobjectifs Nokia OZO. (source : photocinerent.com)

$2-360^{\circ}$, vraiment?

Commençons par un peu de physiologie en nous intéressant aux capacités de l'œil humain et notamment à son champ de vision. Chacun de nos yeux couvre un champ horizontal d'environ 170°, ces deux cônes de vision se recoupent au centre formant ce que l'on appelle la vision binoculaire qui couvre un angle d'environ 110°, c'est dans cette zone que l'on possède la meilleure acuité visuelle, que ce soit en terme de dissociation des couleurs, des formes ou de la perspective. Mais de chaque côté de ce champ se trouvent les champs de vision monoculaires, plus communément appelés « vision périphérique », qui correspondent à la lumière perçue par un seul des deux yeux, ils couvrent environ 50°. L'image y est floue, le contraste faible et les couleurs imprécises mais c'est une zone de veille servant à déclencher des réflexes, notamment des mouvements de la tête, afin d'affiner l'analyse d'un objet grâce à la vision binoculaire. Des mécanismes similaires existent pour l'ouïe, nous en reparlerons plus loin. La vision humaine couvre donc en additionnant les champs de vue monoculaires et binoculaires un angle d'environ 210°.

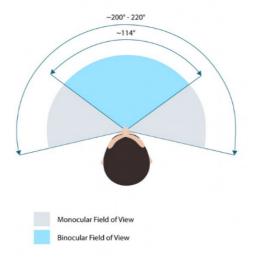


Schéma du champ de vision horizontal humain (source : vr-lens-lab.com)

Parlons maintenant un peu du principe de fonctionnement des visiocasques. Ceux-ci sont constitués de deux écrans, un pour chaque œil ce qui permet une vision stéréoscopique, devant lesquels se trouvent des lentilles destinées à projeter l'image produite sur la rétine tout en limitant la fatigue oculaire. Ces lentilles sont même réglables dans les casques haut de gamme afin de pouvoir corriger les défauts de la vision de l'utilisateur - le port de lunettes étant bien évidemment impossible. En fonction des dimensions et de la forme des écrans ainsi que de la distance focale des lentilles, l'image reconstituée par le cerveau humain sera plus ou moins large. En effet, pour des raisons techniques et économiques il est compliqué de couvrir l'intégralité des 220° du champ de vision humain, pour le moment seul le projet du visiocasque StarVR (soutenu notamment par la société IMAX) propose un champ de vision à 210°. Sur le marché actuel, les leaders que sont Oculus Rift et HTC Vive proposent des champs de vision limités à environ 110°. Leur argument étant qu'il est suffisant de couvrir la vision binoculaire car c'est la seul permettant une stéréoscopie et une acuité visuelle. Ce sujet à notamment été abordé lors de mes discussions avec Philippe Fuchs, pour lui l'immersion ne peut pas être réelle si l'on ne prend pas en compte la vision périphérique. Certes elle n'est pas précise mais elle contribue grandement à l'effet de réel dans notre inconscient, comme peuvent l'être certains sons dans un film, ténus, pas forcément précis mais qui créent « quelque chose ». On se retrouve donc avec les visiocasques actuellement disponibles à ressentir une vision tunnelisée, effet pouvant être aussi crée par la vitesse ou la consommation d'alcool par exemple, ce qui peut être source d'une gêne lors de l'utilisation prolongée de ces dispositifs.

On parle de vidéo à 360° mais pour être plus précis nous devrions parler de vidéo à 360° x 180°. En effet, il ne faudrait pas négliger la liberté verticale de mouvement de la tête qui nous permet de regarder de nos pieds jusqu'au ciel. Il

faut donc un plan horizontal à 360° et un plan vertical à 180° pour nous permettre de couvrir visuellement tout l'espace qui nous entoure. La vidéo 360° peut rarement nous permettre cette liberté de regard, en effet pour des raisons techniques toutes simples, il est impossible de regarder totalement le sol ou alors on verrait le pied sur lequel est posée la caméra qui filme. Cela pose aussi d'autres questions de mise en scène, par exemple si la caméra est à la première personne il faudrait pouvoir voir le corps du personnage que l'on incarne, ses pieds, ses jambes... ce qui est presque impossible dans le cas d'une prise de vue réelle! En image de synthèse cette problématique n'a pas de sens, le jeu vidéo par exemple représente depuis bien longtemps les bras et les mains des personnages mais aussi leurs pieds. Cette question du champ visible de nombreuses possibilités au metteur en scène. En effet celui-ci peut décider par exemple de le limiter à 180° horizontalement afin de créer un hors champ auguel le spectateur ne peut accéder et donc se rapprocher d'une expérience visuelle comme celle de l'Omnimax⁵. Mais on peut aussi imaginer incliner la caméra de 90° ou encore l'accrocher au plafond afin d'offrir des axes de vision non-naturels, une caméra qui nous ferait flotter en l'air au dessus de la scène comme un point de vue omniscient. Des configurations encore plus exotiques sont aussi envisageable avec par exemple un champ visible de 110° horizontalement mais de 180° verticalement, c'est à dire que seul le mouvement zénithal de la tête serait pris en compte.

Ainsi la notion de champ, définie par Almendros comme « [l'] espace déterminé par l'angle de l'objectif de la caméra, c'est à dire la partie du sujet saisie par la caméra »⁶, que l'on pourrait imaginer disparaître avec la vidéo 360°

-

⁴ J'utilise le terme de champ visible, c'est à dire l'espace qui nous entoure que l'on peut parcourir du regard qui est à bien différencier du champ de vision.

⁵ Procédé de diffusion d'images sur un écran hémisphérique, utilisé par exemple à la Géode à Paris.

⁶ Nestor Almendros, *Un Homme à la caméra*, p.176.

existe tout de même dans une forme différente du cinéma traditionnel. Le champ n'est plus défini par un format d'image, un choix de focale et un positionnement de caméra mais c'est un champ de liberté visuelle que l'on peut définir après le tournage soit dans le but de cacher une partie de la scène au spectateur soit de créer de toute pièce une nouvelle façon de regarder son environnement en bloquant ou limitant certains axes de regard.

3 – Utilisations actuelles de la vidéo 360°

Note : afin de ne pas surcharger la lecture, vous retrouverez les liens internet pour consulter les vidéos 360° dont je vais parler en annexe.

Parmi les premiers à s'être approprié la vidéo 360° - comme ils l'avaient fait avec les caméras type GoPro - on retrouve les pratiquants de sports extrêmes comme le snowboard, le parapente ou le rafting dont on retrouve un nombre conséquent de vidéos sur la plateforme YouTube. Très vite sont aussi arrivées des vidéos à caractère plus touristique nous emmenant dans la jungle ou nous faisant descendre en bateau le cours d'un fleuve.

Ensuite, et assez rapidement, c'est la pornographie qui s'est emparée du sujet. Cherchant toujours plus d'immersion pour le spectateur la production de contenu pour adultes s'est depuis plusieurs années mise à la vidéo 360°. C'est actuellement une des industries qui produit le plus de contenus pour visiocasques (hors jeux-vidéo bien évidemment).

Plus qu'une simple immersion et de belles images, la fiction et le documentaire en vidéo 360° se sont aussi développés. En France notamment au travers d'Arte qui a rapidement lancé une plateforme de visionnage et participé au financement de projets. On peut penser notamment à *I, Philip* - réalisé dans

le cadre du cycle Philippe K. Dick début 2016 - qui nous fait incarner Phil, un androïde, et nous plonge dans ses souvenirs réels ou rêvés (mais qui ne nous donne pas non plus la réponse sur la question des moutons électriques) ? À michemin entre documentaire et visite virtuelle, la série Art Stories propose une expérience immersive dans des lieux qui ont marqué l'histoire de l'art en Europe. Pour le moment seul l'épisode sur le château de Fontainebleau est sorti mais d'autres sont prévus, sur l'Alhambra de Grenade notamment.

L'animation aussi a pris part dans la production de vidéos 360°, on peut par exemple citer *The Night café* qui nous permet d'explorer le célèbre tableau de Van Gogh *Le Café de nuit* dans un univers reconstitué en 3D. D'autres créations se rapprochant plus de l'art vidéo ou du domaine de l'installation ont aussi vu le jour comme *Notes on blindness* basé sur les archives de John Hull. L'auteur, professeur en théologie et écrivain, est devenu aveugle au début des années 80 des suites d'une maladie, pour l'aider dans ce bouleversement il a tenu pendant plusieurs années un journal intime audio. C'est à partir de ces enregistrements qu'a été construite la bande son cette œuvre, les images nous plongeant dans un monde fantomatique, afin d'essayer de nous transmettre son ressenti et les épreuves qu'il a dû traverser.

Principalement outre-Atlantique, les journaux et les chaines d'information ont aussi commencé à produire du contenu d'actualités en vidéo 360°, par exemple CNN a réalisé un reportage sur les conditions d'hébergement des migrants arrivés sur les plages de Lesbos en Grèce, ou le New York Times qui a filmé le recueillement après les attentats du 13 novembre 2015 à Paris.

Pour finir sur une note plus gaie on peut aussi penser aux clips musicaux qui exploitent la vidéo 360° comme celui de *Heal Tomorrow* des Naive New Beaters et Izia assez inventif qui nous permet de voir plusieurs scènes simultanément de ce qu'aurait été le clip filmé de manière plus traditionnelle ainsi que les coulisses de ces tournages. C'est donc au spectateur de faire le

montage et de choisir s'il veut regarder le déroulé de l'action normalement ou s'il veut voir ce qui s'est passé entre les prises des différentes séquences.



Vincent Van Gogh, Le Café de nuit.

II – (dé)placement dans l'espace et audition

1 – Écoute binaurale et champ auditif

Avec la vue, l'ouïe est un des cinq sens qui nous permet le mieux de percevoir l'espace dans lequel on se trouve – le toucher aussi mais uniquement pour un environnement proche, l'odorat dans certains cas et le goût, à ma connaissance, jamais. C'est un sens omnidirectionnel, contrairement à la vue qui est dirigée dans une direction : on entend tout ce qui se passe autour de nous, de manière plus ou moins précise bien entendu. On peut tourner le regard afin

de ne pas voir quelque chose mais il nous est en revanche impossible de ne pas entendre⁷ un son, l'ouïe n'est pas un sens sélectif.

Pour comprendre comment nous localisons un son, commençons par revenir à la base. L'onde sonore est une vibration mécanique qui se propage dans un matériau ou un milieu, l'air par exemple, et est à dissocier de la notion de « son » qui n'est que la perception que l'on a d'une onde sonore. Notre oreille capte les ondes sonores au travers du pavillon auriculaire qui va les guider dans le conduit auditif jusqu'au tympan – c'est l'oreille externe. Celui-ci va vibrer et ce faisant faire bouger trois os, parmi les plus petits du corps humain, qui constituent l'oreille moyenne et dont le rôle est de transmettre de manière contrôlée les vibrations acoustiques à l'organe de l'oreille interne, la cochlée. Au sein de celle-ci se trouvent une multitude de cellules cillées qui réagissent chacune à une fréquence donnée et génèrent lorsqu'elles sont stimulées un influx nerveux qui une fois interprété par le cerveau donnera la sensation sonore.

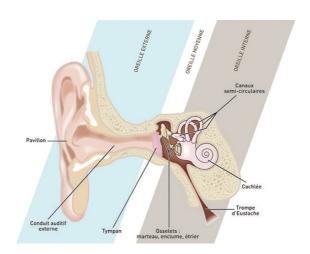


Schéma en coupe de l'oreille humaine. (source : audilab.fr)

-

⁷ Il faut bien différencier entendre et écouter, on entend tous les sons qui nous entourent mais on écoute que ceux auxquels on prête attention.

Pour nous permettre de localiser l'origine d'un son, notre cerveau se base sur deux principes : l'ILD 8 ou différence interaurale d'intensité et l'ITD 9 ou différence interaurale de temps. Notre cerveau fonctionne comme un comparateur entre les deux capteurs que sont nos oreilles, pour l'ILD il va analyser la différence de niveau sonore entre le son perçu par l'oreille droite et celui perçu par l'oreille gauche. Ce faisant il va en déduire sa localisation, plus la différence de niveau sera importante plus il nous indiquera une localisation de la source comme étant à droite ou à gauche. L'ITD utilise peu ou prou le même principe sauf qu'elle se base sur une différence temporelle, si une oreille capte une onde sonore avant l'autre, c'est donc que la source sonore se trouve plus du côté de cette oreille que de l'autre. Bien évidemment dans les deux cas la précision de l'oreille et la capacité d'analyse du cerveau sont telles que toutes les variations sont possibles entre un son localisé totalement à gauche et un son localisé totalement à droite. Ces deux composants sont la base de ce qu'on appelle l'écoute binaurale, grâce à nos deux oreilles nous pouvons localiser la source d'un son, en écoute monaurale il est impossible de faire de même sans bouger la tête! Pour les sons au dessus de 1500 Hz c'est principalement la différence d'intensité qui entre en jeu, en effet la longueur d'onde étant trop courte notre cerveau n'arrive pas à analyser les différences de phase. A contrario, en dessous de 1000 Hz c'est la différence de temps qui nous permet de localiser les sources sonores. Notre cerveau étant capable d'intégrer des écarts temporels de l'ordre de 10ms, cette localisation est fiable jusqu'à environ 100 Hz. En dessous il est très difficile pour l'humain de déterminer la source d'un son.

Maintenant que nous avons vu comment notre cerveau détermine l'azimut d'une source sonore on peut légitimement se demander comment il en déduit l'élévation... De même comment savoir si un son se trouve devant nous

_

⁸ Interaural Level Difference.

⁹ Interaural Time Difference.

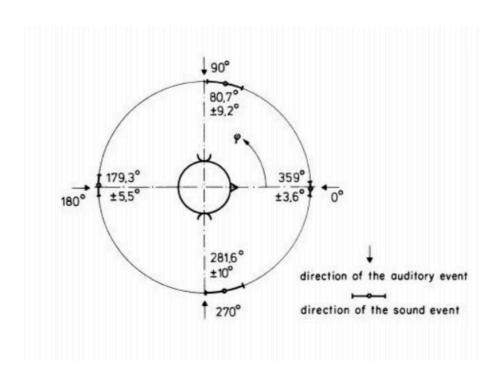
ou derrière nous ? En effet entre un son à 0° ou un son à 180° par rapport à l'axe de notre tête il n'y a logiquement aucune différence de temps ni d'intensité. La réponse à toutes ces questions sont les HRTF – Head-Related Transfer Functions ou Fonction de Transfert Relative à la Tête –, en effet notre corps, nos épaules, notre tête, nos pavillons auditifs sont autant de reliefs qui créent de la diffraction pour les ondes sonores arrivant à nos oreilles. Cette diffraction qui crée un filtrage fréquentiel est intégrée par notre cerveau qui va les interpréter afin d'affiner la localisation de la source. Parmi tous ces facteurs ce sont nos pavillons et conduits auditifs qui sont le plus déterminants dans ces HRTF, fort heureusement sinon nous risquerions de perdre tout nos repères sonores à chaque coupe de cheveux! Les replis de nos pavillons étant petits et diffractant donc les hautes fréquences, ce sont celles-ci – principalement au dessus de 4000 Hz¹⁰ – qui sont les plus importantes dans la localisation verticale. Étant données les différences morphologiques entre tous les humains que nous sommes comme pour les empreintes digitales, il n'y a pas deux pavillons identiques – il y a autant de HRTF que d'Hommes sur terre, c'est une problématique importante et nous y reviendrons lorsque l'on parlera de restitution en synthèse binaurale plus loin.

Mais l'acuité que nous avons dans la localisation des sources sonores est variable, de la même manière que pour la vue, en fonction de leur azimut et de leur élévation. Les deux graphiques ci-dessous tirés du livre *Spatial Hearing* de Jens Blauert nous montrent la précision avec laquelle notre cerveau arrive à distinguer la position exacte d'une source en fonction de son angle d'incidence. Il est clair que comme pour l'acuité visuelle, c'est dans la zone frontale que l'on retrouve la meilleure précision. Mais on peut aussi remarquer que même si on perd beaucoup pour les sons venus des côtés, pour ceux arrivant à l'arrière de

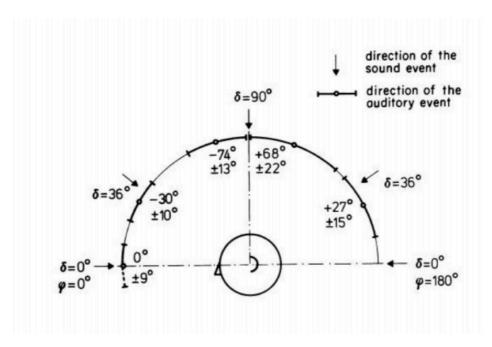
-

¹⁰ cf. Middlebrooks & Green, « Sound Localization by Human Listeners », *Annual Review of Psychology*, 1991, p.144.

notre tête la précision est plus fiable. Cela s'explique simplement comme un mécanisme de veille et de survie, en effet un danger arrivant sur nous par l'arrière ne peut être détecté par l'œil, c'est donc l'ouïe qui prend le relais pour nous en avertir. D'autre valeurs que celles données par Blauert peuvent être aussi trouvées dans la littérature mais elles sont souvent assez proche sauf pour les sons arrivant face à nous où l'angle d'erreur descend parfois à 1°.



Précision de la localisation en azimut. (Blauert, Spatial Hearing)



Précision de la localisation en élévation. (Blauert, Spatial Hearing)

2 – La réverbération, un marqueur important de l'espace

On appelle son direct les ondes sonores qui partent de la source et qui arrivent directement et en ligne droite à l'auditeur. Le son réverbéré quant à lui est constitué de la multitude d'ondes réfléchies par les murs de la salle. Cette réverbération se décompose en deux temps, d'abord les premières réflexions qui sont principalement constituées des ondes n'ayant rebondi que sur une seule paroi avant de nous arriver. Elles arrivent environ entre 10 et 100 ms après le son direct. Si on imagine une pièce parfaitement cubique et sans mobilier il y aurait donc six premières réflexions : le plafond, le sol et les quatre murs. Elles dépendent beaucoup des positions qu'occupent dans la pièce la source et le récepteur ce qui fait qu'elles jouent un rôle important dans la localisation de

l'origine du son, est-elle près d'un mur, loin ou proche... Ensuite vient le champ diffus composé de toutes les réflexions secondaires, qui rebondissent sur plusieurs surfaces avant d'atteindre l'auditeur, sa durée peut atteindre plusieurs secondes. Il est homogène et chaotique¹¹, son niveau sonore est à peu près le même dans toute la pièce et on ne peut déterminer la direction des réflexions qui le composent... C'est ce champ diffus qui est le marqueur le plus important de la sensation de volume de la pièce ainsi que de sa « couleur » sonore. L'acoustique d'une salle est un phénomène complexe et difficile à maitriser pour l'ingénieur du son notamment lorsqu'il faut le recréer artificiellement mais qui nous donne beaucoup d'indication sur l'espace dans lequel on se trouve :

« L'importance déterminante de la réverbération dans la perception de la perspective sonore ne peut pas être évaluée uniquement de façon quantitative. On perçoit plus ou moins de réverbération, mais c'est par sa durée, par la forme de sa décroissance dans le temps, par sa densité à des instants plus ou moins rapprochés du son direct, que la réverbération détermine la sensation de la perspective. »

Jean Chatauret, « La perspective sonore », Communications, 2009, n°1, p.142

La réverbération, par l'apprentissage que l'on fait durant notre vie, nous permet immédiatement de savoir dans quel espace on se trouve et d'imaginer son volume ainsi que les matériaux qui le constituent. Qui ne reconnaît pas dès le premier coup d'oreille l'immensité de la réverbération d'une cathédrale ou l'étroitesse et la faïence de celle d'une salle de bains. L'acoustique des édifices religieux était d'ailleurs tout à fait pensée, son but à la fois de porter la voix du prêtre durant l'office mais aussi de profiter du champ diffus très important pour que l'église dans son entièreté résonne comme pour porter la voix de Dieu.

-

 $^{^{11}}$ cf. Antonio Fischetti, *Initiation à l'acoustique*, p.83

Tatiana Weber dans un article qui parle du parallèle que fait Chateaubriand entre les églises gothiques et les forêts utilise l'axe de l'écho et de la réverbération pour développer son propos¹². L'acoustique d'une pièce nous entoure donc et nous baigne comme une ambiance lumineuse, c'est un facteur important de la sensation d'immersion.

Une dernière réflexion sur les réflexions : certaines personnes arrivent à percevoir non seulement la pièce dans laquelle ils se trouvent mais aussi les objets et obstacles qui se trouvent devant eux à l'aide de réflexions sonores. Ce phénomène connu sous le nom d'écholocalisation, et pour lequel on cite souvent les chauve-souris et leur radar à ultrasons, est développé principalement par des aveugles qui à l'aide d'un son impulsionnel – souvent claquement de doigts ou de langue – et des réflexions qui leur reviennent en déduisent les volumes se trouvant autour d'eux. Chez ces personnes ce sont les aires du cerveau normalement dédiées à la vue qui s'activent dans ce cas là pour leur permettre en quelque sorte de voir grâce à leurs oreilles.

3 – Équilibrioception et proprioception

Notre oreille interne nous permet de transformer les ondes sonores en impulsions électriques mais elle joue aussi un rôle important dans notre sens de l'équilibre. Le liquide situé dans la cochlée fonctionne comme une sorte de niveau à bulle qui nous indique l'inclinaison de notre tête et donc la position de notre corps par rapport au sol. Cette problématique de sensation est importante en vidéo 360°, pour le moment il n'existe pas de solution pour tromper l'oreille

_

¹² Tatiana Weber, « Gothique et expérience du sacré : l'écho et la profondeur », *Bulletin de l'Association Guillaume Budé*, 1996, n°4, pp. 383-394.

interne et nous immerger encore plus mais on pourrait imaginer par exemple une plateforme qui s'inclinerait en fonction de la pente du sol.

On pourrait se dire que si ce n'est que pour un peu plus de sensations, l'équilibrioception n'est pas forcément très intéressante. En réalité elle est aussi un facteur important dans ce que l'on appelle le mal de la réalité virtuelle. En effet, l'utilisation d'un visiocasque peut donner lieu à des vertiges ou des nausées – en fonction de la durée d'utilisation et de la sensibilité des personnes. Ces maux sont dus à plusieurs facteurs, d'abord la différence entre les informations données par la vue et l'oreille interne quand à la position que l'on a par rapport au référentiel terrestre (point de vue d'une personne allongée alors que l'on est assis par exemple), ensuite des lentilles optiques et des écrans de visiocasques de mauvaise qualité peuvent fatiguer l'œil et créer une sensation de malaise on considère qu'une fréquence de rafraîchissement de l'image de 60 Hz est le minimum pour éviter trop de gêne. Enfin et la plus importante, la latence qu'il peut y avoir entre un mouvement de la tête et son rendu sur l'écran du visiocasque, si l'image est trop en retard le cerveau est trop perturbé et déclenche des nausées. Ces désagréments sont bien évidemment plus présents sur les dispositifs d'entrée de gamme comme les cardboards par exemple. C'est un mal qui se rapproche de celui des transports ou de la mer, les personnes y étant plus sujettes dans la vie réelle le seront du coup plus aussi dans la réalité virtuelle!

On apprend toujours à l'école que l'Homme possède cinq sens : la vue, le toucher, l'ouïe, le goût et l'odorat sont les principaux mais il en existe en réalité d'autres. L'un d'entre eux est la proprioception, parfois aussi appelée sensation profonde, c'est un sens qui contrairement aux cinq principaux ne réagit pas à des stimulis extérieurs mais qui nous donne conscience de notre propre corps. Nos muscles, nos os et nos articulations sont remplis de terminaisons nerveuses dédiées à cet usage que l'on appelle propriocepteurs. Ils permettent un retour

d'informations sur l'état de notre position physique au cerveau. Mais ce qui est très intéressant c'est que notre ouïe est liée à la proprioception, notamment pour la localisation des sons que nous produisons nous même. Si on pose un objet sur une table, même sans le voir, on sait très précisément où celui-ci se trouve alors que la seule sensation que l'on a pu avoir est celle du bruit produit par le choc. En réalité c'est la proprioception qui dicte au cerveau la localisation de la source sonore. Une expérience amusante à faire à ce sujet : fermez les yeux, frottez deux de vos doigts doigts l'un contre l'autre et tournez votre main autour de votre tête, vous pouvez très précisément localiser la position de la source du son autour de vous. Maintenant demandez à quelqu'un d'autre de produire ces frottements et de les déplacer autour de votre tête, la localisation est bien moins précise! C'est la preuve de l'interaction entre ces deux sens.

III – Vidéo 360° et mise en scène sonore, quelques pistes

1 – La question cruciale des points de vue et d'écoute

Les possibilités de montage et de cadrage disparaissant avec la vidéo 360°, la question du point de vue prend encore plus d'importance. C'est l'écriture du film qui va déterminer la position de la caméra, on peut imaginer celle-ci se trouvant entre deux personnages pendant un dialogue et nous permettant de monter notre propre champ / contre-champ en temps réel comme nous le permet Jean-Luc Godard utilisant le principe de fonctionnement de la projection stéréoscopique dans Adieu au langage. Mais d'autres types de point de vue sont bien évidemment aussi possible, par exemple imaginons un point de vue omniscient, où la caméra flotterait dans les airs au dessus de l'action nous

empêchant de nous identifier à un hypothétique personnage et créant ainsi un fort effet de distanciation. Bien entendu, le point de vue le plus évident pour la vidéo 360° est la première personne. C'est le point de vue de la réalité virtuelle et de la plupart des jeux vidéos pour visiocasques, mais l'espace est libre pour être exploré par les metteurs en scène afin de trouver de nouvelles idées.

Venons en maintenant à la question du point d'écoute, pendant auditif du point de vue. Le plus logiquement celui-ci est le même que la position de la caméra. En effet, comment produire un effet immersif autrement ? Mais tout de même il existe des possibilités pour ne pas forcément tomber dans cette évidence. Par exemple pour reprendre l'idée d'un point de vue omniscient on pourrait imaginer de même pour le son bien qu'il risque de n'en ressortir qu'une joyeuse cacophonie. On peut aussi imaginer utiliser le montage vertical pour créer un effet Koulechov sonore ¹³ par exemple, et même à cette occasion dissocier le point de vue et le point d'écoute : nous sommes dans une maison, un personnage en sort, la caméra reste à l'intérieur mais le point d'écoute suit le personnage. Certes l'effet serait peut-être perturbant mais pas inintéressant à mon avis.

2 – Comment guider le spectateur dans la narration grâce au son ?

En perdant son pouvoir de choisir le sujet à montrer, il faut que le réalisateur trouve d'autres moyens d'orienter le spectateur vers ce qu'il tient à montrer. Pour ce faire, le plus simple est d'user d'effets psychoacoustiques et de jouer avec la curiosité du spectateur.

-

¹³ Cf. Michel Chion, *Le Son au cinéma*, p.57.

On pourrait par exemple imaginer un dispositif technique dans lequel on brouillerait la compréhension des voix des personnages auxquels on ne fait pas face créant ainsi une frontalité auditive comme celle de la vision. Le spectateur forcément curieux de savoir ce que se raconte, orientera donc son regard dans l'axe voulu par le metteur en scène.

Le réflexe dit « tête et cou » notamment théorisé par Claude Bailblé peut aussi nous être d'une grande aide. En effet lorsque l'on entend un son impulsionnel et auquel on ne s'attend pas derrière nous, automatiquement nous allons tourner la tête afin d'analyser la source du bruit grâce à la vue. C'est un réflexe qui date des débuts de notre espèce où il était une condition sine qua non à la survie. Sachant cela on peut imaginer toutes sortes de stratagèmes afin d'attirer l'attention du spectateur avec par exemple une chute d'objet, un tintement de verre, un claquement de doigts... Ce même genre de réflexe existe aussi avec la voix humaine, si l'on entend quelqu'un parler on va naturellement se retourner vers lui. De la même manière, dans le silence, une musique, un son répétitif ou qui n'a pas forcément sa place dans l'univers dans lequel on se trouve va aussi attirer l'attention du spectateur.

« L'œil sollicité seul rend l'oreille impatiente, l'oreille sollicitée seule rend l'œil impatient. *Utiliser ces impatiences*. Puissance du cinématographe qui s'adresse à deux sens *de façon réglable*. »

Robert Bresson, Notes sur le cinématographe, p.63.

3 – Au sujet de la voix

Avec un point de vue et un point d'écoute à la première personne, la question de la voix du personnage que l'on incarne devient cruciale. Premièrement, est-ce que ce personnage doit parler? Faire incarner au spectateur un personnage muet serait la solution la plus simple, mais pas forcément la plus intéressante on en conviendra. Il faut donc imaginer une manière de faire ressentir une voix intérieure au spectateur. Depuis longtemps le jeu-vidéo est confronté à ce problème et des solutions sont à aller cherche de ce côté là. Par exemple on peut penser à une voix présente dans toute la scène sonore, avec peu d'acoustique et très chargée en graves. Mais aussi, pourquoi pas imaginer un dispositif qui pourrait générer des vibrations comme celles que l'on ressent dans la boîte crânienne lorsque l'on parle afin d'augmenter l'effet d'immersion mais attention au risque schizophrène tout de même !

On peut aussi imaginer d'autres formes voix off qui n'étaient pas possible jusqu'à présent, par exemple en exploitant les possibilités de la synthèse binaurale (nous y reviendrons plus loin) on pourrait avoir une voix venant du ciel et libérant ainsi tout le plan sonore horizontal afin de créer l'immersion.

IV – Les solutions techniques actuelles pour le son en vidéo 360°

1 – La synthèse binaurale

La synthèse binaurale est un procédé de spatialisation qui exploite dans leur entièreté toutes les fonctions de l'écoute binaurale. Elle permet la restitution

de scènes sonores en trois dimensions au travers d'un casque audio et de ce fait est le procédé de restitution idéal pour la vidéo 360°. Le principe est simple mais sa mise en œuvre complexe et encore un peu hasardeuse. Tout repose autour de filtres binauraux qui vont appliquer des HRTF à des sons, le plus souvent monophoniques, afin de les positionner dans l'espace. A chaque position de la source dans l'espace correspond une paire de HRTF, une pour l'oreille droite et une pour l'oreille gauche.

La grande problématique de la synthèse binaurale, ce sont justement ces HRTF... Comme dit plus haut, nos morphologies sont uniques et donc les fonctions de transfert qui y sont associées de même. Le calcul des HRTF d'un individu nécessitant une chambre anéchoïque ainsi que plusieurs centaines de mesures, il n'est pas envisageable d'enregistrer les HRTF de tout un chacun. Plusieurs organismes ont donc décidé de créer des banques de HRTF « moyennes » issues de la pondération des HRTF de plusieurs individus. C'est le cas par exemple du projet Bili auquel participent notamment l'IRCAM et le CNRS. Les HRTF fournies par exemple par l'IRCAM fonctionnent assez bien une fois que l'on a trouvé celle qui nous correspond.

Il est aussi possible d'enregistrer directement un signal sonore en binaural, à cet effet il existe des têtes artificielles sur lesquelles les oreilles ont été remplacées par des micros. Bien évidemment la précision de la localisation n'est pas idéale avec ce procédé, la tête artificielle étant modelée sur les caractéristiques physiques d'un Homme « moyen ». Il est aussi possible, et c'est là que l'on obtient les meilleurs résultats pour soi-même de placer de petits micros dans nos propres oreilles afin d'enregistrer exactement les ondes sonores qui arrivent sur notre tympan. Mais la première expérience d'enregistrement et de restitution en son binaural date de 1881 avec le Théâtrophone de Clément Ader. À l'époque tous ces concepts n'avaient pas encore été théorisés mais l'inventeur avait eu l'idée de place deux micros de chaque côté de la cavité du

souffleur de l'opéra Garnier afin de retransmettre le signal via une double liaison téléphonique dans tout Paris. On pouvait donc écouter le concert sans avoir à se déplacer.



La tête artificielle KU-100 de Neumann.

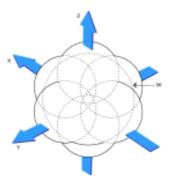


Un système de prise de son binaurale intra-auriculaire.

(source: soundprofessionals.com)

2 – L'ambisonie

Habituellement lorsque l'on veut faire une prise de son multimicrophonique, on utilise autant de micros que de directions que l'on souhaite enregistrer et de haut-parleurs pour les restituer. Le principe de l'ambisonie conçu dans les années 70 en Grande-Bretagne est quelque peu différent : on enregistre le signal à l'aide d'une sphère composée de plusieurs microphones (leur nombre est variable et donnera « l'ordre » de l'ambisonie, c'est à dire sa précision dans la localisation des sources) qui vont être dématricés par la suite afin de diffuser sur un nombre de haut-parleurs que l'on choisit l'espace sonore capturé. C'est à dire qu'un signal enregistré en ambisonie peut ensuite être écouté en monophonie, en stéréophonie, en 5.1 ou même en binaural en casque suite à une simple conversion des données captées par les microphones. La prise de son étant omnidirectionnelle, on peut décider de placer le centre sonore où on le souhaite lors de la restitution. Des formats d'encodage spécifiques comme le format B permettent de transporter les signaux en ambisonie sur quatre pistes audio. Le principe n'est pas d'enregistrer l'onde sonore en un point mais le champ sonore dans sa globalité, ensuite un simple plug-in audio comme le Harpex-B permet de décoder le signal dans le format souhaité.



Modélisation du champ sonore ambisonique.

(source : audiosonica.com)



Le microphone ambisonique du premier ordre Core Sound Tetramic.

3 – Logiciels spécifiques

Au départ les systèmes permettant de créer du contenu sonore pour les visiocasques étaient ceux du jeux-vidéo comme par exemple Unity. Ces contenus étant énormément tournés vers l'interaction avec le joueur, les logiciels sont complexes et nécessitent des connaissances en programmation pour pouvoir les utiliser.

Depuis environ un an, les plateformes de diffusion comme Facebook ou Youtube ont intégré un système de décodage du format B vers le binaural en temps réel dans leurs lecteurs vidéo. Pour ce faire, ils convertissent le signal ambisonique en une scène sonore à huit haut parleurs dont ils créent un rendu en synthèse binaurale.

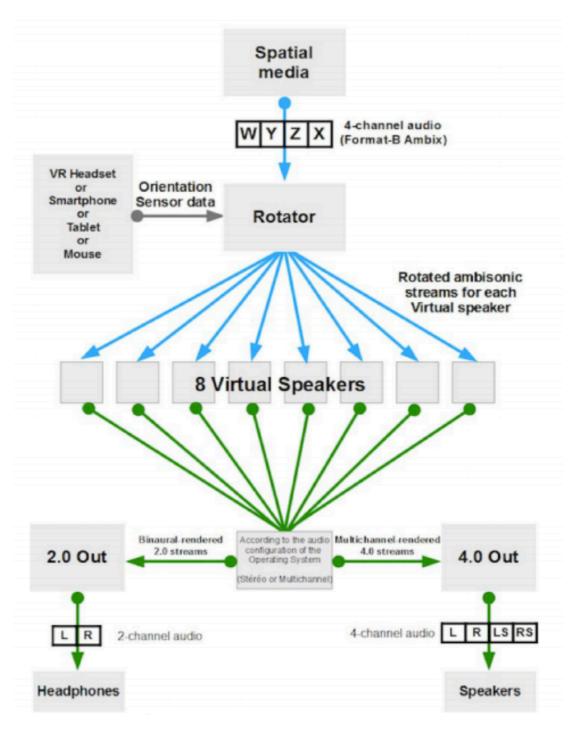
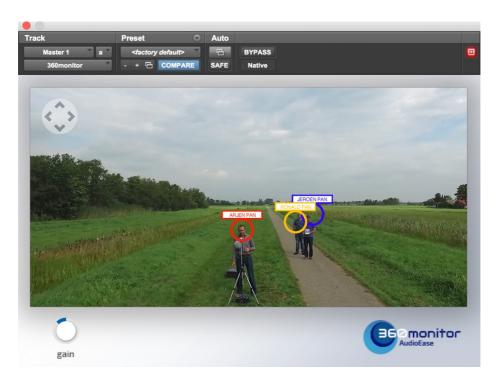


Schéma de principe du décodage Format B vers binaural.

(source : bili-project.org)

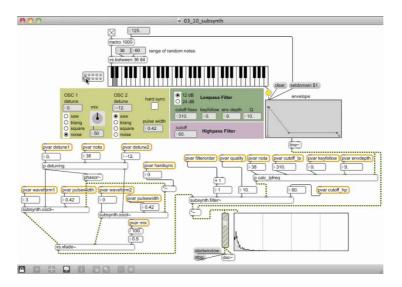
Cette solution est bien évidemment plus simple pour la personne chargée de la bande sonore d'une vidéo 360° étant donné qu'il existe des plug-in audio chargés de gérer une spatialisation à 360° dans un espace ambisonique. On peut donc utiliser nos stations de travail habituelles en ne convertissant le signal en format B qu'au moment du mixage. Divers éditeurs se sont penchés sur le sujet comme Noise Makers qui propose à la fois des logiciels permettant d'encoder un signal en ambisonique et de le décoder en binaural pour l'écoute. AudioEase, concepteur de la célèbre réverbération à convolution AltiVerb propose aussi le même genre d'outils permettant même de visualiser les positions des sources en superposition de la vidéo directement dans ProTools. Facebook de son côté à lancé sa Spatial Workstation dont le principe est le même et qui, de plus, est gratuite! Il n'y a donc plus aucune raison d'avoir peur de faire du son pour une vidéo 360°.



Capture d'écran du 360monitor d'AudioEase.

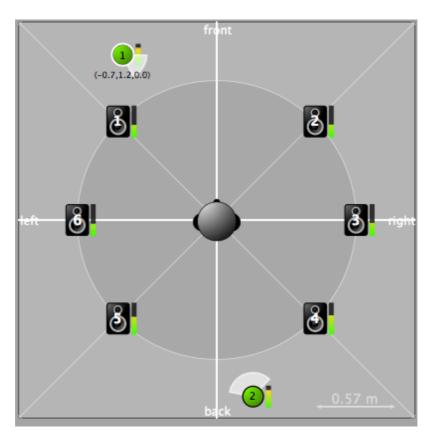
1 – Max/MSP et le Spat

Max est un logiciel développé à l'origine par l'IRCAM dans les années 1980, c'est un système de programmation visuelles où l'on connecte des objets entre eux afin de faire circuler les données, chaque objet ayant une fonction bien précise. Au départ conçu pour permettre de gérer, notamment grâce au protocole MIDI, les créations sonores « live » des réalisateurs musicaux de l'IRCAM il a été étoffé plus tard grâce à l'intégration en son sein du logiciel MSP lui offrant des possibilités de traitement du signal audio. Il permet de concevoir des systèmes de contrôle et de traitement du son complexes et sur mesure. Max/MSP est aujourd'hui développé par la société américaine Cycling'74 qui lui a ajouté de possibilités de traitement de la vidéo avec Jitter.



Capture d'écran d'un synthétiseur additif sous forme de patch Max.

Dans la série des outils développés par l'IRCAM on retrouve aussi le Spat, pour Spatialisateur, qui est un système dédié au traitement de la spatialisation en temps réel. Disponible à l'origine sous forme d'objet pour Max/MSP il a été porté en plug-in afin de pouvoir être utilisé sur toutes les stations de travail audionumériques. Son principe de fonctionnement est simple mais c'est un outil très puissant : on place des sources sonores dans un espace 2D ou 3D autour d'un point d'écoute puis le Spat se charge de gérer la réverbération, la simulation de l'atténuation de l'air et le panoramique afin de créer un rendu d'espace sonore cohérent et réaliste. Le Spat nous permet de choisir le format d'écoute souhaité en sortie, notamment par la synthèse binaurale ce qui va nous intéresser ici.



Capture d'écran du Spat dans Max/MSP avec deux sources et un rendu sur six enceintes.

2 – Mon projet

L'idée de départ de mon travail était de réfléchir à un système simple, se rapprochant au plus des méthodologies de travail que nous avons en cinéma, pour la gestion du son en vidéo 360°. Entre le moment où j'ai commencé à travailler dessus et aujourd'hui les choses ont bien changé avec notamment le développement des lecteurs vidéo décodant de l'ambisonique. Mais je pense qu'il y a encore moyen d'améliorer la manière de créer la bande sonore immersive à 360°.

Le principe sur lequel je m'appuie pour mon format est très largement inspiré de celui du Dolby Atmos (Dolby qui a d'ailleurs sorti entre temps, bien que peu usitée pour le moment, une version VR de son système Atmos...). La base de l'univers sonore est un environnement 5.0 qui reste fixe et ne bouge pas avec les mouvements de tête du spectateur. A ce bed¹⁴ s'ajoutent des pistes objets, elles spatialisées à 360° et suivant les mouvements de la tête. Je suis parti de l'idée que toute la scène sonore n'a pas besoin de tourner avec le spectateur ce qui permet d'économiser de la puissance de calcul et d'améliorer la qualité du rendu des objets. Un petit test m'a permis de confirmer en partie cette idée, j'ai mis une ambiance fixe de forêt dans laquelle j'ai intégré des oiseaux, eux spatialisés en temps réel. Le rendu est tout à fait acceptable et la sensation de mouvement bien présente même si toute la scène sonore ne bouge pas. On peut même imaginer sur ce principe de pouvoir convertir une création sonore d'un visiocasque vers une salle de cinéma Atmos.

Du côté de la technique, j'utilise Max/MSP qui récupère les données de mouvement du visiocasque afin de faire bouger la scène sonore générée par un Spat conjointement avec les rotations de la tête, le rendu du Spat étant généré

¹⁴ Pour reprendre la terminologie Dolby.

en binaural il permet la restitution d'un son 3D. Autour du bed 5.0 il y a la possibilité de placer autant de sources en mode « objet » que l'on souhaite ainsi que de les faire se déplacer dans le temps. À l'heure où j'écris ces lignes le système n'est malheureusement pas encore tout à fait fonctionnel mais j'ose espérer qu'il le sera au moment de vous le présenter!

Conclusion

Nous avons donc exploré ensemble l'univers de la vidéo 360° et plus particulièrement les problématiques ayant trait à la construction d'une bande sonore pour un tel objet. Il n'est pas aisé de travailler le son pour un tel média, les outils étant pour l'instant encore en développement et une bonne connaissance des mécanismes physiologiques de l'audition nécessaires afin d'exploiter pleinement le potentiel offert. Mais il y a tout de même des idées à chercher et des concepts à inventer au niveau de la mise en scène – sonore certes, mais aussi de manière plus globale. C'est un nouvel outil de création et de divertissement avec ses avantages et ses défauts.

Mais la vidéo 360° pose une autre problématique vis à vis du cinéma, c'est son mode de consommation. En effet, l'individualisation de celui-ci pose de nombreuses questions du point de vue de l'exploitation, en effet, plus besoin de salles pour projeter, plus d'horaires de séances. MK2 a réagit en lançant la première salle de réalité virtuelle à Paris, le MK2 VR, même si pour l'instant les vidéos disponibles sont plus du ressort du divertissement, du grand frisson et de l'amusement que d'une véritable recherche artistique. Ne serions-nous pas en train de retourner aux prémices du cinéma avec le Kinétoscope, imaginé par Edison en 1888, dans lequel il suffisait d'insérer quelques cents pour profiter d'un petit film à regarder à travers un binoculaire...



Kinétophone, successeur de Kinétoscope qui permettait la lecture synchrone d'un son sur rouleau de cire. (source : wikimedia.org)



Une autruche portant un visiocasque. (source : publicité Samsung Gear VR)

Bibliographie

Ouvrages

ALMENDROS Nestor, *Un Homme à la caméra*, Hatier, 1980.

BARBOZA Pierre, *Fiction interactive « métarécit » et unités intégratives,* L'Harmattan, 2006.

BLAUERT Jens, *Spatial Hearing : The psychophysics of Human sound localization*, The MIT Press, 1996.

BRESSON Robert, Notes sur le cinématographe, éd. Gallimard, 1995.

CHION Michel, Le Son au cinéma, éd. de l'Étoile/Cahiers du cinéma, 1994.

FISCHETTI Antonio, *Initiation à l'acoustique*, Belin, 2003.

FUCHS Philippe, *Les casques de réalité virtuelle et de jeux vidéo*, Les Presses des MINES, 2016.

ROBERT Paul, Le Petit Robert, Dictionnaires Le Robert, 1967.

ROSSI Mario, Audio, PPUR Presses Polytechniques, 2007.

Articles de revue

CHATAURET Jean, « La perspective sonore », *Communications*, Vol. 85, n°1, 2009, pp. 139-154.

MIDDLEBROOKS John C. & GREEN David M., « Sound Localization by Human Listeners », *Annual Review of Psychology*, 1991, pp. 135-159.

WEBER Tatiana, « Gothique et expérience du sacré : l'écho et la profondeur », Bulletin de l'Association Guillaume Budé, Vol. 55, n°4, 1996, pp. 383-394.

Filmographie

(collectif), Fantasia, 1940.

GANCE Abel, Napoléon vu par Abel Gance, 1927.

GODARD Jean-Luc, Adieu au langage, 2014.

GOLDEN Joseph A., The Power of Love, 1911.

LUMIÈRE Louis, L'Arrivée d'un train en gare de La Ciotat, 1896.

NOÉ Gaspard, Love, 2015.

NURIDSANY Claude & PÉRENNOU Marie, *Microcosmos : Le Peuple de l'herbe*, 1996.

ANNEXE I: Liens vers les vidéos 360°

Conservation International, Valen's Reef

https://www.youtube.com/watch?v=MC_ILxVIuT4

360 Labs, Rogue River Rafting

https://www.youtube.com/watch?v=iv4lHXz2XIQ

Marc Dorcel VR

https://www.dorcel.com/VR/?lang=fr

ZANDROWICZ Pierre, I, Philip

http://creative.arte.tv/fr/episode/i-philip

COGITORE Clément, Art Stories – Le château de Fontainebleau

http://sites.arte.tv/360/fr/art-stories-le-chateau-de-fontainebleau-360

Borrowed Light Studios, The Night Café

https://www.youtube.com/watch?v=Si-pImnlFZs

MIDDLETON Peter & SPINNEY James, Notes on Blindness

http://www.notesonblindness.co.uk/

CNN VR, Refugee: 'We are not animals'

http://edition.cnn.com/2017/03/05/vr/lesbos-migrants-vr/index.html

The New York Times, Paris Vigil

https://www.youtube.com/watch?v=o3yvjEMFbnM

CHASSAING Romain, Naive New Beaters & Izia, Heal Tomorrow

https://www.youtube.com/watch?v=JxVVNm35rJE

ANNEXE II : Comparatif technique de différents visiocasques

	Oculus Rift	HTC Vive	Playstation VR	StarVR
Marque	Oculus VR	HTC, Valve	Sony	Starbreeze
	(Facebook)	Corporation		Studios
Date de sortie	2016	2016	2016	/
Résolution	2160 x 1200	2160 x 1200	1920 x 1080	5120 x 1440
Fréquence de	90 Hz	90 Hz	120 Hz	90 Hz
rafraîchissement				
Champ de vision	110°	110°	100°	210°
Poids	470 g	600 g	610 g	380 g
Prix	589 €	899 €	399 €	Entre 3000 et
				4000 \$