Ecole Nationale Supérieure Louis Lumière
La Cité du Cinéma
20, rue Ampère
BP 12 - 93213 La Plaine Saint-Denis
33(0) 1 84 67 00 01

Mémoire de master

Spécialité son, promotion 2013-2016
Soutenance de Juin 2016

Immersion sonore et interfaces transparentes, les enjeux d'une vidéo interactive

Lucien Richardson

Ce mémoire présente l'élaboration de la partie pratique intitulée :
Autoscopie (2016)

Sous la direction de Thierry Coduys et Guillaume Jacquemin
Rapporté par Gérard Pelé

Remerciements à

Thierry Coduys
Guillaume Jacquemin
Gérard Pelé

Zoé Moineaud

Louise Pagès
Juliette Le Monnyer
Maxime Gourdon
Irène Dieudonné
Arnaud Parenty

Antoine Martin
Cedric Payet
Maxime Rifad
Sylvain Lambinet
Alan Blum
François Salmon
Clémence Reliat
Celsian Langlois
Fabrice Loussert
Gilles Wolff
Florent Fajole
Agnès Hominal

Résumé

Ce mémoire présente, à travers la proposition d'une expérience immersive entre cinéma et multimédia, un questionnement sur la place du spectateur dans l'art interactif.

L'omniprésence de l'écran dans le contexte socioculturel et l'immersion dans les contenus proposés interroge les évolutions technologiques dans le domaine de l'image et du son. Il en vient une disparition progressive des interfaces pour une translation directe des sens du spectateur dans son espace virtuel. On s'intéressera aux techniques et technologies permettant d'augmenter la relation sensorielle entre le spectateur et l'œuvre dans la cadre d'une installation d'art vidéo interactif dont une proposition sera présentée comme partie pratique de ce mémoire.

Mots-Clefs : Art participatif, Art video interactif, Interface, Transmodalité, Binaural, Head-Tracking, Eye-Tracking, Objets sonores, Compositing, Scénario à tiroir.

Abstract

In this Master's thesis, a proposition of an immersive experience, between cinema and multimedia, questions the place of the viewer in interactive art.

The sociocultural context of the ever-present screen and the immersion of the screen user in the proposed contents interrogates the technological evolutions in the fields of sound and image. The interfaces between viewer and virtual content gradually disappear to be replaced by a direct translation of the viewer's senses into his own virtual space. The focus is therefore placed here upon the techniques and technologies enabling greater sensory relationship between the viewer and an interactive video art piece. A practical demonstration of the technological installation, where the spectator is confronted to one of the proposed artworks, will be detailed in part three of this essai.

Key-Words: Participative art, Interactive video art, Interface, Transmodality, Binaural, Head-Tracking, Eye-Tracking, Sound objects, Compositing, Script writing.

Sommaire

INTRODUCTION

La manipulation d'interfaces numériques telles que l'ordinateur ou le téléphone portable permet à leurs utilisateurs d'être constamment connectés par le biais d'un écran. Nous pouvons associer cette pratique massive, devenue usuelle depuis le début du XXIe siècle, à une notion de proximité faisant de l'objet un élément personnel, une extension de son être. Bien que le rapport à l'écran semble identique, le cinéma ne développe pas chez ses spectateurs ce sentiment, notamment parce qu'il invite à une expérience en collectivité et en temps différé. Cette dernière notion prend toute son importance lorsqu'on considère le critère d'immersion dans ce qu'on l'invite à regarder. Au cinéma, le film est écrit dans un premier temps, puis visionné ; ce processus de temps différé ne laisse pas de zone d'incertitude ou de hasard, imposant au spectateur un état de perception passive. Si celui-ci peut influencer la production cinématographique par son adhésion ou son rejet des propositions filmiques, il ne peut en aucune manière participer à l'élaboration d'un produit finalisé.

Cette tendance vient néanmoins à s'inverser : si la 3D au cinéma apparaît dès la fin des années 1890, elle connait actuellement une résurgence qui tend à devenir le standard des plus grosses productions. Ce regain d'intérêt s'intègre dans un climat général d'immersion toujours plus grande dans les médias qui nous sont proposés. L'intégration du spectateur des ces univers virtuels (que ce soit la 3D et l'IMAX au cinéma ou les dispositifs de réalité augmentée dans les jeux vidéos) passe par le développement accéléré des techniques visuels, mais aussi des technologies du son, facteur clef de l'immersion.

On remarque cette évolution dans l'histoire du cinéma, de l'apparition de la stéréophonie en 1977 avec *Star Wars* de George Lucas à l'emploi aujourd'hui de systèmes de diffusions 22.2 et des mixages dits *orientés objet*, utilisant des canaux zénithaux et plusieurs dizaines de sources tout autour du spectateur de cinéma. Le monde du jeu vidéo et notamment les FPS, ou *First Person Shooter* (dans lesquels le joueur incarne un personnage et découvre l'univers du jeu depuis son point de vue) ainsi que l'avènement des casques de réalité virtuelle comme l'Oculus Rift alimente une individualisation de l'immersion audiovisuelle. La question de la retranscription réaliste d'une scène sonore au casque, où les sons nous parviennent de toute part, est alors de première importance.

À quelques exceptions près, le cinéma demeure au premier stade du développement de la 3D et d'autant plus de la réalité augmentée. Néanmoins, le domaine de l'art dit « interactif » a su se saisir de ces nouvelles techniques, autant à l'image qu'au son, dans le but de créer une perméabilité maximale entre la réalité et la virtualité. Faisant davantage appel à ses sens qu'à sa psyché, cette nouvelle forme d'art tend à placer le spectateur au centre d'oeuvres évolutives.

Partant de ces réflexions, j'ai choisi d'étudier cette forme d'art singulière qu'est l'art vidéo interactif. Ainsi je porterai l'accent dans un premier temps sur l'état actuel de l'art interactif, les modalités de son apparition, et les questions qu'il soulève sur la place du spectateur dans l'art. Prenant nécessairement en compte son lien étroit avec les nouvelles technologies de l'image et du son, j'interrogerai l'état de l'art des techniques mises en oeuvre pour une immersion sonore et une transparence des interfaces. Puis, dans le prolongement de ces acquis, je proposerai l'élaboration d'une installation d'art vidéo interactif.

CHAPITRE 1 : Un art de l'interaction

I / APPARITION D'UN ART PARTICIPATIF

A / L'évolution de la place du spectateur

La place du spectateur dans l'art fait l'objet d'interrogations de tout temps. Certaines réponses parmi d'autres sont apportées lors de l'apparition d'un théâtre participatif, comme le théâtre de marionnettes ou le théâtre forum lors duquel le public était invité à intervenir.

Dans les années cinquante et soixante, la place du spectateur face à un art de musée vient à évoluer. Les artistes et théoriciens de l'art s'intéressent ainsi à la notion de participation du public au sein même de l'oeuvre, faisant passer celui-ci d'une position dite « passive » à une position « active ». Ce désir d'intégration a notamment pour origine les mouvements hippys, visant globalement à rejeter les règles bien trop rigides d'un système jugé archaïque, et donc à redéfinir les bases classiques de l'œuvre d'art. En 1969, Jack Burnham écrit que toute chose qui « traite/assimile des données artistiques,… est un composant de l'oeuvre d'art »[1]. Insistant ici sur l'importance d'appréhender une oeuvre au sein de son environnement et de son contexte de création, le public devient par définition partie intégrante de l'oeuvre.

Dans le cadre des premières participations du public, voici l'oeuvre de John Cage : 4'33. L'artiste fait le constat qu'il n'entendra jamais de silence pur après en avoir fait l'expérience dans une chambre anéchoïque. Ainsi, il décide de composer une partition pendant laquelle viennent se jouer des paramètres sonores tels que les bruits liés au spectateurs dans la salle, la respiration de l'interprète, etc. L'oeuvre basée sur le principe du hasard est ainsi composée de sons que l'artiste lui-même n'a pas choisi, et vient redéfinir la notion de silence par le calme ou l'agitation des spectateurs.

Par la suite, le happening (de l'anglais « to happen »: arriver, se produire) inventé par Allan Kaprow à la fin des années 50, correspond à une performance, ou à un événement pendant lesquels le public est invité à participer et est même considéré comme un intervenant en tant que tel. L'exemple du happening *Eat*[2], de Allan Kaprow, est décrit par Micheal Kirby en

---

[1] **BURNHAM** Jack, Real Time Systems. Artforum, Vol. 7, Septembre, 1965.

[2] Smolin Gallery, New-York. 1964.

1965 [3]. L'objectif était de placer le spectateur dans un Environment (terme utilisé par l'artiste), au sein duquel il était libre d'interagir ou non. En l'occurence, choisissant de décrocher une des pommes pendues au plafond, ou simplement de croquer l'une d'elle, voire même de ne pas y toucher, le spectateur s'intégrait dans le système du happening. Ici le fait même d'interagir devient alors l'oeuvre en elle-même.

Le groupe Fluxus, initié par George Maciunas dans les années soixante, fait également parti d'un mouvement visant à redéfinir le statut de l'oeuvre d'art et de l'artiste lui-même en proposant aussi bien des créations d'art visuel que musicales ou littéraires par la réalisation d'événements, concerts, livres et revues, objects, etc. auxquels le public est invité à participer dans un désir sous-jacent de questionner la place de l'art dans la société de l'époque. L'aspect sarcastique et dérisoire de l'identité du groupe est indéniable et tend à créer un « anti-art » à l'inverse d'un art conventionnel, un art dit de distraction. Les artistes du groupe se font notamment un nom grâce à des performances où le spectateur revêt un rôle primordial. Par exemple pour *Cut Piece*[4], Yoko Ono invite les personnes présentes à constituer pleinement sa performance en lui coupant des bouts de vêtements.

[3] **KIRBY Micheal,** Allan Kaprxw's Eat, Tulane Drama Review. Vol. 10, No. 2, 1965.

[4] Carnagie Hall, New York, 21 mars 1965.

B / Vers un art dit « Interactif »

Dans ce développement d'un art participatif, vient à apparaitre un art dit interactif dont l'objectif n'est plus seulement de développer l'activité psychologique du spectateur mais aussi d'inviter ce dernier à réagir grâce à l'utilisation de ses sens. L'oeuvre va ainsi interagir de façon dynamique à son public et/ou à son environnement à l'aide de la technologie, et instaurer un dialogue avec son interlocuteur en temps réel.

Pour mieux envisager cet art dont la manière de percevoir est différente de ce que l'on connaissait jusqu'à présent, je me suis intéressé à la question d'une identité de l'art interactif, que va soulever Erkki Huhtamo en 2004, lorsque le Nicas d'Or[5], pour la catégorie art interactif est décerné à Ben Rubin et Mark Hansen pour *Listening Post*.[6] L'œuvre consiste en 231 petits écrans électroniques suspendus sur une grille incurvée capturent des fragments de textes issus de Chatrooms (salons de discussion en ligne) en continu. Certains fragments sont énoncés par une voix synthétique. Si l'œuvre vit une métamorphose continuelle provoquée par les flux d'informations en provenance du web, Huhtamo argumente qu'elle élude complètement la question de l'interaction avec le spectateur qui n'est finalement que dans une position de réception du contenu. Il propose ainsi de réserver le terme « interactif » aux oeuvres dont le contenu ainsi que le fonctionnement nécessite une participation active et répétée d'un spectateur-utilisateur.

Dans notre environnement contemporain, les nouvelles technologies en constante évolution sont un point d'ancrage important dans le contexte de l'interactivité. Si *Listening Post* s'est développée au moment de l'essor de ces technologies de la communication, les œuvres interactives récentes ont pour bénéfice de disposer d'un public averti et pleinement utilisateur de ces outils d'échange.

---

[5] Prix décerné à l'Ars Electronica, festival consacré à la création numérique depuis 1979

[6] **HUHTAMO Erkii**, Trouble at the Interface, or the Identity Crisis of Interactive Art, dans Framework, The Finnish Art Review, 2004.

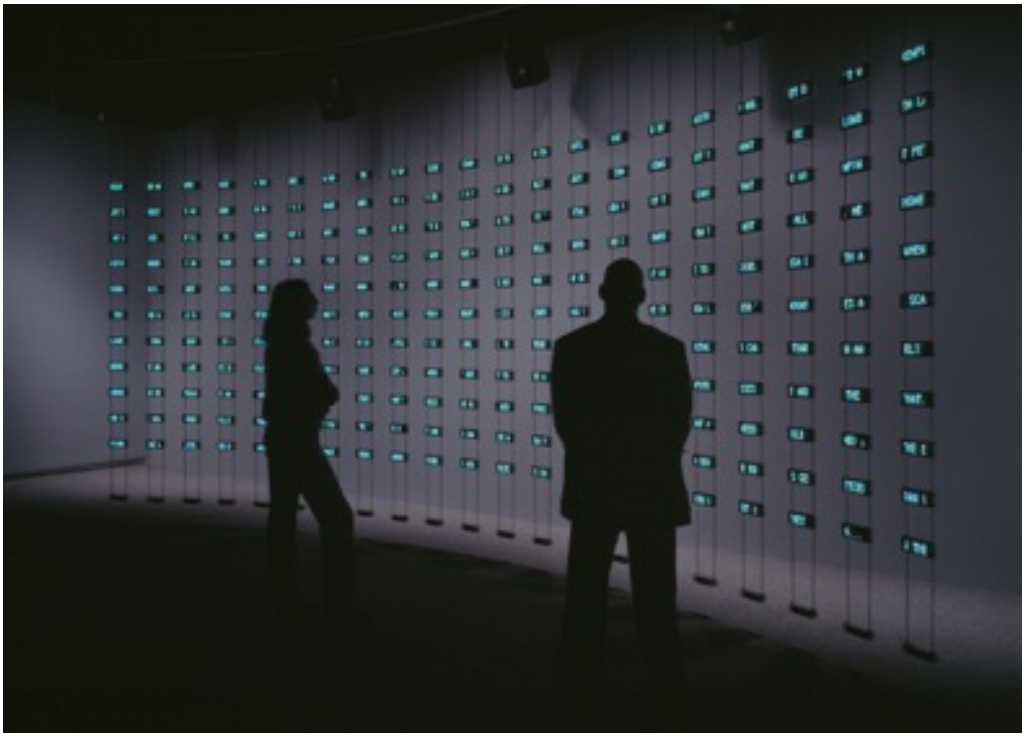

Figure 1.1.1 : Listening Post, Ben Rubin et Mark Hansen, Victoria and Albert Museum, Londres, 2010.
http://www.d-load.de/blg/?p=116

Samuel Bianchini, enseignant-chercheur à l'ENSADLab[7] et artiste de renom dans son domaine, a proposé en février 2016 une œuvre interactive appelée « Surexposition ». L'artiste a créé une installation urbaine sous la forme d'un grand monolithe noir, proposant aux passants de lui envoyer des SMS qui seront ensuite retranscrits en Morse, et projetés dans le ciel au moyen d'un faisceau lumineux. Ainsi le caractère interactif de cette proposition semble évident. Pour aller plus loin, nous pouvons émettre l'hypothèse qu'aucun spectateur de l'oeuvre n'est réellement passif. Dans le cadre de cette installation, l'universalité du téléphone portable et de son utilisation montre que même si le spectateur décide de ne pas participer, ce choix correspond à une interaction en soit. Le contexte contemporain paraît cependant nécessaire pour justifier la possible passivité du spectateur.

Ainsi nous pouvons nous interroger sur la nécessité de la volonté d'interaction du spectateur. Peut-on parler d'oeuvre interactive si le spectateur refuse la participation voire n'a pas conscience ou n'intellectualise pas cette interaction? Dans *Village Green*[8], créé en 2008, Vaughn

[7] Laboratoire de recherche de l'Ecole Nationale Supérieure des Arts Décoratifs de Paris.

[8] Chemical Heritage Foundation, depuis le 8 janvier 2014.

Bell nous apporte une réponse. Il propose aux visiteurs de l'exposition d'insérer leurs têtes dans des « biosphères de poche », de minuscules écosystèmes suspendus, dont le contenu est représentatif de la faune et la flore locale. Or l'échelle de l'expérience rend l'apport du spectateur, par exemple en dioxyde de carbone ou avec son souffle, infiniment plus important que dans l'état de nature. Ainsi, ce système évolue au fil des rencontres avec l'humain, tout comme le spectateur est lui-même influencé par l'expérience inconsciente de la personne l'ayant précédé.

II / DES REFLEXIONS SUR L'ART INTERACTIF

A / Catégorisation de l'art interactif selon Ernest Edmonds

L'ensemble des questions vues précédemment sur le rôle et la dynamique de l'art interactif tend à le rendre complexe à définir et à appréhender. Dans le but de mieux le comprendre, revenons en 1973, moment pendant lequel Ernest Edmonds va proposer une catégorisation des oeuvres d'art, finalement applicable à l'art interactif. Ces catégories tentent de caractériser les oeuvres selon les relations mises en jeu entre l'oeuvre d'art, l'artiste, le spectateur et l'environnement. Il les définit comme suit : statique, dynamique-passive, dynamique-interactive et dynamique-interactive (variable).

Les oeuvres statiques sont les plus communes. Elles représentent des formes d'art dites classiques. L'œuvre n'évolue pas quand elle est observée par un spectateur et il n'y a pas d'interaction possible à proprement parler puisqu'elle fait appel à une appréhension psychologique et émotionnelle de la part du spectateur. L'oeuvre n'est pas sensible à son environnement ou son contexte. Dans cette optique et pour exemple, un album enregistré correspond à une oeuvre statique. D'après ce constat, Alan Licht considère davantage, dans son livre Sound Art, que le compositeur Glenn Gould « conceptualise ses albums comme des installations sonores interactives » (p.40) en intégrant dans son concept d'enregistrement l'influence de la transformation effectuée par le système d'écoute du spectateur.

Dans le cas des oeuvres dynamiques-passives, l'objet d'art contient un mécanisme interne qui rend l'œuvre sensible à son environnement. Ainsi, elle se modifie ou est modifiée par un facteur extérieur comme du son, de la lumière, des variations climatiques ou bien encore la présence d'animaux et leurs actions . C'est l'exemple du film *A Way in Untilled*[9], réalisé durant l'été 2012 par Pierre Huygue, dans lequel nous pouvons voir à la tombée de la nuit des processus organiques ou de putréfaction, autour d'une sculpture composée d'un corps de femme à demi allongé et d'un essaim d'abeilles à la place de la tête. C'est également l'exemple de l'oeuvre *From Here to Ear*[10] de Céleste Boursier-Mougenot réalisé en 2012, dans laquelle l'on

[9] Marian Goodman Gallery, New York, 2012.

[10] Biennale de Venise, 2015.

peut voir 70 Diamants mandarins perchés sur des guitares électriques, le tout produisant des mélodies aléatoires mêlées aux chants des oiseaux.

Figure 1.2.1 : A Way in Untilled, Pierre Huygue, MOMA, New-York, 2015.
http://www.moma.org/calendar/exhibitions/1537?locale=en

Ainsi le moteur génératif est construit et déterminé à l'avance par l'artiste de telle façon que toutes les variations et modifications possibles de l'œuvre soient prévisibles ; elle évolue dans un système prédéfini. Dans ce cas, le spectateur est un observateur passif de l'interaction entre l'œuvre et son contexte direct. Par exemple, les mobiles monumentaux d'Alexander Calder[11] sont limités dans leurs mouvements par l'amplitude permise par leur construction, mais l'oeuvre réagit au vent et conserve une qualité d'improvisation. Pour ses Sculptures Kinétiques, George Rickey[12] va encore plus loin en tablant sur un équilibre très fin entre des objets au centre de gravité désaxé qui réagissent à la plus petite des brises. L'effet de gravité rend ses oeuvres imprévisibles d'autant plus envoutantes qu'on ne peut pas en prédire exactement la trajectoire.

[11] Spirale, 1958, fondation UNESCO

[12] Column of Four Squares Eccentric Gyratory III, Var. II

Les oeuvres dynamiques-interactives impliquent une influence directe de la part du spectateur dans les changements que subit l'oeuvre. Il acquiert ainsi le statut d'utilisateur. En effet, des capteurs sont utilisés pour quantifier ou simplement jauger une valeur définie, modifiée par le spectateur, et ces informations ont des répercussions directes et prédéfinies sur la performance de l'oeuvre. De cette manière, il participe à sa création dans le sens où elle n'existe que dans son interaction avec l'homme. C'est dans cette catégorie que se développe la plupart des travaux sur l'art interactif. De plus, la majorité des oeuvres dites dynamiques-interactives tendent vers la mise en place d'une analogie entre deux sens. Par exemple, un capteur d'image utilise les mouvements perceptibles d'un utilisateur pour jouer sur le son, c'est le cas du *Iamascope*[13] crée par Sidney Fels, ou sur une représentation visuelle comme nous pouvons le voir dans *Boundary Functions*[14] de Scott Snibbe réalisé en 1998, consistant en une plate forme au sol sur laquelle les spectateurs peuvent marcher. À ce moment précis apparaissent sur la plate forme des lignes délimitant l'espace de chaque marcheur et évoluant en fonction du nombre de personnes et de leur déplacements. On pourrait encore décrire les constructions d'Edward Ihnatowicz où plusieurs microphones sont utilisés pour localiser une source sonore et cette information est utilisée pour mouvoir une sculpture et sembler lui donner vie (Edward Ihnatowicz - SAM, *Senster*). A l'inverse de ces analogies prédictives mais toujours dans cette catégorie, Daniel Rozin, s'intéresse davantage à une translation directe au sein du domaine physique. Son travail se caractérise par l'utilisation de miroirs interactifs réagissant de façon cinétique aux mouvements du ou des spectateurs.

[13] Dept. of Electrical and Computer Eng., University of British Columbia, Vancouver, BC. 1998

[14] Ars Electronica 1998 (prize), Linz, Austria.

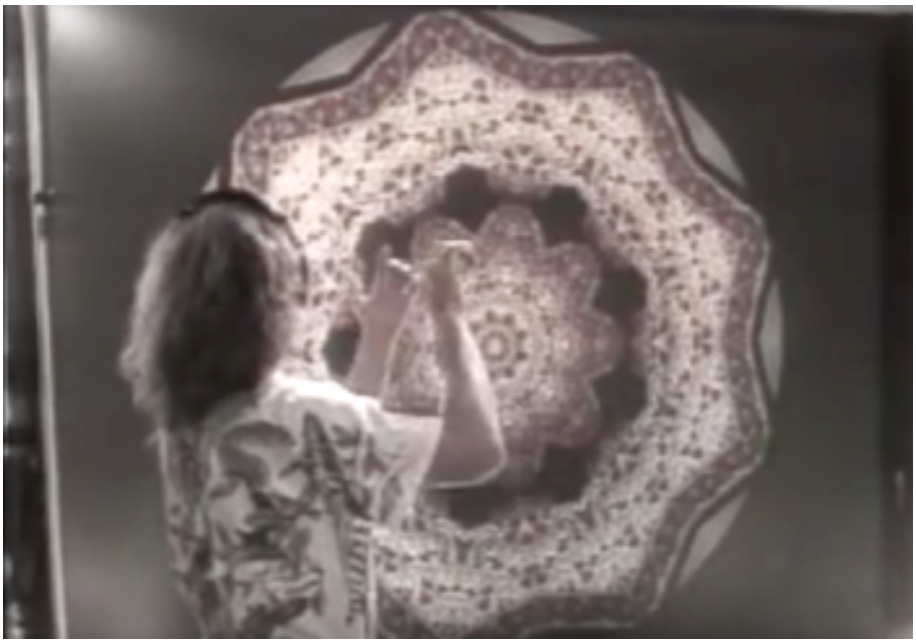

Figure 1.2.2 : Iamascope, Sidney Fels, 1998.
https://www.youtube.com/watch?v=yIIZMO9xPE8

La dernière catégorie définie par Ernest Edmonds est celle des oeuvres dynamique-Interactives variables. Ici, les caractéristiques originelles de l'oeuvre sont modifiées au cours du temps, soit par un programme informatique, soit par l'artiste. De cette façon, le caractère prévisible de la performance, engendrée par le système que forment l'oeuvre et le spectateur, n'existe plus. Elle dépend alors de l'historique des interactions au sein de ce système.

À la suite de ces recherches, Ernest Edmonds développera au début des années 1990, le concept de Learning Interactive Video Construct. Il décrit son travail dans le rapport de conférence de 2004, comme étant le développement successif des catégories sus définies. Ainsi, il propose une vidéo générative faisant apparaitre séquentiellement des formes géométriques de couleurs variées. Le programme obéit à un certain nombre de règles auxquelles le spectateur n'a pas accès. Celui-ci ressent néanmoins un certain contrôle dans le fait que le programme n'est pas aléatoire. Edmonds implémente ensuite l'aspect interactif à certaines de ses vidéos, en les faisant réagir notamment à la proximité des spectateurs grâce à des caméras. La dernière étape consistait à inculquer une forme d'apprentissage à ses oeuvres.

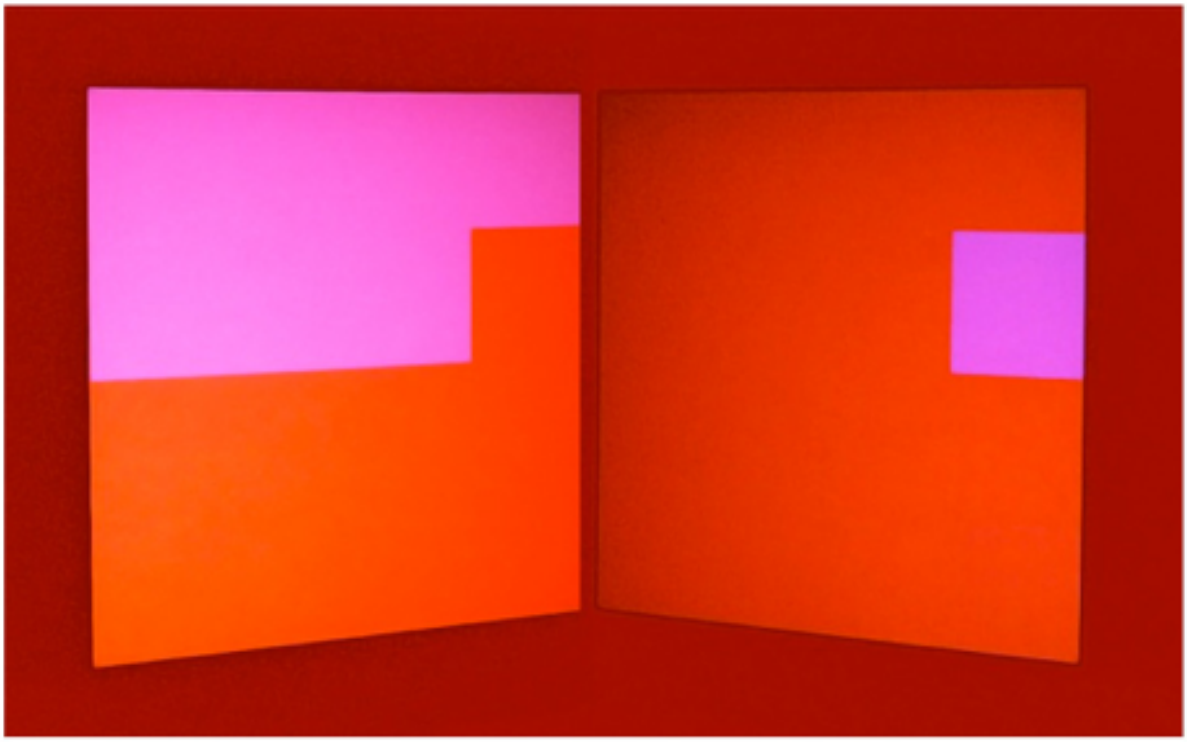

Figure 1.2.3 : Shaping Form, Ernest Edmonds, Site Gallery, Sheffield, 2012.
http://www.sitegallery.org/archives/4849#.UJPP7I6W6ME

Par la suite, il proposera en 2002 son premier *Shaping Form*[15], un alignement projeté de bandes verticales colorées et mobiles. Une caméra dissimulée détecte à la fois la proximité des spectateurs et la quantité de mouvement dans son champ de vision. Lorsque quelqu'un s'approche, les bandes semblent fuir et s'affinent, et la vitesse de cette fuite est proportionnelle à la quantité de mouvement détectée. L'oeuvre n'est jamais immobile, et l'artiste lui prête une appréhension de son public dans deux paramètres supplémentaires : si la quantité de mouvement dépasse un certain seuil, la vitesse de recul revient à son état initial, semblant indiquer que l'oeuvre résiste aux mouvements frénétiques. De plus, l'oeuvre accumule les données de vitesse, de quantité et dynamique de mouvement au cours de la journée. La règle d'apprentissage va ensuite utiliser ces données pour, par exemple, abaisser le seuil auquel l'oeuvre ne réagit plus. De cette façon, Edmonds confère à son oeuvre la possibilité d'être « fatiguée » de toute l'agitation perçue, et l'expérience de chaque spectateur dépend de ce qu'a « vécu » l'oeuvre au cours de la journée.

[15] Site Gallery, Sheffield, 2012.

Par ailleurs, des ponts entre plusieurs catégories sont possibles, c'est le cas de la salle verte dans l'exposition *MindFuck*[16] de Bruce Nauman, qui se situe étrangement entre la catégorie de l'oeuvre statique et celle de l'oeuvre dynamique interactive. Un espace clos est éclairé en vert, il existera comme tel, qu'il soit ou non observé par un spectateur, mais l'oeuvre en elle-même consiste à produire une expérience sur la perception qu'à le spectateur sur son propre corps. L'oeuvre n'existe ici qu'à travers son interaction ou du moins sa relation avec le spectateur.

L'oeuvre de Tania Mouraud, *One More Night*[17] se situe également dans cet entre-deux. Des marches sont installées au centre d'une pièce blanche vide de tout autre objet, le spectateur déambule dans cet espace clos où une fréquence sonore continue est diffusée depuis plusieurs enceintes invisibles. En se déplaçant, le spectateur passe par des noeuds et des ventres acoustiques (où les sons sont respectivement en inversion de phase et en phase) causant la disparition et apparition de sons. Le propos de l'oeuvre étant dans la réception et la perception de zones de méditation liées à des fréquences pures. Ainsi l'oeuvre n'est pas dans la construction spatiale mais dans la relation entre le spectateur et le lieu.

Pour finir, Ernest Edmonds insiste sur l'importance de ne plus parler d'oeuvre d'art, mais bien de système d'art, terminologie qui intègre l'oeuvre, mais aussi son environnement, et en particulier le spectateur. Le rôle de l'artiste n'est plus de construire une oeuvre, mais de concevoir et de modifier les règles et contraintes qui définissent la relation entre l'oeuvre et son public.

---

[16] Hauser & Wirth London, 2013.

[17] Initiation Room, Galerie Rive Droite, Paris, 1970.

B / Dans le processus de réflexion de Simon Penny

À la suite de cette tentative de classification de l'art interactif et d'en comprendre le rôle et le statut, Simon Penny tente de le définir de façon plus approfondie en 1996. A l'époque il n'y a pas d'esthétique de l'interaction en temps réel encore culturellement établi,[18] il avance alors que le sens d'une oeuvre d'art est créé dans le dialogue entre le spectateur et l'oeuvre. Il me semble cependant qu'une barrière s'établit face à cet échange et tient dans le fait que l'utilisation des nouvelles techniques et dynamiques liées à l'art interactif ne sont pas, à l'époque, connues du spectateur : il ne partage donc pas encore le langage de l'artiste, l'oeuvre ne peut ainsi faire sens.

Partant de ce constat et contrairement à Ernest Edmonds et Alan Licht, Simon Penny considère qu'on ne peut pas parler d'art interactif sous prétexte qu'une oeuvre a un effet interne sur un spectateur, comme c'est le cas pour une sculpture ou une photographie. Il définit le terme interactif comme tel : « Un système interactif est une machine / est un système-machine qui réagit dans l'instant, en vertu d'un raisonnement automatisé basé sur des données issues de son apparat sensoriel / de ses capteurs. Un tableau est une instance de représentation. Un film est une séquence de représentations. Les oeuvres d'art interactives ne sont pas des instances de représentation, ce sont des machines virtuelles qui produisent elle-même des instances de représentation basées sur des entrées de données (input) en temps réel. […] Ce qui est perçu chez l'homme comme temps réel est le laps de temps (time frame) lié à une réaction physiologique.»

D'après les remarques faites pat Simon Penny, nous pouvons identifier qu'il existe deux types d'interactivité, celles d'espace et celles d'écran. Les premières partagent les caractéristiques sensorielles de la sculpture et de la danse, tandis que les secondes proposent une perception via une fenêtre, un point de vue guidé par l'artiste, un espace virtuel qu'on nous donne a naviguer. Dans le cadre du cinéma interactif, en développement depuis le début des années soixante, des projections spécialisées imprégnées de cette notion d'espace-écran, permettent au public

[18] **PENNY Simon,** From A to D and back again: The emerging aesthetics of Interactive Art, dans Leonardo Electronic Almanac, Avril 1996.

d'influencer un scénario à tiroir par voix de vote, c'est l'exemple de *Kinoautomat*[19], réalisé en 1967 par Radúz Činčera, au sein duquel les dilemmes moraux du personnage principal sont remis à un vote du public (via des boutons rouge et vert). Le film I'm Your Man réalisé en 1992, par Bob Bejan, utilise également ce principe: les spectateurs se servent d'un joystick pour choisir entre trois options à six moments du film. Enfin, *Last Call*[20], film d'horreur réalisé en 2010 par Jung Von Matt, met en place un scénario dans lequel la protagoniste appelle un spectateur dans la salle de cinéma et lui demande de l'aider. Les propositions sont toujours binaires et un système de reconnaissance vocale en temps réel permet une sensation d'interactivité plus accrue que dans les systèmes par vote.

Nous pouvons remarquer que de manière générale l'expérience du cinéma interactif est limitée par le concept même du film dont la réalité est le temps-différé. La réponse du système en temps réel aide à la sensation d'interaction du spectateur, mais bien souvent les choix sont contraints, car les possibilités doivent être déterminées et tournées à l'avance par le réalisateur du film. Cependant un film se démarque légèrement de cette tendance, en proposant non pas de modifier le scénario au cours du film, mais plutôt d'en influencer le montage. Lors des 30 premières projections de *Twixt* en 2011, Francis Ford Coppola change ainsi le montage de son film en direct, en fonction des réactions de la salle, de l'appréciation du public et ce que le réalisateur lui-même en ressent. Dans ce cas il s'agit davantage d'une performance d'artiste, dont le moteur est le spectateur inconscient de son interaction avec le film.

Quel que soit le type d'interaction, je me pose néanmoins la question de l'interface, dont le rôle est de réaliser une translation entre espace physique et virtuel. La translation se fait généralement de deux manières, soit de façon « cognitive » dans le cas de *Deep Contact*[21] de Lynn Hershman Leeson: une vidéo dans laquelle une femme séduisante invite le spectateur à la toucher par le biais d'une interface tactile, soit de façon linaire comme nous pouvons le constater dans l'oeuvre *Legible City*[22] de Jeffrey Shaw où le spectateur pédale un vélo dont le déplacement s'exécute dans un espace virtuel.

---

[19] Présenté au pavillon tchécoslovaque de l'Exposition universelle de 1967 de Montréal.

[20] 13th Street, chaine TV allemande, 2010.

[21] ZKM | Museum of Contemporary Art, Allemagne, 2015.

[22] ZKM | Museum of Contemporary Art, Allemagne, 1991.

Il semble donc difficile d'établir un canon esthétique pour l'art interactif, ceci est dû à la combinaison des technologies de l'image et du son, d'une utilisation sculpturale de l'espace et d'une automation numériquement coordonnée qui traite des données d'utilisateurs. C'est aussi cette surabondance de la technologie et de son aspect de cause à effet qui tend à soulever certaines interrogations sur la place de l'oeuvre lorsqu'il s'agit d'art interactif.

Figure 1.2.4 : Legible City, Jeffrey Shaw, première exposition au musée voor Hedendaagse Kunst, Antwerp, Belgium, 1988.
http://www.jeffreyshawcompendium.com/portfolio/legible-city/

III / LA PLACE DE L'INTERFACE DANS L'ART INTERACTIF

A / Interrogation sur le statut de l'oeuvre d'art interactive.

La place de l'oeuvre au sein de l'art interactif peut parfois être questionnée. Au cours des recherches, nous avons pu constater l'importance des moyens sensoriels, permettant la translation entre un espace réel et un espace virtuel, et plus particulièrement le rapprochement entre un geste physique et son équivalence sonore numérique, comme c'est le cas du *Iamascope* de Simon Fels, ainsi que du *Very Nervous System*[23] plus connu sous le nom de *VNS* de David Rokeby.

Ces artistes proposent en effet des systèmes qui présentent une interface d'interaction corporelle, par le biais d'une analyse vidéo en temps réel des mouvements que produit le spectateur. L'oeuvre n'existe pas sans une action directe. Nous sommes donc bien dans un système tel que décrit par Ernest Edmonds, où l'artiste met en place le contexte de règles au sein desquelles l'action du spectateur sera déterminante. Si *Iamascope* propose une représentation visuelle, elle aussi interactive, VNS, ne tient qu'à la représentation, la performance de l'utilisateur pour créer/jouer avec de la musique. Nous pouvons alors nous interroger, existe-t-il une différence entre ces oeuvres et un instrument de musique classique? Un instrument devrait-il s'éprouver de manière tactile? Pourrait-on qualifier un Theremin [24] d'oeuvre d'art interactif? Si c'est le cas, tout instrument de musique électronique n'en est-il pas un aussi? Le *Iamascope* est-il différent d'un synthétiseur programmé pour jouer dans une gamme unique? Les règles imposées par l'artiste ne sont peut-être pas si éloignées des choix effectués par un luthier ou un concepteur de synthétiseur. À quoi tient alors cette dénomination d'art sonore interactif?

Premièrement, les artistes définissent leurs créations comme telles. Certes ce point est réfuté par Simon Penny [25], mais il est communément admis qu'une production artistique se définit autant par la volonté de l'artiste que par la réception du spectateur. C'est d'ailleurs la

---

[23] Prix de l'Ars Electronica pour la catégorie Art Interactif, 1991.

[24] Instrument de musique se jouant sans contact physique, le déplacement des mains contrôle tantôt la hauteur tonale, tantôt le volume sonore.

[25] **PENNY Simon,** From A to D and back again: The emerging aesthetics of Interactive Art, dans Leonardo Electronic Almanac, Avril 1996.

définition qu'André Breton donne du concept de Ready-Made de Marcel Duchamp : « Objet usuel promu à la dignité d'objet d'art par le simple choix de l'artiste. ». [26] D'autre part, il ne faut pas oublier le poids du lieu d'exposition dans la perception du spectateur. Brian O'Doherty théorise l'espace d'exposition qui rend art ce qui y est exposé[27].

Finalement, il semblerait que toute oeuvre d'art interactif tient à sa magie, au sentiment de dépassement du spectateur. Et pour cela, on ne peut pas oublier la question du contexte socioculturel, ni l'époque de production d'une oeuvre. En effet cette dernière joue pour beaucoup dans la perception que nous avons de l'art, et de par son assimilation des nouvelles technologies, l'art interactif est peut-être l'une des premières victimes du temps. Lorsque Simon Penny propose sa classification, il définit la « télé-interactivité » et évoque des oeuvres comme *Ornithorinco* (1989, Eduardo Kac et Ed Bennett) ou *TerraVision* (1994, Art+Com), aujourd'hui comparables à des versions simplifiées de *Curiosity* et *GoogleEarth*. Si au départ ce sentiment de dépassement était souvent limité à une incompréhension face à l'élaboration technique d'une oeuvre, aujourd'hui l'artiste tend à laisser apparaitre l'infinité des possibles que celle-ci permet, ou alors il vise à changer complètement la perception du spectateur par rapport à un domaine que lui-même pensait maîtriser.

[26] **BRETON André,** Dictionnaire Abrégé du Surréalisme, Galerie des Beaux Arts, Paris, 1938.

[27] **O'DOHERTY Brian,** Inside the White Cube: the Ideology of the Gallery Space, University of California Press, Janvier 2000.

B / La possible transparence de l'interface.

Dans le but de combler mes interrogations quant aux moyens d'échanges et de mise en place de l'illusion de perception, je fais la supposition que l'intérêt principal que trouve le spectateur dans le travail notamment de Daniel Rozin est lié à l'utilisation d'un interacteur inhabituel, le corps entier. En effet, si aujourd'hui l'art interactif s'est répandu et ses codes sont généralement connus du public des espaces d'exposition, les artistes ont souvent recours à ce qui doit sembler naturel à l'homme pour retranscrire une sensibilité organique dans leurs oeuvres. C'est-à-dire que le spectateur a pour habitude de percevoir avec ses yeux et ses oreilles et de communiquer avec l'oeuvre par un interfaçage de ses mouvements et plus particulièrement de ce qu'il choisit de faire de ses mains. Dans le cas des miroirs de Daniel Rozin, d'une part le spectateur interagit avec tout son corps et d'autre part, il est générateur de l'oeuvre qu'il le veuille ou non. On retrouve d'ailleurs dans ce deuxième point ce qui a fait le côté magique de *Senster* [28] de Edward Ihnatovicz : une sculpture articulée aux mouvements organiques dont la tête se rapproche d'une source sonore, et donc donne l'impression de suivre une femme dont les talons hauts attirent sont attention. Concernant la réaction du public, la réactivité et les limites des oeuvres de Daniel Rozin vont naturellement être testées. Le spectateur tente de s'en approcher ou de s'en éloigner pour trouver l'astuce de cette relation évidente.

Il se dégage alors que deux conditions primordiales doivent être remplies pour le système interactif, qu'il soit à la fois sensible et naturel pour le spectateur, tout en conservant son attention et son intérêt : L'interface proposée doit être différente de ce avec quoi on nous propose généralement d'explorer des oeuvres interactives, mais aussi être naturellement utilisée par l'homme dans sa perception et son exploration du monde réel, de façon à rapidement atteindre une forme d'affordance[29]. Voici l'exemple Kyoto Two[30], décrit par Ernest Edmonds. Il évoque les points forts d'utiliser la position de bout du doigt de la main droite comme interface entre le spectateur et la vidéo-projection. Selon lui, ce choix permet « une réponse imprévisible et impulsionnelle de la présence du corps dans l'espace filmé par le système, de façon à attirer l'attention du spectateur » tout en lui certifiant un « contrôle intuitif et expressif » de l'oeuvre.

---

[28] University College, London, 1970.

[29] La qualité d'un objet à suggérer sa propre utilisation.

[30] **EDMONDS Ernest et al.** Approaches to Interactive Art Systems, GRAPHITE'04, 2004.

Figure 1.3.1 : Senster, Edward Ihnatovicz, University College, London, 1970.
http://www.tate.org.uk/context-comment/articles/gallery-lost-art-edward-inhatowicz

D'autre part, si la perception principale du système se fait par le biais d'un écran de projection, il est primordial d'investir un espace physique à la fois réellement, mais aussi dans l'illusion, de façon a rassembler les différentes sensibilités du spectateur et pour tenter de réduire la distanciation que l'on a intuitivement avec une oeuvre dont l'espace de développement n'est pas le nôtre.

La fin des années 80 voit apparaitre deux grandes réussites dans le domaine de l'art vidéo interactif, des installations qui font aujourd'hui référence dans la réflexion sur le rapport entre spectateur et oeuvre. *Je sème à tout vent* [31], conçue autour d'un microphone dissimulé et d'un pissenlit virtuel diffusé sur un petit écran, invite le spectateur à souffler sur la base du téléviseur avant de retranscrire en temps réel l'effet de ce souffle sur l'envol des akènes. Le même procédé est employé dans *La Plume*[32] où la durée et la puissance du souffle définissent la vitesse de montée d'une plume virtuelle posée en bas d'un écran d'ordinateur. Dans les deux cas, les trajectoires de retombées sont définies par des règles mathématiques et un certain aléatoire, proposés par les concepteurs de l'installation, mais la vie est littéralement insufflée par le spectateur devenu acteur de l'oeuvre. La fascination des spectateurs devant ces oeuvres est liée à l'invisibilité de l'interface. C'est à dire que le mouvement des objets virtuels semble lié de manière transparente à l'action physique du spectateur. La réactivité de l'oeuvre pousse à ne pas remettre en question ce lien de cause à effet, ni à questionner son fonctionnement et fait disparaitre le fait qu'un microphone capte le niveau de saturation de la capsule liée à l'effet du vent et que cette information est ensuite numérisée pour que ces datas contrôlent le programme à l'origine du mouvement de la plume ou des pissenlits.

Comme nous l'avons décrit plus haut, l'intégration du corps dans un domaine virtuel, au moyen de capteurs vidéo par exemple, est le sujet de nombreuses installations interactives. Mais ici la magie est produite par une relation haptique qui ne fait pas appel au sens du toucher. Ce qui fait la force des travaux de Couchot et de Bret est la dissociation entre l'effet produit et le moyen de fonctionnement.[33] La vidéo *De-Viewer*[34] réalisée par Joachim Sauter et Dirk Lüsebrink[35], est un bon exemple de cette dissociation: le spectateur regarde un tableau, là où sont regard se pose vient progressivement déformer l'image.

Pour rebondir sur l'idée d'une relation haptique, Simon Penny oppose deux types de travaux : les oeuvres où l'interaction n'est conçue que comme un moyen pour révéler un contenu, à la manière d'un pinceau ou d'un film et les oeuvres où les dynamiques de

---

[31] COUCHOT, Edmond et BRET, Michel. 1988

[32] COUCHOT, Edmond et BRET, Michel. 1990

[33] Commentaire énoncé par Gérard Pelé lors d'un cours à l'ENS Louis Lumière le 19 novembre 2015.

[34] Ars Electronica Center, Linz, Österreich, 1992.

[35] Art+Com, 1992.

l'interaction sont elles-mêmes le sujet. L'interactivité des commodités modernes de la haute technologie se doit d'être « complètement transparent, intuitive et instrumental » et « devrait générer un comportement qui existe dans le territoire liminal entre une perception du prévisible et une perception de l'aléatoire. Une zone de surprise, de poésie. » [36]

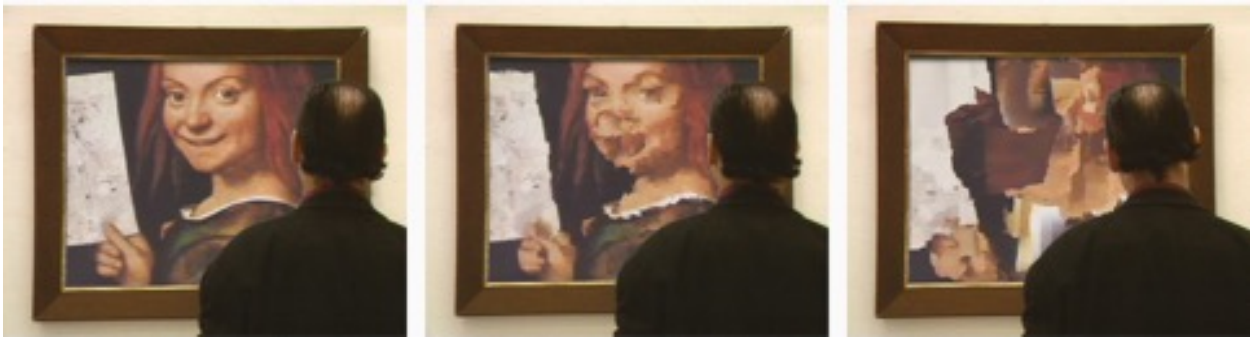

Figure 1.3.2 : De-Viewer, Joachim Sauter, Dirk Lûsebrink, Ars Electronica Center, Linz, Österreich, 1992.
https://artcom.de/project/zerseher/

Néanmoins, l'avènement de l'interaction numérique a imposé le développement d'un paradigme de l'interaction basé sur des structures et des symboles familiers. Aujourd'hui, l'éducation au clavier, à la souris et au curseur fait partie d'un contexte culturel généralisé qui conditionne l'implémentation de l'interactivité dans l'art et ailleurs. Une grande partie des oeuvres interactives répond au même parallèle entre la souris et le curseur, où une translation du corps humain s'opère dans le domaine virtuel. Dans le cas contraire, l'artiste a le défi d'introduire des nouvelles modalités propres à l'oeuvre sans donner l'impression d'un tutoriel. Même si ce protocole semble plus compliqué, il demeure cependant nécessaire dans certains cas de figure. Reprenons l'exemple de « Je sème à tout vent » réalisé par Edmond Couchot et Michel Bret en1988, puis représenté en 2006 sous le nom de Les Pissenlits[37], avec une nouvelle interface physique et une virtualisation de l'écran. L'effet de transparence de l'oeuvre d'origine proche du naturel, est amoindri par la présence d'un tube dans lequel le spectateur doit souffler. Il en résulte une expérience plus limitée, car le dispositif est visible.

[36] **PENNY Simon,** Interactivity - who cares?, Forthcoming Fiberculture, 2011.

[37] Exposition « Artificial Emotion », Centre Culturel ITAU, Sao Paulo, Juillet 2006.

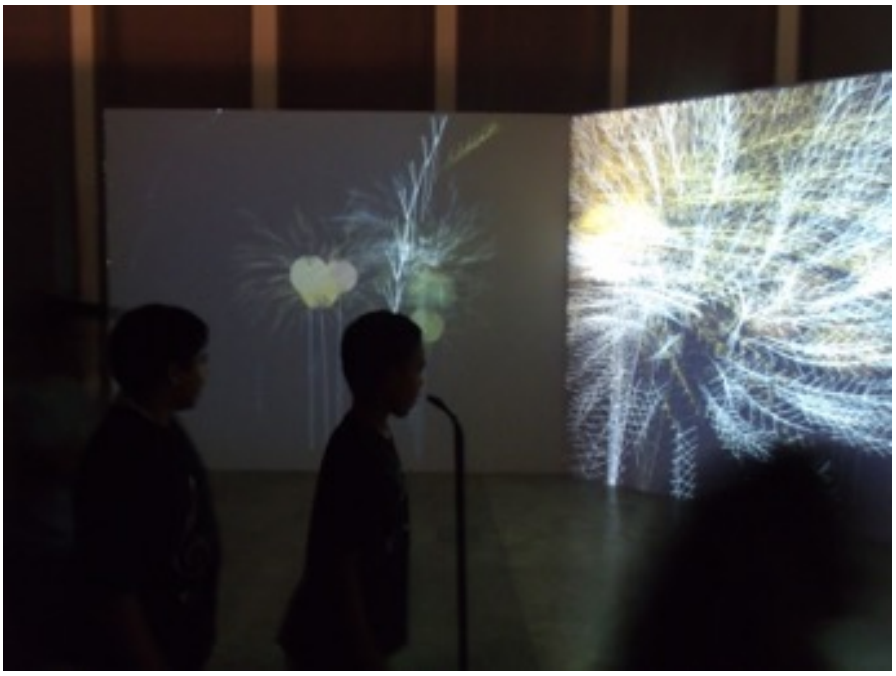

Figure 1.3.3 : Les Pissenlits, Edmond Couchot, Michel Bret, exposition Art Cybernétique, Université de Sao Paulo, 2012.
http://primaparaiba.blogspot.fr/2012/11/arte-cibernetica.html

Cet état de l'art a soulevé un certain nombre de questions sur la place du spectateur dans l'art interactif, et notamment sur son rapport à l'espace virtuel. Je souhaite maintenant approfondir mes recherches dans le cadre de l'art vidéo interactif. Je m'intéresse à la notion d'immersion sonore ainsi qu'aux interfaces transparentes dans un contexte de vidéo-projetée impliquant donc une rupture spatiale entre le spectateur et l'espace virtuel.

Ainsi je souhaite poursuivre mes recherches dans deux directions. D'une part la création d'une illusion sonore: l'immersion dans une écoute réaliste au moyen d'un casque audio. D'autre part j'étudierai la translation directe des sens perceptifs dans le domaine digital par le biais de deux dispositifs d'augmentation : le Eye-Tracking et le Head-Tracking, autrement dit les outils permettant l'intégration transparente de l'ouïe et de la vue dans le système d'une oeuvre interactive.

CHAPITRE 2 : État de l'art des techniques

Dans cette partie nous dressons un état de l'art des techniques d'immersion sonore pour une diffusion avec un casque audio, ainsi que celles des dispositifs de suivi des mouvements de la tête et des yeux, ou plus communément, Head-Tracking et Eye-Tracking.

I / L'IMMERSION SONORE ET LE CASQUE AUDIO

Tout d'abord, revenons sur la question de la diffusion sonore au casque. Actuellement nous avons peu de possibilités dans ce domaine : on peut diffuser le même son sur deux canaux, diffusion monophonique, ou utiliser les deux canaux de manière différente, diffusion stéréophonique. Ces deux techniques présentent néanmoins le même problème dans le rapport du son à l'image : elles ne permettent aucune externalisation des sources. C'est-à-dire que tous les sons entendus paraissent provenir de l'intérieur de la tête de l'auditeur. Si pour la musique c'est devenu un standard, il est difficile d'être en immersion dans l'univers sonore d'un contenu projeté devant nous quand celui-ci est contraint aux limites de notre crâne. Nous allons donc avoir recours à la technique du binaural. Il s'agit d'une illusion sonore basée sur les principes d'une écoute naturelle, c'est-à-dire en champ libre, elle permet une externalisation des sources. En effet, le système auditif a la capacité de localiser la provenance d'un son grâce à l'analyse de ce qui est perçu par chaque oreille, et redonner ces mêmes informations au cerveau par le biais d'un casque lui fait croire que la source est extérieure à sa tête.

A / La localisation des sons par le système auditif.

Pour localiser une source sur le plan latéral, le cerveau se sert des indices interauraux. L'ITD (*inter-aural time difference*) correspond au laps de temps écoulé entre la réception du son par chacune des deux oreilles espacées d'environ 17 cm. L'ILD (*inter-aural level difference*) désigne la différence des niveaux ou intensités sonores perçus par les deux oreilles, liée au filtrage effectué par la masse crânienne. Ces deux critères sont complémentaires: l'ITD permettant de localiser par analyse de la différence de phase est plus adapté aux moyennes fréquences, notamment dans le cas où la demi-longueur d'onde est supérieure à l'écart entre les deux oreilles (2 kHz environ). Dans le cas des hautes fréquences, plusieurs fronts d'ondes peuvent séparer les deux oreilles et créer une confusion dans le recalage temporel. C'est alors l'ILD qui est utilisé pour localiser une source.

Reste alors le problème de l'azimut. En effet, si l'ILD et l'ITD permettent de localiser l'origine d'une source sonore sur le plan horizontal, les indices interauraux ne donnent aucune information sur l'élévation de la source ni sur sa situation à l'avant ou l'arrière de la tête. Autrement dit, en supposant horizontal l'axe des deux oreilles, il existe une infinité de points avec les mêmes attributs d'ITD et d'ILD : ceux-ci forment un cercle de rayon orthogonal à l'axe interaural, et donc sont dans un plan vertical. (voir figure 2.1.1)

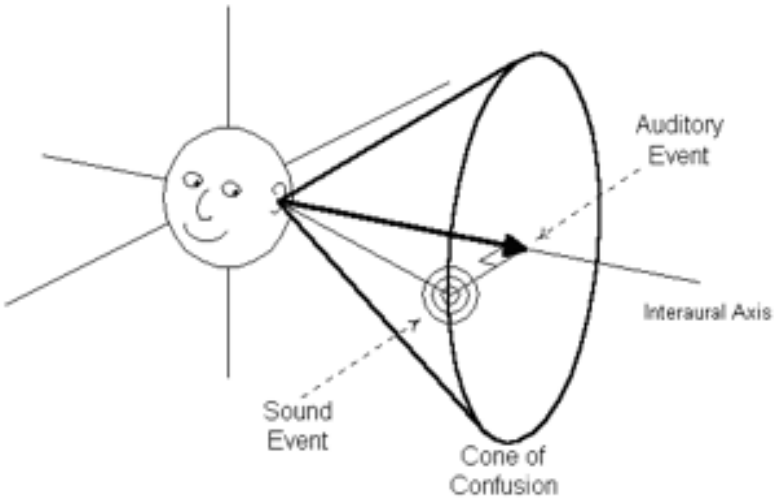

Figure 2.1.1 : Illustration du cône de confusion

http://music.miami.edu/programs/mue/Research/jwest/Chap_2/Chap_2_Spatial_Hearing.html

C'est ce qu'on appelle le cône de confusion (bien que ce soit techniquement un hyperboloïde). L'oreille utilise alors des indices monauraux pour déterminer la provenance d'un son.

Ceux-ci sont liés à l'anatomie de chaque personne, à la fois la forme de son oreille externe, des cavités formées par la *concha* et le pavillon, mais aussi de son buste ou son épaule. La géométrie précise et unique de chaque être humain génère un certain nombre de réflexions et filtrages fréquentiels que son cerveau comprend alors comme des informations de localisation dans le plan vertical. Plus précisément, le filtrage en peigne provoqué par des réflexions dans des zones très proches du conduit auditif crée un ensemble de creux et de pics dans le spectre du son

perçu qui sont spécifiques à la direction de celui-ci.

L'ensemble formé par ces indices interauraux et monauraux est appelé HRTF[38]. Il s'agit de la réponse fréquentielle et temporelle de la partie externe du système auditif, en tout point d'une sphère entourant la tête, et ce pour chaque oreille. Bien que par définition spécifiques à chaque personne, les HRTFs présentent des traits communs. Il est possible d'exploiter ces informations pour reconstruire l'effet de l'anatomie humaine sur une source sonore. On peut ainsi créer avec un casque audio l'illusion d'une écoute réaliste, où les sons peuvent provenir de partout.

B / Les techniques du son en binaural: le Binaural Natif et le Binaural de Synthèse

On peut dès lors imaginer deux façons seulement de créer du son en binaural : soit intégrer les HRTFs directement à la prise de son, c'est ce qu'on appelle Binaural Natif, soit filtrer des objets sonores avec des HRTFs préalablement enregistrées, c'est ce qu'on appelle Binaural de Synthèse.

Le Binaural Natif consiste à placer des microphones de très petite taille dans un environnement permettant d'intégrer la HRTF directement dans la prise de son. Il s'agit souvent de DPA 4060, microphones omnidirectionnels reconnus pour leur sensibilité et leur fidélité de prise de son[39]. Le preneur de son peut les porter dans ses propres oreilles, et donc apposer sa propre HRTF, ou bien les placer dans une tête artificielle. Cette deuxième méthode nécessite un équipement plus imposant et peut être moins compatible avec certaines conditions d'enregistrement, mais permet de se soustraire des problèmes de bruits internes et aussi à l'ingénieur du son d'écouter ce qu'il enregistre. Qui plus est, ces têtes artificielles présentent des caractéristiques neutres, moyennes de mesures sur de nombreux sujets[40].

---

[38] Head Related Transfer Function, ou fonction de transfère relative à la tête.

[39] Ce sont des microphones « cravate » de référence pour la prise de son au cinéma.

[40] « The size is based on the average of about 5000 males and females from the US Air Force. » kemar.us

Figure 2.1.2 : Têtes artificielles.
A gauche Neumann KU100 (http://www.neumann.com/?lang=fr&id=current_microphones&cid=ku100_description)
A droite KEMAR (http://kemar.us/)

Le Binaural de Synthèse est l'application d'HRTFs préenregistrées à une source monophonique ou polyphonique par convolution, permettant de choisir bien après l'enregistrement la place qu'occupera la source dans l'espace sonore. Ces HRTFs sont les fonctions de transfert établies après l'enregistrement d'un signal de référence (typiquement un Sweep[41]) diffusé en tout point d'une sphère entourant une tête portant des microphones dans une chambre anéchoïque[42]. Comme expliqué précédemment, l'HRTF de chaque être humain est différente, mais les similitudes entre HRTF permettent à un auditeur de percevoir l'externalisation voulue avec le filtrage destiné à une autre paire d'oreilles. Des bibliothèques de HRTF ont été dressées par la MSH Paris Nord, le laboratoire CIPIC Interface en Californie, ou encore l'IRCAM pour leur outil de spatialisation, permettant aux utilisateurs d'essayer plusieurs profils de HRTF et trouver celui qui leur correspond le mieux. Un étudiant de l'université d'Aarhus au Danemark a même développé un script Java permettant de sélectionner entre 43 HRTF de la bibliothèque CIPIC en fonction d'une vingtaine de longueurs que l'utilisateur

[41] Un Sweep est un balayage fréquentiel monophonique à durée pré-définie. C'est un outil commun pour mesurer la réponse fréquentielle d'un système.

[42] Un pièce traité de façon acoustique pour empêcher toute réflexion sonore et dans laquelle aucun son extérieur ne parvient.

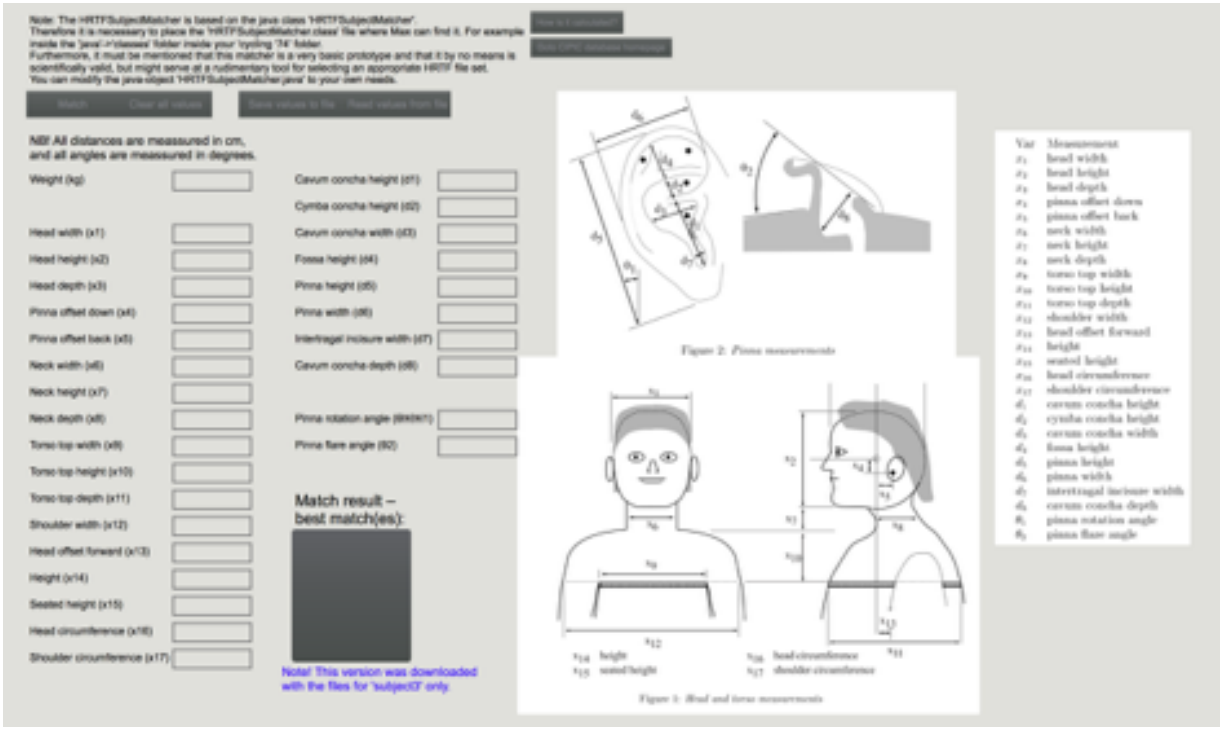

Figure 2.1.3 : Logiciel de Jakob Hougaard Andersen pour la sélection de HRTF à partir de mesures sur le corps de l'utilisateur

mesure sur son propre corps[43].(voir figure 2.1.3) L'utilisateur est alors libre d'implémenter ses propres calculs de convolution dans des environnements de programmation graphique comme PureData ou Max, d'utiliser des objets préexistants souvent mis à disposition avec les bibliothèques[44], ou bien d'utiliser des plug-ins audio[45].

[43] ANDERSEN, Jakob Hougaard. "Realistic mediation of virtual sound sources" Script JAVA pour l'objet Max « FFT-Based Binaural Panner »

[44] FFT Based Binaural Panner, Ambit-Binaural etc

[45] Flux SPAT v3, NoiseMakers BINAURALIZER

C / Etat de l'art du son en binaural

On trouve actuellement sur internet plusieurs types de contenus binauraux que l'on peut séparer en trois catégories. Tout d'abord il y a les vidéos démonstrations, probablement la catégorie la plus fournie, dont l'objectif est de présenter à un auditeur néophyte la « nouveauté » du son en binaural. La vidéo de présentation de The Verge[46] offre une écoute immersive des rues et métros de New York, accompagné d'un commentaire expliquant les principes de l'écoute binaurale. La prise de son est effectuée avec l'étrange *FreeSpace* de 3Dio, qui a pour objectif d'avoir le même rendu qu'une tête artificielle, sans l'encombrement. Si l'impression d'ouverture est bien réelle quand on passe du micro cravate de présentateur à la paire d'oreilles artificielle, l'externalisation est toutefois limitée. Il me semble que les extérieurs d'une ville très bruyante et les intérieurs étouffants de son réseau métropolitain offre un magma sonore indéfinissable dont les sources sonores sont trop diffuses et nombreuses pour donner une réelle illusion de son 3d. Cette vidéo est représentative d'une tendance facilement appréciable sur la toile : une abondance de démonstrations en binaural natif, où l'effet de suggestion d'une voix nous indiquant ce qu'on devrait entendre est plus fort que la réelle sensation d'immersion et d'externalisation voulue.

L'expérience la plus connue sur le web est sans doute le *Virtual Barber Shop* qui fait sa première apparition en 2007. L'auditeur est placé sur un siège de barbier et attend qu'on vienne lui offrir une coupe de cheveux virtuelle. Deux acteurs au grossier accent italien entrent alors en scène et proposent de faire entendre des déplacements sonores, au plus loin (le barbier marche en parlant du fond de la pièce) comme au plus proche (le rasoir électrique semble être appliqué à la tête de l'auditeur). Ici on a une réelle efficacité du binaural, lié à un environnement de synthèse parfaitement contrôlé. Aucun bruit parasite, ni même de réalisme : les acteurs n'ont aucune présence sonore hormis leurs voix, il n'y a donc ni bruit de pas, ni bruissement d'habits. La spatialisation binaurale est faite exclusivement par synthèse par la société Qsounds, comme démonstration de leur moteur 3Dsound. L'avantage de cette production parfaitement irréaliste est le contrôle de l'espace acoustique qui permet de mettre en jeu tous les paramètres liés à une écoute naturelle, et notamment les effets d'éloignement.

La deuxième catégorie adopte un point de vue complètement différent. Il n'est plus question de montrer l'effet magique d'une captation ou d'un traitement binaural, mais de

[46] Journal en ligne à la croisée des nouvelles technologie, de la science et des arts.

partager un point d'écoute, un environnement choisi. C'est ce que Dallas Simpson appelle ses *Binaural Location Performances*, sorte de Land-Art sonore, où l'artiste est soit dans une contemplation passive (Fireworks, Northampton Balloon Festival), une déambulation (Bruges Street Improvisation), ou une improvisation sonore avec les objets trouvés autour de lui (The Adoration of Willow, River Trent). La recherche artistique est dans la performance de Simpson telle qu'il la perçoit lui même, et le binaural natif avec les microphones dans ses propres oreilles et l'unique moyen d'être dans son exacte réception de l'oeuvre. Il décrit son travail comme étant proche d'une improvisation de jazz, où la mise en vibration des éléments sonores dans les espaces acoustiques choisis forme les mélodies de la nature qu'il propose aux auditeurs[47]. Ici le rendu binaural est peu impressionnant, mais l'intérêt est dans la découverte sonore et dans la contemplation de paysages nouveaux.

Finalement, parlons de la présence grandissante du binaural dans le monde de la création radiophonique. J'invite le lecteur à prendre connaissance du mémoire de Léa Chevrier, ENSLL promo 2015 « Expérimentation des techniques binaurales appliquées au documentaire radiophonique » qui détaille l'utilisation à la fois de la prise de son binaural et de ce mode de diffusion au sein même de l'écriture radiophonique. Il peut aussi explorer le site de Radio France sur le Son Multicanal et Binaural, NouvOson, dont les différents contenus pédagogiques (sur la binaural) comme artistiques nous rapprochent d'une utilisation du binaurale qui n'est plus dans la démonstration, mais dans la mise à profit des possibilités offertes par cette technologie pour raconter de nouvelles sensations sonores.

De toutes les productions disponibles sur internet, c'est cette dernière que je trouve la plus intéressante, car il n'est plus question simplement d'impressionner l'auditeur avec des illusions sonores, mais de mettre à profit une nouvelle lisibilité des espaces sonores et une sensation d'enveloppement et d'immersion, d'inclusion dans un espace, qui n'est possible qu'avec ce moyen de diffusion.

[47] **SIMPSON Dallas,** Improvisational Binaural Sound Art : The Foundations of Location Performance, dans Rubberneck n ° 24, 1997.

D / Les limites du son en binaural

Notre consommation audiovisuelle se tourne à la fois vers l'immersion sonore et vers l'isolement de l'auditeur, ce qui devrait favoriser le développement de contenu en binaural. Pourtant nombreux sont les sceptiques à propos de cette forme de diffusion. En effet, que ce soit du binaural natif ou de synthèse, deux problèmes entachent la perception du réel effet d'espace tridimensionnel que recherchent les créateurs sonores. Tout d'abord, comme nous le disions plus haut, l'utilisation d'HRTF neutre ou choisi dans une bibliothèque reste approximative et les résultats ne sont pas constants suivant les auditeurs. On peut aussi citer la thèse de Pierre Guillon[48] qui présente deux méthodes pour produire des HRTFSs individualisées :

« La première solution développée vise à adapter, pour un nouvel auditeur, les HRTFs d'un autre individu. Les transformations à appliquer à ce jeu de HRTFs […] sont contrôlées par le résultat d'une comparaison morphologique entre les pavillons des deux sujets. […] La seconde solution permet de reconstruire les HRTFs d'un nouvel auditeur pour une direction quelconque de l'espace, à partir d'un nombre réduit de HRTFs individuelles mesurées. L'analyse d'une base de données permet de dégager des prototypes, utilisés comme informations dans le processus de reconstruction. »

Le développement d'une méthodologie pour simplifier l'acquisition d'HRTFs individuelles est aussi l'un des sujets du groupe de recherche BiLi[49] qui réunit les plus grands acteurs technologiques audiovisuels en France. L'un des objectifs est de proposer des « processeurs d'écoute binaurale en temps réel »[50], qui devront sans doute intégrer l'utilisation d'HRTF individualisées.

De façon plus importante, le contenu actuellement disponible, ainsi que les techniques de binauralisation évoquées plus haut ne prennent pas en compte un des facteurs les plus importants de la localisation dans l'audition humaine : les indices dynamiques. En effet, dans une situation d'écoute naturelle un auditeur effectue inconsciemment une multitude de petits

---

[48] **GUILLON, Pierre**, Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF, Thèse pour obtenir le grade de Docteur de l'Université du Maine, 2009.

[49] BiLi ou Binaural Listening est un groupe de recherche sur la démocratisation du son en binaural réunissant entre autres Orange Labs, France TV, Radio France, l'IRCAM.

[50] http://www.bili-project.org/le-projet/

mouvements de la tête, modifiant alors dynamiquement la perception de l'ILD, de l'ITD et des Indices spectraux. Il écarte ainsi les confusions possibles entre avant et arrière, et précise les informations d'élévation. A la fin des années 30, le chercheur Hans Wallach publie une série d'articles détaillant ses expériences sur l'augmentation de la précision de localisation en fonction de la perception de mouvements. On en retient que l'appréciation de l'élévation peut être générée uniquement grâce à des modifications de timbre simulant la variation des Indices spectraux, ce qui explique l'efficacité de certaines productions binaurales par rapport à d'autres. Plusieurs autres expériences menées par Thurlow et Runge[51], Pollack et Rose[52], Fisher et Freedman[53], ainsi que Perrett et Noble[54] (détaillées dans le mémoire de Louis Anglionin) confirment l'idée qu'une spatialisation binaurale précise et externalisée repose sur une perception d'indices dynamiques.

Si le Land-Art sonore de Dallas Simpson retient une sensation d'enveloppement par le son, l'externalisation est limité par le manque d'objets dynamiques précisément localisés. C'était aussi le problème de la visite de New York proposée par The Verge : l'environnement avait beau comporter plusieurs éléments dynamiques, l'ensemble était un flou fréquentiel trop diffus pour permettre une perception spatiale précise. A contrario, c'est justement en cela que le Virtual Barber Shop trouve son efficacité par rapport à la majorité des productions disponibles sur internet : les éléments sonores sont constamment en mouvement dans un espace acoustique contrôlé et les indices spectraux varient dynamiquement comme dans une situation d'écoute naturelle. Or on ne peut pas contraindre toute écriture sonore à n'utiliser que des sources en mouvements sous prétexte de conserver l'illusion d'un espace tridimensionnel. En conclusion, l'intérêt du binaural ne semble pas être dans sa réalisation technique, et donc dans la création de contenu mettant en exergue ses prouesses, mais dans les possibilités d'immersion et donc d'implication de l'auditeur dans un univers avec lequel il va pouvoir interagir, que ce soit de manière active ou dans une simple contemplation. Il faut donc que l'illusion soit invisible ou du moins que le spectateur puisse l'oublier. Si les indices dynamiques ne viennent pas de la scène

---

[51] **THURLOW, W. R., RUNGE, P. S.** Effect of induced head movements on localization of direction of sounds, Journal of the Acoustical Society of America n°42, 1967.

[52] **POLLACK, I., ROSE, M.** Effect of head movement on the localization of sounds in the equatorial plane, Perception & Psychophysics n°2, 1967.

[53] **FISHER, H. G., FREEDMAN, S. J.** The role of the pinna in auditory localization, The Journal of Auditory Research n°8, 1968

[54] **PERRETT, S., NOBLE, W.** The contribution of head motion cues to localization of low-pass noise, Perception & Psychophysics n°59, 1997.

sonore, il va donc falloir qu'ils viennent de l'auditeur. Il est dès lors inévitable d'étudier les possibilités de suivi dynamique des mouvements de la tête pour une synthèse binaurale adaptative. De plus, il est alors nécessaire d'intégrer la possibilité de faire varier en temps réel la rotation d'une scène sonore dans le cahier des charges des outils logiciels à utiliser pour une synthèse binaurale.

E / Expériences d'écoute en son binaural

Munis de ces bases sur les différentes techniques de spatialisation en binaural ainsi que leurs limitations respectives, j'ai effectué plusieurs expériences subjectives pour déterminer dans quelles situations leurs points forts seraient les mieux exploités.

Tout d'abord, je voudrais revenir sur mes différentes écoutes de Binaural Natif. Nous avons établi que cette technique n'est pas compatible avec un Head-Tracking, et donc que les indices dynamiques doivent venir du contenu lui-même. Or cela implique des sons qui se développent dans la durée, difficiles donc à mixer avec du son en binaural de synthèse avec suivi des mouvements de la tête. Il faut alors exploiter les possibilités des enregistrements avec tête artificielle dans d'autres situations, basées sur les ITD et ILD ainsi que les indices spectraux, mettant donc l'accent sur le plan azimutal. Après avoir écouté le documentaire sonore de Thomas Claire[55], il semble évident qu'il faut mettre à profit le passif d'une construction empirique de l'espace acoustique chez l'auditeur, ainsi que l'impact d'une image projetée. C'est-à-dire, d'une part, qu'un son que l'auditeur a l'habitude d'entendre en élévation sera plus facilement situé en hauteur (un avion ou un oiseau est naturellement entendu au-dessus de la tête), et il en va de même pour les sons liés au plan horizontal (un démarrage de voiture par exemple). Et d'autre part, comme la longue expérience de l'effet ventriloque au cinéma le prouve, que les indices visuels sont des outils très persuasifs en termes de localisation. Cela implique par exemple que le son d'ouverture d'une porte, d'autant plus lorsqu'il est attaché à un indice visuel, est externalisé avec aisance. L'utilisation d'une tête artificielle pourra donc être préconisée pour des sons de courtes durées visant à rendre le naturel d'un environnement.

Une autre utilisation préconisée pour le Binaural Natif concerne les sources sonores très proches de la tête. En effet les systèmes de synthèse binaurale sont basés sur des HRTFs enregistrés avec des *stimuli* diffusés typiquement à 1m de la tête[56] en chambre anéchoïque. Le placement synthétique de sources très proche de la tête est souvent limité, et dans mon expérience le rendu est peu convaincant. Si le traitement annule l'audition de l'oreille gauche pour un son synthétisé très proche de l'oreille droite, il ne parvient pas à recréer la sensation de proximité (notamment dans la perception des plosives , ou de la dynamique instantanée de

---

[55] Disponible en écoute sur le site de BiLi : http://www.bili-project.org/category/ecouter-du-binaural/

[56] **ALGAZI V.R et al.** THE CIPIC HRTF DATABASE, UC Davis, Californie, 2001.

manière générale) que l'on a avec un enregistrement effectué avec une tête artificielle.

Suite à mes essais, l'outil qui me semble le plus adapté à une spatialisation dynamique en binaural est l'outil de spatialisation de l'IRCAM, commercialisé par FLUX, le SPAT v3. (voir figure 2.1.4) Celui-ci permet de placer 8 enceintes virtuelles dans un espace à 360° à n'importe quelle hauteur par rapport au point d'écoute. Le plug-in intègre une bibliothèque de plusieurs dizaines de HRTF et un outil de monitoring qui effectue la convolution nécessaire à la synthèse binaurale en temps réel. Comme tous les paramètres de placement des sources sont automatisables, il est aisé de développer un patch Max permettant à la fois de contrôler leurs positions choisies par rapport à un plan de projection, et de moduler ces positions en fonction du suivi des mouvements de la tête[57]. Cet outil convient donc parfaitement au placement de sources sonores discrètes, mais qu'en est-il de l'ambiance sonore ?

Figure 2.1.4 : Interface graphique du SPAT v3 de FLUX, spatialisation d'une source sonore dans une configuration d'écoute en binaural

[57] Le patch sera détaillé dans la 3ème partie

Pour mes derniers essais j'ai cherché comment synthétiser le liant, la matière de fond qui va immerger le spectateur dans une scène sonore et visuelle qui lui serait proposée. Il s'agit de reconstituer un environnement sonore le plus proche possible d'un contexte d'écoute réaliste. L'idée est de dépasser une restitution binaurale d'un système de diffusion multicanal classique du cinéma comme le 5.1 ou même le 22.2, qui, au mieux, donne l'impression d'être dans une salle de cinéma. De plus, quelle matière serait alors diffusée dans ces enceintes virtuelles? Quel système de prise de son permettrait une restitution en trois dimensions? La réponse était toute proche et accompagnait généralement les bibliothèques de HRTF que j'avais à ma disposition : l'Ambisonique, ou même l'HOA[58]. L'Ambisonique est une technologie développée en Angleterre au début des années 70, basée sur l'idée que l'on peut décrire l'intégralité d'un champ sonore en un point avec des sommes pondérées de signaux issus de quatre capsules cardioïdes coïncidentes montées en tétraèdre régulier. (voir figure 2.1.5) Ces signaux sont encodés en ce qu'on appelle le B-Format et décrivent alors non plus des canaux de diffusion, mais la globalité d'un environnement, qui est ensuite décodable et adaptable à n'importe quel système de diffusion.

Figure 2.1.5 : Microphone tétraédrique SoundField

L'HOA (ou Higher Order Ambisonics) est une extension de l'Ambisonique, dans le sens où il s'agit de décomposer l'espace en une série d'harmoniques sphériques, dont l'ordre 1 correspond au B-Format. L'enregistrement se fait alors avec beaucoup plus de capsules (il faut 16 canaux dès l'ordre 3), mais le concept reste le même.

Pour davantage de détails techniques sur les systèmes Ambisoniques, le lecteur est invité à lire le mémoire de Clément Cerles, ENSLL promotion 2015.[59]

Dans la présentation de Clément Cerles on trouve une liste de plugins Open Source pour la traitement de l'ambisonique. L' outil de binauralisation de la suite AmbiX, développée par Matthias Kronlachner et récompensée au « AES Students

[58] La Fondation MSH-Paris Nord met à disposition avec sa banque de HRTF un ensemble d'outils pour le traitement et la spatialisation ambisonic, développés par Julien Colafrancesco, Pierre Guillot et Eliott Paris pour le laboratoire CICM de Paris 8.

CERLES, Clément, Caractérisation objective et subjective d'une chaine de traitement HOA, Mémoire sous la direction de Frank Gillardeaux et Jérôme Daniel, ENS Louis Lumière, 2015.

Design Competition 136th AES Convention Berlin » a attiré mon attention. En effet, il intègre des réponses impulsionnelles pour de nombreuses dispositions de haut-parleurs, notamment des cubes et autres formes prenant en compte l'azimut, et son outil de rotation ambisonic qui permet très simplement d'implémenter un suivi des mouvements de la tête. La scène ambisonic est représentée comme une sphère que l'on peut faire tourner selon nos désirs. Sa simplicité d'utilisation et sa compatibilité avec n'importe quel enregistrement, B-format ou autre, en font un outil flexible et puissant. (l'utilisation de ces outils sera détaillée dans le Chapitre 3)

Au cours de mes tests j'ai pu apprécier la qualité d'englobement et d'externalisation qu'offre l'Ambisonics sur un environnement enregistré en champ diffus, mais aussi l'intérêt d'encoder une réverbération 5.1 avec les objets AmbiX et de pouvoir obtenir une réponse de salle évolutive en fonction des mouvements de ma tête, ce qui permet de mettre à profit des effets de *reverb* pour des objets monophoniques, peut-être plus intéressants que ceux proposés dans le SPAT.

Pour résumer, dans le cadre d'une utilisation de son en binaural pour une vidéo projetée, on se servirait du SPAT pour synthétiser des objets monophoniques précisément spatialisés et les déplacer dynamiquement dans l'espace en fonction des interactions avec l'utilisateur. Si l'environnement dans lequel se situe la scène le nécessitait, on pourrait placer ces objets dans des modules de réverbération encodés avec les objets AmbiX. L'ambiance générale de la scène serait enregistrée en B-Format et diffusée de manière dynamique en suivant les mouvements de la tête de l'auditeur, formant un canevas d'immersion. Finalement, pour des objets sonores courts (de l'ordre de la seconde) qui ne sont pas affectés par le suivi des mouvements de la tête, ainsi que des sources très proches de l'auditeur, les enregistrements seraient faits avec une tête artificielle.

II / ETUDE DES CAPTEURS

Nous avons observé qu'un traitement immersif et réaliste du son au moyen d'une écoute au casque était possible grâce à l'utilisation des techniques binaurales. Nous avons présenté les défauts de ces systèmes et établi que l'utilisation de suivi des mouvements de la tête, ou Head-Tracking, pouvait en résoudre un grand nombre. Ce dispositif correspond à la translation direct du sens de l'Ouïe dans le domaine digital. Nous allons aussi étudier les possibilités de translations pour la vue, au moyen de dispositifs de suivi des mouvements des yeux, ou Eye-Tracking.

Cette partie constituant un état de l'art des techniques, nous mettrons à profit les résultats des différentes recherches menées au préalable dans ces domaines. Notamment en ce qui concerne le Head-Tracking, nous nous servirons des résultats du mémoire de fin d'études de Louis Anglionin[60].

A / Pour le suivi de la tête, ou Head-Tracking

Dans son mémoire, Anglionin parle de l'utilisation de trois types de capteurs pour effectuer un suivi des mouvements de la tête : le détournement d'appareils grand-public prévus pour donner des informations d'accélération suivant plusieurs axes, une analyse temps réel d'un flux vidéo de l'utilisateur, une centrale inertielle ou IMU.

Concernant le détournement d'objets industriels, Anglionin arrive à des résultats intéressants, mais peu compatibles avec l'utilisation que nous avons prévu d'en faire. La souris gyroscopique délivre les informations nécessaires à un traitement binaural adaptatif dans le plan des oreilles de l'auditeur, de l'ordre de la rotation de la tête autour du cou donc, mais ne permet pas de travailler avec l'élévation et donc l'inclinaison du plan de la tête. De plus, les informations reçues pendant les tests effectués arrivaient de la souris toutes les 40ms, valeur supérieure à la recommandation de Yairi pour la plus petite latence perçue[61]. La latence désigne

---

[60] **ANGLIONIN Louis**, De nouveaux outils pour un dispositif de suivi des mouvements de la tête en spatialisation binaurale, sous la direction d'Alan Blum et Jason Cook, Mémoire de l'ENS Louis Lumière, 2014.

[61] **YAIRI, S et al**. Estimation of detection threshold of system latency of virtual auditory display, Applied Acoustics n°68, 2007.

ici le retard du système de captation sur les réels mouvements de l'utilisateur. Bien entendu cette valeur dépend du modèle de souris choisi, mais elle n'est pas donnée par les constructeurs et rend cette technique trop imprévisible.

L'utilisation de la WiiMote s'annonce encore plus compliquée : il n'y a pas de problème de latence, car elle est connectée en Bluetooth et la durée entre deux informations reçues est seulement de 10,2ms, mais les accéléromètres intégrés sont trop peu précis pour donner un suivi satisfaisant. Anglionin a finalement utilisé la détection par triangulation d'une LED infrarouge pour déterminer la position relative de la Wiimote dans l'espace et arriver à un suivi précis. Or cette technique est limitée par le « champ de vision » de l'appareil, qui restreint donc les mouvements potentiels de l'auditeur au maximum à une demi-sphère, et ne donne pas d'information d'inclinaison, ou Yaw.

Finalement est étudiée la possibilité d'utiliser les accéléromètres intégrés aux smartphones dont la qualité et le faible niveau de bruit (*de l'ordre de 0,01°)* laissent présager d'un avantage notable sur les précédents candidats. Finalement ici c'est la stabilité du flux d'informations qui n'est pas compatible avec un suivi continu des mouvements de la tête : le téléphone utilisé lors de l'expérience, iPhone 4s d'Apple, se connecte via wifi à l'ordinateur récepteur des données, et la connexion est souvent interrompue.

Si certains des problèmes liés à ces capteurs potentiels pouvaient être résolus de manière informatique (par exemple l'utilisation de la connexion Bluetooth du smartphone), ils sont tous les trois soumis à la question de la transmodalité. Le spectateur est obligé de « faire correspondre deux espaces différents , celui de la main et celui de la tête »[62] et si cette accommodation pouvait être praticable à la longue, elle me semble peu adaptée au cadre d'une installation d'art vidéo interactif.

Anglionin a aussi étudié la possibilité d'effectuer un suivi de la tête via une analyse d'images successives, un flux vidéo issu par exemple d'une webcam. Si ses expériences se sont soldées par un échec, elles permettent tout de même d'étudier les différentes possibilités de ce type de capteur. On détermine généralement deux manières de faire du suivi video en temps réel. La première méthode est celle dite du « Blob Tracking » qui consiste à déterminer une zone de l'image qui se différencie de son environnement par sa forme/couleur/luminosité et de

[62] **ANGLIONIN Louis**, De nouveaux outils pour un dispositif de suivi des mouvements de la tête en spatialisation binaurale, sous la direction d'Alan Blum et Jason Cook, Mémoire de l'ENS Louis Lumière, 2014.

calculer le déplacement de cette zone par différences d'images successives. En plaçant plusieurs marqueurs sur un visage on définit une perception par la caméra des angles les séparant, qui évolue donc en fonction des mouvements de la tête. Ces données peuvent alors être utilisées pour définir le Pitch, le Yaw, et le Roll de l'utilisateur, respectivement les rotations autour des axes *x*, *y* et *z*. (voir figure 2.2.1)

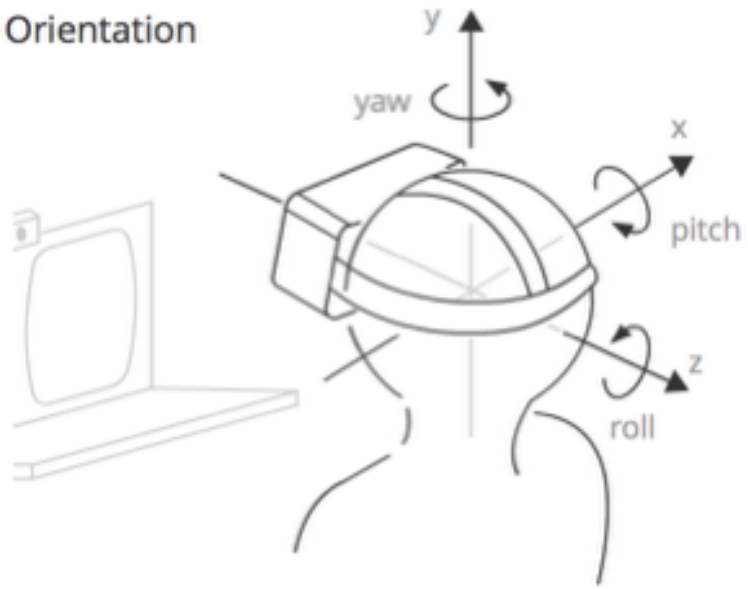

Figure 2.2.1 : Illustration des axes de rotation de la tête.
Site du Mozilla Developer Network, page du WebVR concepts :
https://developer.mozilla.org/en-US/docs/Web/API/WebVR_API/WebVR_concepts

La seconde méthode est d'utiliser un logiciel de reconnaissance faciale, comme FaceOSC ou FaceAPI, qui détecte automatiquement les visages et transmet notamment les valeurs d'angle de la tête par OSC. Nous pourrions rentrer dans le détail de ces deux techniques, mais les défauts soulevés par Anglionin, inhérents à leur fonctionnement, empêche leur utilisation au sein de notre installation.

Il soulève tout d'abord le problème des conditions physiques de la captation, lié à la qualité d'une caméra discrète et peu onéreuse : la luminosité ambiante doit être élevée pour effectuer un suivi colorimétrique précis, ce qui pose un problème dans le cas d'une installation avec une projection. On pourrait imaginer l'utilisation d'une caméra avec une plus haute définition, un capteur plus sensible et un framerate[63] plus élevé, mais la latence de traitement en serait multipliée. Une variante serait d'opter pour des LEDs placées sur le visage, s'apparentant

[63] Nombre d'images par seconde pouvant être enregistrées par la caméra.

alors à des sources lumineuses blanches. Il n'y a alors plus de problème de luminosité, mais les sources identifiées ayant la même taille et la même signature spectrale, le tracking des Blobs est instable et compromis.

Dans le cas des logiciels de reconnaissance faciale, il semblerait qu'il n'y ait pas de problème de luminosité. L'unique frein à notre utilisation de cet outil est sa limite en ce qui concerne l'angle de l'utilisateur par rapport à la caméra. Dans un article écrit au Laboratoire de l'Informatique Musicale de Milan[64] trois étudiants mettent au point un système de spatialisation binaurale avec suivi de mouvements de la tête, basé sur la reconnaissance faciale FaceAPI de *Seeing Machines*, et sur la banque de HRIR CIPIC[65]. Les résultats de leurs travaux corroborent les conclusions d'Anglionin : le suivi ne fonctionne que pour des variations d'azimut et d'élévation limitées (respectivement ± 90° et -30°/+60°) et lorsque le visage quitte ces limites, plus aucune donnée n'est envoyée. Cette méthode n'est donc pas compatible avec une liberté de mouvement à 360°, ce qui est une des contraintes que nous souhaitons lever.

Pour conclure ces réflexions, si nous devions utiliser une caméra pour suivre les mouvements de la tête, nous imaginerions un capteur sans filtre infra-rouge, et cinq émetteurs infrarouges de longueurs d'onde différentes fixées sur le casque audio. Si on ne se limite pas au visage pour le placement des sources lumineuses, et qu'elles sont facilement identifiables spectralement, il devrait être possible d'effectuer un suivi précis et continu pour tout mouvement de la tête.

B / Le cas particulier de l'IMU ou centrale inertielle

La piste la plus prometteuse des recherches d'Anglionin, et celle qui visiblement fait l'unanimité dans le monde du head-tracking pour l'audio binaural, est l'utilisation d'une centrale inertielle. Aussi appelée IMU, pour Inertiel Measurement Unit, ce dispositif a pour but de délivrer des informations de rotation autour de trois axes (le Yaw, le Pitch et le Roll) en continu. Le modèle qui a été retenu pour les essais est le 9DOH Razor IMU vendu par Sparkfun (voir figure 2.2.2), pour son utilisation de trois capteurs précis et sa compatibilité avec le protocole Arduino, ainsi que la communauté Open Source proposant des outils

[64] **LUDOVICO Luca A et al.** HEAD IN SPACE: A HEAD-TRACKING BASED BINAURAL SPATIALIZATION SYSTEM, dans le cadre de LIM - Laboratorio di Informatica Musicale, Universita degli Studi di Milano, Italy, 2010.

[65] **ALGAZI V.R et al.** The cipic hrtf database, IEEE Work- shop on Applications of Signal Processing to Audio and Acoustics, Octobre 2001.

informatiques qui nous permettent de récupérer les données qui nous intéressent.

Le 9-Degrees-Of-Freedom est équipé de trois capteurs : un accéléromètre tri-axial (ADXL345), un gyromètre tri-axial (ITG-3200) et un magnétomètre tri-axial (HMC5883L).

L'accéléromètre mesure l'accélération linéaire inertielle selon trois axes, « Ce type de capteur ne mesure pas l'accélération absolue, mais plutôt les variations de la gravité »[66]. Le capteur donne sa valeur maximale selon l'axe normal à l'horizon lorsqu'il est posé sur une surface horizontale, mais donne des zéros en cas de chute libre. Le circuit intégré ADXL345 propose une précision de variation d'inclinaison de 0,25°.

Le gyromètre mesure la vitesse angulaire par rapport à un référentiel fixe. L'ITG-3200 discrétise alors ces valeurs grâce à trois convertisseurs analogique-numérique 16-bit.

Un magnétomètre mesure l'intensité du champ magnétique perçu en un point donné. Le HMC5883L étant équipé de trois capteurs disposés de façon orthonormée permet d'évaluer la direction du champ magnétique dans l'espace avec une résolution de 5 milligauss.

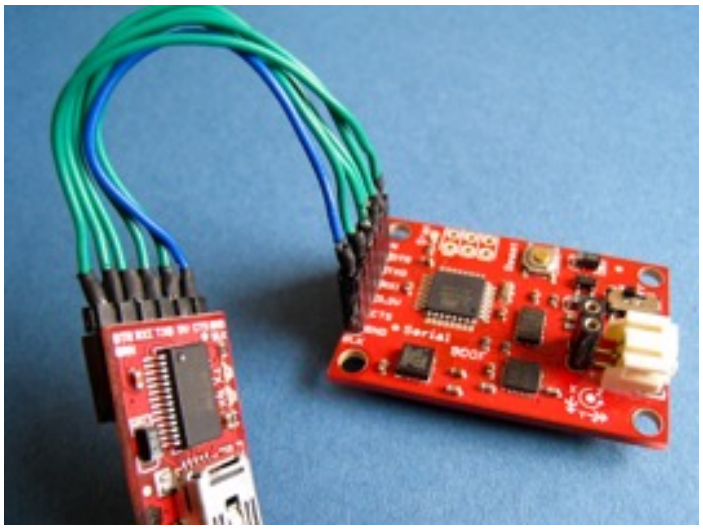

Figure 2.2.2 : 9DOH Razor IMU avec la carte FTDI permettant une connexion par USB.
Site du tutoriel de Peter Bartz indiquant comment mettre en place le suivi des mouvements : https://github.com/ptrbrtz/razor-9dof-ahrs/wiki/Tutorial

Les informations provenants des 9 capteurs sont ensuite traitées par le micro-contrôleur ATmega328 qui possède une sortie série permettant via une carte FTDI basic-breakout de

[66] **ANGLIONIN Louis**, De nouveaux outils pour un dispositif de suivi des mouvements de la tête en spatialisation binaurale, sous la direction d'Alan Blum et Jason Cook, Mémoire de l'ENS Louis Lumière, 2014.

récupérer des informations en continu par un port USB. Le micro-contrôleur est livré avec le Bootloader Arduino pré-installé, qui lui permet d'être reprogrammé postérieurement via un logiciel de codage, comme le logiciel Open Source Arduino. L'intérêt est de pouvoir facilement reprogrammer la carte pour que la sortie série fournisse directement des informations de Yaw, Pitch et Roll. C'est d'ailleurs la motivation derrière le tutoriel de Peter Bartz publié sur le site GitHub.com qui accompagne l'utilisateur novice de l'assemblage du tracteur hardware jusqu'à la calibration logicielle des capteurs. Le firmware permettant de récupérer les données de suivi de la tête sera disponible en annexe du mémoire. Le tutoriel indique comment se servir du logiciel de calcul graphique *Processing* pour observer une représentation 3D temps réel des informations provenant de l'IMU, celle-ci nous permet de vérifier simplement le fonctionnement de la carte.

Coûtant 70€, les circuits intégrés présents sur la carte sont relativement bon marché et présentent des faiblesses auxquelles nous pouvons pallier informatiquement.

La première étape est d'équilibrer les 3 axes de l'accéléromètre. Lorsque la carte est placée parfaitement à la verticale et immobile, le capteur donne en sortie la valeur maximum de l'effet de la gravitation terrestre selon l'axe qui est en direction du centre de la terre. On remarque que chacun des axes ne donne pas les mêmes valeurs maximales et minimales, ce qui déséquilibre la perception de l'angulation de la carte. Renseigner ces valeurs dans le firmware permet de rééquilibrer la lecture de l'accéléromètre.

Le second problème à régler est celui du bruit perçu par le gyromètre, au sens des valeurs qu'il transmet lorsqu'il est immobile et qu'il ne devrait que lire des zéros. Renseigner les valeurs de bruit moyen selon les trois axes permet de pondérer sa sortie série et d'avoir les valeurs réelles du gyromètre.

La calibration du magnétomètre revient à lui faire ignorer l'influence magnétique de tout ce qui est dans son entourage immédiat (ordinateur, aimants présents dans les casques/enceintes, etc.) au profit du champ magnétique terrestre, de manière à toujours avoir une lecture précise du nord magnétique selon les trois axes et donc de constituer un repère absolu.

Une fois calibrée la carte est prête à délivrer des informations de Yaw, Pitch et Roll stables et précises par le port série connecté en USB. Il suffira alors de récupérer ces valeurs dans le logiciel de traitement du signal et de les appliquer aux plug-ins de rotation ambisonique et de spatialisation binaurale évoqués précédemment.

A la suite de cette étude, je conclus que le système de centrale inertielle semble le plus à même de fournir des résultats satisfaisants dans l'optique d'une interaction sans interface ou à l'interface invisible dans le cadre de l'art vidéo interactif. Le système de programmation Arduino est léger et permet avec très peu de latence de recevoir des informations stables de rotation selon les trois axes. De plus, sa taille réduite lui permet d'être monté sur un casque audio et une fois calibré son fonctionnement est indépendant de l'utilisateur et des conditions de projection. Qui plus est, ce placement embarqué évite d'imposer à un utilisateur un système physique de captation (comme une caméra) dans son champ de vision, contrainte qui empêchait l'utilisation des autres méthodes dans le cadre de nos recherches.

Maintenant que la question du Head-Tracking est résolue, je m'intéresse aux techniques permettant d'effectuer un suivi des mouvements des yeux.

C / Pour le suivi des yeux, ou Eye-Tracking

L'histoire du Eye-Tracking remonte à la fin des années 1870, quand Louis Emile Javal observe que la lecture d'un texte ne se fait pas de manière linéaire, mais par une série de saccades et fixations.[67] Il se développe alors un intérêt pour l'analyse et la compréhension du circuit de l'oeil, notamment devant des tableaux, et avec celui-ci la mise au point de plusieurs techniques plus ou moins intrusives de Eye-Tracking.

La technique reconnue comme étant la plus précise est aussi la plus intrusive. Young et Sheena décrivent le placement d'une extrémité d'un fil métallique sur une lentille de contact. L'autre extrémité se déplace alors de la même façon que l'oeil dans un champ électromagnétique contrôlé dont les variations sont ensuite mesurées.

La méthode la plus utilisée est basée sur une captation vidéo des mouvements de l'oeil préalablement éclairé par une source infrarouge. Le reflet cornéen, ou première image de Purkinje, sert alors de repère fixe pour l'élaboration d'un vecteur dirigé par l'épicentre de la pupille, dont on peut tirer la direction du regard. On peut alors choisir de travailler avec des capteurs au niveau de l'écran qui sert au test, ou avec un équipement embarqué sur le sujet, chacune de ces méthodes présentant des avantages et inconvénients. Représentatif de la première catégorie, le RED 250 conçu par SensoMotoric Instruments annonce une précision de 0,4° et une calibration extrêmement rapide (< 3 sec.), mais opère uniquement à une distance de 60cm à 80cm adaptée pour une expérience de bureau. (voir figure 2.2.3) De plus, comme tout système basé sur une captation vidéo, il nécessite une certaine quantité de lumière pour fonctionner, ce qui risque de poserait un problème de la cadre d'une projection. Le principal avantage de ce système est qu'il n'est pas du tout intrusif. Qui plus est, le détecteur ne suivant pas les mouvements de tête du sujet, le système établit directement la zone regardée.

A contrario les systèmes embarqués, en plus d'être particulièrement intrusifs, nécessitent une forme de Head-Tracking supplémentaire pour atteindre les mêmes résultats. Plus précis que les systèmes déportés ils sont généralement bien plus onéreux, bien que plusieurs travaux de

[67] **JAVAL Émile**, Essai sur la physiologie de la lecture, Annales d'Oculistique 80, 1878.

recherche tentent de rendre ces technologies plus accessibles. [68] [69]

Figure 2.2.3 : Technologies de la société SMI, à gauche le RED250, à droite le Natural Gaze ETG.
http://www.eyetracking-glasses.com/

Plus récemment la grandissante qualité des webcams a permis le développement de techniques utilisant uniquement un tracking de pupilles.[70] Bien que peu précises, elles ont le mérite de ne nécessiter aucun hardware supplémentaire que l'ordinateur sur lequel fonctionne le logiciel qui exploite les données mais nécessite souvent beaucoup de lumière[71].

Finalement une solution unique subsiste pour effectuer une Eye-tracking sans nécessiter une luminosité ambiante élevée : l'électrooculographie, ou OEG. L'oeil humain peut être considéré comme un dipôle dont la borne positive est la cornée et la borne négative est la rétine. En plaçant des électrodes très sensibles sur le visage du spectateur, il est alors possible de mesurer les variations du potentiel électrique aux électrodes lors des déplacements de l'oeil. En effet, lorsque

[68] **MANTIUK Radoslaw et al.** Do-It-Yourself Eye Tracker: Low-Cost Pupil-Based Eye Tracker for Computer Graphics Applications, 2012.

[69] **KASSNER Moritz et PATERA William,** PUPIL Constructing the Space of Visual Attention, MIT, 2012.

[70] **SAN AUGUSTIN Javier et al.** Evaluation of a Low-Cost Open-Source Gaze Tracker, IT University of Copenhagen, 2009.

[71] La société xLabs propose une extension pour Google Chrome permettant de substituer à la souris les mouvements de la tête et des yeux. On trouve sur leur site la recommandation : « Make sure you are sitting at a desk, facing the window in daytime, or in a brightly-lit environment. » Assurer vous d'être assis à un bureau, face à la fenetre, ou dans un environnement lumineux.

le spectateur regarde sur sa droite, la cornée se rapproche de l'électrode placée à la droite de son oeil et en augmente le potentiel électrique, tandis qu'une électrode placée de l'autre côté verra son potentiel se réduire. La détection de la direction de l'oeil par rapport à la tête peut donc s'effectuer avec trois électrodes autour de celui-ci. En prenant en compte les informations d'un Head-Tracker il est ensuite possible de déterminer la direction du regard.

Figure 2.2.4 : Exemple d'ElectroOculographie.
http://www.metrovision.fr/mv-eo-notice-fr.html

Si le placement des électrodes sur le visage du spectateur nécessite une intrusion préalable dans son intimité, leur présence se fait ensuite rapidement oublier. Celui-ci peut alors profiter sans encombre de l'expérience de l'écran de projection devant lui. Je retiens donc cette proposition comme étant la plus apte à correspondre à une interface transparente : que ce soient les lunettes, ou un système avec caméra déportée devant être à moins de 80cm du visage, les autres techniques interfèrent visuellement avec l'expérience du spectateur. Les mouvements des yeux deviennent alors moteurs de l'interface et il serait beaucoup plus difficile pour le spectateur de ne pas intellectualiser sa relation à l'expérience.

III / QUELS LOGICIELS POUR TRAITER CES DONNES?

Il reste à décrire les logiciels qui traiteraient les données issues des capteurs dans le cadre d'une installation vidéo interactive. Il s'agit donc de répondre à deux cahiers des charges. L'un d'entre eux aura pour charge l'acquisition des données et le traitement du signal. L'autre devra être capable de traiter plusieurs flux vidéo en temps réel.

A / Pour l'acquisition des données et le traitement du signal

Nous avons besoin d'un logiciel permettant la réception en temps réel des données sérielles émises par l'IMU et l'Eye-Tracker, avec une grande liberté dans le traitement du signal, qui intègre l'utilisation de plug-ins audio pour le traitement binaural du son et qui sera capable de communiquer avec le logiciel de traitement de l'image.

Je me suis donc tourné vers le logiciel Max7, distribué par Cycing '74. Max est un langage de programmation visuelle développé au sein de l'IRCAM au milieu des années 80 par Miller Puckette (aussi le créateur de PureData en 1996), pour donner aux compositeurs un outil d'écriture pour la musique interactive. Si les premières versions de Max traitaient uniquement des informations MIDI pour la communication avec des synthétiseurs et des samplers, l'arrivée de MSP (Max Signal Processing) en 1997 l'ouvre au traitement du signal audio et en fait un outil aujourd'hui synonyme de multimédia interactif et de la composition temps réel.

Max est un environnement de programmation visuelle, ce qui signifie que plutôt qu'écrire des lignes de code, l'utilisateur manipule des Objets représentant des opérations ou des fonctionnalités. Ces objets prennent la forme de boîtes avec des entrées et des sorties que l'utilisateur connecte ensuite à sa guise. Le logiciel est extrêmement polyvalent dans son utilisation grâce à sa facilité d'interfaçage avec le monde extérieur. En effet, il peut accéder à toutes les connectiques informatiques de l'ordinateur pour récupérer ou envoyer des données, et peut aussi accéder aux chemins internes comme le réseau.

B / Pour le traitement du signal vidéo

En ce qui concerne le traitement des images à projeter, le principal critère de sélection de logiciel est la gestion du temps réel. Il faut impérativement un logiciel dit de VJing pour

manipuler et générer des effets visuels en direct, de façon à ce que le contenu visuel de l'installation réagisse aux interactions avec le spectateur. Il faut donc pouvoir contrôler les réglages du logiciel par un patch Max, réactif aux mouvements des yeux et de la tête. De plus, la façon la plus commune d'implémenter des interactions avec un flux visuel est le compositing, c'est-à-dire le fait de travailler sur plusieurs couches vidéos avec des effets de masques. Le logiciel étudié doit donc gérer la lecture simultanée de plusieurs flux HD.

Pour son côté user-friendly, et sa liste de tutoriels exhaustive j'ai choisi de m'intéresser à VDMX 5, l'outil de VJ-ing de la société Vidvox. (voir figure 2.3.1) Son interface modulaire et son fonctionnement en couches ou *Layers* correspondent parfaitement aux besoins évoqués. Il met à disposition de l'utilisateur un grand nombre d'effets vidéos en temps réel calculés avec Quartz Composer et OpenGL. Ces librairies ouvertes et modulaires permettent à une communauté Open Source d'échanger des effets et autres traitements. Certains de ces effets sont des manipulations qui dépendent d'un point du plan projeté, ce qui donne des pistes intéressantes de traitements liés à la direction du regard. De plus, tous les paramètres de compositing et d'effets sont contrôlables par OSC ce qui assure un fonctionnement parfaitement compatible avec Max.

L'OSC ou *Open Sound Control* est un protocole de communication entre machines passant par le réseau, et donc extrêmement rapide. Les données sont échangées dans les deux directions ce qui assure dans le cas de notre étude la possibilité d'une communication triangulaire entre l'acquisition des données, le traitement du signal audio, et le traitement du signal vidéo.

Figure 2.3.1 : Interface graphique de VDMX, utilisation simultanée de plusieurs Layers
vidvox.net

CHAPITRE 3 : Partie Pratique

Proposition d'une installation d'art vidéo interactif

I / CONCEPTION DE L'INSTALLATION

A / Considérations préliminaires

Après avoir dressé un état de l'art interactif et dégagé les pistes de réflexions qui m'intéressaient le plus, à savoir la question de l'immersion par le son et celle de la transparence des interfaces, j'ai étudié les techniques et technologies permettants de mettre en place une installation d'art vidéo interactif répondant à ces réflexions. La partie pratique de ce mémoire sera donc dédiée à la proposition d'une installation d'art vidéo interactif dont les principaux attraits sont un interfaçage transparent et une immersion par le son.

Je souhaite placer le spectateur dans des conditions de projection lui rappelant le confort passif d'une salle de cinéma, mais utiliser la technique de diffusion sonore au casque en mode binaural pour l'immerger dans une illusion de réalisme de la scène. Le casque sera par ailleurs équipé d'une centrale inertielle, comme celle étudiée dans le Chapitre 2, pour assurer un Head-Tracking. Le suivi des mouvements de la tête aura un double objectif. Il devra d'une part assurer une spatialisation cohérente des sources sonores, et ainsi augmenter l'expérience de l'immersion proposée. En effet, les outils logiciels étudiés dans la partie précédente, à savoir le Spat v3 de chez Flux ainsi que les plugins AmbiX de Matthias Kronlachner permettent de positionner des sources dans un espace virtuel binauralisé mais aussi d'intégrer une rotation de la scène sonore en fonction des mouvements de la tête de l'auditeur. D'autre part il devra servir de moteur d'interaction de l'ordre du sensible et du réflexe entre le spectateur et l'œuvre proposée. Pour cette installation je veux aussi mettre à profit les résultats de nos recherches sur l'Eye-Tracking, ou suivi des yeux. Equiper le spectateur d'un ElectroOculographe embarqué permettra de déterminer en temps réel ce vers quoi tend son regard et ainsi mettre en place une réactivité de l'œuvre par rapport à celui-ci. Cette translation directe des sens du spectateur dans le domaine digital devrait instaurer une relation sensorielle et inter-réactive entre ce dernier et l'œuvre.

Les séquences devront être filmées dans l'optique de ces interactions avec le spectateur et le processus d'écriture devra en prendre compte. Nous avons vu que le logiciel VDMX permettait de lire en simultanée plusieurs flux vidéo HD et traiter chacun de ces flux de manière indépendante en assignant des effets et des ordres aux différents *Layers.*

Mon intérêt pour l'immersion sonore au moyen du binaural implique une contrainte pour l'image des séquences : pour garder un référentiel stable dans l'espace virtuel qui permet le

réalisme de l'immersion, il faudra dans un premier temps que tous les plans soient fixes. Cependant, le sujet de cette proposition pratique est entre autres d'interroger le rapport qu'entretient le spectateur avec l'œuvre dans un cadre proche du cinéma, et le plan fixe n'a jamais été jugé comme une contrainte pour celui-ci.

Néanmoins je souhaite pouvoir proposer plusieurs scènes différentes au spectateur, chacune mettant en œuvre des logiques d'interaction variées. Il est donc de mise de trouver des mécanismes d'écriture permettant de passer d'une scène à l'autre.

B / Proposition finale ?

Après une phase d'écriture de plusieurs séquences interactives potentielles, j'ai abouti à une proposition « finale » qui, il me semble, me permettra de mettre en œuvre ce qui est détaillé dans la partie précédente. Le point de départ est un plan fixe d'une pièce à vivre dans laquelle se trouve un grand écran de télévision éteint. Aux murs, plusieurs tableaux qui semblent disparaitre lorsque le spectateur y promène son regard. L'ambiance sonore est plutôt calme, mais on y perçoit une fenêtre en hors champ donnant sur une cour intérieure dans laquelle chantent des oiseaux ou jouent des enfants. Rapidement un son mécanique de rotation en provenance de la télévision attire l'attention du spectateur. Lorsque celui-ci la regarde elle prend vie. On prend alors en cours une scène de film dont le sujet fait étrangement écho à l'un des tableaux. Plus le spectateur passe de temps à regarder la télévision, plus son immersion dans la scène sonore augmente. Si au départ il s'agit d'une source spatialisée en binaural qui semble provenir de l'écran, rapidement le son remplit l'espace et finit par faire rentrer le spectateur dans une retranscription binaurale du mixage du film.

Lorsque l'utilisateur détourne les yeux du poste de télévision celui-ci s'éteint. Il remarque alors qu'un des tableaux accrochés au mur ne varie plus en fonction des mouvements de ses yeux. Ce tableau devient alors une passerelle vers une séquence explorant un autre type d'interaction, que l'utilisateur emprunte en prolongeant un regard fixe sur celui-ci. J'imagine trois séquences subsidiaires, chacune d'entre elles interrogeant une facette des relations que nous avons aux mondes virtuels, et mettant en place son propre *gameplay.* Le choix des séquences subsidiaires ainsi que leur écriture est un *work in progress*, et dépendra entre autres des résultats

d'une séquence prototype que je mets en place dans la suite de ce chapitre. Néanmoins, voici où en sont actuellement ces séquences.

***Le Paysage Voilé*, notions d'exploration, de contemplation et de défi :**

Au premier abord, on a affaire à un mur blanc, filmé d'assez près, accompagné d'une ambiance sonore enveloppante d'un paysage de forêt ou de campagne. Les mouvements des yeux du spectateur effacent progressivement le mur pour laisser place à la scène correspondant à l'ambiance sonore. Peu à peu des éléments sonores de ce nouveau décor se réveillent et leur position est précisément définie grâce à la synthèse binaurale. Un élément sonore particulièrement bruyant se fera entendre à un endroit précis du mur. C'est le bouton reset. Si le regard du spectateur passe sur la zone définie par ce son, le mur se reforme dans sa globalité et la position initiale du bouton reset est aléatoirement changée.

Le spectateur s'y prend alors à plusieurs reprises avant de réussir à voir l'image dans sa globalité, sachant qu'il ne verra jamais ce qui se cache sous la zone du mur qui déclenche la remise à zéro.

Ici, le suivi des mouvements de la tête est primordial dans la précision de la localisation des sources par l'utilisateur. Il aura le loisir de fixer un point tout en faisant des mouvements amples de la tête de manière à être sûr de la localisation d'origine de la source pour mieux l'éviter et dévoiler tout le tableau. Alors seulement il retournera dans la pièce à vivre du début de l'installation.

***Réseaux Sociaux*, notions de frustration et de voyeurisme :**

L'utilisateur devient un personnage au travers des yeux duquel il voit et vit la scène. Il se trouve devant un écran d'ordinateur et la direction de son regard permet de faire des choix de navigation (les options présentées sont des informations-types des réseaux sociaux comme observer la liste des contacts, explorer les photos de quelqu'un). Il est encouragé par un personnage qui lui parle dans l'oreille droite et lui donne des informations complémentaires. Le but est de découvrir quelque chose sur un second personnage qui quitte la scène au début de la séquence avant que ce dernier ne revienne.

L'idée est de créer un sentiment de frustration chez le spectateur allant jusqu'au besoin de savoir ce qui se cache dans ce profil de réseau social.

***Prière De Ne Pas Ecouter*, notion de dissociation des sens :**

Sur une paroi vitrée, une affiche demande au spectateur « Prière de ne pas lire ». De toute façon, lorsqu'il essaye de lire, le texte se brouille. Il se concentre alors sur la scène qui se déroule derrière la vitre. Deux personnes se font face et se parlent. Lorsque le spectateur regarde l'un des protagonistes, celui-ci se met à parler mais le son qui est entendu n'est pas synchrone avec le mouvement de ses lèvres. Le spectateur se rend compte que l'orientation de sa tête dicte lequel des deux il entend, et son regard détermine lequel il voit parler. Il doit alors décorréler son regard de sa posture naturelle pour percevoir les personnages en train de parler normalement, et se rendre compte qu'ils ne dialoguent pas du tout mais sont chacun dans un monologue effréné.

Si le spectateur choisit d'ignorer les deux personnages et de tenter de lire l'affiche, ceux-ci apparaissent tout proche de la vitre et lui demandent « ça ne t'intéresse pas ce qu'on dit ? » et « tu ne sais pas lire ? ». Le spectateur doit finalement écouter ce que disent les personnages pour comprendre comment se sortir de cette mauvaise situation.

Ces scénarios sont encore très perfectibles mais permettent de se rendre compte un peu plus précisément de ce que pourrait être la proposition finale de l'installation.

C / Notes scénographiques

Nous avons précisé ce que devra être la proposition de vidéo interactive. Décrivons maintenant la scénographie. En effet, je situe mes recherches dans le cadre de l'installation, et nous avons évoqué à plusieurs reprises notre intérêt pour les conditions de projections cinématographiques. Il est donc important de définir comment nous allons mettre en place cette installation.

Tout d'abord, définissons les dimensions à respecter. Le spectateur doit être suffisamment proche d'un écran suffisamment grand pour devoir effectuer dans certaines situations des

mouvements de la tête. Après plusieurs plusieurs essais j'ai déterminé qu'avec une image de 3m de base, le spectateur devait se situer à 1,5 m de l'écran. L'écran utilisé sera soit une toile de 3 m sur 2,5 m, soit l'écran de projection déroulant de la salle 14 de l'ENS Louis Lumière, où aura lieu la présentation de l'installation. Pour un format d'image en 16/9, 3 m de base équivaut à 1,7 m de hauteur. Pour atteindre une projection de cette dimension il faudra placer le vidéo-projecteur à environ 5 m de l'écran, à 2,5 m du sol.

Le spectateur sera assis sur un siège de cinéma isolé fixé au sol. Le casque audio embarquant la centrale inertielle et l'électrooculographe sera pendu au dessus du siège par un cable torsadé. Eclairée uniquement par l'écran, l'installation sera plongée dans la pénombre au moyen de lourds drapés obscurcissant. Le sentiment d'isolement du spectateur doit être maximum pour favoriser l'immersion dans la proposition artistique.

II / TOURNAGE DE LA SEQUENCE PROTOTYPE

Pour pouvoir explorer les possibilités des capteurs et des logiciels, ainsi que pour expérimenter avec des interactions et déterminer quelles pistes devront être suivies pour ma production finale, j'ai décidé de tourner une séquence prototype. La première étape était de définir les éléments avec lesquels le spectateur allait pouvoir interagir et dans quel but.

A / Choix des plans à tourner

Les « premiers pas » du spectateur au sein de l'installation doivent lui permettre de comprendre comment il communique avec l'œuvre et quels sont les mécanismes auxquels il a accès. En premier lieu, je choisis donc de focaliser l'attention sur le suivi du regard et deux types d'interacteur : les Convergences et les Hotspots. Il s'agit donc de déterminer deux zones dans l'espace filmé, la première doit mettre en exergue une progression jusqu'à un point clef dans l'espace, la seconde a la fonctionnalité d'un interrupteur.

Dans les faits, j'ai filmé l'intérieur d'un salon pourvu d'un écran de télévision et de plusieurs tableaux accrochés au mur. La caméra utilisée est un appareil photo numérique type DSLR, le Nikon D810, équipé d'un objectif cinéma 35 mm. J'ai effectué deux prises de plusieurs minutes dans un laps de temps restreint, de façon à assurer une continuité lumineuse (la scène est tournée en lumière naturelle). Les deux prises sont cadrées de la même façon pour simplifier l'usage d'effets de masques et de compositing, c'est à dire que je peux superposer les deux flux vidéos et contrôler des effets sur certaines zones de l'image. En pratique ce sont justement les différences entre les deux vidéos qui nous donnent les possibilités d'interaction.

Dans la première vidéo le salon est dans son état naturel. La télévision est éteinte et tous les tableaux sont en place, c'est ce que nous appellerons l'état A. (Voir Figure 3.2.1). La seconde vidéo présente deux modifications : d'une part le retrait du tableau central (une reproduction du *Baiser* de Gustav Klimt, 1909) laisse place au mur blanc cassé. D'autre part, l'écran de télévision est maintenant allumé et un court métrage (*Le Gros et la Pute*, Antoine Paley, 2015) y est diffusé depuis un lecteur de DVDs. Nous appellerons cette vidéo l'état B. (Voir Figure 3.2.2)

Deux modifications, donc, issues de deux volontés d'interactions. La présence ou non du tableau me permet de ne sélectionner que cette partie de l'image en état A, et de gérer manuellement son niveau d'opacité en superposition avec l'image B au moyen d'un fader. Nous

détaillerons dans la partie suivante comment ce fader sera contrôlé par la direction du regard. De cette façon nous pourrons mettre en place une interaction simple permettant par exemple de faire disparaitre le tableau quand le spectateur essaye de la regarder, et de le faire apparaitre quand il détourne le regard. L'allumage de la télévision sert donc d'initiation au concept de Hotspot. En fonction du moment où le spectateur regarde la télévision, il prend en cours ce qui y est diffusé. Pour le moment cette interaction est quelque peu statique mais elle pourra s'étendre comme décrit dans la partie précédente, de façon à ce que des éléments de la scène soient dépendants de ce que le spectateur regarde à la télévision.

B / Choix de la méthode d'enregistrement sonore

Ma conception du son dans cette séquence prototype est double. D'une part il s'agit de construire un environnement immersif qui situe l'auditeur dans l'espace de la scène. D'autre part le son doit servir à attirer l'attention du spectateur sur le Hotspot que constitue la télévision, avec laquelle l'interaction sera d'origine sonore. Il faudra alors construire un Hotspot sonore dont le positionnement perçu sera précis.

En pratique, j'ai positionné un enregistreur portable ZOOM H6, équipé d'une paire de microphones cardioïdes en disposition XY, sur un pied un milieu de la pièce dans la direction de l'axe caméra. Un couple de microphones[72] était placé dans la même direction au niveau de la caméra. Finalement, un microphone à large membrane cardioïde[73] était placé très proche du lecteur de DVD et des enceintes, juste en dessous du bord-cadre.

L'idée derrière cette prise de son multicanale était d'enregistrer de manière synchrone à l'image toute la matière dont nous avons besoin pour les objectifs décrits : les deux couples de microphones constituent une quadriphonie dont nous pourrons gérer la rotation en fonction des mouvements de la tête grâce aux plugins AmbiX. Celle-ci constituera le canevas sonore immersif de l'installation. Le microphone placé près de l'écran sera la source spatialisée par le SPAT. Lorsque j'ai filmé la séquence à l'état A le lecteur de DVD était allumé et le disque en rotation, provoquant un son mécanique de moteur qui a été capté par le microphone proche, mais trop léger pour gêner la captation d'ambiance des deux couples. C'est ce son qui servira à attirer l'attention du spectateur. Lorsque le regard du spectateur arrivera sur la télévision, celle-ci

[72] Nikon ME-1

[73] Neumann TLM 102

prendra vie et la source sonore deviendra le son enregistré synchrone au plan à l'état B, superposé donc à l'ambiance de l'état A. On pourra aussi imaginer que plus le spectateur retient son attention sur l'écran, plus le son issue du film emplit la pièce, correspondant alors au son provenant des couples à l'état B. Finalement, lorsqu'un second seuil temporel est atteint, le spectateur pourra être immergé pleinement dans le mixage multicanal du film.

Bien entendu, les moyens à ma disposition pour tourner cette séquence prototype ne correspondent pas complètement aux techniques que nous avons décrites plus haut, notamment dans l'utilisation de deux couples plutôt qu'un microphone B-format pour l'ambiance binauralisée. Cependant, l'*Encoder* d'AmbiX permet de convertir en format Ambisonique notre quadriphonie, les autres éléments de la chaine sont alors les mêmes que si nous avions tourné en B-format. Le rendu sera forcément moins convaincant mais nous donnera une idée de ce que nous pourrions effectuer avec les outils réellement prévus.

Si cette séquence prototype constitue une plateforme solide pour expérimenter plusieurs types d'interactions dont les moteurs sont à la fois visuels et sonores, et mettant à profit le Head-Tracking et le Eye-Tracking. Elle va donc servir de support à la prochaine étape de cette partie pratique ; la mise en place technique de ces interactions.

Figure 3.2.1 : Extrait de la séquence prototype, état A.

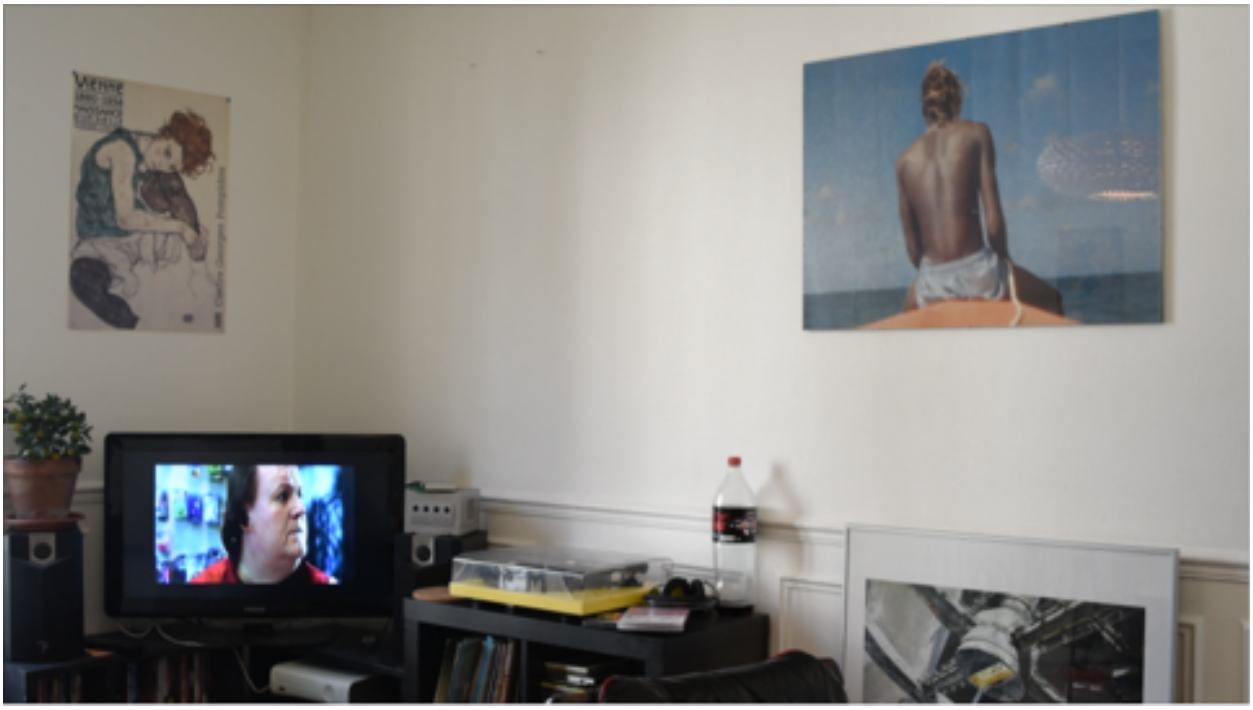

Figure 3.2.2 : Extrait de la séquence prototype, état B.

C / Préparation des fichiers - Derushage

Les fichiers vidéo provenant du Nikon D810 sont encodés en H.264 ce qui est un format de compression en GOP (pour *Groupe of Pictures* ou Groupe d'images) permettant d'avoir des fichiers peu volumineux, mais dont la lecture est plus consommatrice de ressources qu'une compression intra-image comme le Apple ProRes.[74] La première étape est donc de transcoder les deux prises avec un logiciel type MPGE Streamclip, vers un format ProRes. Les fichiers sons synchrones étaient enregistrés sur deux supports. Le couple au niveau de la caméra (Nikon ME-1) était enregistré directement sur l'image, à 48 kHz, 16 bit. Le microphone proche de la télévision et le couple au milieu de la pièce étaient enregistrés par le ZOOM H5 sur une carte SD, à 48 kHz, 24 bit. J'ai ensuite synchronisé les fichiers vidéo et les fichiers sons dans ProTools avant d'en tronquer le début et la fin de façon à avoir des fichiers de même longueur, qui peuvent alors être déclenchés au même instant. Les fichiers sonores sont enfin exportés et renommés de façons à pouvoir les identifier facilement : le préfixe « VIDTEST_B_ » identifie la vidéo à laquelle les fichiers sont synchrones. Le suffixe « MONO » correspond au microphone proche de la télévision, les suffixes « L, R, Ls, Rs » désignent les quatre canaux qui constituent l'ambiance.

[74] Un groupe d'images est codé par rapport à une première image de référence, impliquant qu'un logiciel de lecture utilisant cette video doit garder en mémoire plusieurs images complètes pour décoder chaque image. A l'inverse, dans une compression intra-image, chaque image n'est codée qu'en fonction de son propre contenu. Les fichiers sont alors plus lourds, mais plus apte à une utilisation en post-production, ou en temps réel.

III / MISE EN PLACE TECHNIQUE DES INTERACTIONS

A / Recherche des outils pour l'acquisition des informations des capteurs.

La première étape de cette mise en place technique est de concevoir des outils logiciels, en l'occurence des patches Max, permettant par la suite d'exploiter simplement les informations en provenance des capteurs. En effet, nous avons établi que le module de Head-Tracking, architecturé autour de la carte Razor 9DOH, fournissait à travers un port série les informations de Yaw, Pitch et Roll. Ces informations sont délivrées à 57600 Bauds[75] et dans le format ASCII[76]. Il faut donc convertir ces données en *float*[77], pour retrouver l'angle de rotation transmis selon chaque axe. C'est là l'utilité de l'objet « RAZOR_DIR_OUT ». (voir figure 3.3.1)

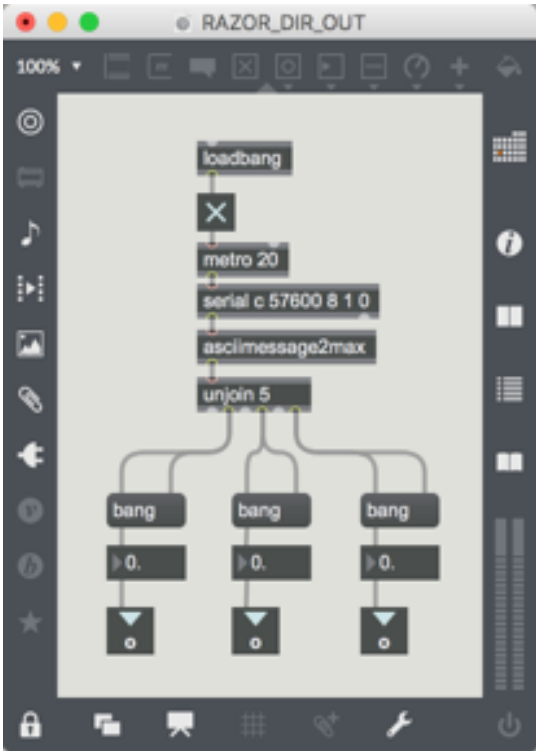

Figure 3.3.1 : Patch Max « RAZOR_DIR_OUT »

Déclenché par un métronome toutes les 20ms, l'objet « serial » récupère les informations arrivant au port C (le port USB auquel est connecté le Head-Tracker). Ces informations sont traduites par l'objet « asciimessage2max » puis séparées en trois *floats*, correspondant donc aux Yaw ,Pitch et Roll. Cet objet donne donc directement accès aux données qui nous intéressent, mais il faut désormais les rendre compatibles avec nos moteurs binauraux.

En effet, que ce soit la suite logiciel AmbiX ou le SPAT, les outils que nous allons utiliser attendent des valeurs entre 0 et 1 pour contrôler leurs paramètres, et non des angles entre -180 ° et +180 °. De plus il est important d'implémenter ces

[75] Le Baud est une unité de mesure qui désigne le nombre de symboles transmissibles par seconde. Lors d'une communication entre deux interfaces, si le récepteur n'est pas réglé sur la même vitesse de transmission que l'émetteur,

[76] ASCII, ou *American Standard Code for Information Interchange*, est une norme de codage informatiques des caractères de la langue anglaise. C'est un code simple qui retranscrit des lettres et symboles en binaire.

[77] Terme de programmation qui désigne un nombre à virgule.

valeurs sur une échelle cyclique, c'est à dire que la valeur 0 et la valeur 1 représentent le même angle de ±180 °. De cette façon, le module peut effectuer des tours sur lui-même et les données reçues ne seront pas bloquées. « RAZOR_MOD_OUT » désigne le patch chargé de ces calculs. (voir figure 3.3.2)

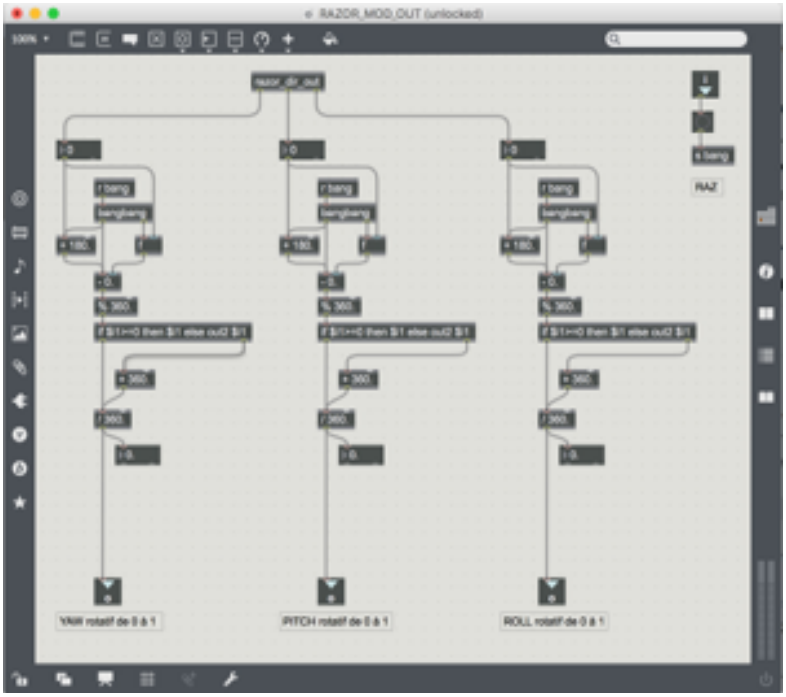

Figure 3.3.2 : « RAZOR_MOD_OUT »

Le dernier outil de ce type est un simple inverseur d'échelle qui permet de ne plus considérer la rotation de la tête mais la rotation relative des objets sonores par rapport à la tête. (voir figure 3.3.3) Par exemple lorsqu'on tourne la tête de 90 ° vers la droite, ce qui correspond à une valeur de 0,75 sur l'échelle cyclique, le nouveau patch transmet une valeur de 0,25 correspondant à l'angle relatif de 90 ° vers la gauche. Cette objet se nomme « RAZOR_ROT_OUT »

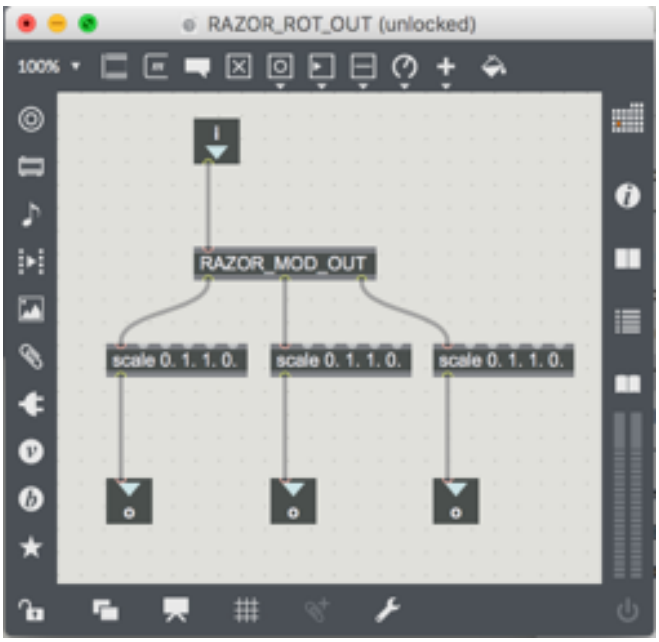

Figure 3.3.3 : « RAZOR_ROT_OUT »

Les informations issues de ce dernier Patch vont pouvoir directement servir pour la rotation des scènes sonores dans AmbiX et dans SPAT, et donc permettre de préserver les angles incidents des sources sonores par rapport à un repère fixe, ici la scène visuelle.

En ce qui concerne le Eye-Tracking, les informations issues de l'électro-oculogramme nous donneront la direction du regard par rapport à la tête de l'utilisateur sous la forme d'un angle horizontal et d'un angle vertical. De la même façon que pour les informations du Head-Tracker, ces données sont ensuite ramenées sur une échelle de 0 à 1, 0,5 représentant donc le fait de regarder droit devant soi. Il suffit ensuite d'ajouter ces valeurs aux données issues du « RAZOR_MOD_OUT » pour avoir des valeurs de direction absolue du regard dans l'espace. La distance entre le spectateur et l'écran, ainsi que les dimensions de l'écran étant connues, la zone regardée sur un plan se détermine alors très simplement par la trigonométrie.

B / Gestion de la spatialisation sonore

Maintenant que nous avons des objets permettant d'exploiter le Head-Tracker, nous pouvons procéder à la spatialisation binaurale du son.

Considérons tout d'abord le son d'ambiance de la pièce. Dans le cas de notre séquence prototype il est constitué de quatre pistes monophoniques. Comme il ne s'agit donc pas d'Ambisonique natif, la première étape est d'utiliser le plugin « AmbiX-Encoder » pour faire de ces quatre pistes une scène Ambisonique dont on pourra ensuite gérer la rotation. L'encodeur AmbiX place des sources virtuelles sur une sphère ambisonique avec un écartement angulaire choisi. Il peut donc reconstituer une quadriphonie en espaçant chacune de nos sources de 90 °. (cf figure 3.3.4) L'ordre des sources demandé par l'encodeur est alors comme suit : Ls, L, R, Rs. A la sortie du plugin on retrouve un signal Ambisonique, de quatre pistes, donc en format AmbiX, qui est alors utilisable par les autres plugin de la suite.

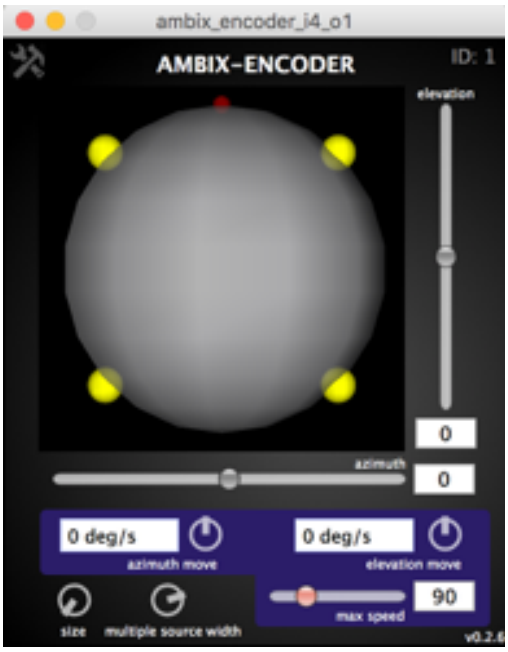

Figure 3.3.4 : AmbiX-Encoder reconstituant une quadriphonie

La deuxième étape est l'introduction des mouvements de la tête au moyen du plugin « AmbiX-Rotator ». Cet outil est directement conçu pour gérer la rotation d'un scène Ambisonique avec des réglages de Yaw, Pitch et Roll. En lui insérant les sorties de notre objet « RAZOR_ROT_OUT » il pourra alors compenser les mouvements de la tête pour conserver une scène sonore stable. (cf Figure 3.3.5)

Finalement, on utilise le plugin « AmbiX-Binaural » pour binauraliser les quatre canaux ambisoniques issues de la rotation préalable. Celui-ci propose des préréglages correspondant à des réponses impulsionnelles enregistrées au moyen d'une tête artificielle pour de multiples dispositions d'enceintes. On compte donc sur le caractère neutre des HRTFs de la tête artificielle, mais surtout sur les indices dynamiques du Head-Tracking pour créer l'illusion d'externalisation et d'espace sonore tri-dimensionnel.

On notera dans cet exemple de patch que les plugins AmbiX sont utilisés au premier ordre pour clarifier le chemin du signal (indiqué par des pointillés verts). Cela a pour

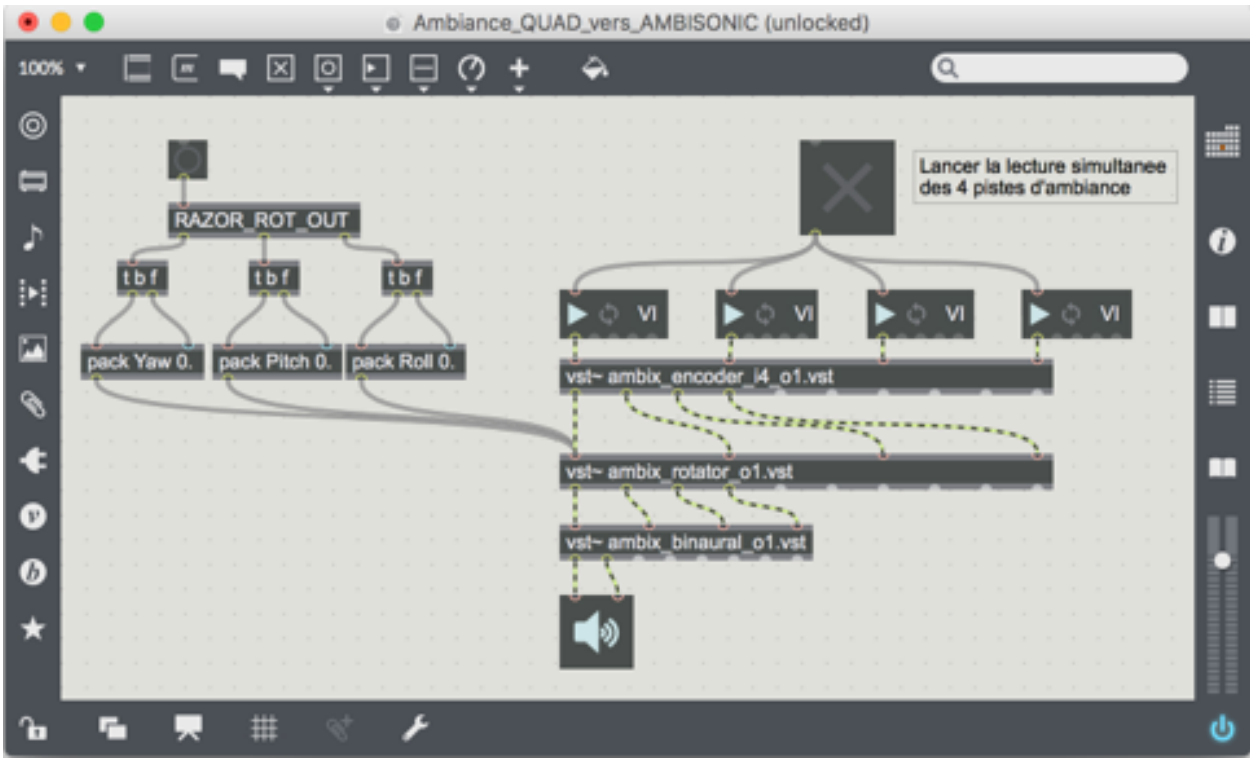

Figure 3.3.5 : Patch Max détaillant la chaîne de traitement du son d'ambiance

conséquence l'utilisation de seulement quatre réponses impulsionnelles pour la binauralisation. Pour assurer une continuité sonore lors des mouvements de la tête, il sera dorénavant préférable de travailler à l'ordre 3 (qui utilise 16 canaux).

Il s'agit maintenant de spatialiser notre source monophonique en binaural avec le SPAT. Pour chacune de ses entrées, celui-ci accepte notamment des informations d'azimut, d'élévation et de distance. Comme l'intérêt de cette spatialisation est de pouvoir précisément localiser des sources présentes à l'image, je veux pouvoir sélectionner des coordonnées sur le plan de l'écran pour en contrôler les indices angulaires. Ainsi, j'intègre dans Max un « pictslider », objet permettant de déplacer un point dans un plan, et d'en fournir les coordonnées relatives à celui-ci. (voir Figure 3.3.6)

Comme notre écran fait trois mètres de base, et les vidéos sont en 16/9, je dessine un plan de 320 points par 180 points, dont les valeurs maximales sont 300 à l'horizontale et 169 à la verticale. De plus je considère que la tête du spectateur se situe à 1,5 m sur la normale du centre de l'écran.

Je peux alors déterminer par des calculs trigonométriques simples les angles d'azimut et d'élévation par rapport au milieu de l'écran ainsi que la distance entre le point du plan et le spectateur.

Il s'agit alors de compenser les mouvements de la tête en soustrayant aux angles ci-dessus les valeurs issues de l'objet « RAZOR_DIR_OUT ». Il faut ensuite convertir ces données pour correspondre aux attentes du SPAT, en ramenant les angles et la distance sur des échelles de 0 à 1. La figure 3.3.6 présente le patch permettant de placer en binaural une source sonore sur un plan, en prenant en compte les mouvements de la tête.

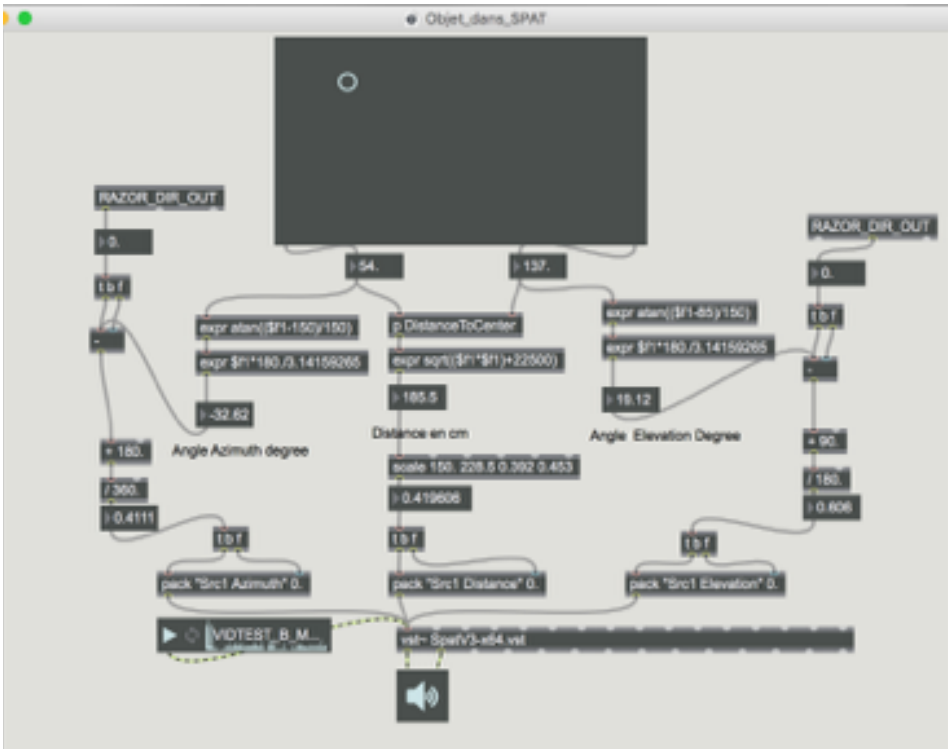

Figure 3.3.6 : Spatialisation d'une source monophonique avec le SPAT

C / Communication entre Max et VDMX

Il est maintenant nécessaire de mettre en place la communication entre Max et VDMX, qui assurera la synchronisation entre l'image et le son. Nous avons établi dans la partie précédente que cette communication se ferait grâce au protocole OSC. L'avantage de celui-ci est qu'il met à profit le réseau, ce qui me permet d'effectuer la mise en place de la communication sur mon ordinateur, mais ensuite d'utiliser un ordinateur plus puissant pour traiter le flux vidéo et conserver la simplicité de communication.

VDMX intègre un outil de détection d'envoi de données qui apparait dès lors que l'on appuie sur l'un des boutons ou l'un des faders du logiciel. Le port de réception pour les données OSC est le port 1234, le port d'émission est le 1235. L'outil repère toute communication OSC arrivant dans le port 1234. Il faut alors passer en « *Hardware Learn Mode* » en utilisant le raccourci (⌘L). Dès lors que ce mode est activé, toute commande assignable sélectionnée dans VDMX enregistrera comme déclencheur la prochaine information qu'il reçoit. Un exemple pour clarifier. Si je sélection le bouton Play du Layer 2, et que je passe en *Hardware Learn Mode*, la prochaine information que je donne à VDMX deviendra un raccourci pour le bouton Play. Cela peut donc être une touche du clavier, une information MIDI venant d'un instrument externe, ou bien encore une commande venant de Max via le réseau.

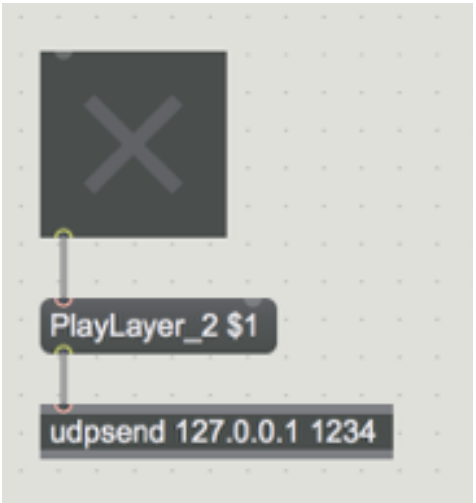

Figure 3.3.7 : extrait du patch Max destiné à contrôler VDMX

Il s'agit maintenant de fournir à VDMX des informations en provenance de Max. Pour cela on utilise l'objet « udpsend » auquel on ajoute comme arguments l'adresse IP de l'ordinateur que l'on cherche à atteindre, et son port de communication. Cet objet envoie sous la forme d'un message OSC toute information qu'on lui donne. (voir figure 3.3.7)

Il s'agit ensuite de donner une dénomination précise à la commande que l'on veut envoyer, car toutes les informations allant vers VDMX empruntent le même port, et celui-ci doit pouvoir identifier chacune d'entre

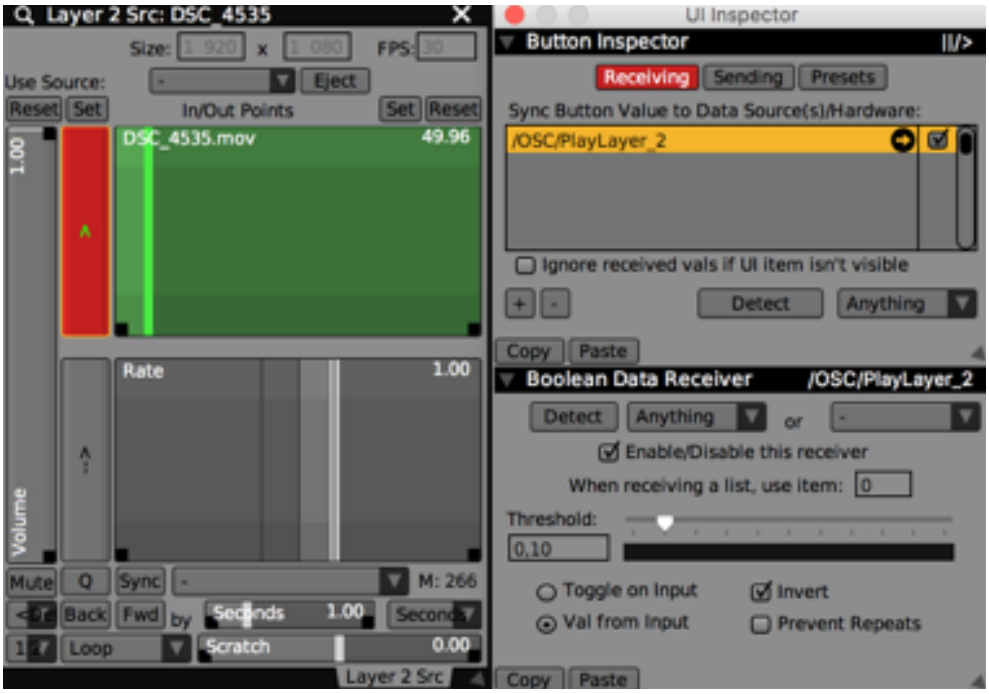

Figure 3.3.8 : interface VDMX - assignation du bouton Play du Layer 2 à une commande OSC

elles. La vidéo de la scène à l'état B est sur le Layer 2 de VDMX. Pour déclencher ou arrêter la lecture, VDMX attend respectivement un 1 ou un 0. J'envoie donc à l'adresse 127.0.0.1 (correspondant à l'ordinateur hôte) sur le port 1234 un message dont le préfixe est « PlayLayer_2 » et l'argument est une valeur contrôlée par un *toggle.*[78]

Dans VDMX, comme décrit précédemment, on enclenche le mode *Hardware Learn* puis on clique sur le bouton Play du Layer 2. Déclencher le *toggle* dans le patch Max assigne alors la commande dans VDMX. (voir Figure 3.3.8)

Cette méthode marche avec toutes les fonctionnalités de VDMX, chacune attendant des valeurs allant de 0 à 1. C'est de cette manière que nous allons pouvoir contrôler des effets progressifs sur l'image à partir des données des capteurs.

Pour les besoins de l'installation, je peux avoir besoin de recevoir dans Max des données venant de VDMX. Dans ce cas le même outil est utilisé dans VDMX, en sélectionnant dans l'onglet « Sending » un envoi par OSC sur le port 1235 et en choisissant un nom pour l'information en question. Dans Max on utilise cette fois l'objet « udpreceive » ouvert sur le port

[78] Toggle est le terme anglais décrivant un interrupteur à bascule. Dans Max il s'agit d'un bouton envoyant alternativement un 1 et un 0. Il est représenté par la croix dans la figure 3.3.4.

1235. Il suffit alors de filtrer les données reçues au moyen de l'objet « route » avec comme argument le nom donné dans VDMX. On récupère alors sur une échelle de 0 à 1 la valeur du paramètre choisi. On peut observer en *figure 3.3.9* l'utilisation de cette méthode pour recevoir l'indicateur temporel de la lecture en cours.

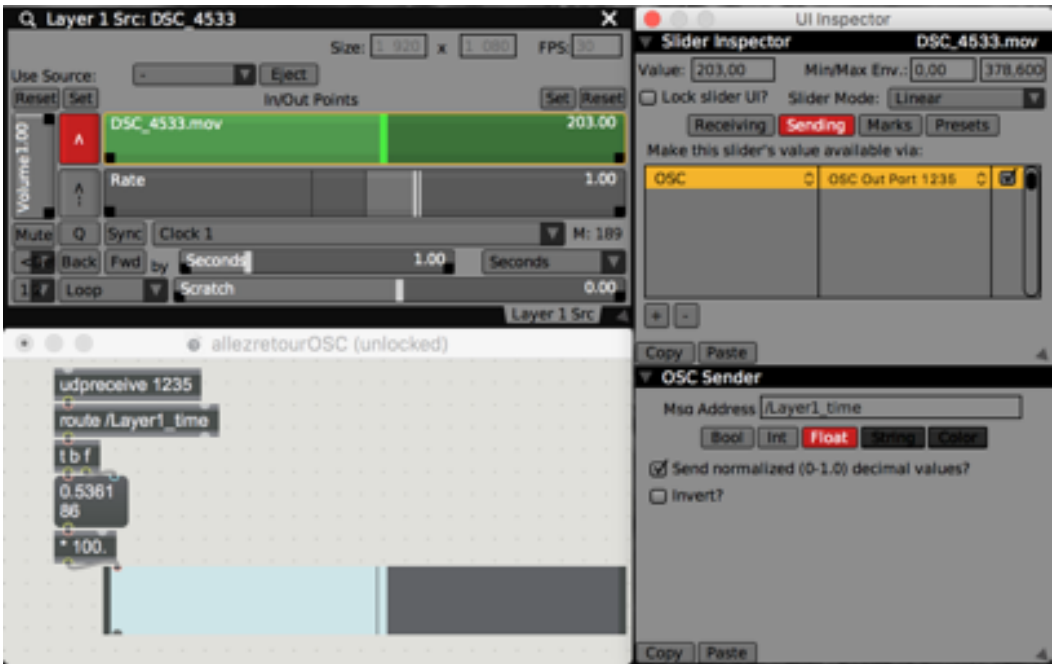

Figure 3.3.9 : Réception des données temporelles depuis VDMX par Max

D / Gestion des zones d'images

Il est important pour les interactions que nous avons décrites de pouvoir définir des zones dans l'image qui contrôleront les interactions. Plus précisément, nous allons maintenant décrire comment définir des *Hotspots*, ou zones de déclenchement, ainsi que des zones de convergence.

La méthode retenue pour définir une *Hotspot* est la suivante. Nous utilisons l'objet « nodes » qui permet de placer plusieurs sélecteurs dans un plan défini (semblable au « pictslider » décrit précédemment). Chacun de ces sélecteurs fournit ensuite son abscisse et son ordonnée. J'effectue ensuite une comparaison de ces données pour extraire les valeurs maximales et minimales selon chaque axe. Celles si définissent alors la zone que l'on a choisie. Je superpose alors à l'objet « nodes » un « pictslider » de même dimension. L'idée est alors de comparer les coordonnées du point sélectionné dans le « pictslider » aux valeurs issues du « nodes », de façon

à émettre des 1 lorsque l'on se trouve dans le Hotspot, et des 0 si ce n'est pas le cas. (voir figure 3.3.10)

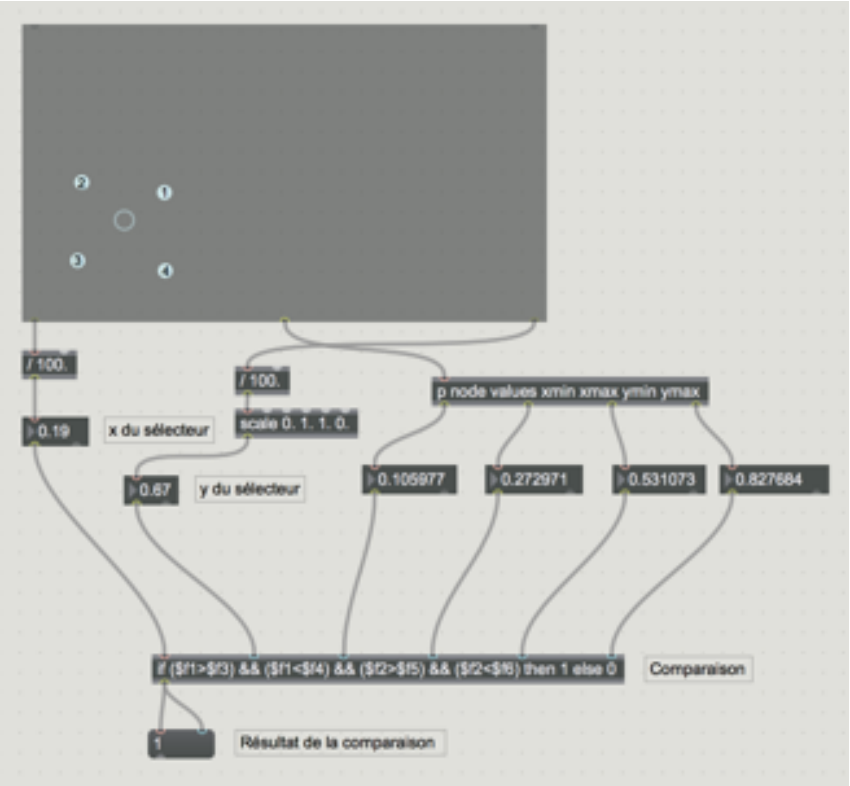

Figure 3.3.10 : extrait du patch Max pour la détermination d'un Hotspot. On observe dans le cadre au format 16/9 la présence du sélecteur dans la zone définie, et en sortie le « 1 » qui en résulte.

On remarquera que cette technique ne permet au final que de définir des zones carrées. Il sera intéressant d'implémenter une fonction permettant de sélectionner n'importe qu'elle forme géométrique dans le plan.

Ensuite il s'agit de mettre en place des zones de convergence. En effet, nous avons parlé par exemple de faire disparaître le tableau de Klimt lorsque le regard du spectateur s'en rapproche. Pour cela, on utilise une autre fonctionnalité de l'objet « nodes ». Donner une taille à un *Node* crée dans un plan carré une zone circulaire dans laquelle le curseur, ou *Knob*, peut circuler. Les valeurs émises par la sortie de l'objet dépendent alors de la proximité du *Knob* au centre du *Node,* à partir du moment où il rentre dans sa zone. Dans l'exemple proposé en *figure 3.3.11,* on voit d'une part que la zone orange du *Node* n'est plus circulaire, mais ovale. Ceci est lié à l'étirement du plan pour correspondre à un format 16/9. D'autre part on observe que la valeur affichée est 0,5 lorsque le *Knob* est à mi-chemin entre l'extrémité et le centre du *Node*.

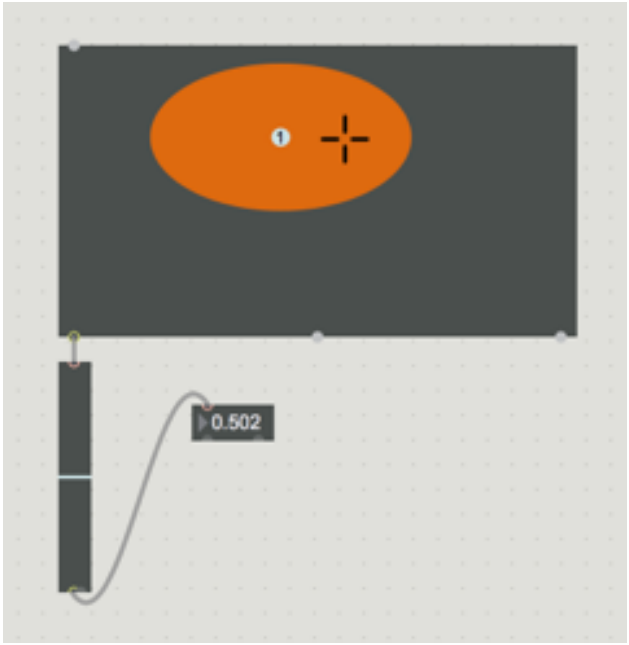

Figure 3.3.11 : Extrait d'un patch Max, démontrant le fonctionnement de l'objet « node »

En utilisant un extrait de la scène prototype comme image de fond de l'objet, on pourra placer le *Node* sur la zone qui nous intéresse et gérer sa taille en fonction de l'effet recherché. Les coordonnées du *Knob* peuvent être insérées dans l'objet, permettant une utilisation dynamique de celui-ci.

Comme pour la gestion des *Hotpsots*, affiner la définition d'un zone de l'image est assez limité avec cette méthode de part la rigidité de la forme de *Nodes*. Notamment, on voudrait pouvoir déterminer un rayon pour le centre du *Node*, c'est-à-dire déterminer une épaisseur de la zone qui donne un « 1 » en sortie. Cependant, on peut ajuster les résultats de cette technique en implémentant un changement d'échelle avec l'objet « scale », qui permet de conserver des valeurs entre 0 et 1, mais de faire varier la courbe de croissance. Techniquement il s'agit d'appliquer une fonction exponentielle ou logarithmique aux résultats, suivant que l'on veut ralentir ou accélérer la croissance liée à un rapprochement du centre.

L'aboutissement de cette étape de recherche pratique est un ensemble d'outils permettant de rapidement et simplement mettre en place des interactions dans le domaine de l'image et du son, dont les moteurs sont le suivi des mouvements de la tête et des yeux. On retiendra que positionner et déplacer des sources sonores par rapport à un plan visuel, et adapter une scène sonore réaliste aux rotations de la tête se fait de manière très précise, et pourra être une notion à explorer davantage dans les séquences de la proposition finale. Par ailleurs, la définition de zones actives dans une image reste contrainte par plusieurs facteurs, comme la forme ovale des *Nodes* ou la forme carrée des Hotpsots. Il faudra prendre en compte ces limites dans la conception des séquences, par exemple en s'efforçant de filmer de manière frontale, limitant ainsi les déformations de perspectives. Néanmoins, je pense pouvoir affirmer qu'il est possible de proposer une installation vidéo interactive basée sur les outils développés dans cette partie.

CONCLUSION ET PERSPECTIVES

Ce mémoire m'a permis de questionner la place changeante du spectateur dans l'art. Partant d'un constat historique sur l'implication grandissante du public au sein même des œuvres, j'ai étudié l'avènement de l'art interactif, notamment dans sa relation aux développements technologiques de la fin du XXe siècle. L'art vidéo interactif interroge des thèmes étroitement liés aux préoccupations du XXIe siècle, à savoir l'évolution d'une consommation immersive des différents médias cinématographiques et vidéoludiques, ainsi qu'une déportation des relations sociales dans le domaine du virtuel. Se pose alors la question de l'interface qu'on tend à vouloir faire disparaitre pour laisser place à une translation directe des sens dans le domaine virtuel.

Je me suis particulièrement intéressé aux notions d'immersion sonore et d'interface transparente dans le cadre d'une installation d'art vidéo interactif. J'ai donc dressé un état de l'art des techniques permettant la mise en place d'une telle installation; les moyens étudiés étant la diffusion sonore au casque en mode binaural, ainsi que l'utilisation de Head-Tracker et de Eye-Tracker comme moteurs d'interaction transparents.

Finalement j'ai mis en application mes recherches par une proposition d'installation d'art vidéo interactif, de l'écriture des scènes, à la mise en place technique des interactions à travers l'exemple d'une séquence prototype.

Il s'agit désormais de poursuivre l'écriture et la mise en place de cette installation, de façon à pouvoir la présenter au moment de la soutenance de ce mémoire. Les conclusions de la dernière étape de recherche tendent à recentrer l'écriture des scènes sur des interactions dans le domaine du son, et sur une cinématographie plus frontale. Il faudra notamment mettre en œuvre les techniques de prise de son Ambisonique décrites dans le chapitre 2.

Plus tard, on pourra se pencher sur la question du plan fixe : on pourrait imaginer, dans un autre contexte, l'adaptation de scènes sonores en binaural à des mouvements de caméra, proposant ainsi une expérience plus proche du cinéma. Les récentes évolutions des technologies de la machinerie *(Motion Control)* permettent aujourd'hui d'effectuer plusieurs fois exactement

les mêmes mouvements de caméra, favorisant ainsi l'élaboration de nouvelles interactions visuelles.

Enfin, je souhaite expérimenter un sytème d'installation interactif capable de gérer les zones de net et de flou en fonction du placement du regard du spectateur. À ce moment là, il ne s'agirait plus seulement d'un interfaçage transparent d'un sens dans le domaine digital, mais un accompagnement, voire même une augmentation du sens par le biais de l'installation.

BIBLIOGRAPHIE

OUVRAGES

BRETON André, Dictionnaire Abrégé du Surréalisme, Galerie des Beaux Arts, Paris, 1938.

O'DOHERTY Brian, Inside the White Cube : the Ideology of the Gallery Space, University of California Press, Janvier 2000.

KIRBY Micheal, Allan Kaprxw's Eat, Tulane Drama Review. Vol. 10, No. 2, 1965.

PUBLICATIONS

ALGAZI V.R et al. The cipic hrtf database, IEEE Work- shop on Applications of Signal Processing to Audio and Acoustics, Octobre 2001.

ALGAZI V.R et al. THE CIPIC HRTF DATABASE, UC Davis, Californie, 2001.

BURNHAM Jack, Real Time Systems. Artforum, Vol. 7, Septembre, 1965.

EDMONDS Ernest et al. Approaches to Interactive Art Systems, GRAPHITE'04, 2004.

FISHER, H. G., FREEDMAN, S. J. The role of the pinna in auditory localization, The Journal of Auditory Research n°8, 1968.

HUHTAMO Erkii, Trouble at the Interface, or the Identity Crisis of Interactive Art, dans Framework, The Finnish Art Review, 2004.

JAVAL Émile, Essai sur la physiologie de la lecture, Annales d'Oculistique 80, 1878.

KASSNER Moritz et PATERA William, PUPIL Constructing the Space of Visual Attention, MIT, 2012.

LUDOVICO Luca A et al. HEAD IN SPACE: A HEAD-TRACKING BASED BINAURAL SPATIALIZATION SYSTEM, dans le cadre de LIM - Laboratorio di Informatica Musicale, Universita degli Studi di Milano, Italy, 2010.

MANTIUK Radoslaw et al. Do-It-Yourself Eye Tracker: Low-Cost Pupil-Based Eye Tracker for Computer Graphics Applications, 2012.

PENNY Simon, From A to D and back again: The emerging aesthetics of Interactive Art, dans Leonardo Electronic Almanac, Avril 1996.

PENNY Simon, Interactivity - who cares?, Forthcoming Fiberculture, 2011.

PERRETT, S., NOBLE, W. The contribution of head motion cues to localization of low-pass noise, Perception & Psychophysics n°59, 1997.

POLLACK, I., ROSE, M. Effect of head movement on the localization of sounds in the equatorial plane, Perception & Psychophysics n°2, 1967.

SAN AUGUSTIN Javier et al. Evaluation of a Low-Cost Open-Source Gaze Tracker, IT University of Copenhagen, 2009.

SIMPSON Dallas, Improvisational Binaural Sound Art : The Foundations of Location Performance, dans Rubberneck n °24, 1997.

THURLOW, W. R., RUNGE, P. S. Effect of induced head movements on localization of direction of sounds, Journal of the Acoustical Society of America n°42, 1967.

YAIRI, S et al. Estimation of detection threshold of system latency of virtual auditory display, Applied Acoustics n°68, 2007.

THÈSES ET MEMOIRES

ANGLIONIN Louis, De nouveaux outils pour un dispositif de suivi des mouvements de la tête en spatialisation binaurale, Mémoire sous la direction d'Alan Blum et Jason Cook, ENS Louis Lumière, 2014.

CERLES, Clément, Caractérisation objective et subjective d'une chaine de traitement HOA, Mémoire sous la direction de Frank Gillardeaux et Jérôme Daniel, ENS Louis Lumière, 2015.

CHEVRIER Léa, Expérimentation des techniques binaurales appliquées au documentaire radiophonique, Mémoire sous la direction d'Alan Blum et Hervé Déjardin, ENS Louis Lumière, 2015.

GUILLON, Pierre, Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF, Thèse pour obtenir le grade de Docteur de l'Université du Maine, 2009.

TABLE DES ILLUSTRATIONS

Chapitre 1

Chapitre 2

Chapitre 3