

Valentin Couineau

# Intégration des techniques binaurales de spatialisation sonore au sein des chaînes de production et diffusion actuelles

Mémoire de fin d'études  
de la Formation Supérieure aux Métiers du Son  
du Conservatoire National Supérieur de Musique et de Danse de Paris

Recherche réalisée en collaboration avec Radio France  
et les membres du consortium sur le binaural : Orange Labs, France TV,  
IRCAM, Arkamys, A-volute, DMS, Trinno audio, CNRS LIMSI et le CNSMDP

Directeur de mémoire : Hervé Dejardin

Soutenance du 25 octobre 2012

## Remerciements

Je souhaite remercier particulièrement Hervé Dejardin, Pascal Besnard et Edwige Roncière, ainsi que toutes les personnes qui m'ont aidé et accueilli chaleureusement durant mon stage à Radio France, et qui m'ont permis de réaliser mes tests d'enregistrements et perceptifs dans les meilleures conditions.

Merci à François Ragenard pour ses mesures de niveaux. Merci à Jean-Marc Lyswa pour les extraits de mixages 5.1 qui m'ont servi durant mes test perceptifs.

Merci à Rozenn Nicol qui m'a mis à disposition son précieux travail de recherche sur le binaural.

Merci aux metteurs en ondes et ingénieurs du son de Radio France et du CNSMDP qui se sont prêtés avec rigueur aux longs et fastidieux tests :

Phillipe Cabon, Hervé Dubreuil, Stéphane Foulon, Bernard Legac, Michel Gacic, Christophe Michou, Pierre Bornard, Sebastien Huel, Philip Merscher, Cédric Chatelus, Vincent Villetard, Laure Jung Lancrey, Catherine Derethe, Cyril Becue, Antoine Viossat, Alain Joubert, Julien Carton.

Une attention particulière à tous mes chers collègues et amis FSMS !!!

Et surtout un énorme merci à ma famille et tous mes proches d'être à mes côtés.

## Guide de lecture (numérique)

Ce mémoire est doté d'un lien à la Table des Matières grâce au «[T.M](#)» situé au bas de chaque page. Un système de liens est également présent sur le chapitrage détaillé.



## TABLE DES MATIERES

<b><u>INTRODUCTION</u></b> .....	6
<b><u>CONVENTIONS</u></b> .....	7
<b>1. <u>CHAPITRE 1 : Le binaural, historique.</u></b> .....	8
1.1. <u>Physiologie, fondements.</u> .....	8
1.1.1. <u>Système auditif.</u> .....	8
1.1.2. <u>Structure oreilles, tête, buste.</u> .....	9
1.1.3. <u>Naissance de la restitution binaurale.</u> .....	10
1.2. <u>Captation et restitution binaurale.</u> .....	11
1.2.1. <u>Têtes artificielles.</u> .....	11
1.2.2. <u>Couplage casque-cavité du tympan.</u> .....	13
1.2.3. <u>Contraintes dues au casque.</u> .....	14
1.3. <u>La localisation sonore.</u> .....	16
1.3.1. <u>Indice perceptif de localisation.</u> .....	16
1.3.1.1. <u>Interaural Level Difference.</u> .....	16
1.3.1.2. <u>Interaural Time Difference.</u> .....	19
1.3.2. <u>Bases de la synthèse binaurale.</u> .....	20
1.4. <u>HRTF (Head Related Transfer Function).</u> .....	22
1.4.1. <u>Obtenir des HRTF.</u> .....	22
1.4.2. <u>Individualisation.</u> .....	25
<b>2. <u>CHAPITRE 2 : Le binaural dans la chaîne de production et de diffusion actuelle.</u></b> ..	26
2.1. <u>Rappels sur les différents modèles de restitutions sonores.</u> .....	26
2.1.1. <u>Modèle stéréophonique.</u> .....	26
2.1.2. <u>Modèle multicanal 5.1 et ses extensions.</u> .....	27
2.1.3. <u>Modèle binaural et son extension «transaurale».</u> .....	30
2.2. <u>Du 5.1 au binaural à Radio France.</u> .....	31
2.2.1. <u>Contexte.</u> .....	31
2.2.2. <u>Contrôle au casque d'une prise de son multicanale.</u> .....	33

2.3. <a href="#">Objectifs des diffuseurs et moyens d'y parvenir : le consortium</a> . . . . .	35
2.3.1. <a href="#">Radio France et France TV</a> . . . . .	35
2.3.2. <a href="#">IRCAM et Orange Labs</a> . . . . .	36
2.3.3. <a href="#">Formats</a> . . . . .	37
2.4. <a href="#">Outils numériques de post-production binaurales actuels</a> . . . . .	40
2.4.1. <a href="#">Ircam Tools et Flux : Spat</a> . . . . .	40
2.4.2. <a href="#">Longcat : H3D</a> . . . . .	42
2.4.3. <a href="#">Orange Labs</a> . . . . .	44
2.4.4. <a href="#">Autres</a> . . . . .	45
2.4.5. <a href="#">Problèmes soulevés</a> . . . . .	46
2.4.5.1. <a href="#">Niveaux</a> . . . . .	46
2.4.5.2. <a href="#">Limites du panoramique d'intensité</a> . . . . .	49
3. <a href="#">CHAPITRE 3 : Un nouvel espace de composition et de mixage : comment l'aborder, comment est-il perçu ?</a> . . . . .	51
3.1. <a href="#">Esthétique</a> . . . . .	51
3.1.1. <a href="#">Genre, style musical et espace sonore</a> . . . . .	51
3.1.2. <a href="#">Création et système de restitution sonore</a> . . . . .	53
3.2. <a href="#">Différents type de productions</a> . . . . .	55
3.2.1. <a href="#">Productions bêta</a> . . . . .	55
3.2.1.1. <a href="#">Franck Avitabile</a> . . . . .	57
3.2.1.2. <a href="#">Chronique : <i>Le réveil</i></a> . . . . .	57
3.2.1.3. <a href="#">EleKtripp</a> . . . . .	57
3.2.1.4. <a href="#">Andrew Bird</a> . . . . .	58
3.2.2. <a href="#">Le mixage binaural</a> . . . . .	59
3.3. <a href="#">Perception tridimensionnelle</a> . . . . .	60
3.3.1. <a href="#">Présentation des Tests perceptifs</a> . . . . .	60
3.3.2. <a href="#">Contexte</a> . . . . .	60
3.3.3. <a href="#">Résultats</a> . . . . .	61
3.3.3.1. <a href="#">Perception arrière et élévation</a> . . . . .	61
3.3.3.2. <a href="#">Externalisation</a> . . . . .	62
3.3.3.3. <a href="#">Apprentissage</a> . . . . .	62
3.3.4. <a href="#">Conclusion Résultats</a> . . . . .	63
<a href="#">CONCLUSION</a> . . . . .	64
<a href="#">BIBLIOGRAPHIE</a> . . . . .	65
<a href="#">ANNEXES</a> . . . . .	68

## INTRODUCTION

Le développement des ventes de casques et les nouveaux modes d'écoute sur *smartphone*<sup>1</sup> laissent penser que les productions à venir gagneront à être spatialisées au casque. La démocratisation de l'écoute binaurale, une technique de spatialisation basée sur un échantillonnage de la perception de la localisation, est la voie qui semble la plus appropriée. On se propose de réfléchir au développement d'une chaîne de production pour une diffusion de type *podcast*<sup>2</sup> destinée à une écoute au casque.

Jusqu'à aujourd'hui seule la technique du *downmix* permettait de réduire sur deux canaux une production multicanale. Aujourd'hui nous pouvons améliorer ces sensations spatiales avec une diffusion en binaurale.

La recherche présentée dans ce document est née d'un projet de consortium entre les entités que sont l'IRCAM, France Télévisions, le CNSMDP, Radio France, Arkamys, A-volute, DMS, Trinnov audio et le CNRS LIMSI afin d'exprimer la situation et les perspectives de chacun en matière de techniques binaurales. Ce projet a pour but de développer l'accès à ces techniques en production, et en diffusion grand-public. Il ouvre trois ans de recherches sur le sujet.

L'objectif de ce mémoire est d'explorer les différents moyens d'aboutir à une restitution binaurale, de les tester sur différents types de productions, de mettre en exergue les problèmes existants liées aux HRTF<sup>3</sup> encore non personnalisées, d'amener un nouveau mode de production sonore aux diffuseurs et informer les compositeurs de ce nouvel espace de composition.

---

<sup>1</sup> terminal de poche

<sup>2</sup> baladodiffusion

<sup>3</sup> *Head-Related Transfer Function*

## CONVENTIONS

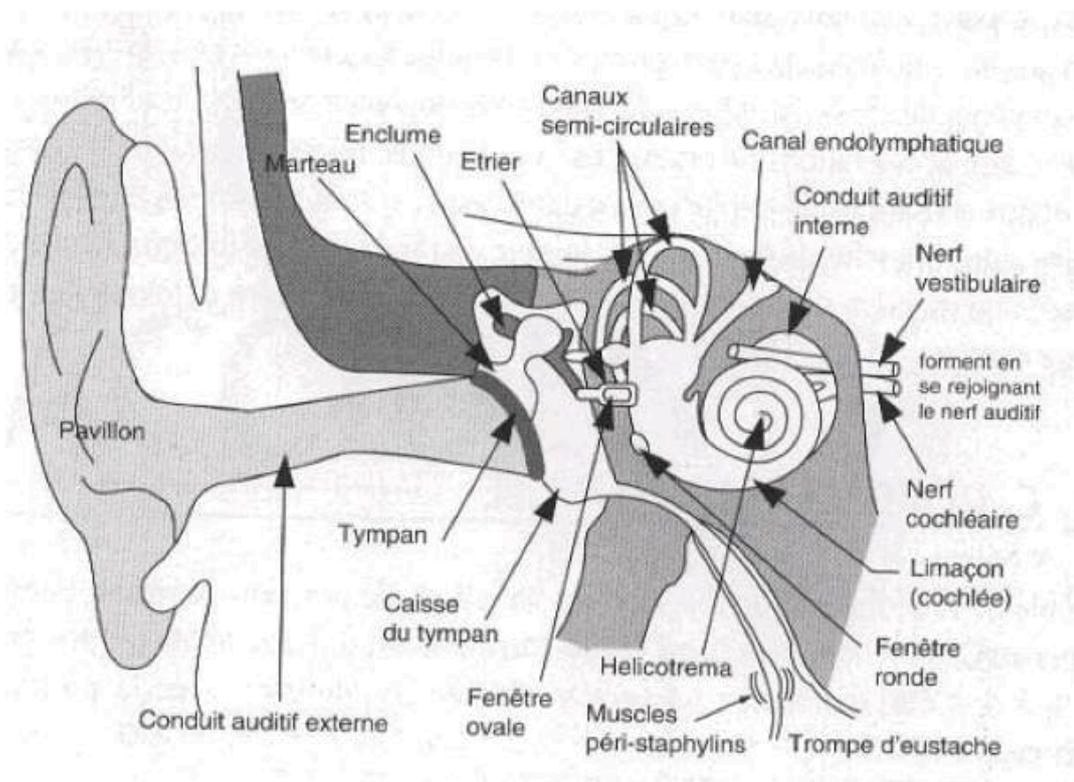
HRTF	: Head Related Transfer Function
HRIR	: Head Related Impulse Response
HpTF	: Headphone Transfer Function
BF	: Basses fréquences (20 Hz à 200 Hz)
MF	: Fréquences médium (200 Hz à 2 kHz)
HF	: Fréquences aiguës (2 kHz à 20 kHz)
ITD	: Différences interaurales de temps (Interaural Time Differences)
ILD	: Différences interaurales d'intensité (Interaural Level Differences)
FTA	: Fonction de transfert anatomique
HP	: Haut-parleur

# 1. CHAPITRE 1 : Le Binaural, historique

## 1.1. Physiologie, fondements

### 1.1.1. Système auditif

«Les ondes reçues et enregistrées par les deux oreilles sont l'entrée physiologique adéquate d'un système sensoriel spécifique, le système auditif. Il transforme chacune des deux ondes en trains d'impulsions nerveuses - après avoir effectué une décomposition spectrale en multiple canaux fréquentiels - parmi d'autres prétraitements. Les trains d'impulsions nerveuses multi-canaux de chaque oreille sont ensuite combinés d'une manière sophistiquée pour générer finalement un "motif d'activité binaurale" quelque part dans le système auditif.» [Bla09]



**Fig 1 : Système auditif**

On parle d'audition binaurculaire et la perception des sons dans l'espace est essentiellement reliée à la perception binaurculaire. Il est important de noter que cette localisation des sons est complémentaire des autres sens, notamment la vision et la perception de son corps dans l'espace (proprio-perception). Elle comprend trois dimensions spatiales : l'azimut (gauche-droite), le plan vertical (haut-bas), et la distance.

### 1.1.2. Structure oreilles, tête, buste

D'une manière extrêmement simplifiée, les délais et résonances au niveau du pavillon aident le cerveau à décoder la position de la source sonore. Les circonvolutions du pavillon provoquent des échos du son direct atteignant le tympan. La conque, cavité profonde du pavillon, est le lieu de résonances dont l'amplitude dépend de la direction des ondes incidentes. Ces délais courts agissent comme un filtre en peigne<sup>4</sup> sur le signal original. Des petits mouvements de la tête correspondent à des changements importants dans la forme du filtrage. Nous effectuons inconsciemment ces petits mouvements afin d'obtenir différentes informations de filtrage d'une source donnée ; en les combinant, la localisation devient plus précise.

Nous allons partir du postulat suivant démontré à plusieurs reprises : la localisation avant/arrière et haut/bas est possible grâce à l'interprétation par le cerveau du filtrage acoustique du son par la Fonction de Transfert Anatomique (FTA), correspondant au filtrage par le pavillon, la tête, les épaules et le haut du tronc.

Physiquement, ce comportement s'explique par des résonances dans le système oreille externe - conduit auditif externe - tympan. Les résonances sont excitées différemment lorsque l'onde sonore arrive de différentes directions et/ou avec différentes courbures du front d'onde. Les fonctions de transfert résultantes sont généralement différentes pour les deux filtres, causant ainsi des différences interaurales des signaux de pression sonore aux deux tympans. Le procédé naturel humain de la perception de la localisation des sources utilise des indices de plusieurs types pour déterminer la position des sources sonores dans l'espace : différences interaurales d'intensité et de temps, ainsi que des indices spectraux (phase et amplitude) qui sont spécifiques à l'angle d'incidence de la source et à l'auditeur, causées par des phénomènes de réflexion, d'absorption et de diffraction liées à la morphologie du sujet dont on vient de parler. Toutes les manipulations que le système oreilles-tête applique

---

<sup>4</sup> Filtre en peigne : C'est la conséquence de la sommation d'un signal avec le même signal affecté d'un retard, ce qui entraîne des annulations ou des accentuations de niveau à certaines fréquences dues aux déphasages provoqués par le délai.

aux signaux sonores sont purement physiques et linéaires. Il est alors évident qu'elles puissent être simulées. (cf. [1.3.2 Bases de la synthèse binaurale](#)).

### 1.1.3. [Naissance de la restitution binaurale](#)

La reproduction auditive authentique est achevée lorsqu'un auditeur entend exactement la même chose dans une situation de reproduction que ce qu'il ou elle entendrait dans un champ sonore original. *Puisque les signaux de pression sonore aux deux tympans sont l'entrée physiologique adéquate du système auditif, ils sont souvent considérés comme la base de mesure et d'évaluation auditivement adéquate dans un sens physique et/ou auditif.*

Dans le cas physique, des procédures basées sur la physique sont utilisées, dans le cas auditif, des auditeurs humains servent comme instruments de mesure et d'évaluation. (cf. [1.5.1 Obtenir des HRTF](#)).

En 1927, on réalise à Chicago le premier enregistrement avec une tête artificielle. Le principe était de construire un mannequin avec des critères morphologiques moyens et de placer deux capsules omnidirectionnelles au niveau du tympan. Barlett Jones créa Oscar. En 1948, Jeffress élaborait une théorie dite « du délai de ligne ». Elle stipule l'existence de neurones binauraux qui reçoivent une information provenant de chacune des deux oreilles via des neurones monauraux. « Ces neurones binauraux ne déchargent que lorsque l'information ipsilatérale<sup>5</sup> et controlatérale<sup>6</sup> leur parvient au même instant, d'où leur appellation de *neurones détecteurs de coïncidence temporelle*. [...] Le temps de propagation des potentiels d'action est proportionnel à la longueur axonale<sup>7</sup>, de sorte que la coïncidence temporelle perçue par les neurones binauraux est le fait d'une compensation très précise du délai acoustique interaural<sup>8</sup> par le délai neuronal interaural. Les décharges des potentiels d'action des neurones monauraux provenant d'une oreille et de l'autre ont la même fréquence, et sont synchronisées avec les vagues sonores ; les neurones binauraux se comportent donc comme des détecteurs de coïncidence de phase. Un détecteur de coïncidence temporelle est activé et décharge quand la somme du délai temporel neuronal et acoustique de l'information qui lui parvient d'un côté et de l'autre est la même. Les neurones binauraux forment une carte des délais acoustiques ; ils ont une sensibilité maximale à un délai acoustique particulier, qui représente un angle azimuthal particulier... »

<sup>5</sup> Information ipsilatérale : Information reçue par l'oreille dite en zone éclairée. (la plus proche)

<sup>6</sup> Information controlatérale : Information reçue par l'oreille dite en zone d'ombre. (la plus loin)

<sup>7</sup> L'axone, ou fibre nerveuse, est le prolongement du neurone qui conduit le signal électrique du corps cellulaire vers les zones synaptiques

<sup>8</sup> interaural : entre les deux oreilles

Avant de parler de captation et de restitution binaurale, Il convient de revenir à l'étymologie du mot binaural et donc de rappeler quelques définitions essentielles. Un stimulus présenté à une seule oreille s'appelle un stimulus monaural ou monodique. Par extension, un stimuli identique présentés aux deux oreilles s'appelle un stimuli binaural ou diotique. On parle également parfois de stimuli dichotiques, ce sont des stimuli différents présentés aux deux oreilles.

## 1.2. [Captation et restitution binaurale](#)

### 1.2.1. [Tête artificielle](#)

LE RELIEF SONORE À LA RADIO



LE RELIEF SONORE À LA RADIO



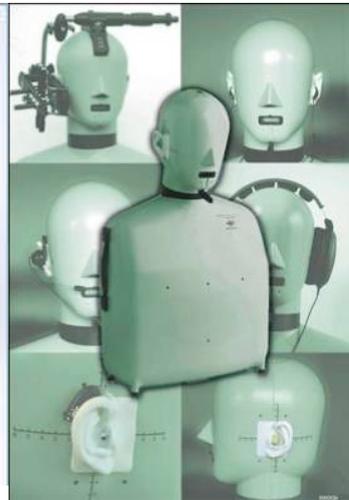
*Fig 3 et 4 : Danielle Delorme et Gérard Philippe*

Les têtes artificielles sont des répliques de têtes naturelles en termes acoustiques, elles réalisent deux filtres auto-ajustés comme des têtes naturelles. Ces systèmes reposent sur un principe simple : il devrait être possible de recréer l'illusion acoustique perçue par un auditeur lors d'une écoute naturelle, au moyen de capteurs omnidirectionnels placés à la place de ses tympans lors de l'enregistrement, puis de lui restituer par casque. Les têtes artificielles, en combinaison avec des équipements de reproduction ajustés de manière adéquate, peuvent être l'instrumentation de base pour un certain nombre d'applications économiquement séduisantes.

Voici quelques exemples des têtes artificielles les plus utilisées actuellement.



**Fig 5 :** Head Acoustics HMSIV



**Fig 6:** Brüel & Kjær HATS



**Fig 7 :** KEMAR<sup>9</sup> dite Dummy Head



**Fig 8:** Tête artificielle Neumann KU100

Il existe d'autres modèles comme la sphère Schoeps KFM6 ou le MKE 2002 de Sennheiser mais les plus répandus restent les quatre ci-dessus. La tête artificielle KEMAR, notamment, est une approche intéressante. En effet, cette tête a été modélisée à l'aide de données basées sur des moyennes morphologiques de la population mondiale. Les écoutes réalisées avec cette paire d'HRTF tendent à se révéler assez satisfaisantes pour les personnes en ayant fait l'expérience. Nous verrons dans le chapitre 3 son importance pour les étapes de production et de mixage.

<sup>9</sup> Knowless Electronics Manikin for Acoustic Research

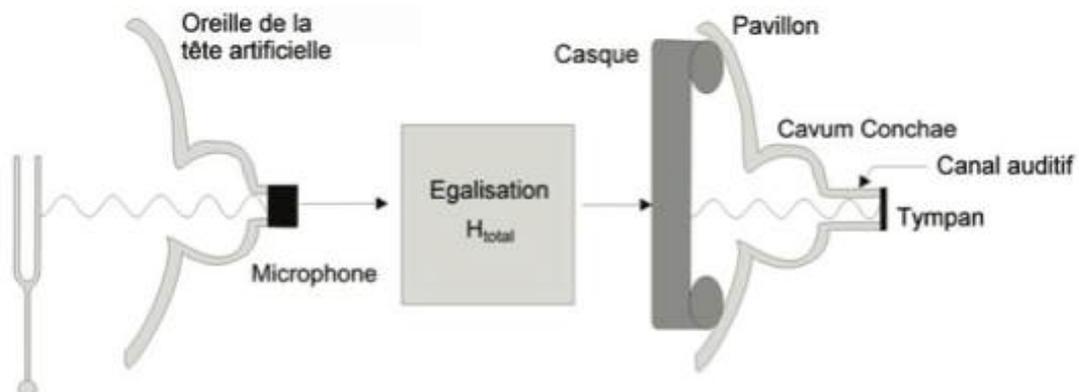
Nous pouvons également aborder le système simplifié conçu par Bernard Lagnol, preneur de son à Radio France, très intéressant pour le reportage, les ambiances ou tout type de production, en fonction de l'inventivité des réalisateurs et preneurs de son...



**Fig 9,10,11 : Système B. Lagnol**

Les microphones utilisées sont des 4060. Il utilise un écartement des oreilles proches de 15 centimètres plutôt que de 17 centimètres d'après les résultats d'un dossier de presse sur une campagne nationale de mensuration en 2006. [CNM06]. Les Genelec de type 8020 ont les dimensions idéales pour simuler une tête et fixer ce système. Il reprend donc l'idée de la tête Neumann KU 100 dans une réalisation plus simple et moins coûteuse.

### 1.2.2. Couplage casque-cavité du tympan



**Fig 12 : Schéma simulant l'intégration d'une HpTF (Headphone Transfer Function)**

A ce stade, le point crucial consiste à intégrer, dans le signal diffusé par un casque, les informations de localisation vues précédemment (tête, torse et forme du pavillon). Mais une analyse rigoureuse du trajet du signal impose de tenir compte d'autres critères spécifiques à l'écoute au casque. En effet, le casque fonctionne sur un principe de couplage acoustique entre la membrane du haut-parleur et le tympan, via la cavité, plus ou moins fermée, créée au niveau du pavillon. On doit donc prendre en compte la fonction de transfert liée à l'oreille externe lors d'une restitution au casque, fonction de transfert qui est indépendante de l'angle d'incidence puisque le haut-parleur du casque reste à la même place. De plus, la fonction de transfert liée au conduit auditif, qui occasionne en champ libre une accentuation maximale vers 3 à 4 kHz, est modifiée dans le cas d'une écoute au casque, à cause du changement d'impédance acoustique d'entrée du conduit produit par la présence du casque...

### 1.2.3. Contraintes dues au casque

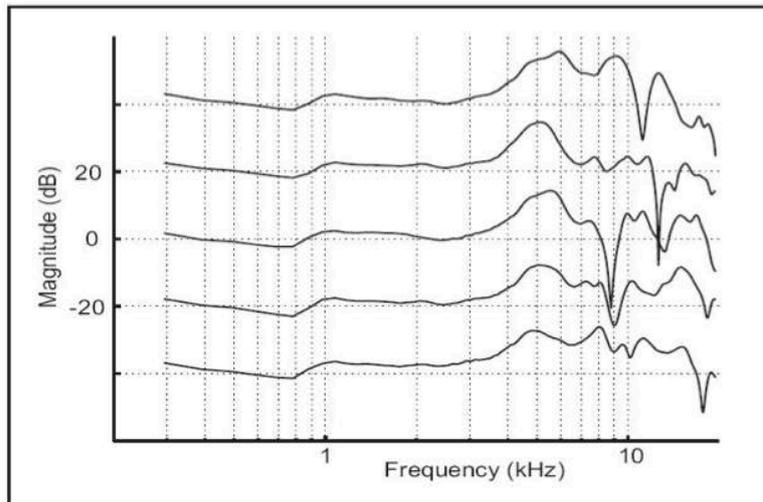
La restitution au casque implique cependant quelques contraintes difficiles à maîtriser. En effet, les perceptions physiques reçues par le corps en situation d'écoute naturelle manqueront fatalement<sup>10</sup> et la fatigue auditive se prononcera plus rapidement... Un deuxième point important est le positionnement de son casque. L'expérience de micros mouvements du casque lors de l'écoute génère habituellement des accidents fréquents conséquents. Le simple fait de déposer son casque et de le replacer modifie d'un point de vue objectif et perceptif le signal.

L'étude de Mathieu Pasquier et Vincent Koehl [PK10] révèle la perception des différences spectrales induites par la variabilité de positionnement des casques. Même si elles diffèrent d'un casque à l'autre, ces variabilités ont toujours été présentes. D'autre part, des études ont été menées et laissent penser qu'un casque ouvert fonctionnerait mieux pour ce type de restitution. Möller recommande un type spécifique de casque ouvert dits de type *Free-air Coupling Equivalent (FCE)* [Möl92]. Ils offrent des conditions similaires au champ libre en terme d'impédance vue par le tympan, comme si l'on ne portait pas de casque. Il proposera même un critère pour évaluer la pertinence d'un casque pour la restitution binaurale : le PDR (Pressure Division Ratio) qui compare les conditions d'impédance vue par le tympan en présence du casque aux conditions de rayonnement en champ libre. Le casque idéal pour un rendu binaural doit satisfaire un PDR égal à 1 et répond alors au label FCE.

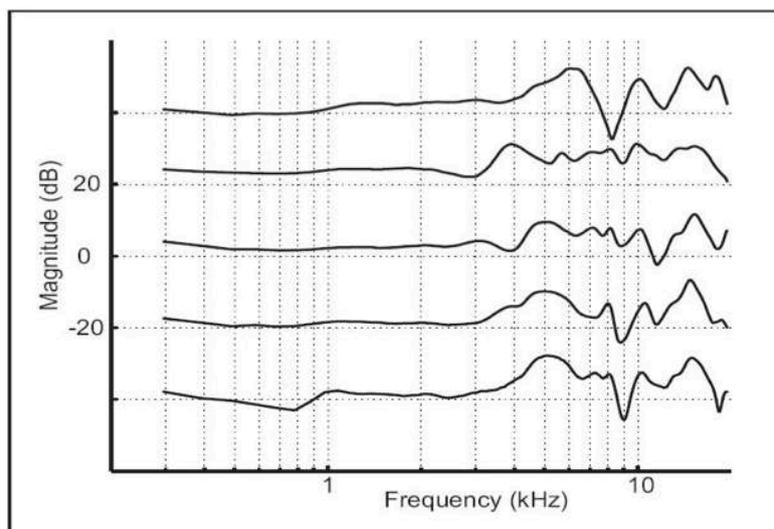
---

<sup>10</sup> L'absence de ressenti des basses

Si l'on admet la croissance malheureuse et magistrale de l'écoute des médias sur oreillettes de *smartphone*, il serait important de réaliser un travail de recherche sur les casques intra-auriculaires. On peut ajouter l'imprécision du constructeur entre chaque même modèle de casque qui soulève une compensation difficile à anticiper.



**Fig 13** : réponses en fréquences mesurées avec le même exemplaire de casque Beyerdynamic DT990 Pro placées sur l'oreille de 5 auditeurs différents. [WK05]



**Fig 14**: réponses en fréquences mesurées avec 5 exemplaires différents du Beyerdynamic DT990 Pro placés sur les oreilles d'un même auditeur. [WK05]

### 1.3. [Localisation sonore](#)

#### 1.3.1. [Indices perceptifs de localisation](#)

Il y a plus de 100 ans, Lord Rayleigh avait déjà compris l'importance pour la localisation de la différence de niveaux d'intensité arrivant aux deux oreilles. C'est en 1907 qu'il publie les conclusions de ses travaux sur la « perception de la direction des sons », aujourd'hui connues sous le terme de « théorie duplex » de la localisation de la source sonore. À cause du phénomène de la diffraction, un son venant de la droite sera plus fort à l'oreille droite qu'à l'oreille gauche, et ce d'autant plus que la fréquence est élevée. En effet, la tête constitue un obstacle à franchir pour les ondes sonores et une "ombre acoustique" apparaît d'autant plus marquée que la fréquence est élevée. Il a donc modélisé la tête par une sphère. (*cf. fig 9 Shoeps reprend l'idée avec sa sphère KFM 6*).



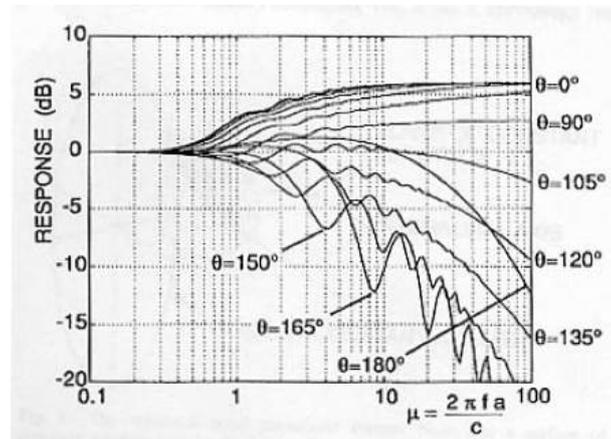
**Fig 15** : Sphère Shoeps KFM 6

Il a résolu les équations de propagation d'une onde autour d'une sphère rigide. Cette approximation de modélisation permet d'expliquer certaines observations en rapport avec les ILD.

##### 1.3.1.1. [Interaural Level Difference](#)

A partir de la résolution des équations, il a obtenu la réponse en fréquence de ce modèle sphérique, illustrée sur la figure ci-dessous, en fonction de la fréquence ( $2\pi fa / c$ ) où  $a$  est le diamètre de la tête,  $c$  la vitesse du son,  $f$  la fréquence. Le modèle est tracé pour plusieurs valeurs de  $\phi$ , angle correspondant à la différence entre l'azimut de la source et l'emplacement de l'oreille.

On peut remarquer que pour des angles  $\phi > 100^\circ$  (point dans l'hémisphère opposé à celle de l'oreille), le modèle se comporte comme un filtre passe-bas.



**Fig 16** : réponse en fréquence du modèle sphérique

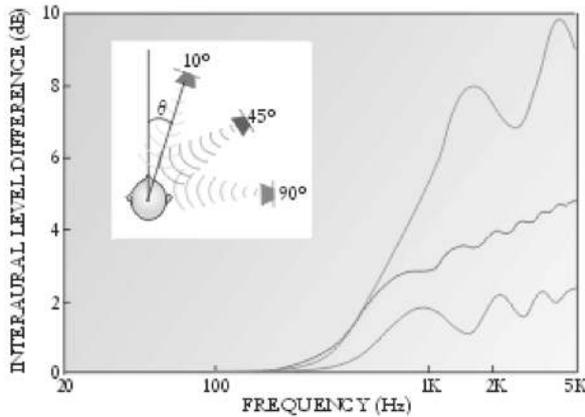
Calcul de la fréquence critique (correspondant aux dimensions d'une tête) :  
 Soit D le diamètre de la tête :

$$D = 15 \text{ cm} = 0,15 \text{ m}$$

$$f = c/l = 344/0,15 = 2293 \text{ Hz}$$

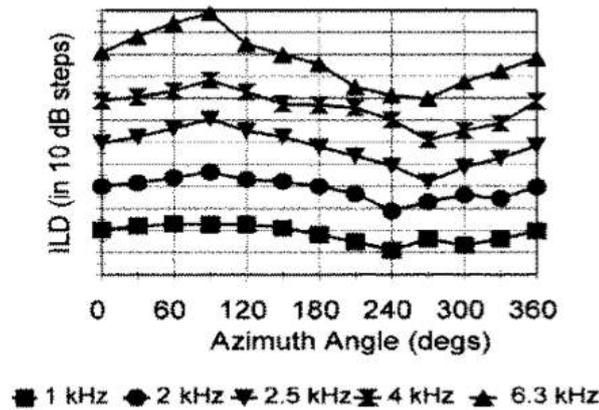
À 500 Hz, la longueur d'onde d'un son est de 68 cm, ce qui représente plus de trois à quatre fois le diamètre d'une tête moyenne. À cause de la diffraction, la ILD est donc petite pour des fréquences inférieures à 500 Hz, si la source est au moins à un mètre de distance. Cela veut donc dire que la ILD est un indice de localisation surtout utilisé pour les hautes fréquences (au-delà de 4000 Hz).

Ci-dessous l'ILD en fonction de la fréquence calculée pour une source dans le plan azimutal (comprenant les deux oreilles et le nez). La source est placée à un azimut de  $10^\circ$ ,  $45^\circ$  ou  $90^\circ$  (résultats d'un modèle sphérique dans lequel les oreilles seraient situées à deux pôles opposés).



A l'aide des figures 17 et 18, on remarque que la différence de niveau s'accroît pour les HF quel que soit l'azimut de la source.

**Fig 17 :** ILD en fonction de la fréquence



**Fig 18 :** ILD en fonction de l'azimut

La plus petite différence de niveau perceptible serait approximativement de 0.5 dB, à toutes les fréquences. Les ILD sont donc des indices de localisation potentiels à n'importe quelle fréquence, à condition que la différence soit supérieure à 1 dB. Ceci explique l'efficacité de la stéréo par différence d'intensité, et son utilisation encore aujourd'hui dans tous les stations de post-production audio. On en verra les limites plus tard.

Conclusion : dans une situation « naturelle » (localisation d'une source unique ponctuelle), les ILD ne sont utilisées que pour la localisation des hautes fréquences. Alors comment expliquer que l'on puisse quand même localiser des sources émettant des basses fréquences ?

L'ITD et l'ILD donnent lieu à une détection latérale des événements sonores, mais aussi à des indices spectraux qui complètent la localisation directionnelle dans le plan médian. Ces indices résolvent les indéterminations avant-arrière et sur la position verticale.

### 1.3.1.2. *Interaural Time Difference*

Pour les BF, c'est-à-dire les grandes longueurs d'ondes, il semble que notre système auditif binaurculaire utilise la différence de phase ce qui correspond à une différence de temps d'arrivée d'où l'appellation ITD.

Il existe cependant une confusion de phase avec les ITD. Effectivement, si la longueur d'onde est plus courte que deux fois le diamètre de la tête, chaque maximum du cycle arrivant à l'oreille droite est précédé d'un maximum arrivant à l'oreille gauche. Ceci crée une confusion, et l'auditeur conclut que la source vient de la gauche.

En d'autres termes :

Si on note  $\Delta t$  le délai entre les deux oreilles, il y a trois possibilités :

- Période  $< 2\Delta t$  : il n'y aura pas d'ambiguïté
- Période  $= 2\Delta t$  : ambiguïté totale
- $\Delta t < \text{période} < 2\Delta t$  : on localise la source du côté opposé

Un autre exemple :

ITD=0,6 ms (l'onde atteint l'oreille droite 0,6 ms avant l'oreille gauche)

Or un son pur de 1666 Hz = 0,6 ms

Des ondes à 1666 Hz peuvent arriver parfaitement en phase aux deux oreilles et donner l'impression que le son vient de devant alors qu'il vient de la droite. Mais le système auditif, parfaitement conçu, semble être beaucoup moins sensible aux ITD justement pour les fréquences où la différence de phase est ambiguë (entre 1000 et 1500 Hz).

Pour des signaux complexes, les ITD peuvent être détectées pour des hautes fréquences si les fréquences de modulations sont inférieures à 1000 Hz.

Un autre phénomène, appelé cône de confusion a été identifié sur des sources sonores situées en tout point de la surface d'un cône latéral, ce qui n'arrange pas l'instabilité en 5.1 des zones situées à environ  $-90^\circ$  et  $+90^\circ$ .

Il est important de signaler que la connaissance timbrale de la source a une influence sur la position de celle-ci car elle renseigne l'auditeur sur le paramètre du filtrage. C'est-à-dire qu'il peut soit reconnaître les caractéristiques sonores de la source elle-même ou bien assimiler ce filtrage à l'acoustique liée à son emplacement.

En champ libre, la répartition des BF et HF se comporte différemment par des effets connus - comme la dispersion des graves, l'absorption des HF par l'air - et renseigne le système auditif sur la notion de distance. Cette analyse spectrale faite par le système auditif est d'autant plus précise que si la source est en mouvement. Dans un lieu réverbérant, l'indice principal sera le rapport entre le champ direct et le champ réverbéré auquel cas les nouveaux indices de réflexion permettent de localiser plus facilement une source que l'on ne connaîtrait pas.

Avec de l'entraînement, (un exercice important chez les ingénieurs et preneurs de son), on se rend compte que notre cerveau peut reconstituer une image sonore virtuelle en séparant le son direct de la source de ses multiples réflexions... Cette faculté aura une importance fondamentale dans la recherche d'un rapport naturel de réverbération quelle que soit l'acoustique du lieu.

### **En résumé**

- Notre système binaurculaire est sensible à des ILD à toutes les fréquences, pour des ondes incidentes planes, bien que des indices de ILD n'existent physiquement que pour des fréquences supérieures à 500 Hz.
- Les ILD deviennent des indices de localisation très efficaces au-delà de 3000 Hz.
- Les ITD sont des indices de localisation pour des fréquences inférieures à 1500 Hz.
- Pour des sons purs ou à bande étroite autour de 2000 Hz, ces deux indices (ILD et ITD) sont moins efficaces et la localisation est imprécise dans cette région.

### **1.3.2. Base de la synthèse binaurale**

Puisque les têtes artificielles ne sont qu'un moyen spécifique de réalisation d'un ensemble de filtres linéaires, on peut penser à d'autres manières de réaliser de tels filtres, par exemple électroniquement, ou numériquement.

[Sig11]

«Les procédés de synthèse binaurale offrent la possibilité de produire le champ acoustique binaural en fonction d'une composition arbitraire de l'environnement sonore virtuel. Ils se basent sur l'utilisation des réponses impulsionnelles binaurales associées à la position d'une source par rapport à l'auditeur dans l'espace virtuel. Un signal monophonique est convolué avec les réponses impulsionnelles binaurales d'un auditeur dans une salle par exemple.»

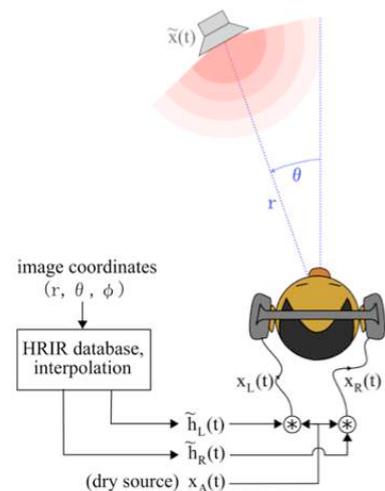
Nous avons vu précédemment l'importance des différents éléments du système auditif. Puis les différents indices d'aides à la localisation sonore, on a donc besoin d'au moins cinq éléments fonctionnels pour décrire le système binaural :

- filtres simulant les fonctions de transfert de l'oreille externe
- filtres simulant les fonctions de transfert de l'oreille moyenne
- éléments simulant les fonctions de l'oreille interne, notamment la sélectivité fréquentielle et la transformation du signal entrant en impulsions nerveuses.
- éléments simulant les différences interaurales de temps, ainsi que l'identification des composantes interaurales cohérentes du signal
- éléments simulant des différences interaurales d'intensité

L'oreille externe et moyenne se modélisent avec une précision acceptable. On a besoin du champ sonore à l'entrée des oreilles et d'une fonction de transfert du système oreilles - cavité du tympan.

La synthèse binaurale offre un avantage considérable dans le positionnement et le contrôle des sources sonores, contrairement à l'enregistrement qui fige sur un support physique toute l'image sonore. Elle permet aussi de coupler le rendu binaural à un dispositif de suivi de mouvements de tête de l'auditeur (*head-tracking*<sup>11</sup>) afin de les compenser et de conserver un positionnement stable des sources quels que soient l'orientation et les mouvements de tête de l'auditeur. On parle alors de synthèse binaurale dynamique. A l'inverse, on se place dans un mode de restitution sonore binaural statique lorsque l'on corrige les transducteurs sans tenir compte des mouvements éventuels de la tête.

**Fig 19** : Principe de la synthèse binaurale



<sup>11</sup> *head-tracking* : système de suivi de tête qui informe sur la position de l'orientation de la tête de l'auditeur en temps réel.

[Nic10]. Un signal monophonique et anéchoïque  $x(t)$  est convolué par la paire de fonctions de transfert  $h_L$  et  $h_R$  associée à la direction dans laquelle on souhaite créer la source virtuelle. Les HRTF utilisées sont soit directement les HRTF mesurées, soit le résultat d'une interpolation si la direction désirée n'est pas disponible dans la base de données.

#### 1.4. HRTF (Head Related Transfer Function)

Comme tout filtre linéaire, une HRTF peut se décomposer dans le domaine fréquentiel comme l'association d'un filtre passe-tout et d'un filtre à phase minimale.

$$HRTF(f) = |HRTF(f)| \cdot e^{j\varphi(f)} \cdot e^{j\phi(f)}$$

Où  $f$  est la fréquence,  $\varphi$  la phase minimale,  $\phi$  l'excès de phase.

Des études perceptives ont montré que l'excès de phase dans le haut du spectre n'a que très peu d'influence. On peut donc modéliser la composante passe-tout par un délai pur indépendant de la fréquence. La composante à phase minimale porte l'information spectrale donc des informations de niveaux.

##### 1.4.1. Obtenir des HRTF



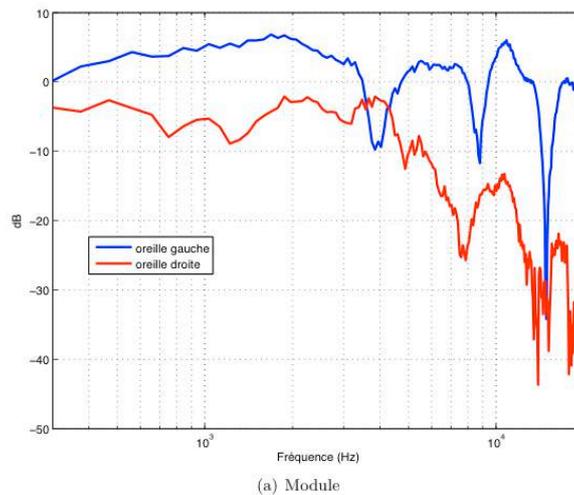
**Fig 20,21** : Mesure d'une HRTF du TNO utilisé pour l'acquisition de la base de Jean-Marie Pernaux<sup>12</sup>

<sup>12</sup> Base de Jean-Marie Pernaux : 8 sujets dans 965 directions

Plusieurs méthodes existent pour obtenir des HRTF. Je les énumérerai seulement.

- La mesure acoustique qui reste la plus utilisée
- La modélisation par éléments finis (*Boundary Method Element* ou BEM). Elle est très performante mais reste limitée aux basses fréquences.

Pour plus d'information, le lecteur pourra se reporter à [Nic10].



**Fig 22** : Exemple de HRTF d'une source sonore située à gauche de l'auditeur

Des jeux assez complets de HRTF et HRIR<sup>13</sup> ont été mesurés par Martin et Gardner sur le mannequin KEMAR. Ils sont largement utilisés par la communauté scientifique. Disposant des réponses impulsionnelles, la réalisation la plus directe du filtrage binaural consiste en une opération de convolution du signal monophonique avec chaque HRIR.

Le coût de calcul d'une convolution dans le domaine temporel étant rapidement rédhibitoire, il est généralement fait appel à des algorithmes de convolution rapide qui consistent à effectuer la transformation dans le domaine fréquentiel à l'aide de FFT.<sup>14</sup>

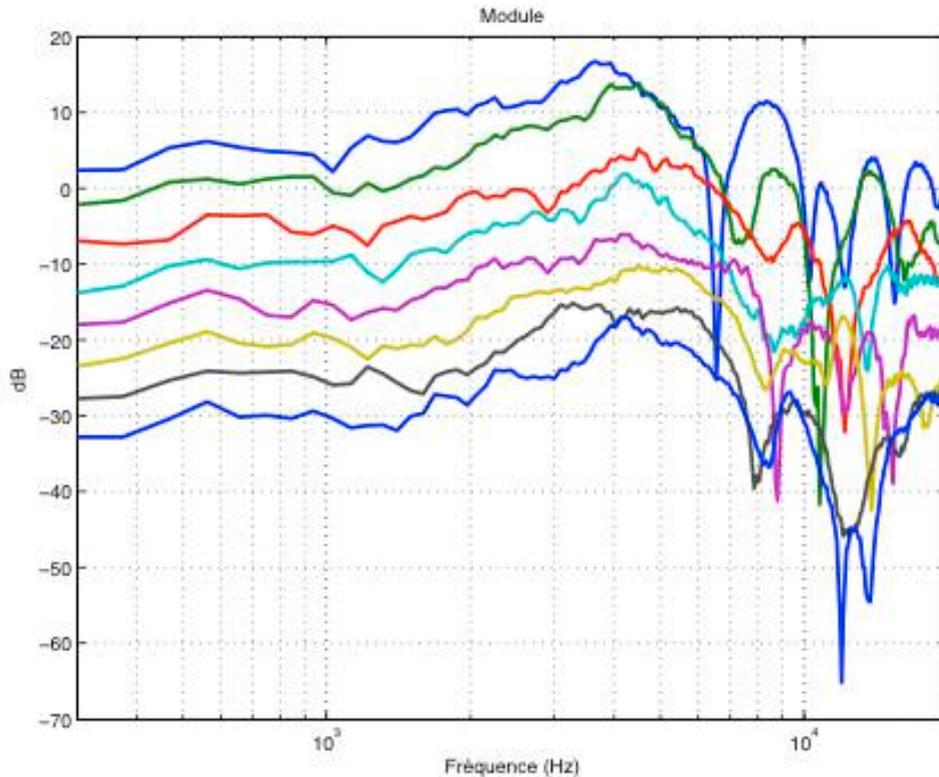
<sup>13</sup> HRIR : équivalent temporel des HRTF

<sup>14</sup> FFT : *Fast Fourier Transform* ou Transformé de Fourier Rapide

Propriétaire	Nombre de sujets	Nombre de directions	Lien
IRCAM (France)	51 sujets	187 directions 	<a href="http://recherche.ircam.fr/equipes/salles/listen/">http://recherche.ircam.fr/equipes/salles/listen/</a>
CIPIC (UC Davis, USA)	45 sujets	1250 directions 	<a href="http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm">http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm</a> [Algazi et al., 2001d]
E. Grassi, University of Maryland (USA)	7 sujets	1093 directions 	<a href="http://www.isr.umd.edu/Labs/NSL/">http://www.isr.umd.edu/Labs/NSL/</a>
Pr. Suzuki, Tohoku University (Japon)	3 sujets	454 directions 	<a href="http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf/index.html">http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf/index.html</a>
Pr. Itakura, Nagoya University (Japon)	96 sujets	72 directions (plan horizontal) 	<a href="http://www.itakura.nuee.nagoya-u.ac.jp/HRTF/">http://www.itakura.nuee.nagoya-u.ac.jp/HRTF/</a>

**Fig 23:** Base de données publiques de HRTF

### 1.4.2. Individualisation?



**Fig 24** : Illustration des variations inter-individuelles des HRTF (base Jean-Marie Pernaux) : Module des fonctions de transfert pour 8 individus dans la même direction. Les HRTF ont été décalées de 5 dB pour une meilleure lisibilité des courbes

Ces courbes sont parfaitement représentatives de la nécessité d'individualiser, de personnaliser les HRTF. Effectivement, d'un individu à un autre on observe des variations très importantes dans le haut du spectre. (creux ou pics). Nous verrons les solutions proposées par les chercheurs dans le chapitre [2.3.2 IRCAM et Orange Labs](#).

Se pose maintenant la question de l'intégration dans le monde professionnel de la production de cette technologie...

## 1. CHAPITRE 2 : Le binaural dans la chaîne de production et de diffusion actuelle

### 1.1. Rappels sur les principaux modèles de restitutions sonores « envisageables »

#### 1.1.1. Le modèle stéréo

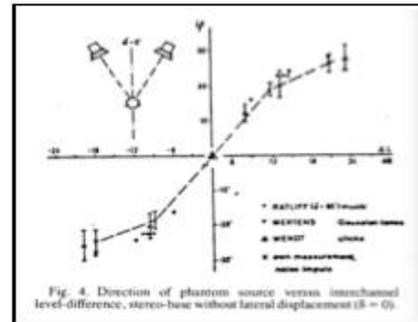
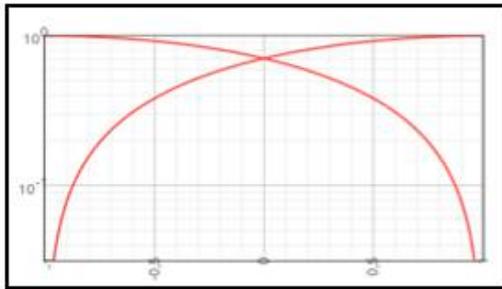
Le modèle stéréophonique est basé sur les indices de latéralisation ILD et ITD vus précédemment. La différence de temps et ou d'intensité permet de localiser une source sonore virtuelle entre les deux haut-parleurs. Ce système permet une restitution sonore sur deux dimensions avec une largeur maximale du plan frontal définie par le triangle équilatéral stéréophonique selon la recommandation ITU-R BS. 775-1. La prise de son est réalisée à l'aide de couple(s) de microphones dit(s) : coïncident(s) pour une stéréophonie par différence d'intensité, distants pour une stéréophonie par différence de temps ou mixte pour une combinaison des deux. Il existe de nombreux systèmes que je ne détaillerai pas dans ce mémoire. On utilisera le panoramique d'intensité selon la loi des sinus ci-dessous. [Car12] :

$$gauche(x) = \cos\left(\frac{\pi}{4}(x+1)\right) \quad droite(x) = \sin\left(\frac{\pi}{4}(x+1)\right)$$

où  $x$  est la valeur, comprise entre -1 et 1, sur laquelle l'ingénieur du son peut intervenir directement par l'intermédiaire d'un potentiomètre.

(1)  $gauche(0) = droite(0) = \frac{\sqrt{2}}{2}$  : 3 dB de moins que le niveau obtenu à gauche pour  $x = -1$

(2)  $gauche^2(x) + droite^2(x) = 1$  : l'énergie globale est conservée



**Fig 25, 26** (à gauche) : Loi de panoramique d'intensité à -3dB en stéréophonie à 2 canaux : atténuation des canaux gauche et droit en fonction de  $x$ .  
 (à droite) : direction perçue de l'image fantôme en fonction de la différence d'intensité entre les deux canaux [Th77]

En agissant sur les contrôles panoramiques, sur les fréquences, sur les retards, ou alors sur d'autres facteurs comme le temps de réverbération, on arrive à placer virtuellement la source sonore à l'endroit désiré. Par exemple, si nous voulons qu'un son paraisse lointain, on peut atténuer les hautes fréquences. Par leur nature même, elles s'atténuent davantage que les basses fréquences au fur et à mesure que la distance augmente. On peut augmenter la réverbération (pour donner l'impression que le son a parcouru un certain chemin avant d'atteindre l'auditeur). Il existe d'autres techniques de mixage permettant à l'auditeur de placer un son dans l'espace sonore frontal qu'on lui propose. Le défaut principal de ce modèle provient du «trou» dans l'espace frontal qui donnera une impression d'image sonore instable au centre. Notamment parce que l'auditeur pourrait ne pas être en permanence au *sweet spot*, le point d'écoute idéal.

C'est un format compatible 5.1 par des processus d'*upmix* ou de *downmix*. C'est-à-dire, respectivement, la conversion d'un flux stéréo en un flux 5.1 ou la conversion d'un flux 5.1 en flux stéréo.

## 1. [Le modèle multicanal 5.1 et ses extensions](#)

Le 5.1 est lui-même une extension du modèle stéréophonique auquel on ajoute un canal central qui stabilise des sources frontales, deux canaux arrières pour les ambiances et effets de salles, et un canal dit LFE<sup>15</sup> pour la gestion des infra basses. C'est dans ce domaine que d'autres modèles émergent comme le 7.1, le 10.2, le 14.1 ou encore le 22.2 développé par la NHK au Japon pour accompagner le «Super High Vision».

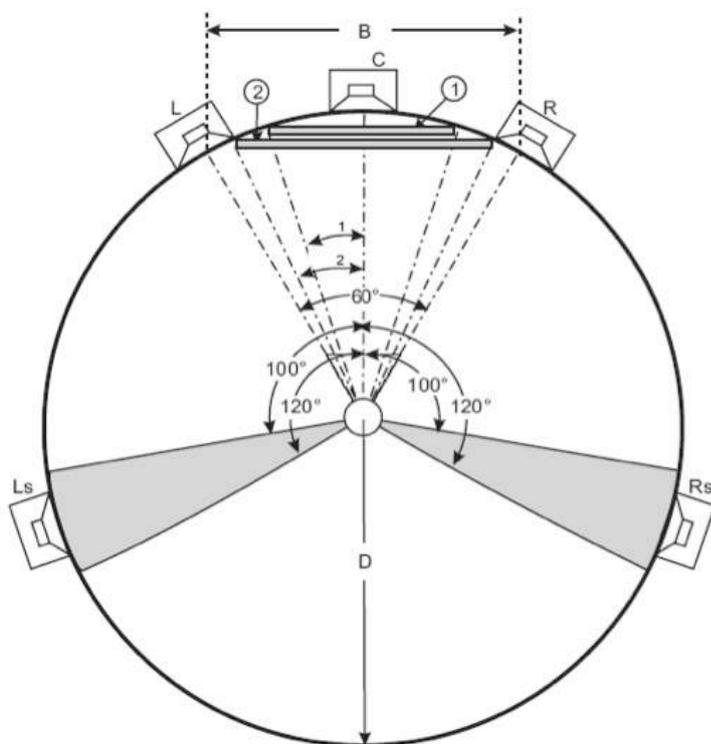
<sup>15</sup> LFE : Low Frequency Effects

Pour ce faire, on reproduit une différence de temps et ou d'intensité entre les oreilles de l'auditeur permettant de localiser des sources virtuelles dans le plan horizontal complet, c'est à dire 360°. Toujours dans deux dimensions, l'ingénieur pour la prise de son pourra utiliser différentes combinaisons microphoniques appelés arbres multicanaux. (INA 5, Fukada-Tree, OCT *Surround*, IRT-Cross, *Hamasaki-Square*) [The01].

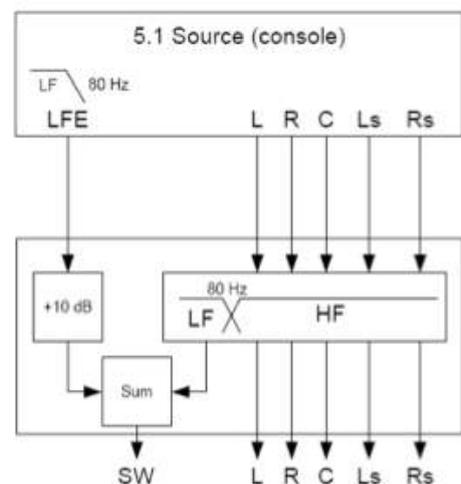
La restitution impose un système de diffusion à six canaux discrets répartis autour de l'auditeur selon la Recommandation ITU-R BS. 775-1. Cinq canaux ont une pleine largeur de bande (20-20000Hz) et le dernier a une bande passante réduite (20-200Hz). Ce canal demande l'utilisation du *Bass Management*<sup>16</sup>.

Le multicanal nous approche d'une d'écoute plus réaliste et naturelle. L'enveloppement et l'immersion sont des paramètres à prendre en compte et il y a un impact non négligeable sur l'espace sonore du mixeur qui offre une plus grande lisibilité de localisation des sources.

[Sig11]. «Le haut-parleur de *Sub Woofer* (ici SW) diffuse ce qui est prévu dans le canal dit LFE et prend éventuellement en charge la gestion des graves des cinq autres haut-parleurs. Il est, en général, possible de régler la fréquence de coupure à 80 Hz ou 120 Hz».

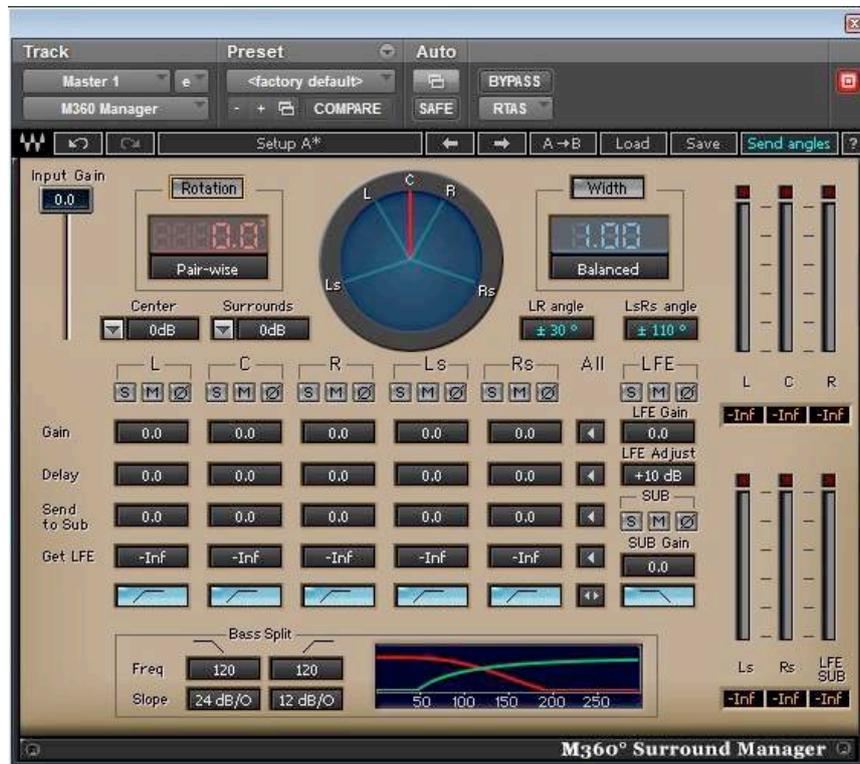


**Fig 27** : Recommandation ITU-R BS. 775-1



**Fig 28** : Bass Management

<sup>16</sup> *Bass Management* : gestion du canal de basse



**Fig 29** : Plug-in utilisé à Radio France dans une configuration de Bass Management (Bass Split : 120 Hz)

Le format 5.1 est compatible avec la stéréo par des procédés d'*upmix* ou de *downmix*. Il est compatible binaural par l'intermédiaire d'un autre procédé de *downmix* appelé *downmix* binaural. Il est également compatible avec des systèmes plus complexes (Ambisonics<sup>17</sup>, HOA<sup>18</sup> ou WFS<sup>19</sup>).

Les technologies Ambisonics, HOA et WFS sont encore majoritairement à l'état expérimental actuellement et ne sont pas envisageables au sein de productions à Radio France. Mais ce sont des techniques de spatialisation 3D à suivre de près.

Le *sweet spot* reste une contrainte en multicanal, tout comme en stéréophonie et nous ne disposons toujours pas de la spatialisation en élévation en 5.1. A noter également un phénomène bien connu des mixeurs qui est l'instabilité des sources

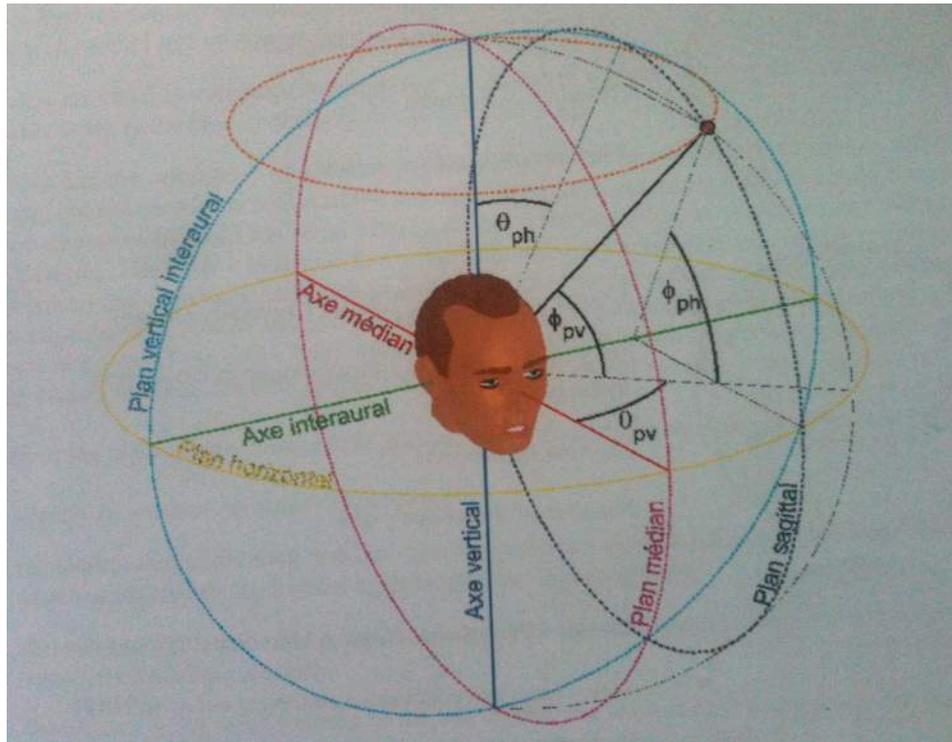
<sup>17</sup> Ambisonic : Modèle basé sur une décomposition mathématique de l'onde acoustique utilisant la base des harmoniques sphériques (fonctions propres de l'équation des ondes acoustiques en géométrie sphérique). Il correspond à la décomposition limitée à l'ordre 1, tandis que HOA en est la généralisation aux ordres supérieurs

<sup>18</sup> HOA : *High Order Ambisonic*

<sup>19</sup> WFS : *Wave Field Synthesis* : Modèle basé sur le Principe de Huygens : recombinaison d'une onde acoustique par superposition d'ondelettes (décomposition physique de l'onde acoustique)

latérales. Les travaux d'Alexis Baskind, Xavier Meunier et Jean-Marc Lyzwa tiennent compte de cette difficulté. Nous développerons cette idée dans le paragraphe suivant. [LB09].

### 1. Le modèle binaural et son extension «transaurale».



**Fig 30:** Coordonnées sphériques

Cette figure de coordonnées sphériques servira à aider les auditeurs à se positionner dans l'espace 3D lors des tests perceptifs et permettra de se familiariser avec les axes et les plans importants. Comme on l'a vu précédemment, ce modèle est basé sur l'imitation de la perception auditive et cherche à reproduire au niveau des tympans de l'auditeur les indices de localisation perçus en situation d'écoute naturelle. Une dimension supplémentaire apparaît dans ce modèle de restitution sonore.

La prise de son est réalisée à partir d'une tête artificielle ou d'un système microphonique se rapprochant des caractéristiques physiques et morphologiques d'un torse et d'une tête. Nous étudierons plus tard les outils de spatialisation virtuelle. La restitution de ce modèle binaural implique donc l'utilisation d'un casque. Néanmoins, «en 1963, Atal et Schroeder proposent une méthode pour diffuser des signaux binauraux via des haut-parleurs [SA63], puis en 1989, Cooper et Bauck reprennent et approfondissent ce procédé qu'ils appellent «transaural stéréo» [CB89].

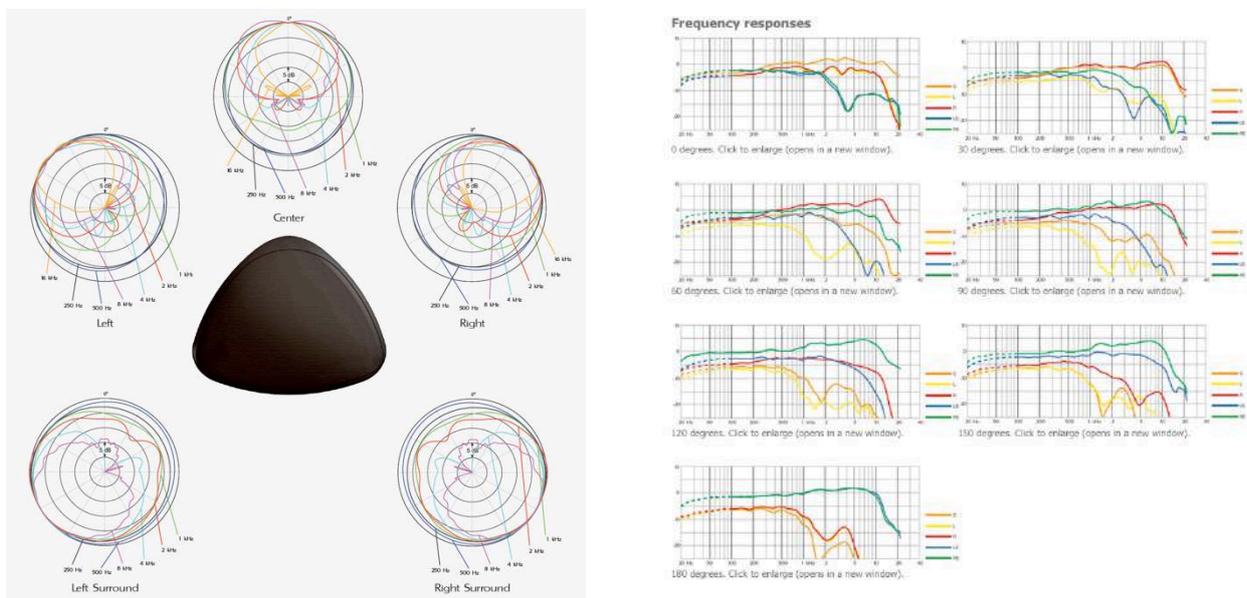
## 1. [Du 5.1 au binaural à Radio France](#)

### 1.1. [Contexte](#)

A l'origine, Antoine Lehembre, délégué à la Direction de la Production et des Antennes de Radio France souhaitait remonter un groupe de travail sur le multicanal. L'objectif était de pouvoir travailler en 5.1 tout comme il le faisait en stéréo avec un surcoût de production de 10 à 15 % maximum.

L'essentiel de la production multicanale étant des opéras sur DVD, l'idée est de développer le 5.1 sur l'ensemble de la production de Radio France avec comme condition principale un apport esthétique par l'espace sonore multicanal sur le sujet traité.

On présentera quelques exemples de productions réalisées à la radio afin de situer l'arrivée du multicanal, et du binaural. Le reportage est un sujet assez simple à mettre en production. «Interception» produira une émission sur le thème de la «Transhumance». Il faut tenir compte de l'aspect pratique du système d'enregistrement dans un contexte comme le reportage. Il sera donc réalisé avec le système de microphones DPA 5100<sup>20</sup>, les 5 capsules sont regroupées de façon compacte et la légèreté permet une maniabilité très intéressante. Le *SoundDevice 788T* servira d'enregistreur.



**Fig 32,33** : Caractéristiques techniques du DPA 5100

<sup>20</sup> 5 microphones omnidirectionnels. Les 3 frontaux sont coïncidents avec un bafflage acoustique entre les capsules, lesquelles sont chargées par des tubes à interférences (DiPMic), les micros arrières sont distants du triplet frontal. Le canal LFE est issu de la somme des capsules L et R (atténué de 10 dB)

Un deuxième exemple très bien reçu par les auditeurs fut la diffusion en public des Folles Journées de Nantes. Le concert était retransmis le lendemain sur un système multicanal Genelec calibré par Trinnov Audio<sup>21</sup>. L'audience avait une jauge limitée pour permettre à l'auditeur de profiter un maximum de l'espace sonore multicanal. Le programme est présenté en [annexe 1](#) de ce mémoire.

Thomas Baumgartner, producteur sur France Culture, travaillera dans ses «Créations Radiophoniques» sur un sujet extrêmement intéressant à réaliser en multicanal, le «Colin-Maillard». Un système de prise de son spécifique sera mis en place. Hervé Dejardin ajustera sur la tête d'un comédien un bandeau auquel il fixera cinq Sennheiser cardioïdes HF. Une capsule sera positionnée au centre, deux capsules à l'avant respectivement à 45° à gauche et à droite et deux dernières capsules à l'arrière respectivement entre 110° et 120°.

La première binauralisation de cette production sera réalisée par le *Smyth-Research* avec la fonction de transfert du mixeur dans son home studio. Une remarque importante, Hervé Dejardin n'utilisera aucune composante acoustique.

Quel que soit le type de production, le monitoring de celui-ci est indispensable afin de pouvoir réaliser des corrections d'éventuelles situations non maîtrisées pendant l'enregistrement. Malheureusement, il n'est pas toujours possible de disposer d'une cabine de monitoring 5.1 en situation extérieure sauf cas particulier, comme ici à *La Gaîté Lyrique*.



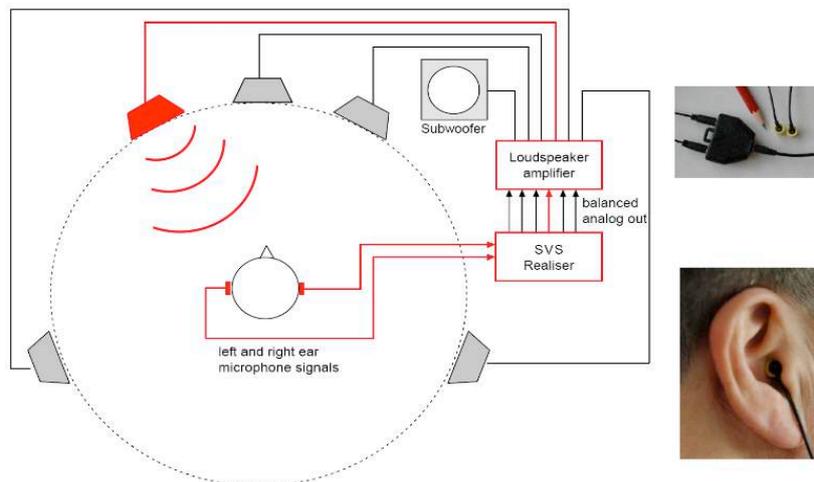
**Fig 34** : Cabine mobile 5.1 à la Gaîté Lyrique.

<sup>21</sup> Trinnov audio : Société qui développe des solutions audio innovantes.

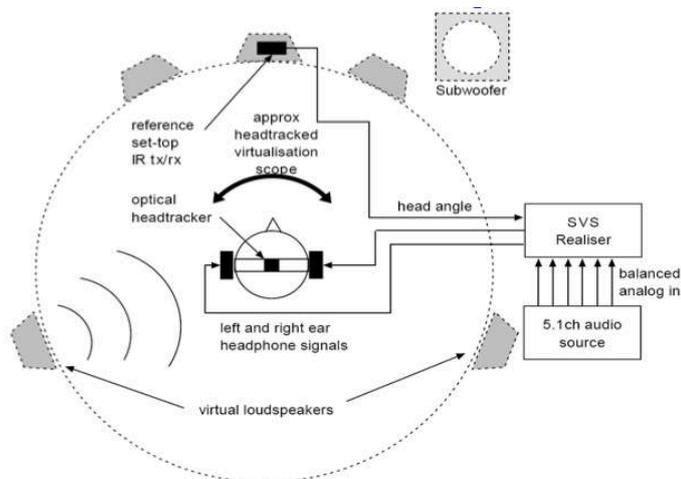
## 1. Contrôle au casque d'une prise de son multicanale

Le système Smyth Research ou Smyth Virtual Surround (SVS) est un algorithme de virtualisation audio pour casque stéréo standard conçu pour reproduire le son d'enceintes dans un environnement connu. Elle tient compte des effets de filtrage de chaque composant modifiant le signal (enceintes, salle, auditeur et casque).

C'est un système «réduit» de mesures binaurales avec un moteur de convolution multicanal temps réel et un système de *head-tracking*. En relevant la fonction de transfert personnalisée de l'ingénieur du son dans sa cabine de mixage, il lui permet d'emporter sous forme de données la réponse impulsionnelle du studio. En effet, le principe est de poser 2 micros dans les oreilles de l'ingénieur et d'enregistrer ce qu'il entend en chaque point de son système multicanal.

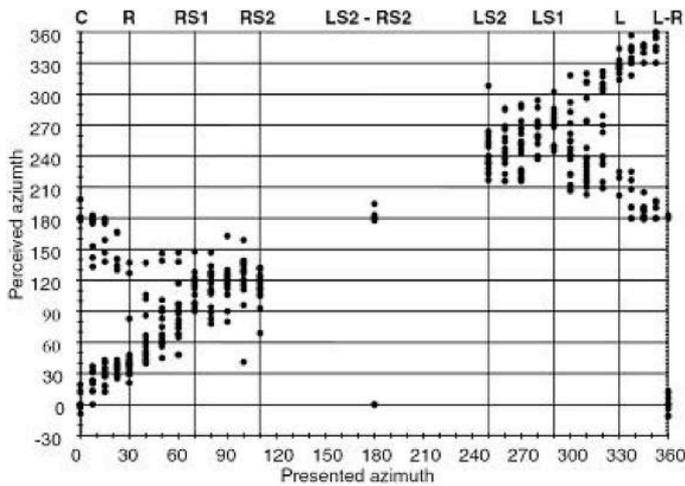


**Fig 34:** Schéma explicatif du SVS

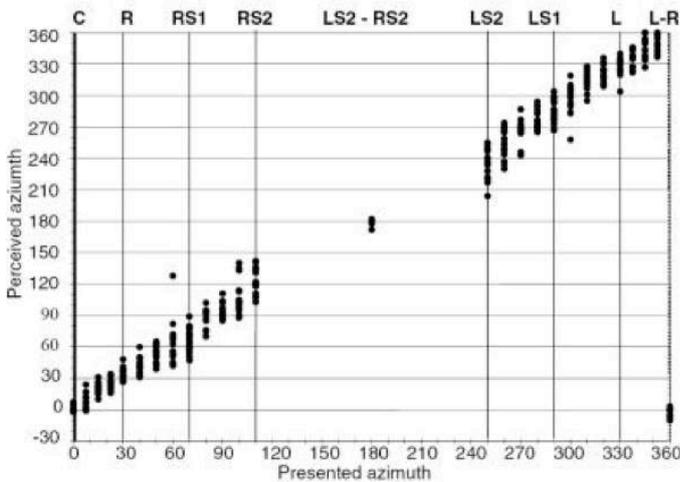


**Fig 35:** Technologie «head tracking» du SVS

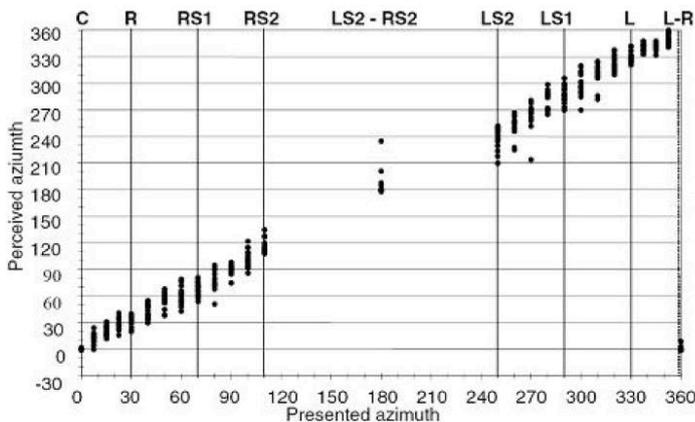
Ici quelques résultats expérimentaux tirés de l'article présenté à la 122<sup>ème</sup> Convention AES à Vienne [PKG07] concernant le «*head-tracking*». Des études ont montré qu'il était nécessaire d'avoir une latence de suivi d'au moins 60 millisecondes pour percevoir une image sonore stable.



**Fig 36** : Localisation à l'aide d'une tête de mannequin fixe sans head-tracking dans le studio



**Fig 37** : Localisation en utilisant une tête de mannequin avec un suivi en tête dans le studio



**Fig 38** : Localisation en utilisant leurs «*propres oreilles*» dans le studio

Une expérimentation intéressante a été réalisée avec le système «SVS». Pascal Besnard, ingénieur du son, mixeur au DPS de Radio France sera le cobaye de cette expérience.

«Un mixage multicanal 5.1 sera considéré comme référent sur enceintes de studio. L'idée est de construire le même mixage avec des casques de différentes conceptions (ouverts, fermés, électro-statiques, dynamiques). Les mixages ont été effectués volontairement à des périodes espacées, afin d'avoir si possible une approche nouvelle. Comparer le résultat sur enceintes par rapport au mixage référent sur haut-parleurs. L'évaluation portera sur les différences de timbres, la profondeur des mixages, la localisation des sources, l'homogénéité de la balance, le placement des effets, le rendu du grave, et le dosage des fréquences sub au casque. Mesurer la valeur moyenne du niveau acoustique lors du mixage sur une tête artificielle, ainsi que le temps de travail au casque. Analyser la fatigue du mixeur, et comparer le temps de récupération par rapport à une écoute classique.»

Cependant on note les limites d'une telle technologie par le peu de mesures existantes : cinq points de diffusion du système de restitution 5.1. Et on en imagine la richesse en intégrant les propres HRTF du mixeur tirée de mesures «dignes» de ce nom. (cf. [1.4.1 Obtenir des HRTF](#)).

## 1. Objectifs des diffuseurs et moyens d'y parvenir

Plusieurs entités ayant au préalable travaillé sur le sujet se sont rencontrées afin d'établir un groupe dans le domaine des techniques binaurales et de collaborer au développement de celui-ci. Afin de développer l'accès à ces techniques en production et en diffusion pour le grand public, un projet de consortium a pris forme. Les travaux de recherches sont principalement tournés vers l'individualisation des HRTF.

### 1.1. Radio France et France TV

Dans la mesure où l'écoute multicanale chez soi est mal répandue et surtout mal calibrée, Antoine Lehembre exprime l'intérêt porté par Radio France à l'écoute binaurale, en particulier en situation de mobilité qui correspond bien à ce media radio. La ressource que Radio France peut apporter au projet concerne sa capacité à créer des éléments sonores diversifiés à tester. Elle dispose également de lieux d'écoute appropriés et d'un large panel d'auditeurs experts. Cette ressource est coordonnée au sein de la Cellule Méthode et Qualité de la DPA. Les expériences et productions réalisées dans le cadre de ce consortium sont détaillées dans le chapitre 3, avec l'ensemble des tests perceptifs liés et l'analyse des résultats.

Matthieu Parmentier, représentant la Direction Innovation et Développements, définit les deux types de populations d'exploitants concernées par l'écoute en binaural à France Télévision :

- ceux qui travaillent en direct sur des productions extérieures type Roland Garros, avec un monitoring HRTF ;
- les techniciens vidéo, qui doivent valider très rapidement, dans un environnement bruyant, des éléments qui arrivent du monde entier, avant d'être poussés vers la régie finale.

France Télévisions a le projet de diffuser huit chaînes en multicanal en 2012. Dans la mesure du possible, la production doit être 5.1 norme ITU. Le home-cinéma restant peu développé, le binaural est porteur d'espoir pour toucher le grand public sur des vecteurs d'écoute individuels, notamment les *smartphones*.

16 canaux de production sont en général disponibles, et on peut imaginer alimenter en diffusion :

- Le départ antenne « premium » 5.1
- Le départ stéréo
- Un départ décliné « mobile » binaural issu du 5.1 « premium »
- VOD : promouvoir des solutions astucieuses en maîtrisant un player propriétaire.

Dans la réalité d'aujourd'hui, le casque Smyth Research utilisé à France Télévisions pose des problèmes de connectiques, considérées comme non-professionnelles. Mais la qualité audio est satisfaisante. Des contacts avec la société A-Volute ont permis d'évaluer des produits grand public (HRTF standard) jugés pour l'instant trop pauvres pour une exploitation professionnelle. L'étude est néanmoins poussée pour le déploiement d'un tel traitement sur un départ décliné "mobile" binaural. Un plug-in VST de HRTF personnalisable est en cours de développement, financé dans le cadre du projet monitoring 5.1 en vue d'un usage professionnel.

## 1. [IRCAM et Orange Labs](#)

L'IRCAM et Orange Labs travailleront en parallèle sur l'idée d'individualisation des HRTF. Olivier Warusfel, qui dirige l'équipe Acoustique des salles de l'Ircam (rattachée à l'Unité mixte de recherche du CNRS "Sciences et technologies de la musique et du son») propose lors de la première réunion des alternatives à la mesure des fonctions de transferts. Par exemple la modélisation physique par un scanner précis de la tête de chaque individu, une solution certes idéale mais guère plus accessible que les mesures des HRTF. Des travaux ont également porté sur le recours à des photographies de la tête sous plusieurs angles et d'en déduire une modélisation des filtres binauraux.

Il évoque ensuite plusieurs systèmes d'adaptation ou de corrections individuelles :

- L'adaptation discrète où l'on sélectionnerait parmi plusieurs possibilités la tête qui nous convient le mieux, à partir de sons tests spatialisés dans plusieurs directions.
- L'adaptation anthropomorphique où à partir de certaines dimensions caractéristiques de la tête (rayon, profondeur et orientation du pavillon), on pourrait estimer une fonction paramétrique reliant grandeur acoustique et anthropomorphisme (exemple de calcul de l'ITD Woodworth [WR54] et ses extensions). On pourrait effectuer un morphisme sur des HRTF d'une tête générique ou une tête individuelle en fonction des paramètres anthropomorphiques (cf. Middlebrooks) et faire une mise en correspondance statistique de l'espace morphologique et de l'espace des HRTF (ITD ou spectre). Cela permet de déterminer un ensemble de têtes à priori voisines du sujet.

Une étudiante New-Yorkaise, Areti Andreopoulou, développe un travail de recherche sur le regroupement de HRTF selon des paramètres particuliers.

La contribution d'Orange Labs vise également une méthode de personnalisation de HRTF s'appuyant sur les travaux de recherches déjà existant du mémoire HDR de Rozenn Nicol. Aujourd'hui la piste qui lui semble la plus intéressante est basée sur une méthode d'acquisition de la morphologie du pavillon. C'est à dire faire une comparaison avec des familles de fonctions de transferts, et optimiser cette fonction de transfert sur la base du différentiel entre le pavillon d'origine et celui de la base de donnée.

Des partenaires privées comme Trinnov Audio, Sonic Emotion, A-Volute ont été contactés pour le développement de l'application. Une autre réunion avec Fraunhofer a eu lieu en ce qui concerne son codec HE-AAC+MPS<sup>22</sup> (cf **fig 40**).

## 1. [Les formats](#)

Pour la production des Folles Journées de Nantes, citées précédemment, un moyen de conserver la compatibilité était de lire par VLC dans le format de compression audio wma<sup>23</sup> (192 kbits / 24 bits / 48 kHz) où il est possible de paramétrer le downmix dans les métadonnées.<sup>24</sup>(cf **fig 39**).

La question d'un double mixage s'est posée dans un premier temps mais l'idée fut vite abandonnée. En revanche un bus 5.1 et un bus stéréo ont été réalisés afin de relever en fin de live «l'offset» trouvé par l'ingénieur du son. En quelques sortes, ils géraient le *downmix* eux mêmes. Une étude semblerait montrer que l'HE AAC 160 kb/s, serait équivalent à un débit de 448 kb/s Dolby Digital. En dessous de ce chiffre, le

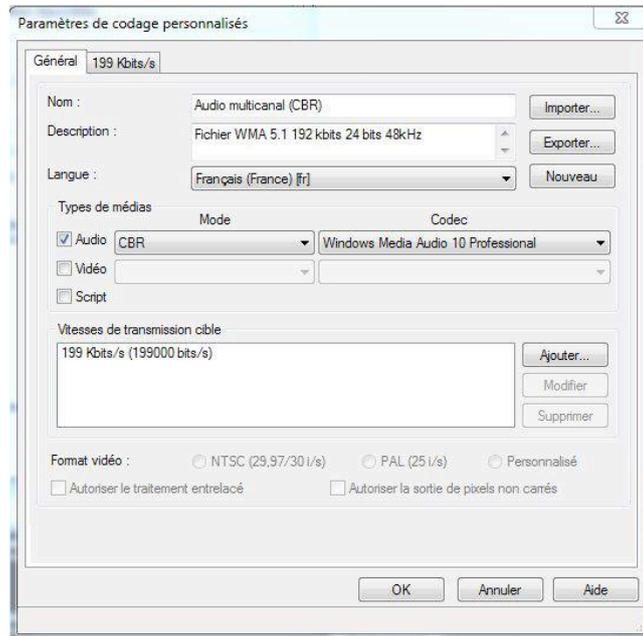
---

<sup>22</sup> HEAAC + MPS : *High Efficiency Advanced Audio Coding + Mpeg Surround*

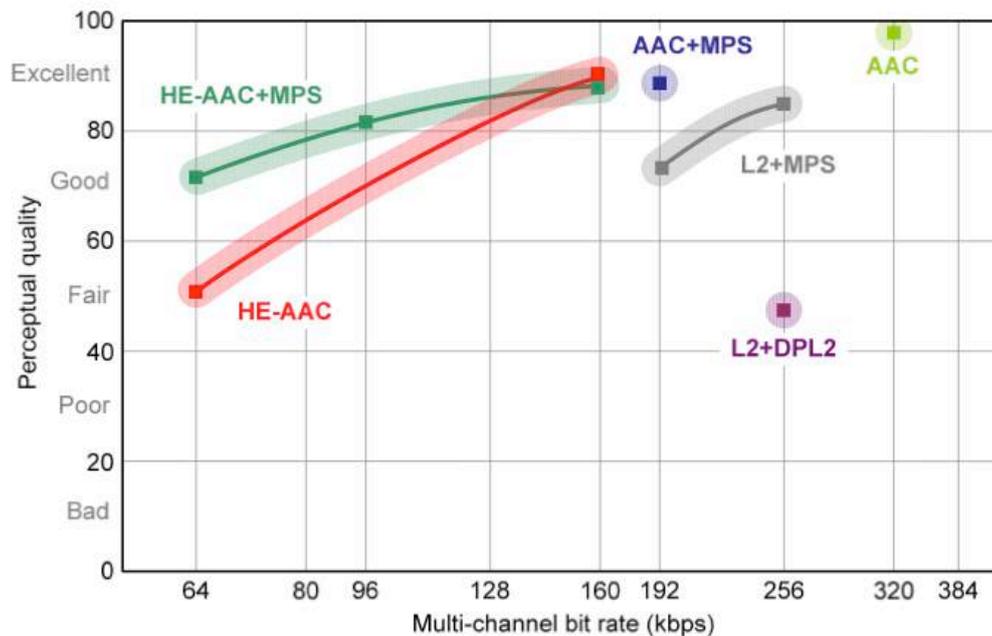
<sup>23</sup> WMA : Windows Media Audio

<sup>24</sup> Métadonnées : donnée servant à définir ou décrire une autre donnée.

codec HE AAC + MPS serait meilleur, et au dessus, 5.1 HE AAC porterait le chapeau. [RBH07] (cf **fig 40**).



**Fig 39:** Fenêtre de paramètres de codage audio wma



**Fig 40 :** La qualité subjective (en moyenne sur extraits) en fonction du débit binaire pour différents codeurs

L'HTML 5 par défaut devrait pouvoir envoyer du binaural (l'idée serait que chacun puisse insérer sa fonction de transfert). A terme, il serait intéressant de travailler directement avec les développeurs de navigateurs. Jusqu'à l'iOS 5 (sur tablette), Apple décode l'HEAAC, mais fait directement un downmix, qu'on récupère par HDMI. Etrange...

Si on pouvait décoder dans les appareils, et récupérer des données PCM en 5.1 par la sortie HDMI en «*pass thru*», tous le parc des amplificateurs home cinéma de décodages serait ouvert.

A titre indicatif, aujourd'hui, une Livebox (Orange) sait décoder de l'HEAAC mais le fait en stéréo. Elle ne gère que les flux AC3 ou AC3+, etc. La Freebox sait décoder l'HEAAC en 5.1. La box de chez Bouygues ne gère que la stéréo sur l'HEAAC et la SFRbox n'a toujours pas donné de réponse. Orange Labs travaille actuellement sur la création de son propre player.

L'objectif final serait un système extrêmement simple et automatisé pour que l'auditeur n'ait rien à faire. On envoie un signal multicanal codé chez l'auditeur. Ce signal s'adapte à son mode d'écoute : s'il possède un système multicanal, les voies audio sont correctement affectées. S'il n'en possède pas, le flux passe en binaural (ou son extension transaurale si l'auditeur veut écouter sur haut-parleurs). Le but du consortium, à ce stade est d'intégrer dans cette chaîne de codage, les HRTF personnalisées...

### **Perspective Européenne ?**

Radio France a rencontré la BBC sur son site R&D de Manchester. La BBC procède à des études en partenariat avec des universitaires. Elle reste essentiellement axée sur le son à l'image, et a pour l'instant produit moins de programmes binauraux purement audio que Radio France. Une collaboration et des échanges entre les deux entreprises sont envisagés pour l'avenir.

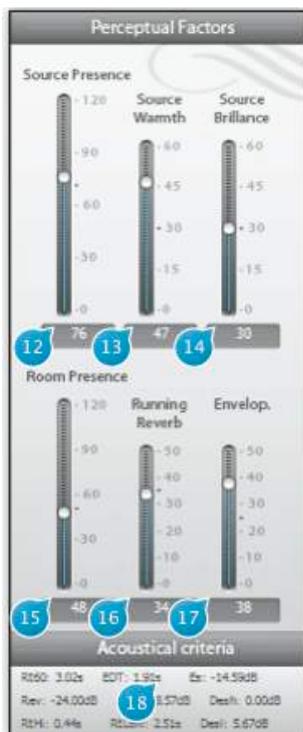
A noter également une énorme avancée technologique des japonais qui ont mis au point une chaîne de production multicanale 22.2 avec la création de panoramiques 3D inclus dans les consoles de type Lawo et un système de monitoring sur casque binaural.

## 1. Outils numériques de post-production binaurale

Plusieurs implémentations de techniques binaurales pour le placement de sources sonores ont vu le jour depuis les années quatre-vingt-dix.

### 1.1. Ircam Tools et Flux : Spat

Le Spat est un des précurseurs du projet Européen LISTEN<sup>25</sup>. Ce logiciel a été créé par l'IRCAM sur la plateforme MAX/FTS puis Flux a continué le développement graphique du plug-in. En des termes simples, il permet de contrôler la localisation d'une source dans un espace virtuel, en jouant à la fois sur l'incidence de l'effet de salle, au moyen de critères acoustiques précis tels que le montrent les figures ci-dessous. Mais il peut être utilisé sur un nombre important de configurations multicanales. Dolby Stereo, binaural, quatre, six ou huit enceintes. La densité des paramètres existants demande une certaine relecture du manuel. Je définirai les plus importants concernant les manipulations liées au binaural.



**12** : Présence de la source : importance du son direct en ce qui concerne le son réverbéré. Ce n'est pas seulement équivalent à un rapport sec / humide. La distance, le rayon entrent en jeu également.

**13** : Présence des BF de la source.

**14** : Présence des HF de la source.

**15** : Importance de la réverbération par rapport à la source.

**17** : Enveloppement correspond à la notion de la perception de la façon dont l'auditeur immergé

**18** : Critères acoustiques comme le RT60, EDT, Es, Rthi, Rtlow...<sup>26</sup>

**Fig 42** : Facteurs perceptifs

<sup>25</sup> LISTEN : L'objectif est d'immerger les visiteurs dans une scène sonore qui prolonge l'espace réel dans lequel ils évoluent.

<sup>26</sup> RT60 : Reverberation Time / EDT : Early Decay Time / Es : Early Sound / Rthi : Reverberation Time in HF / Rtlow : Reverberation Time in LF

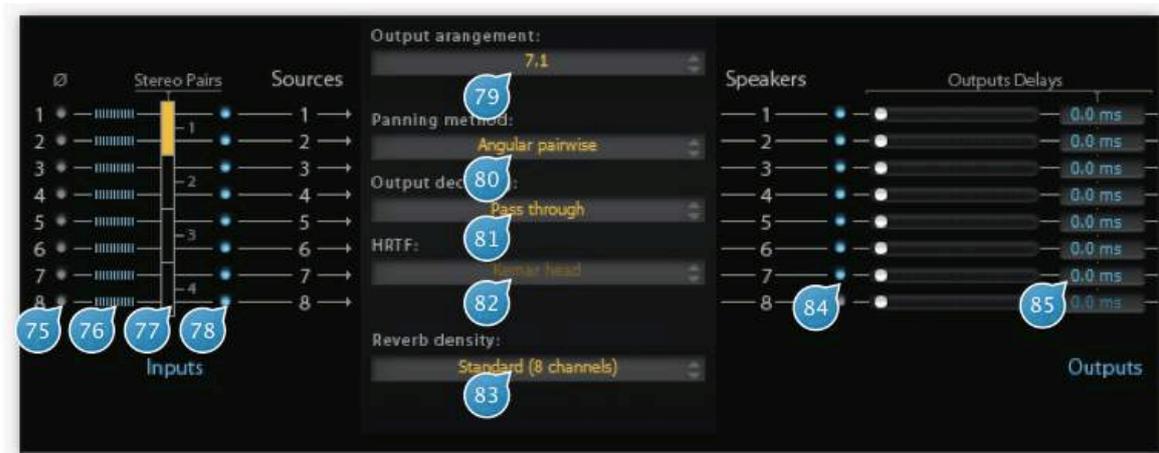


Fig 44 : Setup spat

(Configuration utilisée : Sources : 6 / Output Arrangement : Binaural / HRTF : à choisir)

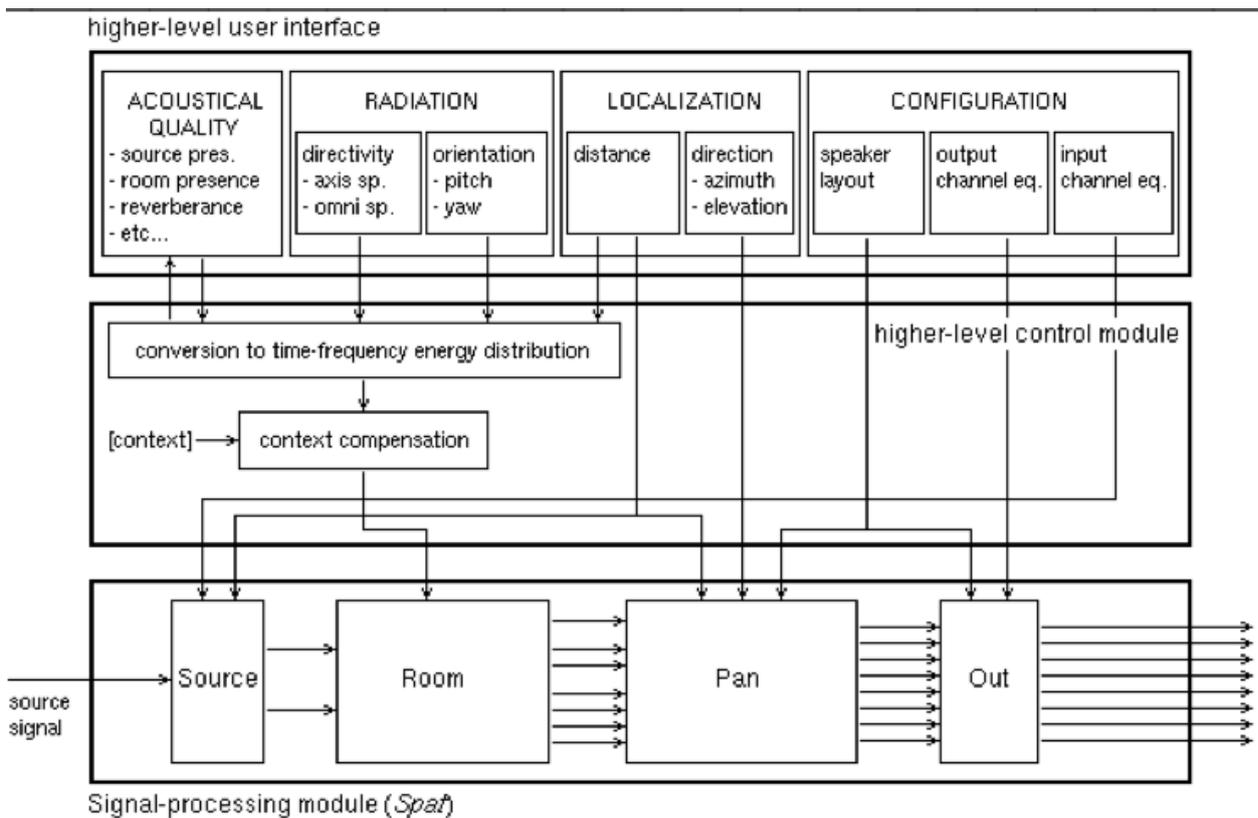


Fig 45 : Schéma synoptique du spat.

La précision des paramètres font de ce plug-in un outil extrêmement puissant. Un autre point très intéressant est la possibilité de choisir les HRTF et même d'en importer, (si le format correspond...) La banque déjà existante est peut être un peu trop étendue. En revanche, nous avons été confrontés à de gros problèmes d'instabilité (graphique ?) sur la station de travail Protools HD 9 du studio 132. Le passage de la fenêtre du plug-in à la fenêtre d'edit ou de mixage de Protools engendrait fréquemment un bug graphique qui nous obligeait à redémarrer le séquenceur. La source du problème reste encore énigmatique. L'expérience nous a conduit à ne pas utiliser la *Virtual Room* dont la coloration est réellement gênante dans un mix. C'est une observation que l'on retrouve souvent quel que soit le plug-in. Ce qui malgré l'astuce d'ajouter une réverbération artificielle 5.1 reste un souci car l'absence de virtualisation d'acoustique limite la sensation d'externalisation...

### 1. [Longcat : H3D](#)

Développé par Longcat, voici un exemple de la présentation du plug-in ou pour être plus précis, six plug-in<sup>27</sup> ouvert sur une session protools. Le H3D est simple d'utilisation et stable, mais il ne dispose que de peu de paramètres. On choisit le type de sources d'entrée (Input Mode), on peut manipuler la largeur de la stéréo (*Stereo mode* en degré), et sa position dans l'espace 3D (3D position).

Le rendu spatial est cependant très intéressant. L'option 3D View fonctionne parfaitement en automatisation. Il permet d'automatiser des déplacements de sources dans l'espace tridimensionnel. J'utiliserai cette fonction dans la réalisation d'une petite chronique sur «Le réveil» dont je parle dans le chapitre 3 ...



**Fig 45 46** : Configuration d'encodage binaural à partir d'un mixage 5.1

<sup>27</sup> Virtualiser le canal LFE n'a que peu d'intérêt perceptif si ce n'est ajouter une coloration non naturelle. Nous avons fait le choix dans le processus de binauralisation de le muter par défaut.



Fig 47: Zoom sur la fenêtre de paramètres

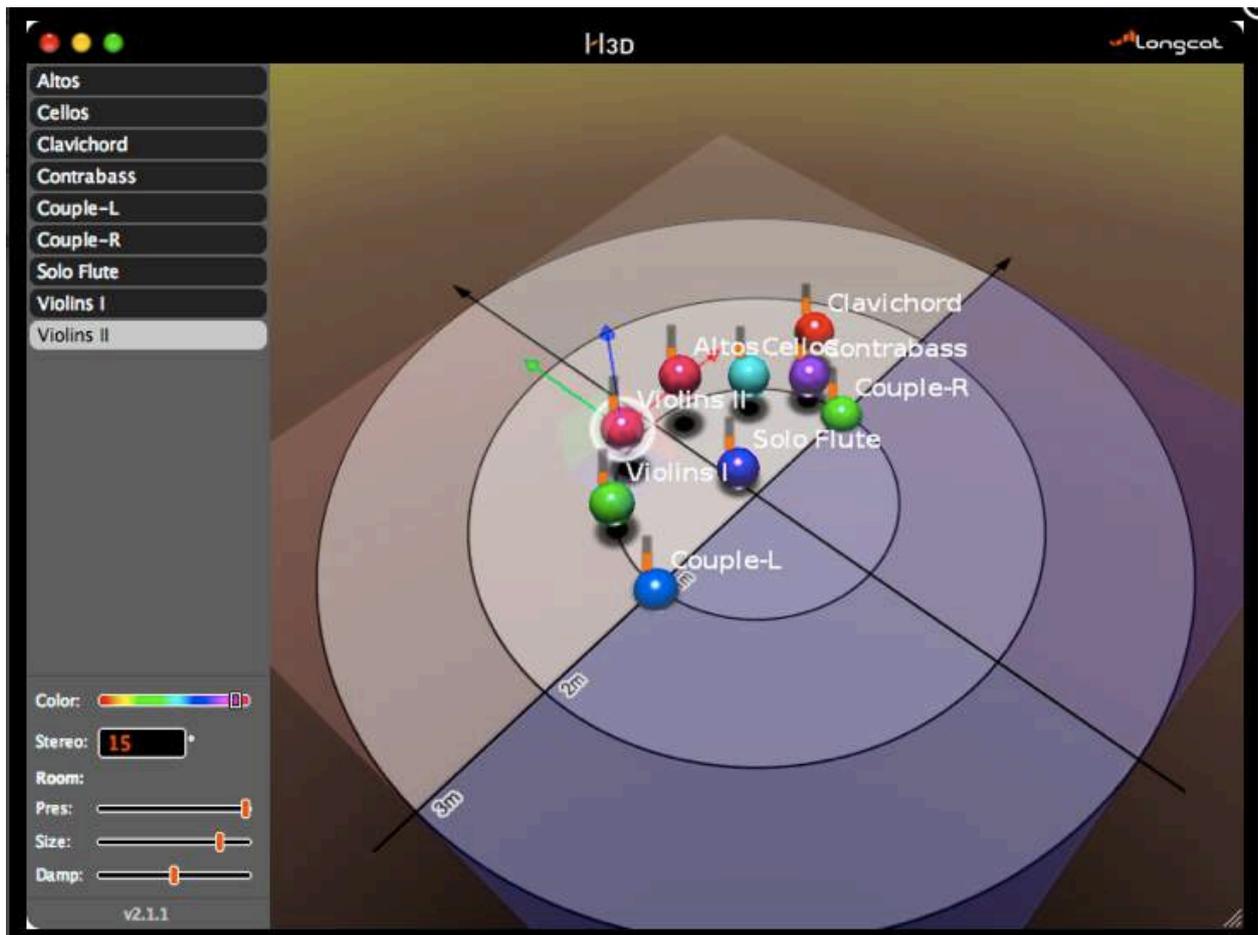
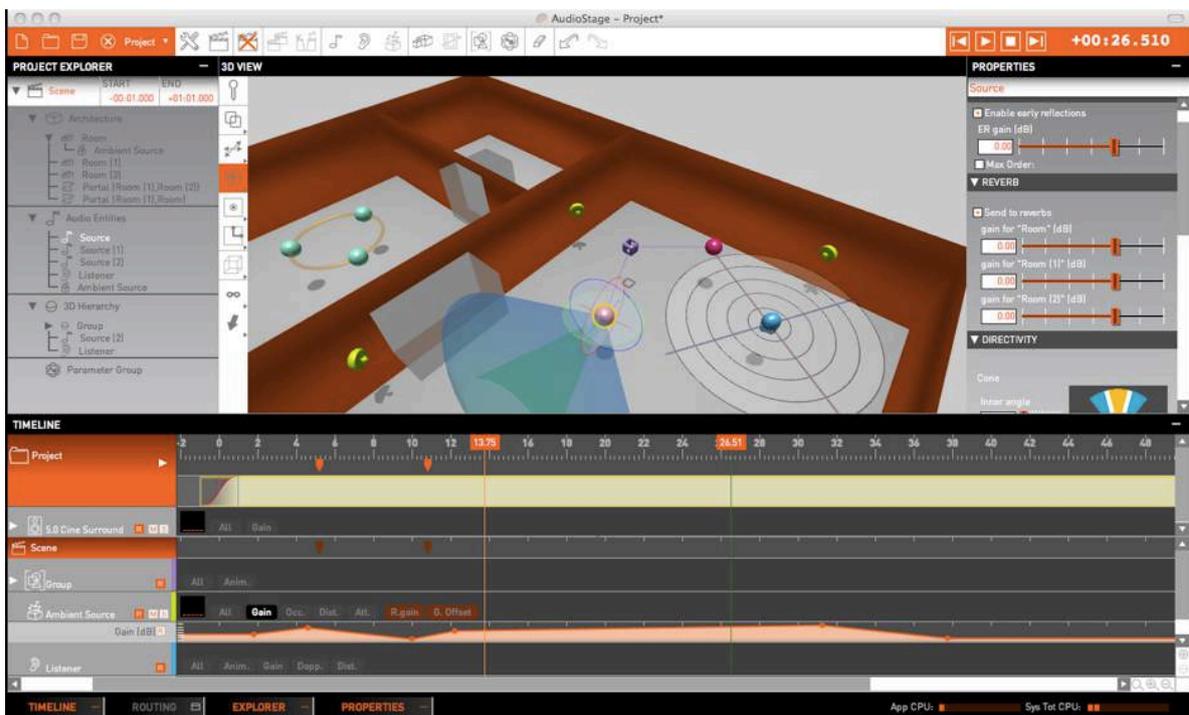


Fig 48: Configuration d'une scène sonore globale

Il est tout à fait imaginable et même conseillé de placer des microphones d'ambiances à l'arrière de la scène sonore pour récupérer une réverbération naturelle et de positionner ces sources microphoniques dans le demi-plan arrière de cet exemple. (cf. **fig 48**). Cela permet de s'affranchir des paramètres de *Room-Virtuel* encore non convaincants dans ce plug-in. Dans cette optique, Longcat développe le grand frère du H3D, plutôt destiné au jeu, nous ne l'avons pas encore testé. (cf. **fig 49**).

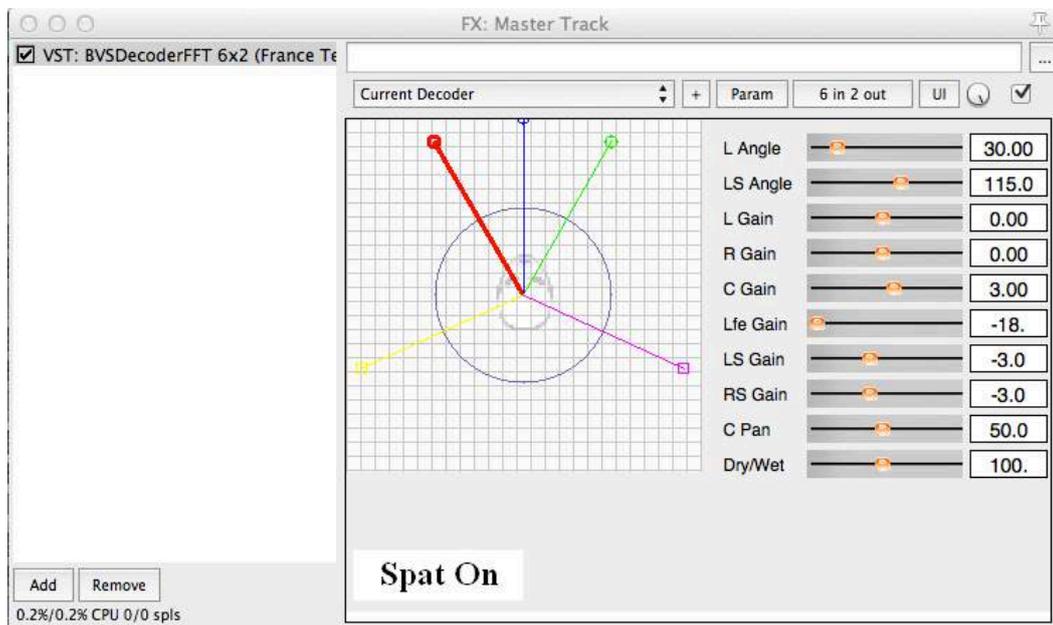


**Fig 49** : Interface AudioStage de Longcat

## 1. [Orange Labs](#)

Marc Emerit qui travaille pour la recherche et le développement chez Orange Labs nous propose de tester ses plug-in. On remarque que l'on peut agir sur les angles et les coefficients d'atténuation. D'après lui ils ont été réalisés pour qu'il y ait un minimum de traitement et de coloration.

Dans les HRTF, il utilise deux filtres, un par oreille. Un filtre côté oreille éclairée qu'on appelle composante ipsilatérale et un autre filtre côté opposé, qu'on appelle zone d'ombre ou composante contralatérale. Pour éviter la coloration, il divise chacun des filtres par la composante ipsilatérale. Par conséquent on obtient juste un dirac côté éclairée soit la contribution du haut parleur sur cette oreille.



**Fig 50** : Plug-in développé par Marc Emerit

Il élimine les composantes monaurales, c'est à dire, les composantes présentes en même temps dans les deux oreilles. L'inconvénient est que l'on ne peut pas mettre un haut-parleur virtuel pour le centre. On a donc la même HRTF à gauche et à droite. Les HRTF utilisées sont celles d'une personne ayant réalisé les mesures de Jean-Marie Pernaux. Elles se révélaient convaincantes pour plusieurs personnes. En mixage, l'intérêt de cette technique est la grande compatibilité avec le *downmix* ITU où il n'y a pas de binauralisation.

En résumé, il s'agit d'un *downmix* avec des gains de 5.1 vers stéréo auquel on ajoute un signal qui correspond à la partie controlatérale. Le résultat ressemble beaucoup à un filtrage passe-bas retardé qui implique une faible coloration mais une spatialisation bien perceptive.

## 1. [Autres](#)

Il existe de nombreux plug-ins ou processeurs de spatialisation tridimensionnelle, que je n'ai pas encore pu tester correctement. Mathieu Aussal, travaille actuellement avec DMS<sup>28</sup> sur un prototype que j'ai eu l'occasion d'essayer sur une très courte durée et dont le rendu perceptif est très prometteur. N'étant pas encore commercialisé, il a tenu à ne pas me dévoiler le fond de son travail basé sur une analyse en composantes principales.

<sup>28</sup> DMS : Digital Media Solution

Connu pour son célèbre extrait «*Barber Shop*» le *QSound*, comme ces confrères est une technique de positionnement du son dans un environnement 3D. Le processeur permet de mettre en œuvre un algorithme chargé de réaliser des effets sonores 3D à l'aide de plusieurs sources monophoniques. En 1996, le processeur *QSound* entre en conformité avec les normes Dolby Digital (AC3) et Pro Logic. Ce processeur sous forme de puce intègre diverses cartes son pour PC puis devient Q3D en version software pour directX. Les albums *Amused To Death* de Roger Waters, *Pulse* de Pink Floyd, *Nothing Like the Sun and The Soul Cages* de Sting, *The Immaculate Collection* de Madonna ont utilisé cette technologie.

## 1. [Problèmes soulevés](#)

### 1.1. [Niveaux](#)

Dans une chaîne de production ou une chaîne de diffusion, on a une parfaite connaissance du diagramme des niveaux. Un problème simple se pose. Le binaural est, de manière extrêmement simplifiée, l'équivalent d'un filtrage propre à l'identité morphologique de chacun. Ce filtrage induit des variations d'énergie par les accidents fréquentiels qu'il implique, et donc des variations de niveaux. Autrement dit, d'une HRTF à une autre, nous n'aurons pas les mêmes niveaux en sortie. Cette idée nous a conduit à réaliser des mesures de niveaux des différents plug-in.

(Cf. [fig 51,52,53,54,55,56,57](#))

Observations :

Signal : Sinus 1000 Hz // -20dBfs

Nous observons avec un signal sinusoïdal alimentant les différents canaux du SPAT des niveaux qui nous semblent incohérents. Quand nous alimentons uniquement le canal gauche nous obtenons un signal plus élevé à droite et inversement avec le canal droit.

Nous observons le même comportement avec les canaux arrières (valeurs en rouges sur le premier tableau). Ce qui est également surprenant c'est le déséquilibre entre les canaux gauche et droit alors que nous alimentons le canal central. Tout comme le résultat de la sommation des 6 canaux avec 10 dB de différence entre la gauche et la droite.

Ces comportements ne se reproduisent pas avec le plug-in H3D. Quand nous alimentons tous les canaux, le niveau de sortie du Spat augmente de 3 dB à gauche et de 13 dB à droite. Dans le cas du H3D le niveau de sortie augmente de 4 dB des deux côtés.

Signal multitons : 31 fréquences 20Hz –20 kHz // -20 dBfs

Les différences entre la gauche et la droite sont cohérentes. Les plug-ins engendrent une atténuation d'environ 10 dB entre l'entrée et la sortie quand nous les alimentons avec un seul canal. Par contre dans le cas où tous les canaux sont alimentés, nous avons 4dB d'atténuation pour le Spat et 7dB pour le H3D.

Une solution au niveau de l'encapsulage serait d'imaginer une compensation possible sur les niveaux de sortie en fonction de la fonction de transfert et HRTF utilisées.

Signal Multitons 31 fréquences  
1/3 octave, 20 Hz-20kHz, -20 dBfs

IN	OUT BINAURAL (dBfs)	
	L	R
L	-37	-42
R	-43	-36
C	-39	-38
LFE	-40	-40
LS	-37	-40
RS	-41	-37
L & R	-34	-34
L & R & C	-30	-30
LS & RS	-33	-33
ALL CHANNEL	-24	-24

Signal Sinus 1000 Hz -20 dBfs

IN	OUT BINAURAL (dBfs)	
	L	R
L	-25	-20
R	-29	-21
C	-26	-20
LFE	-27	-22
LS	-24	-21
RS	-30	-34
L & R	-22	-15
L & R & C	-18	-11
LS & RS	-28	-19
ALL CHANNEL	-17	-7



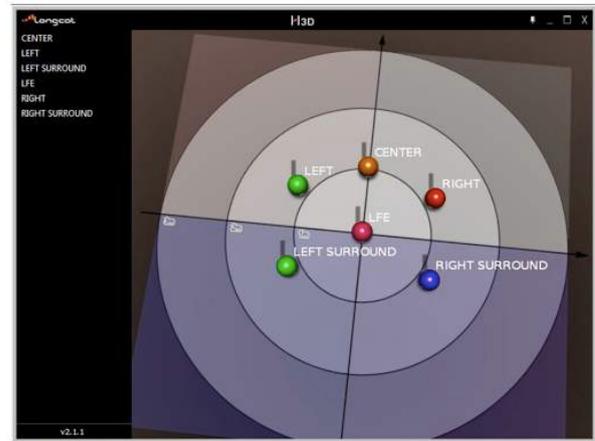
Fig 51,52,53 : Mesures de niveaux du plug-in RTAS SPAT V 1.2.3.0 // HRTF : KEMAR // Protools HD9 V9.0.0

Signal Sinus 1000 Hz -20 dBfs

IN	OUT BINAURAL (dBfs)	
	L	R
L	-28	-37
R	-36	-29
C	-30	-30
LFE	-22	-22
LS	-29	-38
RS	-36	-29
L & R	-31	-31
L & R & C	-25	-25
LS & RS	-33	-31
ALL CHANNEL	-16	-16

Signal Multitons 31 fréquences  
1/3 octave, 20 Hz-20kHz, -20 dBfs

IN	OUT BINAURAL (dBfs)	
	L	R
L	-38	-46
R	-45	-39
C	-39	-39
LFE	-32	-32
LS	-40	-47
RS	-46	-40
L & R	-37	-37
L & R & C	-34	-34
LS & RS	-38	-39
ALL CHANNEL	-27	-27



**Fig 54,55,56,57** : Mesures de niveaux du plug in RTAS H3D V 2.1.1 // HRTF : inconnue // Protocols HD9 V9.0.0

### 1. Limites du panoramique d'intensité

Malgré les questions soulevées dans ce paragraphe, ces outils de spatialisation restent d'une extrême puissance notamment comme alternative aux panoramiques d'intensité qui manquent de stabilité dans le positionnement des sources pour le multicanal. La perspective serait de pouvoir intégrer des panoramiques qui travaillent en même temps sur du  $\Delta t$  et du  $\Delta l$  car comme nous l'avons vu dans les paragraphes précédent, nous ne localisons pas les sources sonores seulement en  $\Delta l$ . Or la stabilité latérale et arrière de l'image sonore est directement liée à l'utilisation de ce type de panoramique.

Jean-Marc Lyzwa et Alexandre Baskind ont travaillé sur une opération de double synthèse binaurale afin de figer et de stabiliser les images sonores latérales. [LB09]. Je me servirai dans mes tests perceptifs de quatre items utilisant cette technologie de mixage auxquels j'appliquerai un encodage binaural.

La troisième partie sera consacrée au développement des productions, innovantes et multicanales. Cette matière servira de production zéro, de démonstration pour proposer des contenus binauraux. Ces contenus seront la source d'une grande partie des tests perceptifs réalisés. Pour optimiser ce type de production, il est important, de se poser des questions sur la chaîne de production dans sa globalité, dès sa composition...

Le compositeur de musique électro-acoustique Ake Parmerud a travaillé sur la création de son propre spatialisateur et propose des productions multicanales dans des formats plutôt étonnants. Sa manière de composer est étroitement liée au support multicanal qu'il choisit...

# 1. CHAPITRE 3 : Un nouvel espace de composition et de mixage : comment l'aborder, comment est-il perçu?

Albert Einstein : « La logique te mènera de A à Z, l'imagination te mènera plus loin... »

## 1.1. Esthétique

### 1.1.1. Genre, style musical et espace sonore

Selon le genre ou style musical, plusieurs critères peuvent être pris en considération. Allan Moore [Moo07] a répertorié quatre façons d'appréhender la relation entre le genre et le style musical. J'en retiendrai une : «Le genre se référerait au contexte donc à l'esthétique et le style à la poïétique» c'est à dire au processus de création. Un critère pourrait faire le lien entre les deux, celui du choix de la source sonore qui à mon sens est déterminant. Elle est la matière première du «Trouver le son, trouver notre son», l'étape créative fondamentale dans les musiques dites «actuelles» ou modernes. Elle est directement liée à la question d'identité sonore. Et on la retrouve quel que soit l'étape de production dans laquelle on se situe : du processus de création au *mastering* en passant par la prise de son et le mixage (cf. mémoire de fin d'études de Martin Fouilleul pour la question d'identité sonore en mixage). Une question supplémentaire pointe son nez et devrait se développer petit à petit en musique actuelle : quel est l'espace de restitution sonore final ?

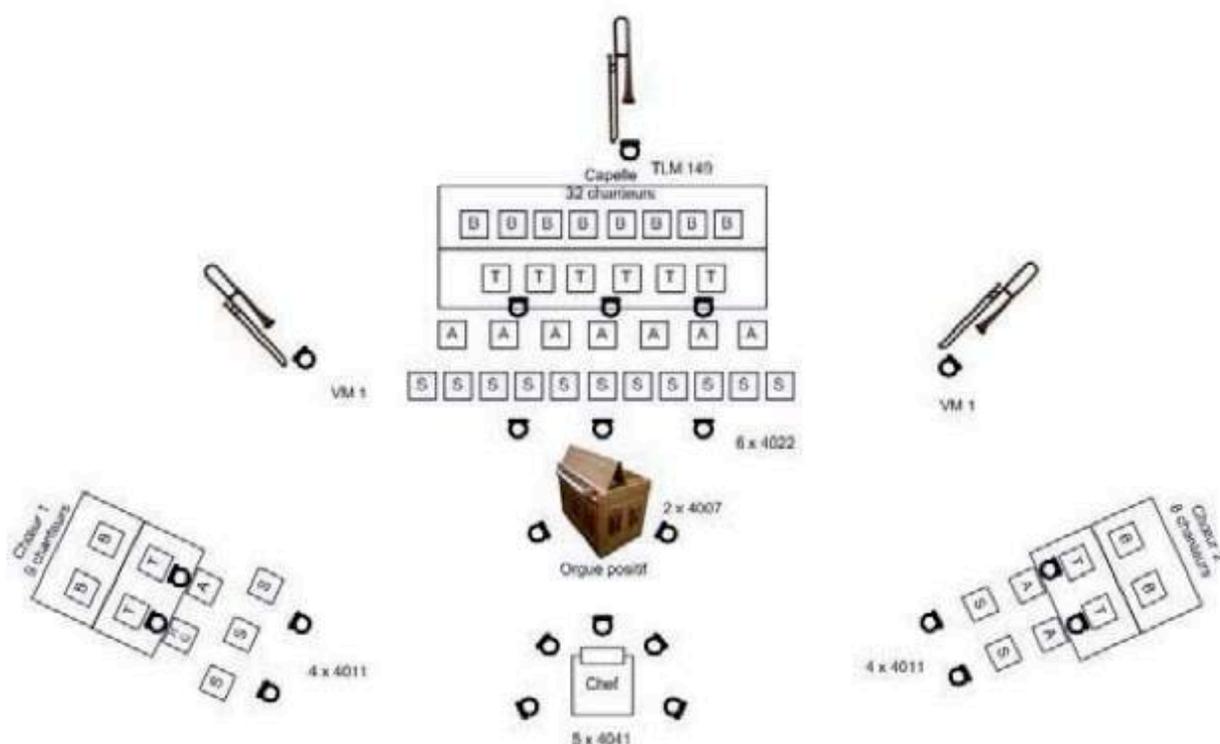
Le multicanal s'est développé dans le milieu du septième art où on l'utilise avec une esthétique bien particulière. La gestion des ambiances, des voix, des musiques, des effets offrent un sentiment d'immersion et d'environnement sonore qu'il est difficile voire complètement impossible d'obtenir avec les possibilités de la stéréophonie. Cette dernière offre, elle, un espace sonore que l'on peut juger limité au regard du potentiel de notre système auditif.

Depuis déjà plusieurs années, une réflexion est menée au CNSMDP<sup>29</sup> sur les techniques de prise de son et de post-production à mettre en œuvre afin de créer une immersion sonore plus réaliste pour l'auditeur. *L'objectif est d'améliorer, de préciser et de stabiliser la sensation d'enveloppement tout en renforçant la lisibilité et la compréhension de l'œuvre restituée.* [Lys09]. Une notion encore trop méconnue dans les musiques dites actuelles pour lesquelles l'espace sonore reste trop souvent restreint à la stéréophonie.

---

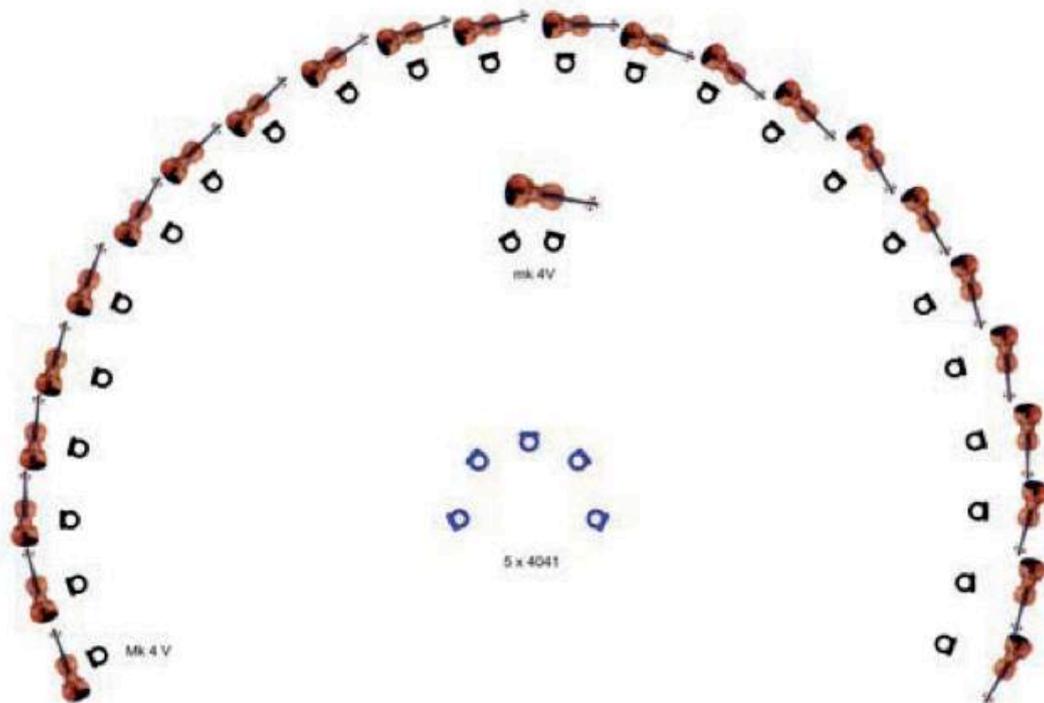
<sup>29</sup> Conservatoire National Supérieur de Musique et de Danse de Paris

Certains compositeurs dits «classiques» se posent la question depuis très longtemps. A titre d'exemple, dans le cadre de l'édition 2011 de la Semaine du Son, la Formation Supérieure aux Métiers du Son a participé à l'enregistrement sous plusieurs formats (Stéréophonie, multicanal 5.1, Ambisonics et WFS) de deux oeuvres destinées à être spatialisées. Le Gloria de Gabrieli pour triple chœur faisait déjà l'oeuvre d'une réflexion sur la perception sonore avec l'apparition du plan médian (élévation) des deux chœurs latéraux.



**Fig 58** : Gloria, extrait de Sacrae Symphoniae, Giovanni Gabrieli

Le deuxième enregistrement a fait l'objet d'une commande à un jeune compositeur du CNSMDP<sup>30</sup>, Benjamin Attahir. Il a composé une pièce pour violon solo et 24 violons disposés en cercle : «Rituales Libri»



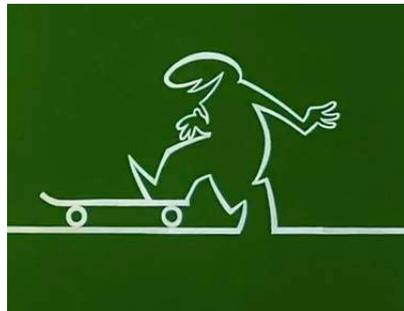
**Fig 59** : *Rituales Libri*, Benjamin Attahir

### 1. Création et système de restitution sonore

On peut raisonnablement penser que le modèle de restitution sonore utilisé influence la composition : un espace sonore tridimensionnel devrait simplifier le positionnement des sources ou offrir une meilleure lisibilité par un phénomène de démasquage ou pour être plus précis de non masquage. La perception auditive nous cache bien des choses. Sachant que rendre les sons mobiles permet de lever certaines ambiguïtés sur la localisation, que ces indices de localisations réagissent différemment en fonction de la fréquence, que la vision a une importance capitale car le cerveau ajuste l'image sonore reconstituée pour la faire coïncider aux sources sonores situées dans son champ visuel : comment aborder une production destinée à mettre en valeur la technologie binaurale ?

<sup>30</sup> Conservatoire National Supérieur de Musique et de Danse de Paris

Nous avons, avec Hervé Dejardin, imaginé un personnage sonore qui pourrait pédagogiquement mettre en évidence l'espace acoustique tridimensionnel. Nous sommes partis de notre expérience perceptive qui vérifiait l'hypothèse d'une bonne localisation des sons riches en transitoires et en HF. C'est de ce postulat que l'on trouvera l'identité sonore du personnage. «*Mister Bin*» sera défini par des sons directement liés aux mines de crayons, de toutes sortes, en contact avec des supports variés. Notre instrument, aux caractéristiques timbrales bien particulière, devra se comporter musicalement en fonction de l'esthétique choisie donc d'un petit scénario. En quelques mots il s'agit d'une interview pédagogique du personnage «*Mister Bin*» ayant comme objectif la prise de conscience de l'espace sonore tridimensionnel au casque. Cette expérience, dans l'idéal, devrait être couplée à un visuel de type «*La linéa*».



**Fig 60** : Impression écran «*La Linéa*»

Une notion de geste sonore est entrée naturellement en jeu afin que le mouvement virtuel recréé en post-production colle à des sensations d'accélération ou de décélération dès la prise de son. Pour l'enregistrement, nous avons utilisé un couple Schoeps de CMC 6 Ug en configuration XY (15 mV / Pa)<sup>31</sup>. Un *Sound Devices 722* servait d'enregistreur. L'idée était de récupérer de la matière brute en ultra proximité. Nous nous sommes donc isolés et affranchis de toutes sources potentielles de bruit de fond (sources électriques etc...).

La matière récupérée servira d'identité sonore au court extrait musical de cette production. Un changement de variable rythmique sera opéré. En effet, j'ai recomposé les éléments d'une batterie, à partir de certains sons percussifs de mines de crayons en ayant à l'esprit que chacun de ses éléments seraient spatialisés de manière extrême dans l'espace 3D, ceci afin d'exploiter au maximum notre espace de restitution. Pour mettre en valeur tout cet environnement immersif, nous avons fait le choix de mixer la voix du journaliste en interaural. Une expérience qui met en relief la partie binaurale.

---

<sup>31</sup> La tête Neumann KU 100 n'était plus disponible

Cette production électroacoustique demandera une grande organisation dans la gestion de la session protocols. En effet, il était important d'avoir à l'esprit que le mixage serait suivi à la fois en 5.1 et en binaural pour vérifier la crédibilité des déplacements en binaural et une certaine compatibilité de format.

## 1. Différents types de productions

### 1.1. Productions bêta

#### 1.1.1. Franck Avitabile

Le premier enregistrement réalisé fut le pianiste de jazz Franck Avitabile. L'objectif était de tester différents systèmes de prise de son multicanale dans des configurations simples afin de les transposer en binaural ensuite.

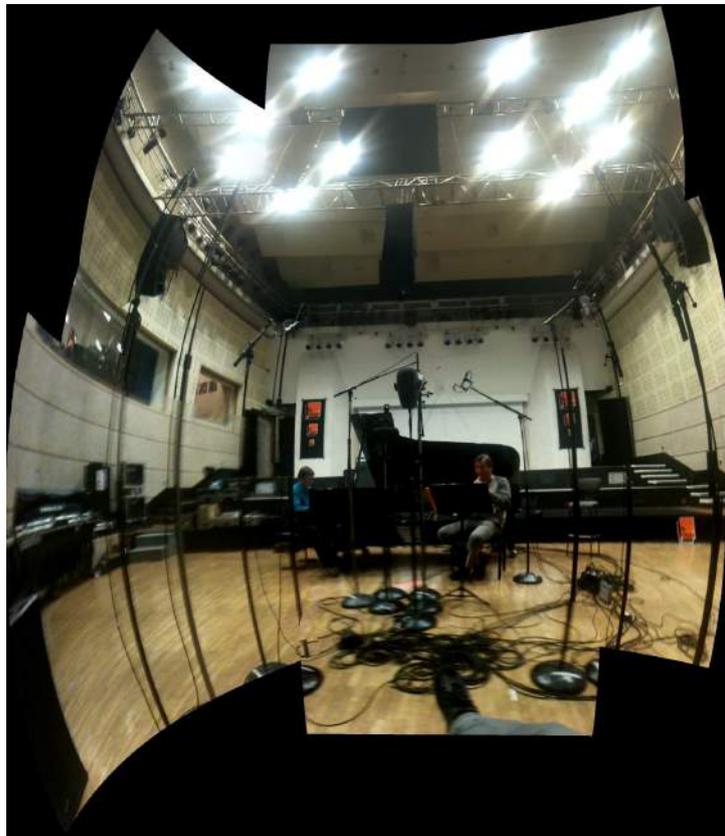
**STUDIO 105 prise 5.1 piano hautbois**

SPLITTER 1 boîtier 1				SPLITTER 2 boîtier 2			SPLITTER 3 boîtier 3						
*	Paire	Systèmes	Micro	*	Paire	Source	Micro	*	Paire				
	1	SYST A	M149		21	TETE ARTIF	KU100		41				
	2		M149	KU100		22	DECCA TREE	KU100		42			
	3		M149	geffel l		23			geffel c		43		
	4		km hautb	geffel r		24			bk lb		44		
	5		km hautb	bk rb		25					45		
	6		bk arriere			26					46		
	7		bk arriere			27					47		
	8	SYST IRT	croix oct l		28					48			
	9		croix oct r		29				49				
	10		croix oct lb		30				50				
	11		croix oct br		31				51				
	12	SYST DPA	dpa 5.1		32				52				
	13		dpa 5.2		33				53				
	14		dpa 5.3		34				54				
	15		dpa 5.4		35				55				
	16		dpa 5.5		36			56					
	17	AMBISONIC	soundfield l		37				57				
	18		soundfield r		38				58				
	19		composant b		39				59				
	20		composant b		40				60				
SOURCES HORS SPLIT				SOURCES HORS SPLIT									
	61				65								
	62				66								
	63				67								
	64				68								

	RV1		
	RV2		
	RV3		
	RV4	Comm	

Fig 60 : Patch de l'enregistrement de Frank Avitabile au Studio 105

J'utiliserai dans mes tests perceptifs un extrait pour piano seul. Les choix de mixages m'ont conduit à utiliser la réverbération artificielle du Spat et à affecter chaque source microphonique comme source d'entrée dans le plug-in H3D.



**Fig 61 :** *Installation microphonique pour l'enregistrement piano-hautbois*



**Fig 62 :** *Tête artificielle KU100*

## 1. [Chronique : Le réveil](#)

Cette chronique a été réalisée dans le but de tester différents moyens d'aboutir à des sons binauraux. Une partie de la matière a été enregistrée à l'aide de la tête artificielle Neumann KU100, et l'ensemble a été mixé à l'aide du plug-in H3D en virtualisant chacune des sources. Il s'agissait de rendre compte des sons présents à notre réveil. Pour aboutir à une bonne externalisation, il était intéressant de laisser un son permanent figeant la scène sonore et permettant à l'auditeur d'avoir un point d'ancrage de localisation. La chronique commence par la sonnerie d'un téléphone à environ un mètre de la tête artificielle, le sujet se déplace et allume une chaîne hi-fi<sup>32</sup>... Cette petite production m'a permis de tester l'automation sous le H3D et mettre à l'épreuve sa fiabilité. Un petit bémol quant à la perception des sources sur le plan vertical.

En guise d'anecdote, lors des tests perceptifs, et de l'écoute de cet item, plusieurs personnes se sont arrêtées et retournées... L'effet de la sonnerie de téléphone enregistrée avec la tête artificielle donne une impression de réalisme remarquable autant en multicanal qu'en binaural.

## 2. [EleKtripp](#)

Cette production visait à tester l'utilisation de la tête artificielle dans un enregistrement d'une formation de musique actuelle, le trio «eleKtripp» : piano, basse et trombone, batterie. L'idée était de trouver un équilibre acoustique intéressant sur le plateau et de placer la tête à la manière d'un couple afin de trouver un équilibre naturel et bénéficier de l'acoustique du lieu. L'ensemble des sources était également appointé par des microphones de proximité pour pouvoir comparer les systèmes. L'équilibre fut difficilement atteignable, les éléments de la batterie ne sonnant pas comme dans la salle avec la tête artificielle. La caisse claire et les cymbales étaient très en avant et la grosse caisse très derrière. L'approche binaurale de cet enregistrement changera au cours de la production pour laisser place à un enregistrement multicanal «traditionnel». Un travail de *sound-design* sera réalisé plus tard en post-production afin de développer l'esthétique musicale du morceau «BlackBach». Cette fois-ci, le Spat nous servira d'encodeur décodeur binaural avec comme HRTF : la tête KEMAR. ([annexe 2](#)). Plusieurs mouvements circulaires seront travaillés afin de comparer le résultat en 5.1 et au casque avec différentes HRTF. Selon les HRTF, les mouvements circulaires, dont le rendu en 5.1 était proche d'un cercle parfait, se déformaient plus ou moins en terme de

---

<sup>32</sup> Point d'ancrage

localisation et de timbre. Au cours des tests, ces mouvements circulaires seront étudiés à partir cette fois-ci d'une prise de son à l'aide de la tête KU100.



**Fig 63:** Installation au studio 105 avec la tête artificielle

## 1. [Andrew Bird](#)

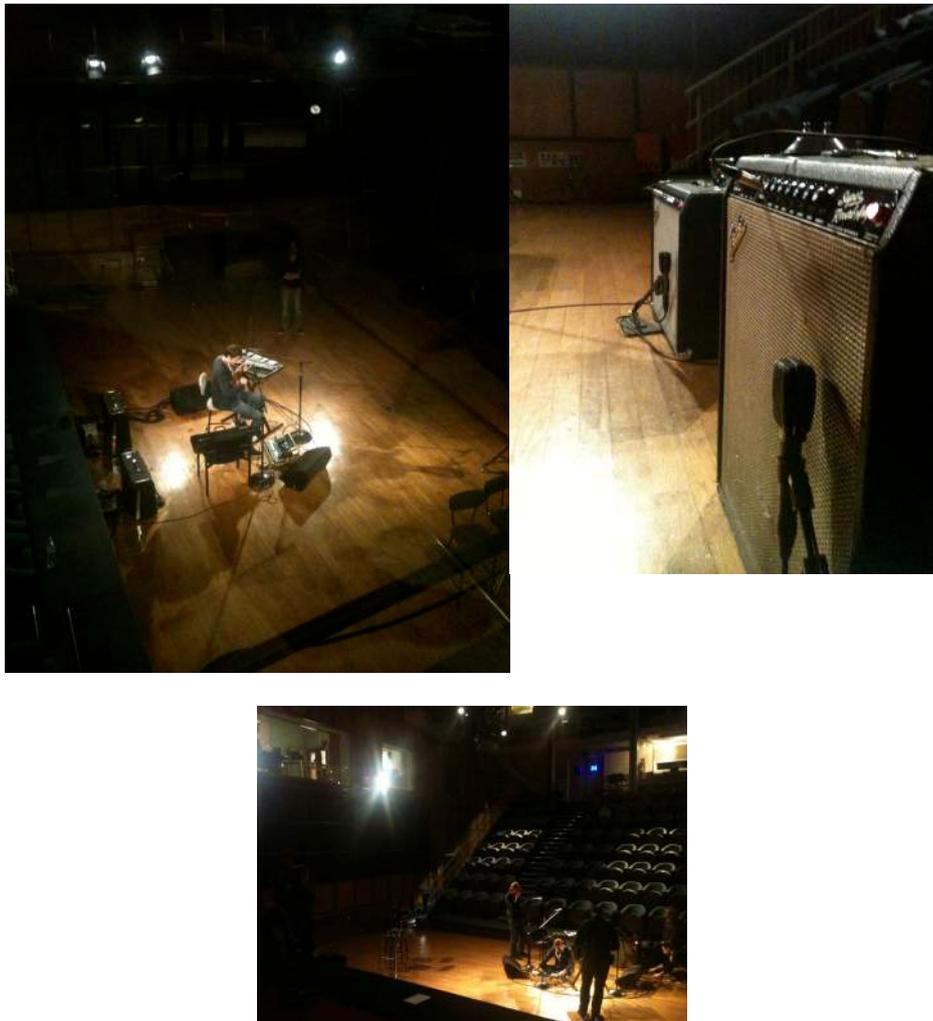
<http://andrewbird.lemouv.fr/>

Le concept de la production sur Andrew Bird est né de l'évolution du navigateur Chrome en HTML 5<sup>33</sup>. La *Blogothèque* s'est inspirée d'un clip interactif de «Ok Go»<sup>34</sup>. Une collaboration entre les *Nouveaux Médias* de Radio France, le *Mouv* et la *Blogothèque* permettra une captation vidéo, un enregistrement en multicanal et sera mixé en 5.1, stéréo et binaural. L'axe esthétique s'est défini assez rapidement. Effectivement le principe de bouclage utilisé par Andrew Bird est devenu l'idée directrice de la réalisation vidéo : faire entrer les images sous forme de fenêtres en fonction des entrées instrumentales dans l'espace multicanal 5.1. Le défi était de récupérer de la matière pour les canaux arrières. Or, comme le montre le schéma du chemin du signal audio de son installation, la tâche se révélait compliquée. (cf. [annexe 4](#))

<sup>33</sup> HTML 5 : *HyperText Markup Language 5*

<sup>34</sup> [http://www.allisnotlo.st/index\\_en.html](http://www.allisnotlo.st/index_en.html) (fonctionne seulement avec Chrome)

En effet, l'utilisation de pédales *hardware*<sup>35</sup> pour boucler son violon ne nous donnait que très peu de matière. Nous avons donc convenu, après écoute du morceau proposé, l'éventualité de lui faire rejouer quelques passages. Dans un premier temps, nous nous sommes concentrés sur une version entière comme base (validée avec Andrew) et ensuite, on lui a demandé de rejouer chaque entrée, seule, afin de récupérer de la matière pour pouvoir spatialiser en 5.1. La difficulté de montage était de reconstruire l'organisation de bouclage d'Andrew Bird, et de respecter les entrées et sorties de ses évènements musicaux à partir de la version entière et des plans vidéos. Pour des raisons musicales, le clic n'était pas envisageable. Cette production sert actuellement de démonstration au cours des réunions aux perspectives européennes et a été remarquée par la BBC, intéressée par le binaural. Deux jours de post-production supplémentaires ont été nécessaires. (cf. [annexe 5](#)).



**Fig 64, 65, 66:** Illustrations pendant la production Andrew Bird au studio 106

---

<sup>35</sup> *Hardware* : matériel

## 1. [Le mixage binaural](#)

Connaissant les outils de post-production, nous pouvons aborder le mixage de plusieurs manières. D'un point de vue pratique, l'idéal serait de pouvoir à partir des pistes séparées d'un mixage 5.1, 7.1, etc., paramétrer un rappel de la configuration du studio et récupérer un flux stéréo binaural. Cela demanderait de connaître sa référence d'encodage parfaitement, et ceci en fonction du contenu. C'est une méthode qui n'est à mon sens pas encore valable. En imaginant un encodage binaural à partir des HRTF personnelles du mixeur, et en ayant à l'autre bout de la chaîne un auditeur qui, par des méthodes d'individualisations d'HRTF écouterait lui aussi à l'aide de ces propres empreintes auditives, il est possible que cela fonctionne. Mais il reste les problèmes des sources latérales instables en 5.1. Des tests sont prévus au studio 132 à partir d'un système 7.1.

La deuxième méthode consiste à mixer directement en binaural. D'un point de vue logiciel, chaque source sonore ouvre un plug-in différent H3D ou Spat. D'un point de vue spatialisation, on s'affranchit des panoramiques d'intensité des consoles ou stations de travail. Malheureusement cela demande des puissances de calculs très importantes. Seules des petites sessions peuvent être imaginées dans cette configuration. Et cela implique de mixer entièrement au casque, entraînant ainsi une fatigue auditive plus rapide.

La troisième méthode, celle du compromis, consiste à jongler entre les mixages 5.1, 7.1, etc. et le binaural en paramétrant sa session en *double monitoring*<sup>36</sup>. C'est assez fastidieux à mettre en place mais c'est une méthode qui permet à la fois de prendre conscience de la couleur des encodeurs et de trouver un équilibre compatible entre le 5.1 et le binaural.

Dans les conditions actuelles, il n'est pas encore possible d'avoir des HRTF personnalisées en bout de chaîne, l'utilisation de la tête KEMAR est donc conseillée afin de correspondre avec un panel d'auditeurs le plus large possible. Plusieurs problèmes ont déjà été soulevés dans la restitution binaurale : le plan frontal reste ambigu, les sources stéréophoniques ont tendance à être perçues à l'intérieur de la tête, l'élévation ne reste que très peu entendue. Il faut alors envisager le recours à des astuces, comme par exemple l'exagération de la distance de la source centrale, le décalage de quelques degrés des éléments habituellement dans le canal central...

Quel est le résultat ?

---

<sup>36</sup> *Combinaison de touches sous Protools : «alt» + «⌘» dans l'onglet Output du «mixer».*

## 1. [Perception tridimensionnelle](#)

### 1.1. [Présentation des tests perceptifs](#)

On se place ici au niveau de l'espace perceptif de l'auditeur. L'objectif est d'évaluer comment la scène audio tridimensionnelle est perçue, analysée et interprétée par l'individu. La méthode d'évaluation consiste à mener des tests de localisation. Conscient de l'apprentissage nécessaire et du temps d'adaptation à l'écoute tridimensionnelle, j'ai choisi et organisé mes tests suivant une chronologie précise, découpée en trois parties. L'ensemble pouvait durer jusqu'à deux heures, maximum.

Les dix premiers tests sont des enregistrements très courts réalisés avec la tête artificielle Neumann KU100. Deux sources sonores possibles : un générateur de bruit blanc<sup>37</sup> ou un iPhone diffusant un mp3. La source se déplace autour de la tête suivant les différents axes et plans formant la sphère. L'objectif de ces tests est de mettre en évidence trois notions nouvelles qui me paraissent importantes : la perception arrière, l'externalisation et l'apprentissage. On peut consulter l'ensemble des tests en [annexe 7](#). Ils sont volontairement nombreux et ne visent pas à être significatifs mais à dégager des grandes tendances afin d'appréhender la perception générale des individus face à ce nouveau mode d'écoute.

La deuxième partie des tests concerne des paramètres plus subjectifs comme l'immersion, l'intégration dans le public, le positionnement de l'auditeur en live...

La troisième partie des tests consiste en une appréciation comparative de chaque système d'encodage au casque, par rapport à la référence constituée par l'écoute 5.1 sur enceinte. Connaissant les conditions de réalisation et de production des items, j'espère voir apparaître quelques concordances et regroupements de résultats. Trois critères d'appréciation ont été retenus, le respect du timbre, la précision de la spatialisation et l'immersion. Il s'agit du test numéro 17. (cf. [annexe 6](#))

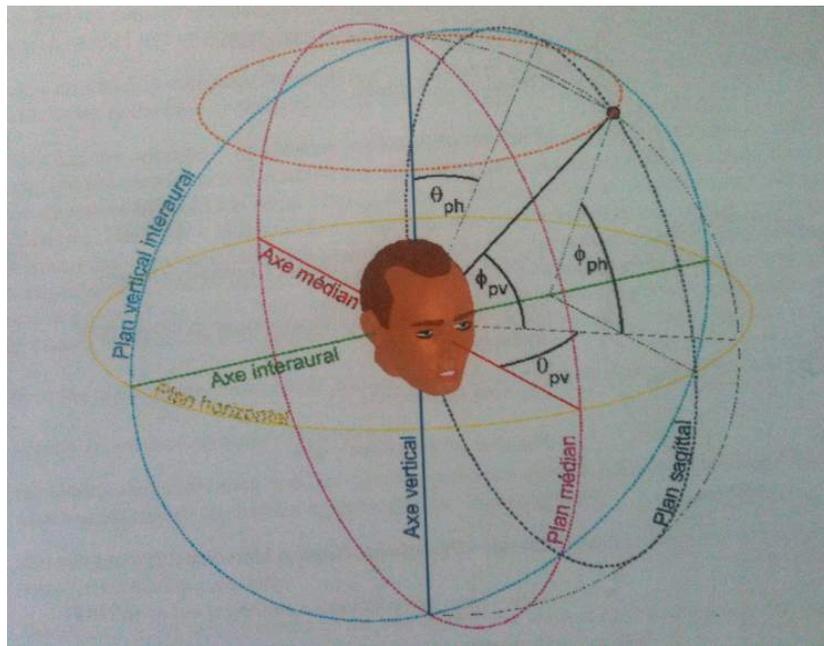
### 1.2. [Contexte](#)

Ces tests se sont déroulés au studio 132 de Radio France du 12 au 22 juin 2012 sur une station Protools 9. Tous les sujets ont utilisé le même casque d'écoute, un Beyerdynamic 990 Pro ouvert. Le système de monitoring comprenait 5 enceintes actives Genelec 8020 ainsi qu'un *subwoofer* 1094A Genelec alimentés par une console numérique Yamaha DM1000 avec une interface RME *Fireface* 800. Il était conseillé de fermer les yeux afin de s'affranchir du lieu, et d'écouter plusieurs fois les extraits. Aucune limite sur le nombre d'écoutes n'était fixée.

---

<sup>37</sup> Calibreur de microphone Mediatech

L'auditoire était constitué de 18 participants, 15 hommes et 3 femmes ; parmi eux, 17 ingénieurs du son professionnels, un étudiant et une réalisatrice. Cette illustration était communiquée à tous les participants.



**Fig 67** : Coordonnées sphériques

## 1. Résultats

### 1.1. Perception arrière et élévation

Sur 17 auditeurs, 14 perçoivent le déplacement de la source sur un demi-cercle arrière quasiment parfait. Cependant 13 entendent une déformation sur le plan horizontal de la source ce qui confirme les problèmes d'instabilités du plan frontal. On retrouve ces mêmes tendances dans des cas plus complexes que des sources seules. Dans la pièce électro-acoustique nommée «Osci», l'arrière semble bien perçu, en revanche l'immersion est parfois critiquée pour son manque de solidité au centre. La demi-sphère supérieure du plan vertical interaural est également bien perçue par les sujets. En effet, un seul n'a pas entendu l'élévation. 5 sujets ont décrit un plan vertical déformé et 12 un plan convenable.

## 1. [Externalisation](#)

Quelques constats intéressants ressortent de ces tests. Une grande majorité des individus, que ce soit pour le test 1a 2a ou 3a perçoit particulièrement bien l'externalisation (source virtuelle située hors de la tête), notamment sur l'axe interaural. L'enregistrement de la moto en extérieur avec la tête Neumann donne des résultats très intéressants, et vient conforter l'efficacité de la spatialisation le long de cet axe dans les techniques binaurales. En effet, 16 sujets sur 18 perçoivent l'externalisation et ont des notions de distance, d'éloignement de la source assez homogène, environ 27m en moyenne. La distance réelle était de 250 mètres. On relève un facteur d'erreur d'environ 10. Il serait intéressant d'essayer de chercher à évaluer une moyenne de ce coefficient de sous-évaluation de distance dans d'autres conditions. D'autres part, 17 sujets sur 18 n'entendent pas l'externalisation s'arrêter avant la fin du bruit du moteur. Ce test amène une notion supplémentaire car on aurait pu croire que le phénomène d'externalisation serait logarithmique et atteindrait une limite.

L'externalisation des sources est favorisée par l'ajout d'un effet de salle, comparativement à une écoute « anéchoïque ».

Dans le cadre de diffusion par HP virtuels, il peut être intéressant de travailler directement à partir d'un format 7.1 ou supérieur de sorte à multiplier (et diversifier) les directions d'émissions réellement synthétisées par HRTF. Il est probable que le fait de travailler initialement en 7.1 au lieu de 5.1 améliore la localisation latérale de la restitution binaurale.

## 2. [Apprentissage](#)

Une étude a montré que les chefs d'orchestre ont développé une capacité exceptionnelle à localiser latéralement. C'est le résultat de la plasticité cérébrale qui fait que leur expertise (via leur expérience) leur permet d'être nettement meilleurs que la population générale sur ce point.

Ainsi, la pratique d'une activité comme l'activité binaurale permet de développer de nouvelles connexions au cerveau, et d'augmenter les ressources dédiées à ce mode. On l'a souvent démontré pour des aspects reliés à l'expertise musicale, entre autres.

Les tests 3 et 4, 6 et 7, et 9 et 10 cachent dans leur continuité un test subjectif d'apprentissage. Dans un premier temps, pour le test 3, il était demandé de dessiner le mouvement entendu sur une sphère sans aucune aide. Le type de trajectoire était ensuite révélé dans le test 4 (de même pour les tests 6 et 7 d'une part, et 9 et 10 d'autre part). Or les stimuli sonores étaient identiques.

Le plan vertical : sur les 18 individus, 11 ont abordé le deuxième dessin différemment pour aller dans la bonne direction et 4 avait déjà dessiné la bonne trajectoire. Une remarque importante, 12 individus n'entendent pas le passage de la source sonore sous la taille et ont perçu un trajet symétrique inverse. (cf. [annexe 6](#))

## 1. Conclusions résultats

Ces embryons de résultats doivent tout de même être nuancés. En effet, les participants n'ont suivi qu'une courte phase d'entraînement préliminaire aux tests. Il faut prendre en compte leur inexpérience relative en matière de techniques binaurales, et leur état de fatigue. Certains avait déjà travaillé plusieurs heures au casque avant de participer aux tests.

Les résultats des enregistrements avec la tête artificielle sont réellement convaincants, dans la limite des objectifs esthétiques du contenu. Il faut cependant émettre une réserve sur la partie basse de la demi-sphère qui a très rarement été perçue par les sujets. Le plan frontal s'est confirmé être instable, par les nombreuses déformations ou confusions mises en évidence. Cependant, la bonne perception du cercle sur le plan sagittal est une observation intéressante car il correspond à l'espace du cône de confusion abordé dans les chapitres précédents.

A titre personnel, les heures travaillées sur les différents plug-in et productions ont réellement développé mon écoute spatiale et surtout tridimensionnelle. Le facteur de l'apprentissage est une étape fondamentale à mon sens, et il est important de former l'auditeur en lui proposant du contenu dans ce format. Ceci peut rappeler le passage de la monophonie à la stéréophonie.

Les différents plug-in de virtualisation spatiale amènent une perception auditive nouvelle et très intéressante, notamment sur l'homogénéité spatiale, et la localisation. En revanche, les colorations se sont révélées encore trop importantes sur de nombreuses productions pour réellement se servir des paramètres de virtualisation d'acoustique de salles. Non sans conséquence sur le sentiment d'externalisation.

## Conclusions

Ce travail de recherche et d'apprentissage a été réalisé à partir de HRTF non individuelles depuis le début. Il a été nécessaire de s'adapter aux fonctions de transfert existantes, avec l'ensemble des problèmes que cela engendre, détimbrage et mauvaise localisation principalement.

Le casque peut avoir également une influence sur la qualité de reproduction spatiale. [BR11]. Cependant on peut admettre qu'il modifie le timbre de façon globale, c'est à dire indépendamment des directions virtuelles des sources présentes dans la scène auditive transmise. On peut donc s'attendre à une possibilité d'accoutumance aux caractéristiques du casque, et par conséquent à une influence finale secondaire.

Il est important de garder à l'esprit également que les imprécisions, par exemple dans le plan vertical, ou les confusions arrières peuvent également exister avec des HRTF personnalisés. Cela nous conduit naturellement aux méthodes de *Head-tracking* qui cumulées à la synthèse binaurale peuvent lever certaines ambiguïtés.

Mais les solutions ne résident pas seulement dans la personnalisation, elles seront un travail sur l'ensemble de la chaîne binaurale. Effectivement, en s'intéressant à la chaîne de production globale, on prend conscience que le mode de restitution est aussi important que le choix de la matière sonore et la manière de l'agencer. L'expérience des quelques productions à la radio et à l'extérieur nous a clairement dirigé vers un type de sources sonores particulières. Une étude intéressante serait de classifier les sources en fonction de leur rendu spatial, à la manière du célèbre écrit de Pierre Schaeffer où il définit la typologie des objets sonores dans son *Traité des objets musicaux*. [Sch66]. Effectivement, cette base de travail permettrait d'aborder la perception suivant la morphologie d'un son, en quelque sorte suivant ses caractéristiques perceptives binaurales.

«Ce format de restitution sonore au casque ouvert n'est pas fermé à se suffire lui-même...»

*Nicolas Canot<sup>38</sup> : " J'ai donc quelques idées en tête : j'ai présenté au mois de mai 2012 une pièce "électroacoustique" s'appuyant sur un dispositif d'écoute un peu particulier, deux moniteurs généraux et un casque (ouvert) pour chaque auditeur (assis dans un fauteuil ou transat) qui joue donc sur la création d'espaces sonores "imaginaires" par la superposition d'une écoute "interaurale - transaurale classique".*

<sup>38</sup> Compositeur électro-acoustique. [http://nikano.free.fr/aka/aka\\_2\\_3.html](http://nikano.free.fr/aka/aka_2_3.html)

# Bibliographie

[Bla83]

J.Blauert. *Spatial Hearing : The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA. 1983.

[Bla09]

J.Blauert. *La technologie binaurale : bases scientifiques et domaines d'application génériques*. Journal de physique IV. 1994.

[Nic10]

R.Nicol. *Représentation et perception des espaces auditifs virtuels*. Mémoire d'Habilitation à Diriger des Recherches. 2010.

[Bom08]

P.Bompy. *Contrôle au casque d'une prise de son multicanale*. Mémoire de fin d'études. Paris. 2008.

[Cor11]

C.Cornuau. *Etude et optimisation de la synthèse transaurale à deux canaux*. Mémoire de fin d'études. Formation Supérieure aux Métiers du Son, Paris. 2011.

[Car12]

J.Carton. *Intégration et exploitation de traitements transauraux pour la production au format multicanal 5.1*. Mémoire de fin d'études. Formation Supérieure aux Métiers du Son, Paris. 2011.

[Sig11]

P.A Signoret. *Le son multicanal*. Conservatoire de Paris. Formation Supérieure aux Métiers du Son. 2011.

[TP77]

G. Theile and G. Plenge. *Localization of lateral phantom sources*. J. Audio Eng. Soc., 25 (4) : 196-200. 1977.

[LB09]

J.M.Lyzwa and A.Baskind. *Utilisation de techniques binaurales et transaurales en prises de son et en post-productions multicanales 5.1*. AES 3rd Latin American Conference, Sao Paulo, Brazil. 2009.

[Lyz03]

J.M.Lyzwa. *Prise de son et restitution en multicanal 5.1. Problématique d'une oeuvre spatialisée : Répons, Pierre Boulez*. Conservatoire de Paris. 2003.

[LM00]

J.M.Lyzwa et J.C.Messonier. *Utilisation d'une rampe microphonique circulaire pour la prise de son et la post-production de sources sonores réparties sur 360° et destinées à un système de restitution multicanal 5.1*. Conservatoire de Paris. 2000.

[Lar01]

V.Larcher. *Techniques de spatialisations des sons pour la réalité virtuelle*. Thèse de doctorat. Université Paris 6. 2001.

[Bur11]

V.Burgun. *Conception et réalisation d'un système microphonique 2D pour la technologie de spatialisation sonore Higher Order Ambisonics (HOA)*. Mémoire de fin d'étude. Formation Supérieure aux Métiers du Son, Paris. 2011.

[SA63]

M.R.Schroeder and B.S.Atal. *Computer simulation of sound transmission in rooms*. Proceedings of the IEEE, 51(3) :536–537. 1963.

[CB89]

D.H.Cooper and J.L.Bauck. *Prospects for transaural recording*. *Journal of the Audio Engineering Society*, 37(1–2) :3–19.1989.

[Gui09]

P.Guillon. *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la re- construction de HRTF*. PhD thesis, Université du Maine, 2009.

[Moo07]

A.Moore. *Style and genre as a mode of aesthetics*. 2007.

[LCR10]

M.Paquier et V.Koehl. *Perception sonore de la variabilité de positionnement d'un casque audio*. 2010.

[PR11]

Y.L.Parodi and P.Rubak. *A subjective evaluation of the minimum channel separation for reproduction binaural signals over loudspeaker*. AES Paper . 2011.

[PKG07]

R.Pellegrini, Clemens Kuhn, and Mario Gebhardt. *Headphones Technology for Surround Sound Monitoring - A Virtual 5.1 Listening Room*. Audio Engineering Society, Convention Paper 7068L. 2007.

[RBHPSKLN07]

J.Röden, J.Breebaart, J Hilbert, H Pumhagen, E Schuijers, J Koppens, K Linzmeier and A. Höltzer. *A study of the MPEG Surround quality versus bit-rate curve*. Audio Engineering Society. Convention Paper 5656. New York. 2007.

[WK05]

F.Wightman, D.Kistler. *Measurements and Validation of Human HRTFs for Use in Hearing Research*, Acta Acustica United with Acustica, vol 91. 2005.

[BR11]

B.Boren, A.Roginska. *The Effects of Headphones on Listener HRTF Preference*. Music and Audio Research Lab, New York University, New York. 2011.

[WS54]

R.Woodworth, H Schlosberg. *Experimental Psychology*. New York. 1954

[Mid92]

J.Middlebrooks. *Narrow-band sound localization related external ears acoustics*. J. Acoust. Soc. Am. 1992.

[Mid99]

J. Middlebrooks. *Individual differences in external-ear transfer functions reduced by scaling in frequency*. J. Acous. Soc. Am. 1999.

[Mid99]

J. Middlebrooks. *Virtual localization improved by scaling non individualized external-ear transfer functions in frequency*. J. Acous. Soc. Am. 1999.

[MJG89]

J.Middlebrooks, J.Makous and D.Green. *Directional sensitivity of sound pressure levels in the human ear canal*. J. Acous. Soc. Am. 1989.

[MMO, 2000]

J. Middlebrooks, E.A. MacPherson and Z.A.Onsan. *Psychophysical customization of directional transfer functions for virtual sound localization*. J. Acous. Soc. Am. 2000.

# ANNEXES

## Annexe 1

Programme des Folles Journées de Nantes. Productions multicanales.

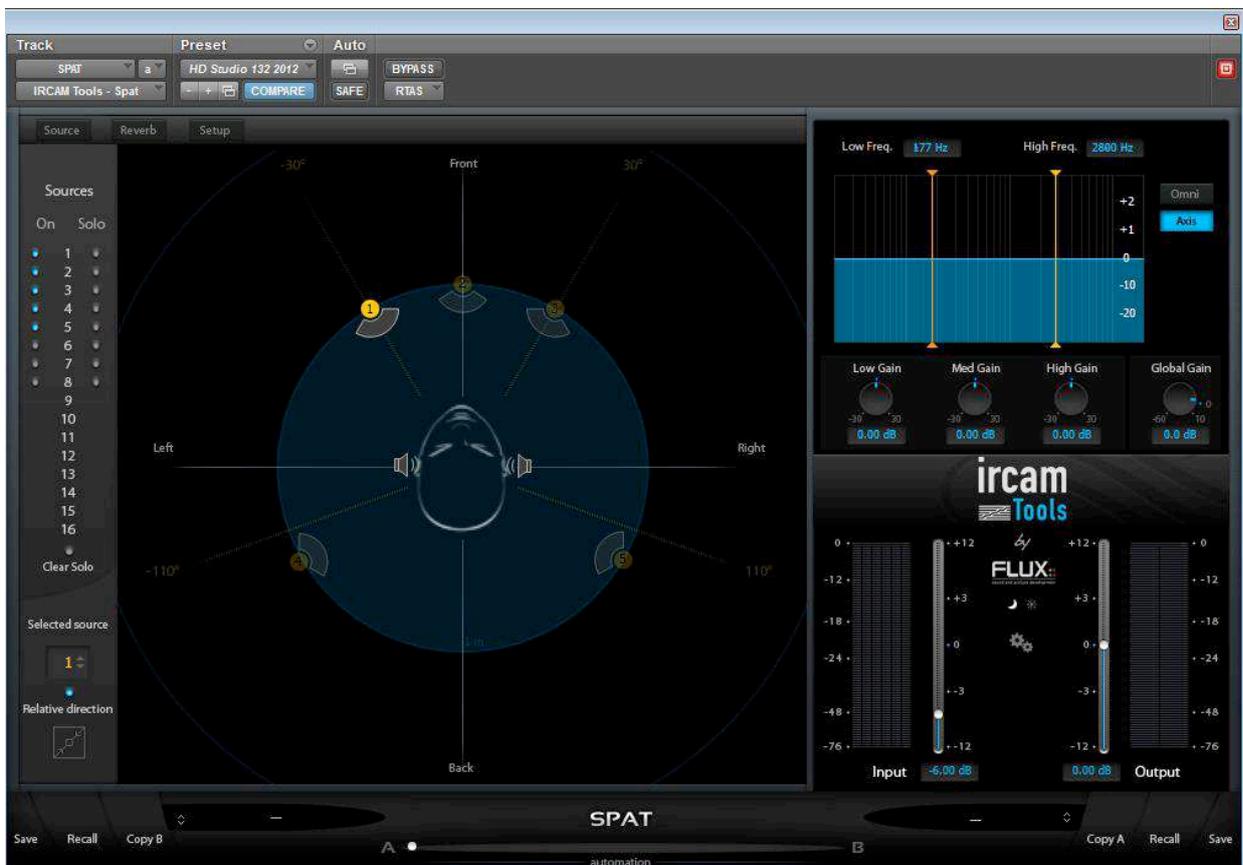
Brahms Symphonie N°2 en ré majeur opus 73  
Folle journée de Nantes 2011  
Enregistrement multicanal  
Orchestre de la résistance de la Haye  
Direction Neeme Järvi  
Soliste Gerhild Romberger mezzo soprano  
Prise de son: Yves Baudry, Pierre Bornard, Loïc Duros, Julien Bourdais, Alain Piet  
Musicien metteur en onde: Vincent Villetard  
Chargé de réalisation: Philippe Petit

Wagner Prélude et Liebestod, extrait de "Tristan et Isolde"  
Wagner Prélude, extrait de "Parsifal"  
Mahler Kindertotenlieder  
Folle journée de Nantes 2011  
Enregistrement multicanal  
Orchestre de la résistance de la Haye  
Direction Neeme Järvi  
Soliste Gerhild Romberger mezzo soprano  
Prise de son: Pierre Bornard, Loïc Duros, Julien Bourdais, Alain Piet  
Musicien metteur en onde: Raffi Kervokian  
Chargé de réalisation: Philippe Petit

Brahms Ouverture pour une fête académique opus 80  
Strauss Ainsi parlait Zarathoustra opus 30  
Folle journée de Nantes 2011  
Enregistrement multicanal  
Orchestre Philharmonique de l'Oural  
Direction Dmitri Liss  
Prise de son: Yves Bourdais, Loïc Duros, Julien Bourdais, Alain Piet  
Musicien metteur en onde: Vincent Villetard  
Chargé de réalisation: Philippe Petit

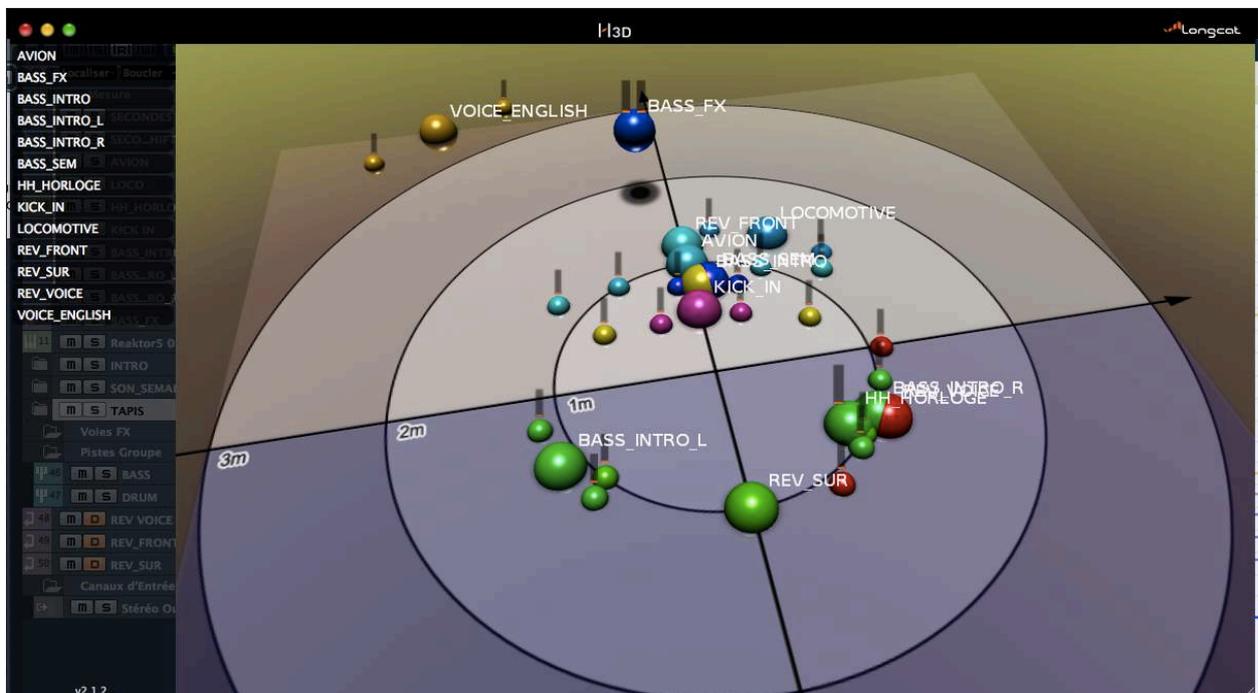
## Annexe 2

Configuration du Spat : 5.0 vers headphones.



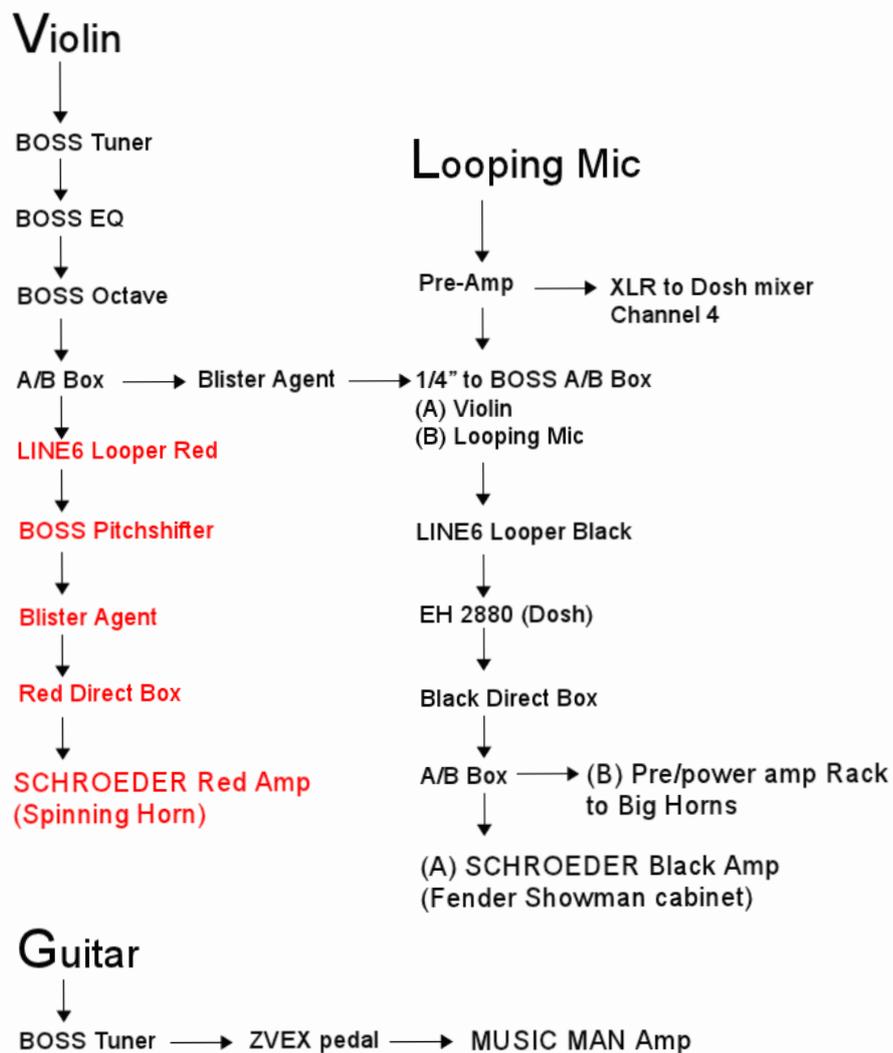
## Annexe 3

3D View du plug-in H3D de Longcat : Session du Générique «Partout Ailleurs»



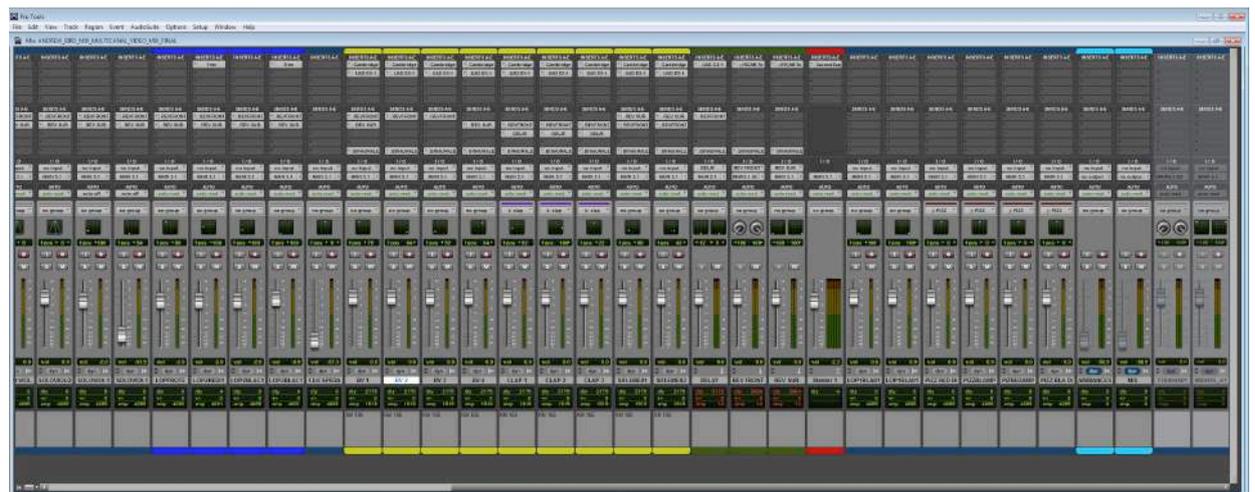
## Annexe 4

### Bird Signal Path



# Annexe 5

Fenêtres des mixeurs de la production d'Andrew Bird : mixage binaural et 5.1 simultanément

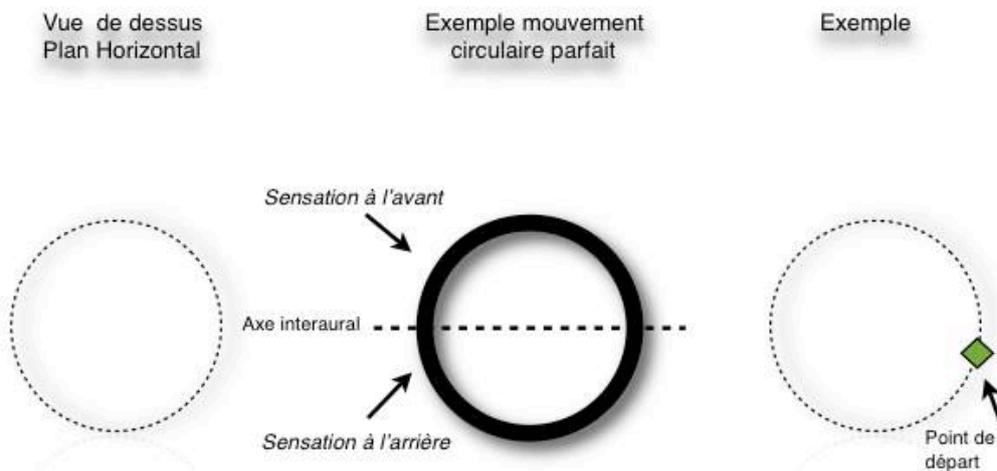


## Annexe 6

### TESTS D'ECOUTES BINAURALES

N'hésitez pas à fermer les yeux

**Exemple :** Mouvement circulaire sur le plan horizontal.

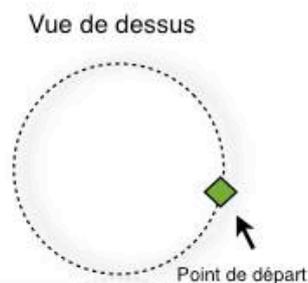


#### TEST 1 : Plan Horizontal

- Entourer la réponse

- 5 : la source passe à l'arrière suivant un demi-cercle parfait
- 4 : la source passe à l'arrière suivant un demi-cercle légèrement déformé
- 3 : la source passe à l'arrière suivant un demi-cercle déformé
- 2 : la source passe à l'arrière suivant un demi-cercle très déformé
- 1 : la source passe à l'arrière suivant un demi-cercle quasiment inexistant
- 0 : la source ne passe pas à l'arrière

- **Dessiner** le mouvement entendu. Indiquez le sens du mouvement par une flèche.
- Avez vous des remarques ? (pas obligatoire)



## TEST N° 2 :

**Axe Interaural :** (cf document : **en vert**)

- 5 : l'externalisation de la source est parfaitement perceptible sur l'axe interaural
- 4 : l'externalisation de la source est bien perceptible sur l'axe interaural
- 3 : l'externalisation de la source est perceptible sur l'axe interaural
- 2 : l'externalisation de la source est peu perceptible sur l'axe interaural
- 1 : l'externalisation de la source n'est mal perceptible sur l'axe interaural
- 0 : l'externalisation de la source est imperceptible

- **Dessiner** la trajectoire sur le schéma ci-dessous. Évaluez la distance ou donnez moi une échelle.
- La source passe devant ou derrière la tête ?



Remarques éventuelles :

## TEST N°3 :

**Axe Interaural :**

- 5 : l'externalisation de la source est parfaitement perceptible sur l'axe interaural
- 4 : l'externalisation de la source est bien perceptible sur l'axe interaural
- 3 : l'externalisation de la source est perceptible sur l'axe interaural
- 2 : l'externalisation de la source est peu perceptible sur l'axe interaural
- 1 : l'externalisation de la source n'est mal perceptible sur l'axe interaural
- 0 : l'externalisation de la source est imperceptible

- **Dessiner** la trajectoire sur le schéma ci-dessous. Évaluez la distance ou donnez moi une échelle.
- La source passe à devant ou derrière la tête ?

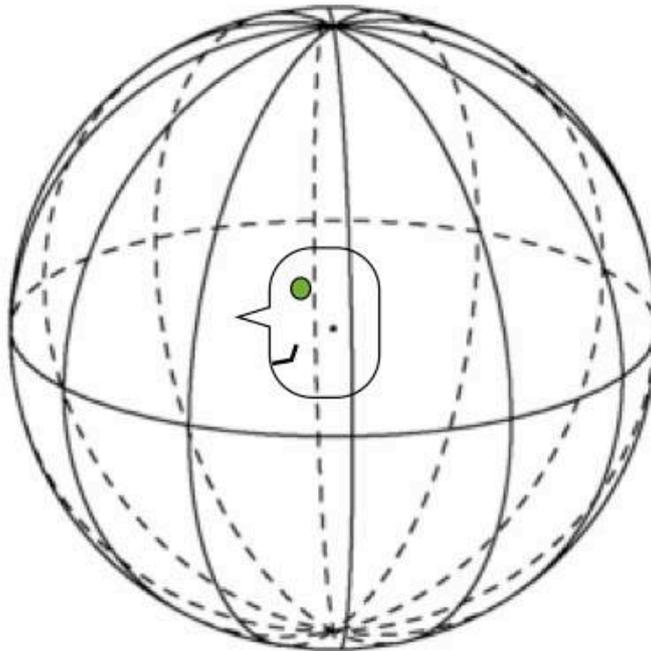


Remarques éventuelles :

**TEST N° 4 :**

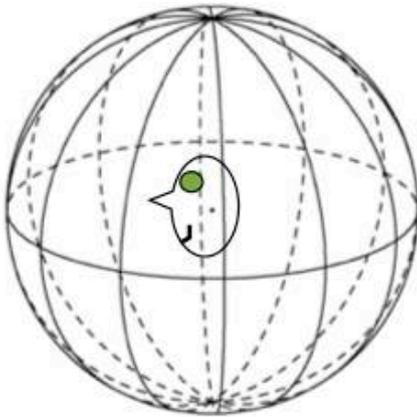
Après quelques écoutes, dessinez sur la sphère le mouvement de la source entendu.

Tête vue de  
profil



Remarques éventuelles :

**TEST N°5 : Axe vertical. (cf document : bleu)**



•Après quelques écoutes, dessinez le mouvement entendu sur cette sphère.

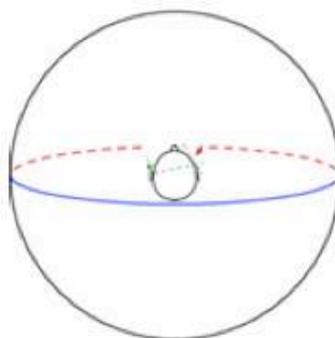
•Sur l'axe, la source passe devant ou derrière la tête ?

• De bas en haut, à combien estimez vous la longueur de l'axe (en m) parcourue par la source ?

**TEST N° 6 :**

Après quelques écoutes, **dessinez** sur la sphère le mouvement de la source entendu.

Vue de dessus



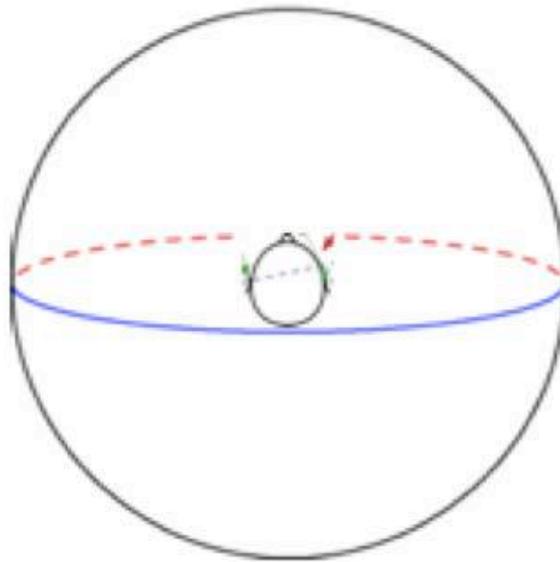
Remarques éventuelles :

**TEST N° 7 :**

**Plan Sagital droit : vue de dessus. Plan sagital ( cf document )**

Après quelques écoutes, dessiner sur la sphère **le plan sagital droit** que vous entendez.

Proposez moi une échelle. (rayon en m)



*Remarques éventuelles :*

**TEST N° 8 :**

D'après le schéma proposé (cf document), quel **axe** est mis en valeur dans ce test ?  
{ interaural (vert) / median (rouge) / vertical (bleu) }

*Réponse :*

(suite **TEST 8**)

- 5 : l'externalisation de la source est parfaitement perceptible sur l'axe médian
- 4 : l'externalisation de la source est bien perceptible sur l'axe médian
- 3 : l'externalisation de la source est perceptible sur l'axe médian
- 2 : l'externalisation de la source est peu perceptible sur l'axe médian
- 1 : l'externalisation de la source n'est mal perceptible sur l'axe médian
- 0 : l'externalisation de la source est imperceptible

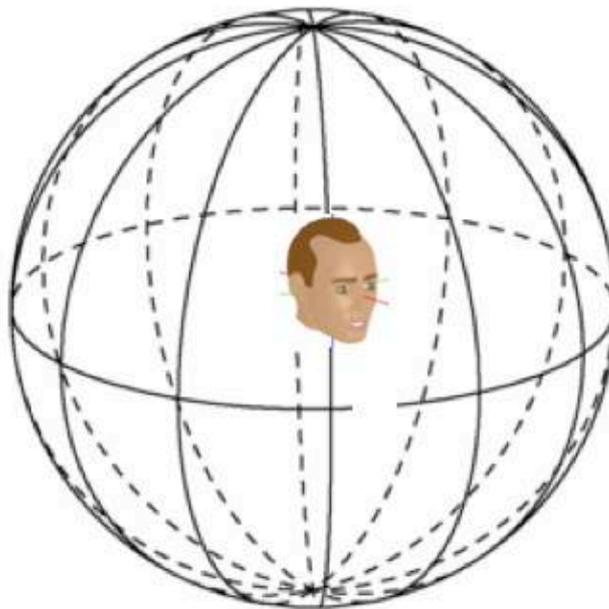
(Entourer la réponse)

- Quelle oreille la source contourne-t-elle ?

Dans quel sens se dirige la source ?

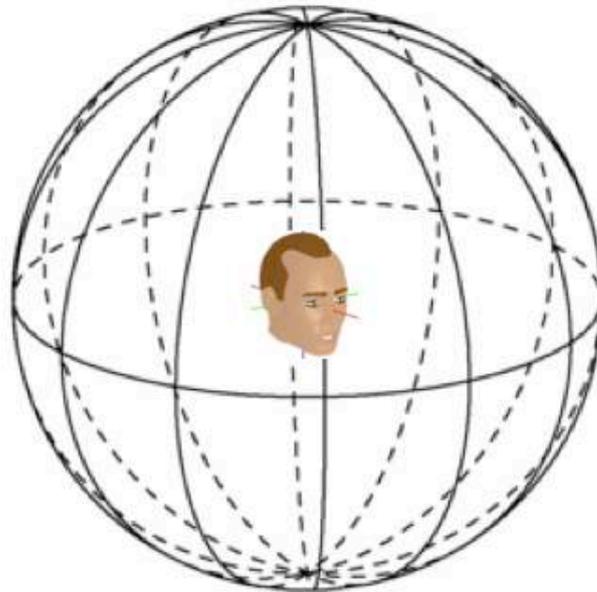
**TEST N°9 :**

- Après quelques écoutes, dessinez le mouvement de la source. (Attention à la position de la tête)



**TEST N°10 : Plan Vertical Interaural. (cf document)**

- Après quelques écoutes, dessinez le mouvement de la source entendu. Vous pouvez éventuellement vous aider des plans méridiens.



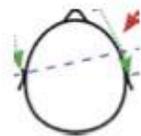
**TEST N°11 :**

- Après quelques écoutes, représentez la trajectoire de la moto à partir du schéma proposé (tête vue de dessus), en essayant de respecter des notions de distances. (donnez moi une échelle).



### TEST N°12 :

- Avec l'aide de ces 2 images, représentez à nouveau la trajectoire entendue.
- L'externalisation s'arrête-t-elle avant l'extinction du bruit du moteur de la moto ?
- Les images vous ont-elles aidé à évaluer la trajectoire ainsi que la distance ?



**TEST N°13 :**

**Extrait de la préparation «Les Papous»**

Vous écoutez les deux versions A et B en switchant de l'une à l'autre (utilisez le solo sur protocols).  
• Comment évaluez vous chaque version ?

Commentaires		
	A	B
+		
-		

**Version préférée ?**

Réponse :

### «OSCI» :

Sensation arrière :

- 5 : parfaitement rendue
- 4 : bien rendue
- 3 : rendue
- 2 : peu rendue
- 1 : presque pas perceptible
- 0 : imperceptible

Sensation d'immersion :

- 5 : l'auditeur se sent immergé totalement.
- 4 : l'auditeur se sent bien immergé
- 3 : l'auditeur se sent immergé
- 2 : l'auditeur se sent peu immergé
- 1 : l'auditeur se sent mal immergé
- 0 : l'auditeur ne se sent pas immergé

### TEST N°15 :

#### «Applaudissements» :

- Vous intégrez-vous dans le public présent ? Oui / Non

- 5 : parfaitement intégré
- 4 : bien intégré
- 3 : assez intégré
- 2 : pas assez intégré
- 1 : mal intégré
- 0 : pas intégré du tout

## TEST N°16 :

### Live R. Bona Marciac

- Vous intégrez-vous dans le public présent ? Oui / Non

5 : parfaitement intégré

4 : bien intégré

3 : assez intégré ou trop intégré pour bien apprécier le concert

2 : pas assez intégré

1 : mal intégré

0 : pas intégré du tout

- Où vous situez vous ?

- Distance par rapport à la scène ?

## TEST N°17 :

Vous écoutez les 8 items sonores successivement sur les enceintes, puis au casque en cliquant sur les solos LFE, A et B du protocols.

Vous donnez pour chaque item au casque, une évaluation par rapport à la référence sur les enceintes. Les notes attribuées sont :

0 : si l'écoute au casque est équivalente en spectre, en image et en immersion à l'écoute sur les enceintes

+1 : si l'écoute au casque est légèrement mieux rendue en spectre, en image et en immersion à l'écoute sur les enceintes

+2 : si l'écoute au casque est mieux rendue en spectre, en image et en immersion à l'écoute sur les enceintes

+3 : si l'écoute au casque est beaucoup mieux rendue en spectre, en image et en immersion à l'écoute sur les enceintes

-1 : si l'écoute au casque est un peu moins bonne en spectre, en image et en immersion à l'écoute sur les enceintes

-2 : si l'écoute au casque est moins bonne en spectre, en image et en immersion à l'écoute sur les enceintes

-3 : si l'écoute au casque est beaucoup moins bonne en spectre en image et en immersion à l'écoute sur les enceintes

ITEMS	Critère	A	B
1 Chronique Réveil	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
2 BlackBach	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
3 Avitabile	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
4 Géode	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
5 Attahir	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
6 Stravinsky	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
7 Jazz Sympho	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		
8 Hearing is believing	<b>Spectre</b>		
	<b>Image</b>		
	<b>Immersion</b>		

# Annexe 7

15		16		17		18		ITEMS	MOYENNE SPECTRE		MOYENNE IMAGE		MOYENNE IMMERSION		MOYENNE PAR ITEM Mix A	MOYENNE PAR ITEM Mix B	
A	B	A	B	A	B	A	B		A	B	A	B	A	B			
-1	0	0	0	-3	0	-1	-2	CHRONIQUE	SPECTRE	-1,0833333	-0,3076923					-1,7388888888889	-0,76923076923
-1	0	0	0	-3	+1	-1	-3	Monitoring Binaural	IMAGE		-2,53333333	-1					
-1	+1	0	0	-3	+1	-1	-2	H3D	IMMERSION				-1,6	-1			
0	-1	-2	-2	-1	-1	0	-2	BLACK BACH	SPECTRE	-0,6923077	-1,33333333					-1,050213675214	-1,28327228327
0	-1	-2	-1	-2	-2	0	-2	Monitoring Binaural	IMAGE		-1,33333333	-1,23076923					
-1	-1	-2	0	-1	-2	0	-2	Scat	IMMERSION				-1,125	-1,28571429			
-1	+1	0	-1	0	-1	-1	-2	F.AVITABILE	SPECTRE	-0,7692308	-0,8666667					-0,923076923077	-0,933333333333
-2	0	-1	-2	+1	-1	0	0	Monitoring Binaural	IMAGE		-0,91666667	-1					
-2	-1	-1	-1	-1	-1	0	-1	Scat	IMMERSION				-1,08333333	-0,93333333			
-2	-1	-2	0	-2	0	-1	0	GEODE	SPECTRE	-1,5833333	-0,9166667					-1,815656565657	-0,88888888889
-1	0	-2	-1	-2	+1	-1	0	Lvswe + transaural	IMAGE		-2,36363636	-1					
-2	0	-1	-1	-2	0	-1	0	H3D	IMMERSION				-1,5	-0,75			
0	0	0	-1	-1	-1	0	0	B.ATTAHIR	SPECTRE	-0,73333333	-0,9230769					-1,29503968254	-1,02847152847
+1	0	-1	-1	-2	-2	-1	-1	Lvswe + transaural	IMAGE		-2,21428571	-1,07142857					
0	-1	0	-1	0	-1	-2	0	H3D	IMMERSION				-0,9375	-1,09090909			
-1	-1	-2	-1	-3	0	-1	-1	STRAVINSKY	SPECTRE	-1,5	-0,6923077					-2,145833333333	-0,74198717949
-2	0	-2	-1	-3	0	-2	0	Lvswe + transaural	IMAGE		-3,4375	-0,6875					
-2	-1	-2	-1	-3	+1	-1	0	H3D	IMMERSION				-1,5	-0,84615385			
-1	-1	-2	0	-3	-1	-1	0	JAZZ SYMPHO	SPECTRE	-1,0666667	-0,8666667					-1,334722222222	-1,061111111111
-2	-1	-1	0	-2	-1	-1	-1	Lvswe + transaural	IMAGE		-1,8125	-1,25					
-2	0	-1	0	-2	-2	0	0	H3D	IMMERSION				-1,125	-1,06666667			
-1	0	-1	0	-1	0	-1	-2	HEARING IS ...	SPECTRE	-0,5384615	-0,6428571					-1,304273504274	-1,14814814815
-1	-2	-2	-3	-2	-1	-1	-2	Monitoring Binaural	IMAGE		-2,06666667	-1,44444444					
-2	-1	-1	-1	-1	0	-1	-2	H3D	IMMERSION				-1,30769231	-1,35714286			
								MOYENNE GENERALE PAR CRITERES		-0,9958333	-0,8186584	-2,08474026	-1,08551778	-1,27231571	-1,04124001		
								MOYENNE GENERALE ITEMS								-1,450963099401	-0,98180540524

Résultats test n°17