



**TENTATIVE D'ÉTABLISSEMENT
D'UNE GRAMMAIRE DE
L'ÉCRITURE SONORE SPATIALISÉE**

Vincent MEURISSE

Mémoire de MASTER 2 – Spécialité Son

Directeur de mémoire interne : Alan BLUM
Directeur de mémoire externe : Frédéric CHANGENET
Examinateur : Jean-Loup PECQUAIS
Référent académique : Laurent MILLOT
Coordinateur des mémoires : Corsin VOGEL

Remerciements

Mes remerciements les plus chaleureux sont adressés à Frédéric Changenet, sans qui ce sujet n'aurait pas existé. Je le remercie pour le temps et la confiance qu'il m'a accordés, ainsi que pour toutes les personnes qu'il m'a fait rencontrer. Il a su me guider lorsque que je m'égarais sur le chemin de ce mémoire. Son expertise et sa rigueur m'ont été plus que précieuses. Ma plus grande appréhension concernant cet écrit est qu'il soit à la hauteur de ses attentes.

Je remercie vivement tous nos professeurs de l'ENS Louis Lumière pour ces trois années de formations qui m'ont tout appris d'un métier dont je ne connaissais rien à mon arrivée. Merci à Corsin Vogel pour son accompagnement académique irréprochable et Alan Blum pour ses retours avisés. Je remercie particulièrement Florent Fajole pour sa connaissance sans limite de notre littérature spécifique, ainsi qu'Agnès Hominal pour sa réactivité dans son suivi administratif.

Merci à tous mes camarades de l'école, avec qui nous nous sommes retrouvés dans différentes bibliothèques parisiennes pendant la rédaction de nos mémoires, dans une période où il était difficile pour l'ENS Louis Lumière de défendre quelconque esprit de cohésion. Je leur souhaite à tous, sincèrement, de s'épanouir dans leurs vies respectives.

Je remercie infiniment toutes les personnes qui m'ont accordé de leur temps pendant mes entretiens : Martin Antiphon, Jérémie Basset, Hervé Déjardin, Alessandra Denegri, Alexandre Garcia, Xavier Gibert, Olivier Goinard, Augustin Muller, Antoine Petroff, Éric Raynaud, William Russell, Gerriet K. Sharma et Lukas Wilke. Je remercie également Abigail Wincott et Adrien Zanni pour nos discussions informelles particulièrement intéressantes.

Je remercie tout spécialement Marion Laporte pour nos longues discussions qui ne cessent de me remettre en question et de me faire avancer.

Mes derniers remerciements sont évidemment adressés à ma famille et à mes proches qui me soutiennent au quotidien.

*Toutes mes pensées se dirigeant maintenant vers Pascal Spitz,
à qui j'aurais aimé présenter ce dernier travail à l'ENS Louis Lumière.*

Résumé

Dans un contexte où l'audio immersif devient un standard des manifestations culturelles à succès et où le format tend à s'imposer dans l'industrie du streaming, la question de la manière dont on conçoit une image sonore immersive se pose tout particulièrement. Le constat de pratiques empiriques employées par les ingénieurs du son et artistes de tous médias incite à la tentative d'établissement d'une grammaire commune.

Ce mémoire a recueilli les témoignages d'acteurs de l'audio immersif et étudié dix d'entre eux à l'aide d'une analyse thématique afin d'en extraire les processus et méthodologies qui semblent faire consensus. Au regard de nos connaissances en psychoacoustique, de l'histoire de l'audio spatial et d'une étude de la littérature spécialisée, ce travail a établi une première ébauche non exhaustive d'une grammaire de l'écriture sonore immersive. Les résultats sont présentés en trois catégories : le mixage, la technique et la philosophie. Chaque catégorie détaille un certain nombre de concepts et règles de grammaire qui seront questionnées, ainsi que les citations qui ont permis de les établir.

Certaines individualités choisies parmi les entretiens réalisés sont mises en avant afin d'expliciter les différentes visions qui se confrontent aujourd'hui. Trois mixages spatialisés ont également été réalisés. Ils tentent d'assumer des choix esthétiques ambitieux qui entrent en résonance avec les résultats de l'analyse présentée au préalable. Un lien vers les fichiers ADM, ainsi que les rendus binauraux, sont mis à disposition.

Ce travail a pour ambition de donner des idées aux artistes et créateurs de contenus, d'offrir un accès à la connaissance de certaines méthodes utilisées dans le processus d'écriture immersive et de partager des outils de communication aux acteurs de l'audio spatial.

Mots clefs : Audio orienté objet, écriture sonore, grammaire, immersion, mixage, perception, psychoacoustique, son immersif, son spatialisé.

Abstract

In a context where immersive audio is becoming a standard for every kind of successful cultural event and where spatial audio is conquering the audio streaming industry, the time has come to question how to build an immersive sound image. The observation that engineers and artists of all disciplines are working on the basis of empirical knowledge is an incentive for us to establish a common sound grammar.

This Master's thesis has collected the testimonies of immersive audio professionals and studied ten of them via a thematic analysis to highlight processes and methods that appear to enjoy a consensus among professionals. In view of the above, this research paper lays down the foundations for an immersive sound writing grammar. The results are presented in three categories: mixing, technical and philosophical. For each of these categories, details of specific concepts and the testimonies that refer to them are provided.

Certain professionals have been chosen to showcase the different visions that exist nowadays in the approach to spatial audio. The process of three mixes in which we attempted to make strong aesthetic choices in accordance with the analysis presented in the previous section is also described. A link to the ADM and binaural files is provided.

The aim of this research is to provide ideas for artists and sound engineers, offering them access to certain methods used in the immersive writing process. It is also a means of sharing communication tools with spatial audio professionals.

Key words: *Audio grammar, immersion, immersive audio, object-based audio, perception, psychoacoustic, sound mix, sound writing, spatial audio.*

Table des matières

<i>Introduction.....</i>	8
<i>1. Rappels sur les principes de l'écoute binaurale</i>	10
1.1 Les indices de localisation d'une source sonore	10
1.1.1 Localisation dans le plan azimutal.....	10
1.1.2 Perception de l'élévation	11
1.1.3 Distance et profondeur	13
1.1.4 L'importance des mouvements de tête.....	14
1.1.5 Sources sonores mobiles.....	14
1.2 Influence d'autres critères sensoriels et subjectifs	15
1.2.1 Le rôle de la mémoire et de l'expérience.....	15
1.2.2 L'importance de la vue	15
1.2.3 L'intégration de plusieurs sens nécessaire à la bonne perception spatiale.....	16
<i>2. L'immersion sonore ; histoire d'un confort d'écoute.....</i>	18
2.1 La relation historique entre musique et espace	18
2.1.1 La spatialisation statique	18
2.1.2 L'électroacoustique introduit la spatialisation dynamique	18
2.1.3 La relation son – espace	19
2.1.4 Se donner les moyens techniques des visions créatives.....	20
2.2 Une catégorisation de l'audio spatial	23
2.2.1 Audio orienté canal	23
2.2.1.1 La stéréophonie.....	24
2.2.1.2 Les systèmes de diffusions à plus de 2 haut-parleurs.....	24
2.2.2 L'audio orienté scène et la reproduction du champ acoustique	26
2.2.3 La <i>Next Generation Audio</i> : l'Audio Orienté Objet et l'ADM.....	26
2.2.4 Le binaural	27
2.3 Les solutions technologiques à disposition pour travailler l'espace	28
2.3.1 Le travail de l'espace sur logiciel.....	28
2.3.2 Les processeurs de diffusion spatialisée	28
2.4 Le temps de la démocratisation	29
2.4.1 Le binaural à la conquête du streaming.....	29
2.4.2 La multiplication des concerts immersifs	30
2.4.3 Des outils accessibles au Home Studio	31
<i>3. Contexte et caractéristiques du mixage immersif.....</i>	33
3.1 L'enveloppement, clef de l'immersion	33
3.1.1 L'impression d'espace chez l'auditeur : ASW et LEV.....	33
3.1.2 L'enveloppement, une notion difficile à définir.....	35
3.2 Des définitions du son immersif	36
3.2.1 Les éléments de définition de l'immersion et de l'audio spatialisé	36
3.2.2 Les différents contextes de l'écoute immersive	37
3.2.3 Nos définitions	38
3.3 Conception d'une image sonore à 360°, des règles à suivre ?	39
3.3.1 Du mixage stéréophonique au mixage immersif : du complexe au naturel	39

3.3.2	Utiliser les critères de perception.....	40
3.3.3	L'importance du mouvement de tête.....	42
3.3.4	L'expérience en audio 3D décisive dans la perception	43
3.3.5	L'importance de l'élévation discutée.....	44
3.3.6	La localisation et le rôle du mouvement	46
4.	<i>Retours d'expérience ; les pratiques immersives actuelles des ingénieurs du son</i>	48
4.1	La pratique empirique de chacun, un accès aux connaissances	48
4.1.1	Le recours aux entretiens individuels	48
4.1.2	Comprendre les visions individuelles de l'audio spatial	50
4.1.3	Les personnes contactées.....	50
4.2	Analyse des entretiens	51
4.2.1	Des entretiens qualitatifs et semi-directifs	51
4.2.2	Guide d'entretien et difficultés	52
4.2.3	Le recours à une analyse thématique.....	53
4.3	Résultats d'analyse	54
4.3.1	Les résultats de l'analyse thématique.....	54
4.3.1.1	L'arbre thématique	54
4.3.1.2	La liste des concepts et la grammaire associée.....	59
4.3.2	Des visions de l'audio spatialisé très personnelles	60
5.	<i>Le détail des concepts, l'établissement d'une grammaire non exhaustive de l'écriture sonore immersive.....</i>	61
5.1	Catégorie 1 : Le mixage.....	61
5.1.1	Thème 1 : La position.....	61
5.1.2	Thème 2 : L'espace.....	67
5.1.3	Thème 3 : Le spectre	68
5.1.4	Thème 4 : La compatibilité.....	69
5.1.5	Thème 5 : L'influence de la stéréo	70
5.1.6	Thème 6 : Les effets et traitements	71
5.2	Catégorie 2 : La technique.....	73
5.2.1	Thème 1 : La prise de son.....	73
5.2.2	Thème 2 : La diffusion.....	74
5.2.3	Thème 3 : L'adaptabilité.....	75
5.2.4	Thème 4 : L'accessibilité	77
5.3	Catégorie 3 : La philosophie.....	79
5.3.1	Thème 1 : L'immersion.....	79
5.3.2	Thème 2 : L'écriture	81
5.3.3	Thème 3 : La culture	84
5.3.4	Thème 4 : La subjectivité	87
5.3.5	Thème 5 : L'histoire.....	89
<i>Discussion</i>		91
<i>Conclusion.....</i>		93

<i>Bibliographie</i>	96
<i>ANNEXES.....</i>	100
Annexe 1 : La stéréophonie.....	101
Annexe 2 : L'ambisonie et la WFS.....	103
A1.1 L'ambisonie	103
A1.2 La synthèse de front d'onde (WFS).....	105
A1.3 Les limites de l'audio orienté scène et de la WFS	106
Annexe 3 : Exemple d'utilisation de l'audio orienté objets.....	107
Annexe 4 : Le binaural.....	110
Annexe 5 : Le travail de l'espace sur logiciel	113
Annexe 6 : Quelques définitions de l'immersion et de l'audio spatialisé.....	115
Annexe 7 : Les processeurs de diffusion spatialisée	117
Annexe 8 : Liste de personnes interviewées et leurs parcours.....	118
Annexe 9 : Formulaire de consentement aux entretiens	121
Annexe 10 : Guide d'entretien.....	123
Annexe 11 : Les philosophies de chacun	124
A11.0 Avant-propos : la vision de Frédéric Changenet.....	124
A11.1 Martin Antiphon.....	127
A11.2 Olivier Goinard	129
A11.3 Augustin Muller.....	131
A11.4 Antoine Petroff.....	133
A11.5 Éric Raynaud	136
Annexe 12 : Retranscriptions des entretiens.....	139
A12.1 Entretien avec Martin Antiphon	140
A12.2 Entretien avec Hervé Déjardin.....	153
A12.3 Entretien avec Xavier Gibert	169
A12.4 Entretien avec Olivier Goinard.....	185
A12.5 Entretien avec Augustin Muller.....	196
A12.6 Entretien avec Antoine Petroff	204
A12.7 Entretien avec Éric Raynaud	220
A12.8 Entretien avec William Russell et Alessandra Denegri	225
A12.9 Entretien avec Gerriet K. Sharma.....	232
A12.10 Entretien avec Lukas Wilke.....	241

Introduction

Depuis maintenant 70 ans, la stéréophonie s'est imposée comme le dispositif idéal d'écoute et est encore le plus répandu. Son règne semble aujourd'hui menacé par le développement du son immersif. Que ce soit en écoute individuelle au casque, dans un contexte de live, au cinéma ou encore à la radio, les contenus immersifs se multiplient. L'immersion s'installe comme la norme des contenus de spectacles vivants, d'expositions et de manifestations culturelles en tout genre, et le terme est largement utilisé aussi bien pour définir des expériences visuelles que sonores.

Les premières expérimentations sonores immersives remontent aux années 50. Le Pavillon Philips, dans lequel est diffusé le Poème Électronique de Iannis Xenakis (1922-2001) et Edgard Varèse (1883-1965) sur 425 haut-parleurs, marque l'histoire en 1958. L'essor de la musique concrète et électroacoustique dans la seconde moitié du vingtième siècle permet à de nombreux compositeurs d'explorer des dispositifs toujours plus innovants, avec l'aide desquels ils posent un regard neuf sur les interactions entre haut-parleurs, musiciens et spectateurs. Des approches compositionnelles inédites voient le jour.

D'un point de vue technique, des dispositifs comportant de plus en plus d'enceintes apparaissent et commencent à trouver leur place, parmi lesquels nous pouvons citer la quadriphonie, le 5.1, le 7.1, ou encore le 8.0¹. Tout est finalement permis en termes de nombre de haut-parleurs et de disposition, seule changera la technologie de traitement du signal utilisée afin de créer une scène sonore.

Depuis quelques années, au-delà de la multiplication des expériences immersives, l'accès aux contenus musicaux spatialisés est facilité. Dès 2021, *Apple* pousse les producteurs à proposer leur musique en audio immersif, notamment via la technologie Dolby Atmos. Les studios s'équipent ainsi de systèmes de plus en plus complexes – jusqu'au 9.1.6² – afin de répondre à la demande des distributeurs.

Tout un espace s'ouvre aux créateurs de contenu et les manières de l'appréhender sont particulièrement variées. Si l'écriture stéréophonique fait preuve de dizaines d'années de technique et de consensus, l'écriture spatialisée ne semble pas avoir trouvé

¹ La notation x.y correspond à une disposition à x haut-parleurs sur le plan azimutal et y LFE pour « *Low Frequency Effect* ».

² Le troisième chiffre correspond ici au nombre de haut-parleurs qui se situent en hauteur.

commune entente. Les professionnels évoluent individuellement, engageant dans l'écriture des processus techniques et narratifs propres à des conclusions empiriques.

Ce mémoire souhaite précisément se concentrer sur la grammaire de l'écriture spatialisée – c'est-à-dire sur tous les éléments et processus, relatifs à la technique aussi bien qu'à la narration, mis en œuvre pour obtenir un contenu immersif – et se pose la question suivante : comment, concrètement, construire une image sonore à 360 degrés ?

Pour répondre à cette question et tenter d'établir une ébauche de grammaire, nous souhaitons recueillir des témoignages d'artistes et d'ingénieurs du son qui produisent des contenus en audio spatialisé. C'est à travers la pratique empirique que nous récolterons le plus d'éléments de grammaire. Nous tenterons également, à travers ces entretiens, de répondre à des questions soulevées par la littérature, qui s'est penchée sur le sujet de l'audio spatialisé depuis de nombreuses années. Nous nous intéresserons notamment aux critères de perception, aux difficultés rencontrées lors du mixage, au recours au mouvement, à l'utilisation des *tops*, des *surround* et des arrières³, à l'impact de la culture sur la production, à l'orientation du contenu et à l'influence de la stéréophonie sur l'écriture sonore immersive.

Nous développerons cinq parties avant de discuter la pertinence de nos résultats et de conclure. La première partie effectuera un résumé de la perception sonore de l'être humain. Nous reviendrons ensuite, dans une seconde partie, sur l'histoire de la spatialisation et des techniques qui en découlent, ainsi que sur son implantation dans l'actuel paysage culturel. En troisième partie, nous évoquerons les questions que se sont posées les chercheurs quant au rendu sonore immersif et donnerons certaines définitions. Nous étudierons ensuite dix entretiens réalisés avec des professionnels de l'audio spatial à l'aide d'une analyse thématique qui nous donnera une ébauche de grammaire de l'écriture sonore immersive. Cette grammaire sera enfin détaillée à l'aide de citations en cinquième partie.

³ Dans ce mémoire, les notions de *tops*, *surround* et arrières font référence aux enceintes respectivement placées en hauteur, sur les côtés et à l'arrière de l'auditeur.

1. Rappels sur les principes de l'écoute binaurale

« L'oreille est l'organe du temps, un paramètre insaisissable, étirable, beaucoup plus complexe à appréhender. »

Solomos, cité par Guiot (2017)

Comme l'explique Rumsey dans son livre *Spatial Audio* (2001, p.21), « [ce chapitre] n'a pas l'intention d'être une revue exhaustive de la perception de l'espace étant donné que cela a été largement documenté dans d'autres contextes, notamment par Blauert (1997) dans son livre *Spatial Hearing* et Moore (1989) dans An introduction to the Psychology of Hearing. »⁴ En nous inspirant du livre de Rumsey, nous synthétiserons dans ce qui suit les principes fondamentaux de l'écoute binaurale afin d'offrir au lecteur les outils nécessaires à la bonne compréhension de ce mémoire sans entrer dans les détails des ouvrages de Blauert ou de Moore.

1.1 Les indices de localisation d'une source sonore

1.1.1 Localisation dans le plan azimutal

La localisation d'une source sonore par l'être humain dans le plan azimutal (plan parallèle au sol et qui passe par les deux oreilles) est réalisée grâce à deux phénomènes essentiels. Lorsqu'une source sonore est émise, les deux oreilles sont capables d'interpréter :

- La différence de temps entre le moment où le son émis par la source arrive à l'oreille droite et le moment où il arrive à l'oreille gauche. On la note Δt ou ITD pour « Interaural Time Difference » en anglais.
- La différence d'intensité sonore perçue par l'oreille droite par rapport à celle perçue par l'oreille gauche. On la note Δi ou ILD pour « Interaural Level Difference » en anglais.

Ces indices étant basés sur une différence de perception entre les deux oreilles, ils portent le nom d'indices binauraux de localisation.

Ces différences sont nulles pour une source située exactement à l'avant l'auditeur, position de la source que l'on considère comme étant à 0° (origine du cercle

⁴ « [This chapter] is not intended as an exhaustive review of spatial perception, as this has been very thoroughly done in other places, notably Blauert (1997) in his book *Spatial Hearing* and Moore (1989) in *An Introduction to the Psychology of Hearing*. » Notre traduction

trigonométrique qui entoure l'auditeur) et exactement à l'arrière, à 180°. Des expériences ont d'ailleurs mis en lumière des confusions de perception avant-arrière dès la fin du 18^e siècle (Venturi, 1796). L'existence d'un cône de confusion, dont l'axe de symétrie est celui des oreilles, est également prouvée à la fin du 19^e (Rayleigh, 1877). La localisation d'une source sonore située dans ce cône est propice à des erreurs de jugement. Enfin, à la fin du 20^e siècle, Blauert (1997) travaille sur le pouvoir séparateur spatial sur le plan azimutal et conclut que notre perception est plus précise à l'avant et à l'arrière (+/- 4 à 5°) que sur les côtés (+/- 10°).

Le Δi est plus performant pour les hautes fréquences tandis que le Δt l'est plus pour les basses fréquences. Cela est dû à l'obstacle formé par la tête et son impact sur la longueur d'onde des sons perçus. Un son dont la longueur d'onde est deux fois supérieure à la distance entre nos deux oreilles (environ 17cm, ce qui donne des fréquences inférieures à 1000Hz) ne sera pas impacté par la tête et n'aura pas un Δi significatif. Pour ces fréquences basses, l'être humain est plutôt en mesure de détecter des différences de temps, qui correspondent à des différences perçues de phase du signal. À l'inverse, lorsque la longueur d'onde est plus petite et la fréquence supérieure à 1000 Hz, la tête jouera son rôle d'obstacle absorbant et réfléchissant, induisant un Δi notable. (Jullien, 2022)

Une des conséquences de la perception du Δi et du Δt réside dans l'effet de précédence, autrement appelé effet Haas ou loi du premier front d'onde. Lorsqu'un même son est émis deux fois d'affilée en dessous de la limite de l'écho (environ 50ms), l'oreille humaine ne sera pas capable de discriminer les deux signaux et n'en entendra qu'un seul. La localisation d'une source sonore unique sera perçue dans la direction du signal que l'oreille perçoit en premier. Dans le cas où la différence de temps entre le signal perçu par les deux oreilles est inférieure à 1 ms, on distingue une image fantôme entre les deux sources. Dans un contexte de son réverbéré, l'effet Haas est fondamental dans le phénomène de perception car il implique que la réverbération de la source ne perturbe pas sa localisation.

1.1.2 Perception de l'élévation

Avant qu'une source sonore ne soit perçue par un individu, son spectre a subi différentes modifications notables. Comme nous l'avons vu précédemment, la tête représente une barrière pour les hautes fréquences. Le haut du corps et la forme du visage vont également filtrer certaines fréquences, tout comme l'oreille externe, dont la forme favorise l'apparition de résonnances dont les fréquences dépendent de l'angle

d'incidence de la source. Tous ces éléments constituent ce qu'on appelle la HRTF pour « Head Related Transfer Function » en anglais.

La HRTF, intimement liée au physique de l'auditeur, est, par définition, propre à chaque individu. Le cerveau est en mesure d'analyser les différences de spectre apportés au son perçu, conséquence de la HRTF. Ces différences sont appelées les indices spectraux. En termes d'élévation, le Δi et le Δt ne suffisent pas à établir la localisation d'une source sonore. Ce sont ces indices spectraux qui le permettent. Il est important de préciser que la HRTF joue également un rôle dans la localisation azimutale. Les différences de timbre apparaissent dans toutes les directions et participent ainsi à une localisation globale.

La précision de localisation en élévation est notablement moins bonne que dans le plan azimutal. Damaske et Wagener (cités dans (Blauert, 1997)) montrent que le pouvoir séparateur dans le plan sagittal (plan perpendiculaire au sol et qui passe entre les deux oreilles), est plus précis à l'avant (+/- 9 à 10°) que sur le dessus et l'arrière (+/- 13 à 22°).

La perception dans le plan sagittal est également dépendante du contenu fréquentiel de la source. Roffler et Butler (1968) ont montré que les sons purs ont tendance à être toujours perçus avec la même élévation, variable en fonction de la fréquence, même si ceux-ci viennent d'une hauteur différente. Les sons complexes, quant à eux, sont localisés à leur véritable hauteur. Dans la continuité de ces recherches, Blauert (1997) met en évidence l'existence de bandes de fréquences directives en termes de perception de sons purs ou à spectre étroit :

- La bande de fréquence 500Hz – 2kHz est perçue à l'arrière
- La bande de fréquence 2kHz – 4kHz est perçue à l'avant
- La bande de fréquence plus étroite autour de 8kHz est quant à elle perçue en hauteur. Ce phénomène est surnommé le « Pitch Height Effect » en anglais (Moélants, 2023)

Nilles (2021) souligne elle aussi l'importance du contenu spectral, particulièrement l'importance des hautes fréquences, dans la précision de localisation. En citant les expériences de Dobreva, O'Neill et Paige (2001), elle explique que le contenu fréquentiel au-delà de 3000 Hz doit être le plus étendu possible afin d'établir une bonne localisation dans le plan sagittal. C'est pourquoi les personnes âgées, ayant perdu leur acuité auditive dans le haut du spectre, auront plus de difficultés à localiser en hauteur.

1.1.3 Distance et profondeur

La distance et la profondeur jouent eux aussi un rôle crucial dans notre perception sonore. Rumsey (2001) définit un certain nombre de facteurs qui contribuent à leur perception :

- une perte de niveau due à la distance parcourue,
- une perte de hautes fréquences due à l'absorption de l'air,
- un rapport son direct / son réverbéré différent,
- une différence de temps entre la perception du son direct et des premières réflexions plus faible
- une atténuation des réflexions du sol.

Ces facteurs, tout comme les indices spectraux et contrairement aux indices binauraux, ne nécessitent pas d'utiliser ses deux oreilles pour être perçus. Une seule oreille suffit. Ils sont appelés indices monauraux. (Moélants, 2023)

Le rapport entre son direct et son réverbéré est directement lié à la sensation de distance. Jullien (2022) nous explique que « *dans une salle qui a une certaine réverbération, le rapport des deux niveaux est directement lié à la distance source-auditeur.* » Il est toutefois difficile de juger de la distance d'une source sonore en utilisant uniquement l'audition, surtout dans un champ libre non réverbéré, ou lorsque l'acoustique du lieu n'est pas connue. L'individu peut s'aider de la vision dans son jugement. Il peut également faire appel à son expérience personnelle et à son apprentissage pour mieux juger de la distance. A. Moélants (2023) le précise : « *Quand on considère un son familier, comme une voix calme, nous avons une certaine idée de la manière dont il sera entendu à une certaine distance.* »⁵ Dans le cas d'une source inconnue, l'erreur d'évaluation de distance peut s'élever à 20% (Nilles, 2018, cite Moore). Nous parlerons plus en détail du rôle de l'expérience sur la perception en 1.2.1.

Moélants (2023) ajoute que si la perception de la distance et de la profondeur est principalement due aux indices monauraux, nous pouvons également ajouter un indice binaural qui complète cette perception : la corrélation entre le signal arrivant à l'oreille droite et celui arrivant à l'oreille gauche. Plus les signaux sont décorrélés, plus la source dont ils proviennent sera perçue proche.

⁵ « *When we take a familiar sound, like a calmly talking voice, we had a certain expectation of how loud it should be at a certain distance.* » Notre traduction.

1.1.4 L'importance des mouvements de tête

Nous avons vu que les différents indices de localisation d'une source sonore dans l'espace peuvent donner lieu à certaines confusions (confusion avant-arrière, cône de confusion, perception de hauteur biaisée). Ces erreurs de jugement peuvent être corrigées par des mouvements de tête de la part de l'auditeur. Comme le souligne Moélants (2023) : « *Les mouvements de tête modifient à la fois des indices binauraux et monauraux.* »⁶ Lorsque nous bougeons la tête, les différents filtres induits par notre corps changent, ainsi que l'angle d'incidence de la source. Tous les indices s'en retrouvent modifiés et analysés par le cerveau. La précision de localisation en sera meilleure.

1.1.5 Sources sonores mobiles

À l'instar du mouvement de tête, le déplacement d'une source sonore implique le même type de modifications des facteurs qui permettent la localisation. Rumsey (2001) le souligne : « *Les sources mobiles ont tendance à donner plus d'informations que les sources statiques, permettant au cerveau d'analyser les différences dans les informations reçues et ainsi de résoudre certaines imprécisions.* »⁷

La perception d'une source mobile est contrainte par certaines limites. Blauert définit en 1997 les constantes de temps de la perception du mouvement (Nilles, 2018). Il écrit qu'un mouvement de gauche à droite est perçu à partir de 172 ms tandis qu'un mouvement d'avant en arrière est perçu à partir de 200 ms. Plus récemment, Féron *et al.* (2010) démontrent qu'une rotation autour de l'auditeur est perçue jusqu'à une vitesse de 2.5 rotations par secondes. Au-delà de cette limite, nous ne sommes pas en mesure de suivre la trajectoire de la source.

⁶ « *Head Movement shifts both binaural and monaural variables.* » Notre traduction.

⁷ « *Moving sources also tend to provide more information than stationary ones, allowing the brain to measure changes in the received information that may resolve some uncertainties.* » Notre traduction.

1.2 Influence d'autres critères sensoriels et subjectifs

1.2.1 Le rôle de la mémoire et de l'expérience

« Il ne faut pas s'attendre à ce que tout le public perçoive les mêmes objets sonores et le même espace. »

Zanni (2020)

Depuis la naissance, nous emmagasinons quantité d'informations issues de notre perception quotidienne. Lorsqu'une nouvelle sensation se présente, notre mémoire nous donne la capacité de la mettre en lien avec des sensations similaires que nous avons ressenties par le passé, ce qui nous permet d'analyser la nouvelle information et d'en tirer des conclusions. Nous l'avons vu précédemment : ce mécanisme joue un rôle primordial dans la perception de la distance (Moélants, 2023 ; Nilles, 2018). Il peut aussi être à l'origine de systématismes de jugement, même si ce dernier va à l'encontre de la réalité physique. C'est ce qui se passe lorsque nous entendons un avion. Depuis que l'avion existe, nous avons l'habitude de l'entendre en hauteur. Rumsey (2001) a fait l'expérience de faire écouter à des auditeurs, en contexte d'écoute binaurale, des sons d'avions venant du bas. Alors même que les indices de localisation prouvent le contraire, l'avion sera entendu en hauteur par les auditeurs.

Nilles (2018) explique, en citant Schmele, que, dans un contexte de concert acousmatique⁸, les auditeurs tendent à s'imaginer ce à quoi les sons entendus correspondent dans le monde réel. Ce sont les représentations mentales des sources qui constituent la scène sonore, représentations issues de notre analyse par le cerveau de ce que nous entendons, et des liens que le cerveau fait entre les sons et notre mémoire sensorielle, qui vont avoir le plus grand impact sur notre expérience.

1.2.2 L'importance de la vue

La vue joue un rôle primordial dans la perception d'une scène auditive. Jullien (2021) le rappelle dans son article : « [...] la vision est capable de corriger de fortes distorsions auditives. Autrement dit, les indices visuels l'emportent presque toujours sur les indices auditifs. » En plus de nous aider à juger de la distance, elle nous aide à

⁸ *Acousmatique : Se dit d'une situation d'écoute où, pour l'auditeur, la source sonore est invisible (Larousse).* Le concert acousmatique met en avant le sonore et non l'image. Ces représentations peuvent être réalisées dans le noir, ou rideau fermé.

localiser de manière plus précise. Il est prouvé que la localisation d'une source sonore est plus efficace les yeux ouverts que les yeux fermés (Nilles, 2018 cite Warren).

La relation entre la vision et l'audition existe jusque dans le cerveau, lors de l'analyse des informations transmises par ces deux sens. Nilles (2018) explique, en citant Auerbach, Sperling et Hollander, que l'élaboration de cartographies spatiales se fait dans une partie du cerveau qui analyse en même temps les données reçues par la vue et par l'ouïe. Les deux sens sont ainsi intimement liés.

La vue peut également être à l'origine de distorsions de la réalité et d'erreurs de perception. Lorsque la vue n'est plus présente, nous remarquons qu'il y a perte de repères et que la scène sonore est perçue différemment. C'est le problème constaté lors d'écoutes binaurales non dynamiques⁹ (c'est-à-dire lorsque les mouvements de têtes ne sont pas suivis d'une rotation de la scène sonore). Dans ces cas-là, l'absence de stimuli visuels entraînent une perception globale de la scène sonore à l'arrière de la tête (Rumsey, 2001). La scène frontale sera mal perçue car il n'y a pas de liens entre la vue et l'audition. L'être humain s'est en effet habitué à interpréter des sons qu'il ne voit pas comme provenant de l'arrière.

1.2.3 L'intégration de plusieurs sens nécessaire à la bonne perception spatiale

« Entendre est une manière de toucher à distance, et l'intimité s'élargit au social lorsque plusieurs personnes se réunissent pour écouter ensemble. »¹⁰

Murray Schaeffer (1977)

Au-delà de l'importance de la vue, tous les sens dans leur ensemble sont nécessaires à une bonne perception spatiale. Nilles (2018) cite Smalley et Schmele pour appuyer cette idée que la perception de l'espace est une expérience vécue par le corps tout entier¹¹. Elle cite également Periaux *et al.* (2015), qui explique que « *la plus forte représentation spatiale implique l'intégration des informations provenant des différents sens.* »

Cela va en faveur de nos développements précédents, notamment en ce qui concerne le binaural qui ne permet pas une image frontale sans *head tracking*. Cette idée

⁹ Dans la suite de ce mémoire, nous utiliserons le terme anglais *head tracking* pour aborder les écoutes binaurales dynamiques.

¹⁰ Traduit de l'anglais (Canada) par Sylvette Gleize

¹¹ « *A full body experience* » Notre traduction

d'expérience multisensorielle peut être interprétée d'une manière plus approfondie en affirmant que même une expérience binaurale avec *head tracking* ne sera pas optimale. Une expérience impliquant le corps tout entier ne peut pas se réduire à une écoute au casque, aussi interactive soit-elle.

2. L'immersion sonore ; histoire d'un confort d'écoute

« [La spatialisation] a pour but essentiel d'améliorer la définition des objets par leur répartition dans l'espace. [...] Il ne s'agit pas d'un luxe qui viendrait s'ajouter à l'audition, mais d'une facilité qui lui est offert. »

Schaeffer cité par Guiot (2017)

2.1 La relation historique entre musique et espace

2.1.1 La spatialisation statique

La spatialisation de la musique est utilisée comme pratique musicale depuis le 16^e siècle. Les compositeurs répartissent alors les musiciens dans l'espace physique. Cela donne lieu à de nouvelles approches compositionnelles parmi lesquelles nous pouvons citer les structures polychorales (plusieurs chœurs, séparés physiquement, chantent en alternance), l'antiphonie (alternance de deux chœurs) ou encore les simulations d'échos. T. Tallis (1505 – 1585), compositeur et organiste anglais, expérimente dans ce sens et compose *Spem in allium*, un motet pour 40 voix, dans lequel les chanteurs sont répartis tout autour de l'audience.

Dans le même temps, le style polyphonique Vénitien travaille l'espace d'une manière similaire. Des compositeurs utilisent l'espace et l'acoustique des grands édifices religieux en positionnant, par exemple, des groupes de musiciens sous les différentes coupoles de la Basilique Saint Marc. Plus tard en France, au 19^e siècle, Berlioz (1803 – 1869) positionne quatre groupes de cuivres aux quatre points cardinaux de l'Hôtel des Invalides et entoure ainsi l'audience de sources sonores. (Féron, 2022)

Ces différents courants et expérimentations s'intéressent à la localisation des sources dans l'espace. Ces sources restent immobiles. Nous parlons ici de spatialisation statique.

2.1.2 L'électroacoustique introduit la spatialisation dynamique

Le 20^e siècle assiste à la naissance de la musique électroacoustique « *en 1948 avec la conceptualisation de la musique concrète par Pierre Schaeffer (1910 – 1995).* » (Zanni, 2020) Dès les débuts du courant dans les années 1950, les musiciens en électroacoustiques commencent à utiliser les techniques de panoramique pour disposer

de sources virtuelles entre les haut-parleurs. De nouvelles possibilités s'offrent alors aux compositeurs à travers les trajectoires sonores. Une nouvelle approche, qui cumule des effets de spatialisation à la fois statiques et dynamiques, voit le jour.

La musique concrète expérimente la spatialisation comme partie intégrante de la performance musicale. Le 6 juillet 1951, au théâtre de l'Empire à Paris, Pierre Schaeffer (1910 – 1995) et Pierre Henry (1927 – 2017) présentent leur pièce *Orphée 51 ou Toute la lyre*. Cette date marque l'utilisation en concert d'un dispositif expérimental permettant de contrôler la distribution sonore sur différents haut-parleurs. Ce dispositif est connu sous le nom de *pupitre potentiométrique de relief sonore* et offre au public une expérience immersive.

En Allemagne et à la même période, à la WDR (Westdeutscher Rundfunk – Radiodiffusion ouest-allemande) de Cologne, des développements similaires voient le jour, notamment à partir de 1953, année d'arrivée de Karlheinz Stockhausen (1928 – 2007) au Studio de musique électronique. En 1956, ce dernier présente *Gesang der Junglinge*, pièce en 4 pistes dans laquelle la position et le mouvement des sources sonores fait partie intégrante de la composition, à importance égale que le pitch, le rythme ou la dynamique. Stockhausen, à travers son parcours au Studio de musique électronique de la WDR, a développé un véritable art de la quadriphonie (Féron, 2022).

2.1.3 La relation son – espace

Zanni (2020) parle longuement des recherches qu'ont réalisé les scientifiques et musiciens pendant cette première phase de développement de la musique concrète. Les années soixante représentent « *l'époque [à laquelle] on remet en question les concepts qui rendent le son et l'espace étranger l'un de l'autre et les compositeurs s'accordent sur une relation forte entre les deux.* » Zanni mentionne notamment le compositeur italien Luigi Nono (1924 – 1990) qui place l'espace au centre de sa réflexion artistique.

Zanni (2020) explique que l'écoute ne serait plus la même depuis le 18^e siècle et qu'elle se serait « *effacée derrière des concepts visuels* ». Il cite Nono qui critique notamment l'uniformisation des salles de concert. Selon ce dernier, une œuvre musicale se doit d'être intimement liée à son espace de diffusion. L'écoute est acte même de composition.

Zanni souligne que deux philosophies de l'écoute se confrontent :

- La première est une philosophie traditionnelle qui donnerait à la musique un rôle sémantique. Elle serait une étape transitionnelle de la transmission.
- La seconde place l'écoute comme partie intégrante de l'œuvre et l'espace ne peut en être détaché.

Quand Nono compose *Prometeo* en 1984, il fait appel à l'architecte Renzo Piano pour construire une structure dans l'église San Lorenzo de Venise. Cette structure a pour objectif de modifier l'acoustique du lieu. Elle sert également de support pour le public et permet de positionner les musiciens à différents niveaux et tout autour de l'audience. Pour le compositeur, il ne s'agit pas ici de rajouter de la réverbération mais de créer de nouveaux espaces sonores. Selon Nono, il est primordial de comprendre l'espace qui accueille le sonore. Ce travail de prise de conscience spatial doit selon lui être effectué pour chaque composition et pour chaque concert.

Iannis Xenakis (1922 – 2001), compositeur et architecte grec, travaille lui aussi avec différentes disciplines. Il associe notamment son, architecture et lumière dans des compositions qu'il appelle des polytopes. Un autre exemple représentatif de ce travail pluridisciplinaire est le Pavillon Philips. En 1958, lors de l'exposition universelle à Bruxelles, la société Philips commande à l'architecte français Le Corbusier (1887 – 1965) un bâtiment qu'il conçoit avec Xenakis. Le pavillon accueille une installation sonore de 425 haut-parleurs, qui, couplés avec des effets visuels, diffusent les pièces *Concret P.H.* de Xenakis et *Poème électronique* du compositeur français Edgard Varèse (1883 – 1965).

On voit bien à travers ces exemples l'intérêt que portent les compositeurs à l'espace qui le positionne comme paramètre essentiel du son.

2.1.4 Se donner les moyens techniques des visions créatives

Différentes innovations technologiques ont vu le jour au fil des courants musicaux et des recherches en musique électroacoustique afin d'accompagner la volonté des compositeurs de travailler l'espace. Il fallait développer les outils capables de se mettre au service des intentions musicales. Nous en avons évoqué un premier, utilisé au tout début des années 1950 pour spatialiser la pièce de Schaeffer et Henry *Orphée 51 ou Toute la lyre* : les pupitres potentiométriques. Mentionnons maintenant d'autres outils techniques qui ont été développés par la suite.

Tout d'abord, il est pertinent de présenter l'Acousmonium, créé en 1974 par François Bayle à l'INA-GRM. Ce dispositif est défini par son créateur comme étant un orchestre de haut-parleurs. Positionnés de manière précise dans l'espace et en fonction du lieu de diffusion, les haut-parleurs diffusent une création stéréophonique « éclatée » en différents points sources. Chaque point possède ses paramètres précis en termes de timbre – on parle de couleur – et de puissance, définis par les caractéristiques du haut-parleur qui le constitue. D'après Bayle, il s'agit d' « *organiser l'espace acoustique selon les données de la salle, et l'espace psychologique selon les données de l'œuvre* » (Bayle cité par Guiot, 2017). L'Acousmonium, dans sa philosophie, place l'acoustique du lieu de diffusion au premier plan, puisque c'est de celle-ci que va dépendre le positionnement des haut-parleur et que va découler l'interprétation acousmatique. Nous pouvons faire un rapprochement avec Nono et Xenakis qui, eux aussi, composent avec l'espace comme l'interprète joue de l'Acousmonium.

« L'intérêt de l'Acousmonium est de pouvoir s'adapter à n'importe quel type de lieu susceptible d'accueillir un concert. [...] L'intérêt est ici de taille car il permet de faire « sonner » une œuvre à travers un lieu, à l'aide d'un lieu, avec ce lieu. »

Guiot (2017)

Dans les années 1970 également, des moteurs de spatialisation sont développés parallèlement en France, en Italie et en Allemagne. Zanni (2020) rappelle ces différentes inventions et nous allons suivre son développement.

Luciano Berio (1925 – 2003), compositeur italien et directeur du département électroacoustique de l'Ircam entre 1974 et 1980, souhaite développer à cette époque une technologie permettant de composer un son, en utilisant la technique de la synthèse additive, avec au moins 1000 oscillateurs en temps réel. La technologie à sa disposition ne permettant à l'époque pas d'obtenir un nombre d'oscillateurs aussi élevé, il décide de faire appel à un ingénieur italien pour répondre à sa demande. Un premier modèle voit le jour en 1975, la 4A, capable de générer 256 sons différents en temps réel. Quelques temps après, en 1981, la 4X voit le jour et permet d'atteindre 2000 sons différents en temps réel. Précisons : « *“temps réel” signifie que si une touche est appuyée vous entendez le son immédiatement. [...] [A l'époque, ndlr] De nombreux compositeurs ont préféré composer en “temps différé” [...] Le son n'arrivait, selon la complexité, qu'après quelques minutes ou heures.* » (Giuseppe Di Giugno cité par A. Zanni, 2020)

Berio n'est cependant pas satisfait de ces sons et souhaite les voir entrer en mouvement. Un premier dispositif, appelé le « spatializer » est développé pour accompagner la 4X. Une version plus élaborée porte le nom de « MATRIX 32 » et permet de gérer des trajectoires grâce à des panoramiques d'intensité en Δi . Ce processeur de spatialisation est développé entre 1983 et 1985.

Ce système de spatialisation a notamment été utilisé pour *Répons* (1981, 1982, 1984), œuvre majeure de Pierre Boulez (1925 – 2016). Dans cette œuvre, Boulez place un orchestre de 24 musiciens au centre du plateau, et 6 solistes entourent le public. 6 haut-parleurs accompagnent cet effectif, disposés en hauteur et entre chaque soliste. Ces haut-parleurs jouent des sons captés par des microphones sur les solistes, transformés et spatialisés par la 4X et le MATRIX 32. Boulez utilise dans cette œuvre une forme de composition qu'il emprunte à l'antiphonie¹² en y décuplant l'importance donnée à l'espace.

Pour les premières versions de *Répons* (1981, 1982), Boulez ne disposait pas encore du MATRIX 32. Il utilisait un Hallaphone. Alors que la WDR (*Westdeutscher Rundfunk*) accueille Stockhausen et ses expérimentations, la SDR (*Südwestrundfunk* – Radio-télédiffuseur de l'Allemagne du Sud-ouest) développe elle aussi des technologies de spatialisation. Zanni (2020) nous explique qu'en 1970, « *la représentation de [Planto por las victimas de la violencia de C. Halffter] est considérée comme le premier travail avec un contrôle électronique du son surround en temps réel.* » Afin d'effectuer ce contrôle de spatialisation, Halffter a utilisé un processeur qu'il a développé et qui porte le nom de Hallaphone. Ce dispositif sera également utilisé par Luigi Nono pour sa pièce *Prometeo* (1984) que nous avons évoquée plus haut. Dans *Prometeo*, l'Hallaphone permet un déplacement perpétuel des sons dans l'espace.

Soulignons également une autre pratique utilisée par Nono dans sa pièce. Si les différents processeurs évoqués jusqu'ici permettent d'effectuer des mouvements, le compositeur italien souhaite aussi créer de nouveaux espaces sonores. Comme le suggère l'Acousmonium, il va positionner certains de ses haut-parleurs à des positions très particulières. Il en place 4 en dehors de la structure de Renzo Piano pour créer du champ diffus. Nous retrouverons notamment cette pratique de couplage point de diffusion – acoustique dans le témoignage d'Antoine Petroff en A12.6 p.204.

¹² Procédé d'exécution en alternance, entre deux chœurs, dans les liturgies chrétiennes. Définition Larousse.

Si Nono attache de l'importance à comprendre les salles qu'il occupe pour ses pièces afin d'en apprivoiser les caractéristiques acoustiques, Stockhausen souhaiterait s'en affranchir. Il exprime le besoin de construire des salles intrinsèquement adaptées à la diffusion de la musique électroacoustique et à ses besoins grandissants. En 1970, il propose un premier espace de diffusion qui sera construit à l'occasion de l'exposition universelle d'Osaka. Cet auditorium prend la forme d'une sphère sur laquelle sont répartis 50 haut-parleurs. Il parvient même à faire parvenir des sons du dessous de l'audience en utilisant une plateforme transparente au son pour accueillir les spectateurs. Stockhausen manifeste également sa volonté de contrôler l'acoustique des salles. Son souhait sera exaucé en 1978 lors de l'inauguration de l'Espace de projection à l'Ircam. Surnommé l'ESPRO, cette salle de 24x15,5 mètres est à acoustique variable grâce à ses murs munis de panneaux rotatifs. Ces derniers permettent de choisir entre une surface réfléchissante, absorbante ou diffusante. Le plafond, quant à lui, est à taille variable, pouvant passer de 1,5m à 10,5m de haut. Cela permet de définir un temps de réverbération sur mesure allant de 0,4 à 4 secondes.¹³

2.2 Une catégorisation de l'audio spatial

Les différentes expérimentations que nous avons évoquées précédemment ont accompagné ou donné lieu à des développements technologiques de spatialisation sonore. Nous allons prendre le temps de les définir à travers une catégorisation de l'audio spatial.

2.2.1 Audio orienté canal

L'audio orienté canal, ou *Channel Based Audio* en anglais, peut être considéré comme le format de base de l'audio. Il consiste à envoyer un signal audio, donc un canal, dans une ou plusieurs enceintes sans lui apporter la moindre modification – mis à part de potentielles corrections de niveaux en fonction de l'enceinte pour compenser des sommations acoustiques ou électriques. C'est ce qui est le plus répandu dans le grand public à travers la stéréophonie et le 5.1¹⁴.

¹³ Ircam. (2022) <https://www.ircam.fr/article/connaissez-vous-l'espace-de-projection>

¹⁴ La notation x.y.z correspond à une disposition à x haut-parleurs sur le plan azimutal à hauteur d'oreille, y LFE pour « *Low Frequency Effect* » en anglais et z haut-parleurs en élévation.

2.2.1.1 La stéréophonie

Nous considérons la stéréophonie comme une technologie connue du lecteur et nous avons, dans un souci de praticité, placé le développement de cette partie en Annexe 1 p.101. Nous retenons que la stéréophonie peut être considérée comme le dispositif d'immersion sonore le plus simple grâce à son champ sonore de 60° entre deux enceintes.

2.2.1.2 Les systèmes de diffusions à plus de 2 haut-parleurs

Afin d'outrepasser les limites de la stéréo et de considérer un espace plus large, différents systèmes à plusieurs haut-parleurs voient le jour, le plus connu étant le 5.1.

C'est à 1940 que remonte la première expérience s'apparentant au 5.1 que nous connaissons aujourd'hui avec les projections de *Fantasia* par *Disney*. Un dispositif appelé *Fantasound* imagine une diffusion avec 3 points derrière l'écran (L, C, R)¹⁵ et deux en fond de salle (LS, RS)¹⁶. Ce n'est cependant qu'en 1987 que le cinéma 5.1 est véritablement normé (Périaux *et al.*, 2015).

La première étape vers le 5.1 réside dans l'ajout d'un centre à la diffusion stéréo. Cette enceinte centrale a pour conséquence de rendre moins persistantes les limites de la stéréo. L'image centrale, contrairement à une image fantôme, sera plus stable et bénéficiera d'une meilleure restitution de son timbre. Le *sweet spot* sera également élargi (Périaux, 2015). La seconde évolution apportée au 5.1 est l'ajout des enceintes latérales dites « surround ». Ces deux nouveaux points de diffusion « permettent de reproduire à 360° les réflexions latérales précoces et le champ réverbéré, lesquels se déploient tout autour de l'auditeur donnant ainsi une meilleure lecture de l'architecture de l'événement sonore » (Périaux, 2015). Périaux emploie ensuite les termes d'« enveloppement et d'immersion dans le champ sonore » pour décrire le rendu 5.1.

Malgré tout, la position des enceintes induit également certaines limites. L'écartement des enceintes *surround* avec le LR (80°) et celui entre les deux *surround* (140°) ne permettent pas la formation de sources fantômes stables et correctement localisées. Le *surround* est ainsi principalement utilisé pour restituer des ambiances et de la réverbération en soutien au champ stéréophonique frontal. Le 5.1 ne semble pas

¹⁵ LCR renvoi à des enceintes positionnées à gauche (L pour « left » en anglais), à droite (R pour « right » en anglais) – ces deux enceintes constituant la paire stéréo – et au centre (C pour centre en anglais).

¹⁶ LS et RS renvoie à des enceintes positionnées latéralement à gauche (LS pour « left surround » en anglais) et à droite (RS pour « right surround » en anglais)

vérifier le concept de localisation dans une image sonore à 360° (Rumsey, 2001). Le 7.1 tend à aller plus loin et à proposer une meilleure localisation en ajoutant 2 enceintes à l'arrière. Les enceintes latérales sont placées à 90° et les arrières à 150°, ce qui donne au total 6 angles à 60°. Cela permettrait de reconstituer des images fantômes à 360° avec la même précision que la stéréo classique.

Au début des années 2000, l'envie se fait ressentir d'ajouter de nouvelles enceintes aux systèmes de diffusion pour ajouter de la hauteur au son. De nouveaux dispositifs sont dévoilés au grand public. Nous pouvons citer l'Auro-3D, initialement développé dans les *Galaxy Studios* en Belgique et aujourd'hui maintenu par *NEWA URO BV*, qui se veut être un système de diffusion qui s'adapte à plusieurs tailles de salles et propose des rendus à nombre de canaux variables grâce à son propre processeur de spatialisation. La NHK (*Nippon hōsō kyōkai* en japonais, société publique gérant les stations de radio et de télévision japonaises) a également développé un système 22.2¹⁷ qui se veut encore plus réaliste. Globalement, il est logique d'affirmer que, à l'image de ces systèmes, plus le nombre de haut-parleurs est élevé, plus le rendu immersif sera cohérent et réaliste (Shim *et al.*, 2010).

Si la stéréophonie est un système de diffusion égocentrique avec un *sweet spot* très restreint, les systèmes que nous avons évoqués ci-dessus, qui proposent une nouvelle forme d'immersion que la stéréophonie ne permet pas, ont été principalement imaginés pour un contexte cinématographique. S'il existe bel et bien une zone d'écoute idéale pour ces dispositifs, le rendu spatial est cohérent sur une plus large surface (Shim *et al.*, 2010). Il serait cependant hâtif de qualifier ces systèmes d'allocentrique, c'est-à-dire à destination d'une zone d'audience et non d'un auditeur unique. Une diffusion allocentrique n'opère qu'en couplant l'utilisation d'un système spécifique avec une approche de mixage bien précise. Tout le processus de mixage et les choix qui sont pris doivent être pensés dans l'objectif de pouvoir offrir une écoute optimale pour le plus grand nombre, le dispositif d'écoute seul ne suffit pas.

¹⁷ La notation *x.y* réfère à un système de diffusion comportant x haut-parleurs à hauteur d'oreilles et y caissons de basses. Notons que pour le système japonais *NHK 22.2*, cette notation peut porter à confusion car il comporte des haut-parleurs en élévation bien que sa notation ne le laisse pas deviner.

2.2.2 L'audio orienté scène et la reproduction du champ acoustique

Après le développement de systèmes orientés canal, les ingénieurs ont voulu aller encore plus loin dans la fidélité de restitution. Pour pourvoir l'auditeur d'une sensation la plus réaliste possible, il s'agit de tenter de reproduire au plus proche un véritable champ sonore tel qu'il se construit physiquement dans la nature.

Le format Audio orienté scène de référence est le HOA, pour *Higher Order Ambisonics* en anglais. Bien que cela ne soit pas répandu dans la littérature et que l'EBU ne la considère pas comme telle (Olivieri *et al.*, 2019), nous allons parler de la WFS dans cette partie dédiée à l'audio orienté scène. Nous nous référons ici à l'article de Jullien (2021) qui les associe dans leur objectif de reconstruire la physique du champ acoustique, et il s'avère que ces deux technologies sont souvent comparées pour leur réalisme.

Dans un souci de facilité de lecture, le lecteur trouvera les explications relatives à l'ambisonie et à la WFS en Annexe 2 p.103.

2.2.3 La *Next Generation Audio* : l'Audio Orienté Objet et l'ADM

Dans le rapport technique de l'EBU de 2019, Olivieri rappelle les 3 fonctionnalités qui définissent la NGA – *Next Generation Audio* en anglais :

- proposer à l'auditeur des contenus sonores immersifs (en incluant l'élévation),
- permettre à l'auditeur de personnaliser son contenu,
- introduire l'audio objet pour faciliter ces deux derniers points.

L'Audio Orienté Objet est ainsi représenté comme la solution à implémenter pour proposer aux auditeurs de contenus broadcast une toute nouvelle expérience.

L'audio orienté objet se base sur le concept d'objet sonore. Un objet sonore est une entité constituée d'un fichier son et de métadonnées. Les métadonnées contiennent toutes les informations nécessaires à la bonne restitution du signal audio lors de l'écoute sur le système souhaité.

Le premier intérêt de l'audio orienté objet réside justement dans cette adaptabilité : le paradigme adopté par l'audio orienté objet est de s'adapter au système de diffusion au moment de l'écoute. Un système informatique que l'on nomme « renderer » sera implanté dans la station d'écoute de l'auditeur et sera à même, grâce à la séparation audio – métadonnées, de lire les métadonnées et de spatialiser l'audio en fonction du

contexte d'écoute. Ensuite, le concept d'objet permet d'ouvrir la voie à une nouvelle forme d'interactivité. À partir du moment où le rendu est effectué au moment de la lecture dans le dispositif de restitution de l'auditeur, il est possible de donner à celui-ci certains choix à travers l'activation ou non de certains objets.

Le lecteur trouvera en Annexe 3 p.107 un exemple concret et illustré de l'utilisation de l'audio orienté objet dans un contexte de diffusion d'un match de foot.

Quand les avantages liés à l'objet se sont confirmés, il a été question de créer un standard pour le transport et l'encodage des objets. Cela a donné lieu à la création de l'ADM pour *Audio Definition Model* en anglais.

L'objectif de l'ADM et de proposer un langage commun de métadonnées que tout le monde pourrait lire. Il se base sur l'XML¹⁸, qui est largement utilisé et compréhensible par l'humain. Au regard de l'histoire de l'audio, l'ADM se veut également facilement adaptable à d'éventuelles modifications et évolutions qui pourraient arriver rapidement à la suite de développements technologiques. L'ADM inclut ainsi les informations audio et les métadonnées associées – qui comportent par exemple toutes les informations de localisation – mais ne comprend pas de données relatives à la restitution. Il ne décrit pas le processus de rendu, il sera uniquement en mesure de donner au *renderer* les informations utiles à la manière dont il va effectuer sa diffusion (quelles sont les informations audio à modifier ou les pistes à conserver telles quelles en fonction du système de diffusion) (EBU, 2019).

2.2.4 Le binaural

L'audio binaural est une technologie qui vise à reproduire la sensation d'écoute réelle en utilisant des écouteurs ou un casque. L'idée est de simuler les indices de localisation d'une source en 3 dimensions par l'oreille humaine et lui donner l'illusion d'entendre une scène sonore réelle. Lorsqu'un contenu stéréo classique est écouté au casque, il sonnera comme étant à l'intérieur de la tête. Le contenu binaural, quant à lui, et grâce à l'illusion donnée à l'oreille de percevoir les indices de localisation, permet une externalisation des sources sonores.

Le lecteur trouvera des explications plus approfondies sur les techniques utilisées pour la création de contenus binauraux en Annexe 4 p.110.

¹⁸ XML (pour *Extensible Markup Language*) : Langage informatique utilisant un système de balisage.

2.3 Les solutions technologiques à disposition pour travailler l'espace

Maintenant que nous avons en tête la catégorisation de l'audio, nous allons évoquer certaines des solutions existantes pour la production.

2.3.1 Le travail de l'espace sur logiciel

Depuis les premiers développements de systèmes immersifs dans les années 1970, il est aujourd'hui beaucoup plus facile de travailler l'espace grâce aux nombreux logiciels qui ont fait leur apparition sur le marché. Le lecteur en trouvera une liste non exhaustive en Annexe 5 p.113. Nous prenons cependant quelques lignes pour présenter le *Dolby Atmos*, que nous rencontrerons plusieurs fois dans ce mémoire.

Le *Dolby Atmos* est une technologie propriétaire de reproduction sonore qui intègre *surround* et élévation pour des restitutions sur différents systèmes de diffusion. La société *Dolby* considère le 7.1.2 comme étant le système de diffusion minimal pour profiter au mieux de l'apport de la spatialisation. Le travail de l'espace dans ce contexte passe par le *Dolby Atmos Renderer*. Il est aujourd'hui le premier outil utilisé dans l'industrie musicale et cinématographique pour exporter un fichier ADM. Il permet de manipuler jusqu'à 128 canaux audio (*bed* et objets). Le travail en Dolby Atmos comporte deux éléments à appréhender : le *bed* et les objets.

Le *bed* est un format *Channel Based* qui peut comporter jusqu'à 10 canaux. C'est le cas spécifique du 7.1.2, la configuration de *bed* la plus élaborée. Les 10 premiers canaux de traitement lui sont ainsi réservés. Dans le cas où un *bed* 5.1 est utilisé, par exemple, les canaux numérotés de 7 à 10 seront vides. Les objets commencent ensuite à partir du canal 11. Ces derniers sont constitués d'un canal audio et de métadonnées, qui comportent par exemple des informations sur la position de l'objet. Un objet aura sa propre fonction de transfert au décodage tandis que le *bed* sera restitué à l'aide d'enceintes virtuelles.

2.3.2 Les processeurs de diffusion spatialisée

Au-delà du logiciel, l'immersif se travaille grâce à des processeurs de spatialisation physique, utilisés majoritairement dans des contextes de concert. Différentes technologies sont utilisées en fonction des constructeurs. Nous citons les plus connues, toutes concurrentes car positionnées sur le même marché, en Annexe 7 p.117.

2.4 Le temps de la démocratisation

En 2002, Rumsey remarque que jusqu'alors, le travail de recherche s'est concentré sur la qualité spectrale et la fidélité de la restitution et fait le constat que l'Homme est allé au bout de ses capacités en termes de qualité audio. Le caractère spatial était quant à lui réduit à un niveau de priorité plus faible et avaient encore « *un long chemin à parcourir avant que leur courbe [de croissance, ndlr] devienne asymptotique d'un idéal.* »¹⁹ Rumsey (2002, p.651) Depuis, l'espace a fait l'objet de nombreux travaux et semble avoir atteint une étape charnière dans son processus d'adoption auprès du grand public.

2.4.1 Le binaural à la conquête du streaming

Avant le développement de l'*Objet*, les contenus binauraux se faisaient plus rares et étaient disponibles que sur quelques plateformes peu connues du grand public. Nous pouvons citer *Hyperradio* (ex. *nouvOson*), site internet regroupant les contenus immersifs de Radio France. Le développement actuel s'inscrit dans la continuité de ces sites, portant la vision de certaines maisons ou ingénieurs du son qui croient depuis déjà plusieurs années que l'immersif va remplacer la stéréo (voir Annexe 12 p.139).

Avec la mise en place de l'ADM, différentes plateformes de streaming demandent aujourd'hui des contenus immersifs aux maisons de disques. Le binaural, rendu à partir du mixage Orienté Objet, principalement réalisé sur le *Dolby Atmos Renderer*, est maintenant disponible sur des plateformes de grande écoute à l'international comme *Apple Music* qui ajoute des contenus *Atmos* dans son catalogue dès 2021. Les contenus se font de plus en plus nombreux. Certains classiques du *Back Catalog* de l'histoire de la musique sont remixés en *Atmos*, des nouvelles sorties stéréophoniques sont accompagnées de version immersive, et l'espace est parfois même pensé dès la composition, bien que cela reste encore trop rare (voir 5.3.3).

La technologie binaurale permet la restitution de contenus *channel based* grâce à la virtualisation de systèmes de diffusion, ce qui la rend d'autant plus pratique à mettre en œuvre pour les diffuseurs. S'il est difficile pour le grand public d'avoir accès à des systèmes de diffusion à plusieurs haut-parleurs, notamment d'un point de vue financier, le binaural permet d'en bénéficier à moindre coût grâce à des casques ou oreillettes de

¹⁹ « *Spatial quality and character [...] have some way to go before the curve could be said to be asymptotic to some ideal.* » Notre traduction.

plus en plus bon marché. De plus, les technologies de *head tracking* ou de réduction de bruit disponibles sur certains d'entre eux offre un confort d'écoute amélioré.

Enfin, les technologies d'*upmix*²⁰ participent activement à l'augmentation du catalogue immersif sans demander un temps de mixage supplémentaire. Le plugin *Halo Upmix* de *Nugen*²¹, que l'on retrouvera en 5.1.2, en est un bon exemple.

2.4.2 La multiplication des concerts immersifs

En 2019, l'artiste Molécule lance sa tournée *Acousmatic 360* en partenariat avec Radio France. Pour ses concerts, Molécule repense le concept de musique acousmatique pour l'intégrer dans un contexte club. Il souhaite que son public soit immergé dans le noir et qu'il se concentre sur les sensations que lui procure la musique. La musique doit venir de toutes les directions pour avoir une immersion sonore totale.

Pour cette tournée, l'artiste est accompagné sur scène d'Hervé Déjardin, ingénieur du son de Radio France spécialisé dans le son immersif. Hervé Déjardin spatialise en direct les sources sonores manipulées par Molécule. Le système de spatialisation choisi est une couronne de 12 haut-parleurs, fournie par L-Acoustics. Hervé Déjardin et Molécule sont situés au centre de l'audience, entourés par le public et les enceintes. La spatialisation se fait grâce au système L-Isa que nous avons évoqué précédemment.

Les retours positifs et encourageant du public confortent Radio France dans la direction prise et une série de concerts en découle, qui portent le nom de FIP 360, et utilisent la même configuration spatiale. Plusieurs artistes de renom ont joué dans ces soirées ; parmi lesquels nous pouvons citer Irène Drésel, Camion Bazar, Pablo Bozzi, Para One ou encore Jean-Benoît Dunckel ; organisées dans des lieux originaux comme le Petit Palais ou la Monnaie de Paris.²²

En France également, l'Ircam réouvre son Espace de Projection en 2022 après plusieurs années de travaux, l'occasion pour l'institut de profiter de son système hors normes et d'y inviter son public. Pendant la saison 2023/2024, la série de concert Electro-Odyssée en est le meilleur exemple. Des pointures de la musique électronique

²⁰ Processus de traitement de signal qui consiste à transposer un contenu audio mixé pour un système de diffusion précis afin de l'adapter à un système de diffusion plus conséquent (de stéréo à 5.1 par exemple)

²¹ <https://nugenaudio.com/haloupmix/>

²² Radio France (2019 – 2023) <https://www.radiofrance.fr/fip/podcasts/fip-360>

tels que Murcof ou Richie Hawtin s'y succèdent en offrant au public des expériences 360 inédites.²³

À l'international, la Sphere de Las Vegas est un bon exemple de l'ambition qu'ont certains acteurs du milieu et de l'intérêt porté à l'immersif dans le renouveau de l'expérience du spectacle vivant. Si l'effort apporté sur le visuel est colossal (54 000 m² de LED), le système son n'est pas pour autant lésé. *Sphere Entertainment Group* évoque « 168 000 haut-parleurs utilisés pour créer des environnements sonores réalistes à 360° grâce à une nouvelle technologie connue sous le nom de “wave field synthesis”. »²⁴ La WFS est donc mise à l'honneur dans ce qui est aujourd'hui sans doute la salle de spectacle immersive la plus grande au monde.²⁵

Cette appétence pour l'immersion dans l'expérience live apparait justifiée au regard de la littérature et de ce qui permet une immersion totale d'un point de vue physique. Nous l'avons vue en Partie 1 (p.10), plusieurs sens sont nécessaires à la sensation d'enveloppement et une bonne représentation de l'espace implique le corps tout entier. Le live semble donc être le moyen le meilleur moyen d'obtenir l'immersion la plus réaliste qu'il soit.

2.4.3 Des outils accessibles au Home Studio

*« Les frontières entre musique acousmatique et IDM [Intelligent Dance Music, ndlr] deviennent floues. »*²⁶

Ramsey (2011)

Depuis les années 1990, l'accès à des technologies toujours plus performantes et accessibles financièrement parlant facilitent le processus de création musicale. Il est aujourd'hui possible de s'équiper d'un ordinateur et de quelques machines ou logiciels à faible coût et de laisser libre cours à son imagination. L'ère du « Home Studio » bat son plein. Ce sont maintenant les technologies permettant le travail de l'espace qui sont de plus en plus accessibles. Ce qu'affirme Ramsey en 2011 (cité ci-dessus), Amélie Nilles l'explique également en 2016, en rapprochant musique savante et musique populaire. Les expérimentations autrefois réalisées dans des studios nationaux sont aujourd'hui

²³ Ircam (2024) <https://www.ircam.fr/agenda/electro-odyssee-plastikman/detail>
<https://www.ircam.fr/agenda/electro-odyssee-murcofsimon-geilfus/detail>

²⁴ « 168,000 speakers [...] to create realistic, 360 audio environments leveraging a new sonic technology known as “wave field synthesis”. » Notre traduction.

²⁵ Sphere entertainment Group. (2024) <https://www.thesphere.com/science>

²⁶ « Boundaries between acousmatic music and IDM are becoming blurred. » Notre traduction.

réalisées dans des studios personnels installés dans les chambres de passionnés. Le logiciel *Max* de la société *Cycling '74*²⁷, est le parfait exemple d'outil abordable et extrêmement puissant qui ouvre toutes les portes de l'expérimentation à son utilisateur. Et si l'utilisation de *Max* n'est peut-être pas à portée de tous, les *devices Max4Live*²⁸ qui intègrent le SPAT de l'Ircam permettent à n'importe qui de manipuler l'espace à sa guise grâce à des interfaces utilisateurs simplifiées.

²⁷ Logiciel musical de programmation graphique développé à l'Ircam au cours des années 80 - <https://cycling74.com/>

²⁸ Les *devices Max4Live* sont des plugins paramétrables par l'utilisateur issus de l'implémentation du logiciel *Max* dans la station audionumérique *Ableton Live*.

3. Contexte et caractéristiques du mixage immersif

Les recherches sur la relation espace – son et les technologies qui en ont découlé permettent aujourd’hui un travail accessible et complexe de l’espace sonore. Savons-nous cependant ce qui provoque précisément l’immersion chez l’auditeur ? La définition de l’immersion est-elle universelle et objective ? De quels leviers disposons-nous pour travailler cette immersion ? Existe-t-il des éléments concrets qui permettent de guider la conception d’une image sonore à 360° ?

Nous tenterons dans ce chapitre de répondre à ces questions en nous intéressant aux publications qui ont questionné les différents systèmes immersifs existants et l’impact perceptif qu’ils peuvent avoir sur les auditeurs. Bien que certaines notions relatives à la perception de l’espace soient clairement établies, nous allons voir que la grammaire et le vocabulaire utilisés dans le domaine de l’audio immersif sont encore sujets à réflexions. En reprenant des concepts de psychoacoustique, nous essayerons de voir si des règles objectives peuvent guider la conception d’une image sonore à 360°.

3.1 L’enveloppement, clef de l’immersion

3.1.1 L’impression d’espace chez l’auditeur : ASW et LEV

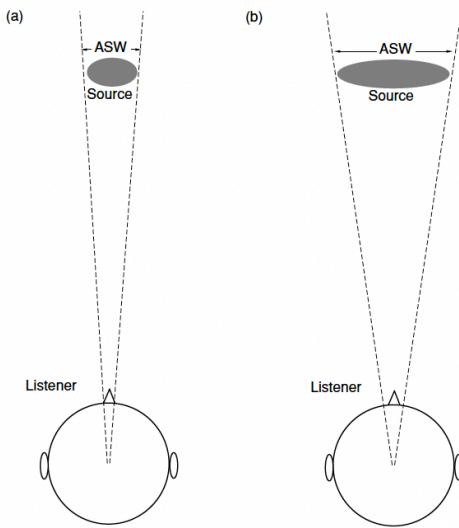
Il est reconnu dans la littérature que deux notions se distinguent comme étant responsable de l’impression d’espace chez l’auditeur en situation d’écoute en acoustique naturelle (Rumsey, 2002) :

- La largeur apparente de la source ou *Apparent Sound Width* (ASW) en anglais
- L’enveloppement de l’auditeur ou *Listener Envelopment* (LEV) en anglais

La largeur apparente de la source peut être visualisée comme la place que prendrait au sol l’élément responsable du son, et qui rayonnerait de toutes parts. L’ASW est intimement liée aux premières réflexions d’une salle (inférieures à 80 ms), qui ont tendance à augmenter la largeur perçue de la source. L’ASW peut également être assimilée à un flou de l’image sonore (Rumsey, 2001). Elle est influencée par le niveau relatif des premières réflexions latérales et leurs angles d’arrivée (Bradley & Soulodre, 1995).

Si l’ASW est relative aux premières réflexions, le LEV, quant à lui, dépend du champ diffus. Il se réfère à l’impression subjective d’être immergé dans un son réverbéré. Il s’avère que c’est plus précisément le champ diffus latéral qui en est à l’origine dans

une salle de concert. Le LEV est influencé par le niveau, la direction d'arrivée et la distribution temporelle des réflexions tardives (Bradley & Soulodre, 1995). Rumsey (2001) ajoute que l'enveloppement est utilisé pour décrire la sensation d'immersion dans un champ sonore réverbéré, avec un son qui semble provenir de toutes les directions. Il assure que cette notion est considérée comme étant gage de qualité d'une bonne salle de concert.



*Figure 1 : Représentation graphique du concept d'ASW
a) ASW petite – b) ASW plus grande (Rumsey, 2001)*

Comme le souligne Nilles (2018), d'après ces définitions, la sensation d'immersion est propre à l'environnement et non à la source. Elle est issue du phénomène de réverbération, qui apparaît comme étant la caractéristique spatiale la plus importante (Sazdov, 2007).

Si ces deux notions (ASW et LEV) font l'unanimité dans la communauté scientifique, Rumsey (2002) pointe du doigt une ambiguïté. Selon lui, l'ASW et le LEV pourraient être le même « attribut spatial » (voir 3.2.2). Il émet l'hypothèse que la distinction entre ASW et LEV ne serait finalement qu'une question d'échelle, et que c'est uniquement dans la taille de la source que résiderait leur différence. Il n'y aurait alors qu'un seul terme caractérisant l'impression d'espace chez l'auditeur.

En se posant la question de la légitimité des deux termes qui définissent la sensation d'espace chez l'auditeur, reconnus par la communauté, Rumsey souligne la fragilité des termes employés pour décrire l'immersion. La question sémantique apparaît pour la première fois comme étant importante.

3.1.2 L'enveloppement, une notion difficile à définir

Rumsey remarque que les termes liés à l'enveloppement et la sensation d'espace sont sujets à des interprétations très variées dans la littérature et qu'ils sont difficiles à concevoir. (Rumsey, 2002, p. 661) L'ambiguïté dans la compréhension du terme d'enveloppement réside dans sa mise en contexte. Si l'ASW et le LEV sont des termes éprouvés dans un contexte d'acoustique naturelle, qu'en est-il dans un contexte de diffusion spatialisé dans lequel un certain nombre d'enceintes entourent le spectateur, disposition de plus en plus fréquente et permise grâce aux avancées technologiques que nous avons évoquées précédemment ? Selon Rumsey (2002) nous avons des conclusions à tirer de ce que nous connaissons de l'impression d'espace en acoustique naturelle, mais que ces connaissances ne suffisent plus dans ces contextes où ce n'est plus uniquement du son réverbéré qui « enveloppe » l'auditeur mais également des sources directes.

S. Adair *et al.* (2008) ont réalisé une étude sur la perception de l'enveloppement en musique électroacoustique et se sont particulièrement intéressés à la validité de la notion d'enveloppement en dehors du contexte d'acoustique naturelle. Ils ont réalisé des tests sur des sujets en leur donnant la possibilité de modifier leur sensation d'enveloppement grâce à un contrôle sur les enceintes du système qui comportait, en plus du LCR, des enceintes *wide* (stéréo élargie), *surround*, et arrières. Les auteurs ont remarqué que les sujets s'intéressaient en premier lieu aux enceintes arrières et *surround*. L'importance des *wide* dans la sensation d'enveloppement est moindre. Ce résultat va à l'encontre de la physique du LEV traditionnel. Dans une salle, il y aurait moins d'énergie sonore à l'arrière et sur les côtés. C'est pourtant là que les sujets ont mis le plus d'énergie pour favoriser leur sensation d'enveloppement.

Les résultats des tests de Rumsey (2002) pointent la même conclusion. Ils montrent que les sujets utilisent le terme d'enveloppement quand ils sont entourés de sources directes, ce qui ne correspond pas non plus au LEV dans sa signification traditionnelle. Selon lui, un nouveau terme est nécessaire pour définir ce nouveau type d'enveloppement.

3.2 Des définitions du son immersif

Les réflexions précédentes nous invitent à nous poser la question : comment parler de son immersif de manière à se faire comprendre par tous ses acteurs ? Quelle grammaire doit être utilisée pour que les termes employés par un ne soit pas redéfinis par un autre ? Comment décrire son ressenti de manière universelle ?

Grâce à l'état de l'art que nous avons commencé à présenter jusque-là et à d'autres résultats de recherche que nous allons partager, nous nous attardons à trouver notre propre définition du son immersif associée au contexte de ce mémoire. Nous ne nous intéressons pas spécifiquement à un contexte d'acoustique naturelle de salle de concert, mais plutôt à un contexte plus large et actuel de diffusion sonore sur système à multiple haut-parleurs entourant l'auditeur, où à une écoute au casque. Notre définition permettra d'avoir un axe d'analyse précis lors de l'analyse d'entretiens effectués auprès d'acteurs du son immersif.

3.2.1 Les éléments de définition de l'immersion et de l'audio spatialisé

Nous avons relevé quelques paramètres qui apparaissent de manière récurrente dans les définitions de l'immersion et de l'audio spatialisé trouvées dans la littérature. Les définitions sur lesquelles nous nous sommes référés sont accessibles en Annexe 6 p.115.

1. La technologie

Dans un premier temps, soulignons que l'immersion, dans ces définitions, passe toujours par la technologie, le système de diffusion, les algorithmes de traitement du signal. Slater (2003) explicite d'ailleurs le fait que ce que délivre la technologie doit pouvoir être décrit de manière objective.

2. L'illusion

La technologie sert d'outil pour créer l'illusion sonore, pour fabriquer du virtuel.

3. Toutes les directions

L'espace sonore, créé artificiellement, permet de disposer de sources dans toutes les directions possibles. L'illusion se fait à travers le positionnement des sources, mais également grâce à l'acoustique, à l'environnement. Ces notions, comme on peut le lire chez Ramsey ou Jullien, sont étroitement liées.

4. Les principes de l'écoute binaurale

Les sources sont localisées grâce aux principes de l'écoute binaurale – voir Partie 1 p.10 – que la technologie sait imiter.

D'autres paramètres moins récurrents interviennent également dans ces définitions, cités de manière ponctuelle. Biocca et Delaney (1995) définissent l'immersion comme capable de submerger le système perceptif de l'auditeur. Cette idée pourrait se rapprocher du paramètre n°4 explicité ci-dessus, mais « système perceptif » peut également faire référence aux autres sens que le système auditif. Deacon (2019), lui, emploie le terme « complexe » pour définir les scènes acoustiques qu'un système spatial est capable de créer. Nous pouvons rapprocher cette idée du paramètre numéro 3, une scène complexe fait cependant appel à des caractéristiques autres que la position seule, parmi lesquels nous pouvons citer par exemple la profondeur, la réverbération, la dynamique. Deacon valorise également le paramètre temporel en estimant que les compositeurs travaillants l'audio spatialisé « doivent [...] manipuler les sources au cours du temps ».

3.2.2 Les différents contextes de l'écoute immersive

L'immersion sonore est particulièrement dépendante des différents contextes d'écoute. Rappelons-les ici.

Nous avons déjà introduit auparavant les notions de systèmes égocentriques et allocentriques – voir 2.2.1.2. Un système d'écoute égocentrique est dimensionné pour un unique auditeur. On parle alors de *sweet spot*, particulièrement étroit. Une stéréophonie est égocentrique. Le système allocentrique, quant à lui, est dimensionné pour couvrir une audience, et se doit d'être couplé à une technique de mix adapté. La zone d'écoute devient plus large, comme ce que l'on retrouve en salle de cinéma.

L'autre paramètre essentiel à la contextualisation de l'écoute réside dans l'orientation de la scène sonore. Une écoute orientée correspond à une situation dans laquelle l'auditeur est invité à se placer dans une direction précise. La création sonore est, dans ce cas, pensée et travaillée en ayant pleinement conscience de l'orientation de l'auditeur. Une musique stéréophonique est orientée, tout comme un film. Une écoute non orientée se présente dans un contexte où l'auditeur est complètement libre de ses mouvements, et où il n'y a pas de raisons de se placer dans une direction précise. Un contenu en réalité virtuelle ou augmentée, ou une installation sonore dans un musée, peuvent être des situations d'écoute non orientée.

3.2.3 Nos définitions

La tentative d'établissement d'une grammaire du mixage immersif ne s'intéresse pas spécifiquement à un contexte particulier. L'objectif est d'établir une grammaire commune à toutes les personnes travaillant l'immersion, et dans les domaines : cinéma, radio et musique. Nous nous intéressons aussi bien à l'écoute sur enceintes comme au casque, à une écoute allocentrique comme égocentrique, à un système 7.1.4 comme à un 5.1.

Notre définition de l'immersion, qui sera questionnée lors de notre Discussion, reprend les termes que nous avons exposés en 3.2.1. C'est une définition qui se veut technique, étant donné que nous n'avons pas eu le temps, dans le cadre de ce mémoire, de nous pencher sur les aspects de psychologie cognitive, qui nous aurait permis de développer plus en détail ce que l'être humain peut considérer comme étant immersif.

Immersion :

Possibilité donnée à un système de diffusion de faire entendre à un auditeur des sources provenant de toutes les directions, en simulant les principes de perception binaurales, afin de créer l'illusion de scènes sonores complexes.

Nous ne reprenons pas l'idée de maîtrise en direct du facteur temporel ici. Une scène sonore immersive peut parfaitement être écrite à l'avance. De plus, l'idée de « scènes sonores complexes » induit la possibilité que les sources soient en mouvement ou que leurs positions soient modifiées au cours du temps.

Nous considérons qu'un système immersif sera d'autant plus pertinent s'il implique plusieurs sens. Le *live* semble donc particulièrement adapté. Nous ne le précisons cependant pas dans notre définition car ce n'est pas la situation d'écoute la plus fréquente aujourd'hui. Et, bien que ce soit le binaural qui soit aujourd'hui le plus répandu, nous considérons également qu'un binaural non *tracké* ne présente pas une forme d'immersion optimale en raison de ses différents défauts.

L'écriture immersive peut être définie à son tour.

Écriture sonore immersive :

Processus narratifs et techniques mis en œuvre dans la création d'un contenu sonore afin de le rendre immersif.

La grammaire est définie comme étant « l'ensemble des règles d'un art »²⁹. De là, nous pouvons déduire notre dernière définition.

Grammaire de l'écriture sonore immersive :
Ensemble des règles de l'écriture sonore immersive.

3.3 Conception d'une image sonore à 360°, des règles à suivre ?

Au regard de notre définition de l'immersion et des informations partagées en Partie 1 p.10 et Partie 2 p.18, nous allons pouvoir avoir un regard critique sur certains résultats publiés dans la littérature concernant le mixage spatialisé.

3.3.1 Du mixage stéréophonique au mixage immersif : du complexe au naturel

Si le mixage stéréophonique est une pratique qui requiert un apprentissage complexe, celle du son spatialisé semble passer outre certaines difficultés et limites.

Le mixage immersif tend à reproduire les sensations d'écoute réelle tout en prenant des initiatives esthétiques qui peuvent totalement s'opposer à une situation naturelle. Toujours est-il que l'espace à appréhender est intégral et n'est pas réduit à un espace limité, comme l'est la stéréophonie dans la zone frontale de 60°. De ce fait, certaines difficultés du mixage stéréophonique sont amenées à disparaître, notamment les problématiques de masquage. Il est souvent affirmé dans la communauté que le mixage immersif est plus facile, ce que nous évoquerons en Partie 4 p.48 et Partie 5 p.61.

Une étude de Kim *et al.* (2021) arrive à une autre conclusion intéressante en faveur du son spatialisé. Il s'avérerait que, dans le cas d'un contenu audio de musique orchestrale, la qualité d'une diffusion immersive ne soit pas dépendante du lieu d'écoute. En d'autres termes, quel que soit le lieu de diffusion du contenu sonore immersif et son acoustique, la perception par l'auditeur sera toujours qualitative. L'environnement d'écoute n'a que peu d'impact sur le LEV.

Si l'étude de Kim *et al.* ne s'intéresse qu'à la reproduction de la musique orchestrale captée dans un contexte traditionnel d'immersion au sens de l'ASW et du LEV, il n'est pas aberrant d'émettre l'hypothèse que ce résultat reste valable pour d'autres contextes immersifs. Si, dans un contexte de scène orchestrale frontale où les *surround*, les arrières

²⁹ Définition Larousse.

et les *tops* soutiennent le son direct par un champ réverbéré, nous pouvons imaginer que, dans un autre contexte où tous les haut-parleurs peuvent être amenés à diffuser des sources directes, l'effet produit sur l'auditeur amènerait aux mêmes conclusions.

La conclusion de la recherche de King *et al.* (2019) sur le niveau sonore nécessaire pour les canaux en élévation dans la reproduction immersive prend une direction similaire. Ils trouvent une plus grande plage d'acceptabilité en niveaux que dans des recherches équivalentes sur des niveaux de réverbération en stéréophonie. Ils se posent alors la question : « *Est-ce que le son immersif possède une plus grande plage d'acceptabilité en termes de balances que le son stéréophonique ?* »³⁰ (King *et al.*, 2019)

Si l'immersif ne semble pas perdre de sa qualité en fonction du lieu d'écoute, il ne semble pas non plus être aussi exigeant en termes d'équilibres que la stéréophonie. Nous pouvons émettre l'hypothèse que cela est dû au fait que les sources, démasquées, sont plus facilement identifiables, même à niveau variable.

3.3.2 Utiliser les critères de perception

Les connaissances que nous avons sur la perception binaurale (voir Chapitre 1) nous indiquent certaines directions à prendre concernant la manipulation de sources sonores en contexte de mixage spatialisé.

Nous savons par exemple que les transitoires nous permettent de localiser plus facilement la source sonore qu'un contenu stationnaire. (Rumsey, 2001, p.31) Dans un contexte d'écriture spatiale où l'on recherche une précision de localisation, certains objets sonores seraient ainsi plus faciles à localiser que d'autres (Moélants, 2023) :

- Comme nous venons de le voir, plus il y a des transitoires ou plus l'attaque est nette, mieux on localise le son.
- Plus un son est large bande, plus il est facile à localiser comparé à un son pur comme une sinusoïde,
- Plus un son est familier, mieux on le localise (l'avion sera localisé en hauteur)

En ce qui concerne le deuxième point sur la composante spectrale du son et de sa localisation (large bande ou pureté), rappelons-nous des résultats de Roffler et Butler (1968), présentés en 1.1.2, qui affirment que les sons purs ont tendance à toujours être localisés à la même hauteur pour une fréquence données, tandis que les sons complexes

³⁰ « *Does immersive audio have a wider range of acceptable balance choices than stereo reproduction?* » Notre traduction.

sont localisés à leur véritable hauteur. Cela est à relativiser par notre troisième point. Bien que l'avion produise un son complexe, il sera toujours localisé en hauteur, même si le stimulus sonore réel de celui-ci vient d'une autre direction. (Moélants, 2023)

Évoquons maintenant le lien entre espace et source sonore et la notion de CSI pour *Continuous Spatial Impression*, définie par David Griesinger. Ce terme fait référence à des sources sonores continues qui ne peuvent pas être séparées en événements ponctuels. Griesinger affirme que l'interaction de ce son avec son environnement et la réverbération associée donne lieu à un sentiment d'enveloppement et d'espace directement lié à la source elle-même. (Rumsey, 2001, p.44)

D'après ces résultats, nous pouvons commencer à émettre des hypothèses de mixage, et associer des gestes de mixage à des intentions en fonction de la matière sonore travaillée.

- Le *pad*³¹, étant défini par sa continuité, possède toutes les caractéristiques de la source sonore à l'origine de la CSI. Dans une volonté de provoquer une sensation d'enveloppement chez l'auditeur, un mixeur pourrait donc avoir recours à un *pad*. Sans transitoires, le *pad* serait donc moins facilement localisable. Il serait pertinent pour le mixeur de jouer avec ce constat, et de le rendre dans plusieurs haut-parleurs, ce qui appuierait la sensation d'enveloppement. Un rendu dans plusieurs haut-parleurs soulève cependant la question du point d'écoute, qui serait fragilisé si les signaux envoyés dans les différents haut-parleurs ne sont pas décorrélés.
- Les sons percussifs, constitués d'une attaque franche et des transitoires marqués, sont ainsi facilement localisables. Ils peuvent permettre à un mixeur d'attirer l'attention de l'auditeur dans une direction très précise. Cela peut notamment permettre d'affirmer fermement une scène frontale dans le cas d'une écoute orientée, ou au contraire, d'inviter à écouter dans d'autres directions de manière explicite.

³¹ Le *pad* est un élément sonore d'un arrangement musical caractérisé par sa continuité. Il est souvent réalisé tenant un accord de synthétiseur tenu sur toute une mesure.

3.3.3 L'importance du mouvement de tête

Le mouvement de tête apparaît comme fondamental dans notre perception de l'espace et dans la sensation d'immersion. Il est d'ailleurs mis en avant dans beaucoup de tests sur l'évaluation de l'immersion d'un système de diffusion.

- Sazdov *et al.* (2007) tentent de définir un terme lié au sentiment d'enveloppement dans le cas d'un système avec élévation. Ils prennent le temps de sensibiliser leurs sujets aux mouvements de tête, en leur expliquant clairement que bouger la tête leur sera utile dans leur évaluation des systèmes étudiés.
- Adair *et al.* (2008) ont travaillé sur la perception de l'enveloppement dans un contexte électro-acoustiques. Ils ont souligné explicitement dans leur analyse le fait que les sujets tournaient la tête de manière récurrente lors de leurs écoutes.
- Kamekawa et Marui (2020) se posent la question de la largeur fréquentielle nécessaire à une diffusion en élévation correcte sensiblement parlant. Ils ont pour cela effectué des mesures à l'aide d'une tête artificielle – donc fixe – de contenus avec et sans élévation. Les résultats ne présentent quasiment pas de différences, ce qui impliquerait, par la mesure, que seuls les mouvements de tête permettraient de détecter l'élévation.

Ces résultats ne vont pas en faveur d'une écoute binaurale non trackée. Nous avons déjà vu précédemment qu'une telle écoute ne permet que rarement une image frontale, que les différences entre HRTF sont telles qu'un même contenu ne sera pas perçu de la même manière entre individus, ou encore qu'il est difficile de différencier l'avant de l'arrière. Le *tracking* permettrait de résoudre ces limitations en ajoutant à la perception un élément de différenciation supplémentaire associé à la perception sonore naturelle. L'efficacité de la HRTF dynamique est démontrée depuis déjà une trentaine d'années grâce notamment à l'étude d'Inanaga *et al.* (1995).

Ces informations sont évidemment à prendre en compte par les ingénieurs du son qui travaillent en immersif et dont le contenu pourra être écouté au casque. Deux questions principales se posent dans ce contexte. Est-ce que le binaural sera *tracké* ou non ? Est-ce que le même mixage sera à la fois écouté au casque par certains et sur enceintes par d'autres ? À ces questions découlent des décisions de mixage décisives. Par exemple, pour un contenu destiné à être écouté en binaural non *tracké*, l'utilisation des mouvements, même de faibles amplitudes, apparaît nécessaire afin d'engendrer des

modifications d'indices qui ne peuvent pas se faire grâce au mouvement de tête. Des compromis devront également être assumés afin d'assurer une compatibilité des systèmes d'écoute.

D'une manière générale, l'intérêt du binaural statique semble se questionner par ses limitations. Il s'avère que la direction prise par les acteurs majeurs de l'industrie audio semble tendre vers une généralisation du *Head Tracking* qui promet un rendu optimal et permettrait une plus large portabilité des mixages.

3.3.4 L'expérience en audio 3D décisive dans la perception

Nous avons remarqué dans la littérature que l'expérience en audio, et plus particulièrement en audio immersif, joue un rôle déterminant dans la perception de la scène sonore immersive. Ce rôle se manifeste dans certains résultats de tests de recherches sur lesquelles nous nous sommes penchés. Souvent, les résultats des experts sont analysés séparément et leurs contenus diffèrent largement des résultats de sujets profanes (Guastavino & Katz, 2004 ; Kim & Howie, 2021 ; Kamekawa & Marui, 2020).

Howie *et al.* (2019) se sont précisément penchés sur la question de l'expérience dans l'évaluation d'un système 3D. Leurs résultats montrent qu'une pratique de production audio et une pratique instrumentale jouent un rôle dans la performance des sujets à analyser un système de diffusion 3D. En l'occurrence, leurs résultats présenteront une variance d'erreur plus faible que des sujets non expérimentés.

Rumsey (2001) se réfère à Komiyama pour expliciter cette idée de différence de perception relative à l'expérience, ici dans le contexte audiovisuel. L'expérimentation de Komiyama cherchait à établir la limite à partir de laquelle la différence entre la position d'un son et la position de sa source affichée à l'écran devient dérangeante. Il a montré que pour des oreilles expertes, une différence de 11° entre l'image et le son est problématique, tandis que pour des oreilles novices, cette différence s'élève à 20°.

Ces différences s'expliquent. Nilles (2018) souligne qu'il est possible d'entrainer l'oreille à percevoir l'espace. « *Nous pouvons rendre plus sensible notre perception auditive de l'espace à force d'écoute multiples, comme un entraînement.* » (Nilles, 2018). Pour cette affirmation, elle se réfère notamment à Rusmey (2001), qui parle de l'adaptation aux HRTF. Si on attribue une HRTF aléatoire à un sujet, il pourra, après un temps d'adaptation, retrouver les mêmes reflexes de perceptions que s'il utilisait sa propre HRTF. Grothe *et al.* (2010) est également cité à travers une expérience réalisée

de pose de prothèse dans des oreilles de patient afin d'altérer la forme du pavillon. Cela a pour but de perturber la localisation dans le plan vertical. Il s'avère que les erreurs de localisation sont corrigées au fil du temps.

Guastavino et Katz (2004) analysent la nature des attributs spatiaux qui apparaissent importants pour l'oreille entraînée et ceux qui sont déterminant pour l'oreille novice. Il semblerait que la localisation et la précision des sources sont importantes pour les oreilles expertes, tandis que ce sont la présence et la distribution spatiale³² qui jouent le plus grand rôle sur la perception des oreilles novices. Kim et Howie (2021) nous présentent d'autres résultats. Selon eux, pour les experts, c'est le LEV qui est considéré comme le premier facteur de différenciation entre différents environnements immersifs, tandis que pour les novices, il s'agit de la fidélité spectrale³³. Ces deux résultats n'apparaissent pas forcément comme équivalents, et peuvent même être considérés comme contradictoires si on associe la notion de LEV à celle de distribution spatiale, qui n'est pas clairement définie dans l'article de Guastavino et Katz.

Pour les ingénieurs du son, ces résultats peuvent exercer une influence sur le mixage en fonction du public visé. La question se pose sur la manière dont il faut les appréhender, d'autant plus face à des résultats différents.

3.3.5 L'importance de l'élévation discutée

Dans les contextes qui nous intéressent en termes d'immersion sonore, les systèmes de diffusion comprennent de l'élévation car ils doivent permettre de diffuser du son provenant de toutes les directions. Dans la littérature, différentes expériences ont été réalisées afin de questionner le rôle des haut-parleurs en élévation. Des résultats obtenus, on peut se poser la question des conséquences sur les pratiques de mixage.

Tout d'abord, l'utilité même de l'élévation est remise en question par Guastavino et Katz (2004). D'après leur étude, l'élévation n'apparaît pas comme indispensable. Elle rendrait même le système de diffusion moins efficace. Les sujets ayant réalisés des tests

³² Selon Guastavino et Katz (2004) :

Localisation : Localisation des sources/précision de l'image (Choix : précis > indistinct)

Présence : Sensation d' « être là », sensation d'être (Choix : à l'intérieur > à l'extérieur)

Lisibilité : Définition spatiale, lisibilité de la scène (Choix : bien définie > mal définie)

³³ Selon Kim et Howie (2021) :

Fidélité spectrale (ou balance) : « The balance in tone color (spectrum) variation of the sound image.”

LEV : “The sense of being enveloped by the sound field (both horizontally and vertically)”

comparatifs entre un système de diffusion en 2 dimensions (*surround*) et un système en 3 dimensions (avec élévation) présentent le système *surround* comme étant plus naturel. Guastavino et Katz (2004) questionnent leurs résultats. Ils estiment que si les sujets n'ont pas trouvé un intérêt particulier dans le système 3D, c'est peut-être car ils ne sont pas habitués à écouter un rendu de ce type. Le contenu *surround*, plus familier, est ainsi perçu comme plus naturel. Cela entre en résonance avec la partie précédente sur l'apprentissage de l'oreille face à un contenu immersif.

Des contradictions apparaissent sur le rôle de l'élévation. Hamasaki *et al.* (2006) assurent, eux, que le 22.2 améliore la présence, le réalisme de la reproduction sonore et agrandit la zone d'écoute. Kim *et al.* (2010) affirment d'ailleurs eux aussi qu'un système comportant de l'élévation permet d'obtenir une zone d'écoute agrandie. Sazdov *et al.* (2007) défendent eux aussi une meilleure qualité de reproduction pour les systèmes comportant de l'élévation. Selon eux, la 3D permet une sensation qui n'existe pas en 2D et qu'ils nomment « *engulfment* », ce qui signifierait être couvert de sons plutôt qu'être simplement enveloppé.

La question se pose également du contenu à diffuser dans les haut-parleurs en élévation. Kamekawa et Marui (2020) nous donnent un premier élément de réponse en affirmant que les informations en dessous de 400 Hz sont inutiles en élévation. Ils séparent le spectre sonore en 2 bandes aux caractéristiques précises :

- De 400Hz à 4kHz : nécessaires pour une sensation d'enveloppement mais pas utile à la localisation.
- De 4kHz à 16kHz : nécessaire à la localisation dans le plan vertical médian.

Cette deuxième bande de fréquence est cohérente avec ce que nous avons en 1.1.2 lorsque nous avons affirmé que les fréquences au-dessus de 3kHz étaient les plus utiles pour une localisation verticale (Nilles, 2018).

Au-delà du contenu fréquentiel, King *et al.* (2019) se posent la question du niveau sonore nécessaire aux enceintes en élévation. Ils se placent dans un contexte précis où les *tops* sont utilisés pour diffuser de l'ambiance ou de la réverbération. Ils donnent ensuite la main à des sujets afin de régler le niveau de diffusion. Il s'avère que l'élévation est toujours mixée plus faible, de 7 à 16 dB moins fort que le LCR. King *et al.* ajoutent qu'il est difficile de savoir s'il y a réellement de l'élévation si les *tops* ne sont utilisés que pour de la réverbération ou de l'ambiance. Et dans le cas où des sources sont localisées en hauteur, il est alors possible que l'image sonore globale soit translatée vers le haut.

La qualité du rendu dépendrait aussi du nombre de haut-parleurs, même si les avis divergent. Kim et Howie (2021) citent Oode et affirment que, quand il y a de l'élévation, la sensation d'enveloppement augmente quand le nombre de *tops* augmente. Kim *et al.* (2010) l'avaient déjà affirmé 10 ans plus tôt. Ces derniers montrent qu'on obtient tout de même de très bons résultats avec seulement 3 *tops* alors que le système 22.2 en contient 9.

Ces différents résultats montrent que l'utilisation des *tops* soulève encore quelques questions. Si leur utilité dans la précision de la scène sonore semble vérifiée pour certains, leur présence même ne semble pas indispensable pour d'autres. Seuls des placements de sources sonores directes dans les *tops* auraient un sens, encore faut-il que cela soit esthétiquement justifié.

3.3.6 La localisation et le rôle du mouvement

Dans la pratique créative de contenus immersifs, l'utilisation du mouvement est fréquente. Cependant, l'importance du mouvement n'est pas perçue de la même manière dans l'histoire. Si, selon Nono, on doit toujours donner l'impression d'être en mouvement (Zanni, 2020), Payre affirme que l'enjeu de la musique spatialisé réside dans l'immersion et non dans le mouvement. (Nilles, 2018) C'est un outil qui reste utilisé dans l'écriture immersive. On peut le définir à l'aide de 3 paramètres : sa direction, sa vitesse et sa prédictibilité. (Moélants, 2023)

Nilles (2018) insiste sur la notion de vitesse et cite MacDonald : « *La vitesse du mouvement est importante à prendre en compte.* » Kim *et al.* (2010) expliquent par exemple que, dans un contexte de cinéma, on ne peut pas distinguer le mouvement quand il est trop rapide. On ne distingue que le point de départ et le point d'arrivée. Nilles ajoute que le mouvement est à utiliser avec modération et qu'il ne faut pas en abuser. Elle se réfère d'ailleurs plus tard à Vande Gorde, qui a expliqué que l'abondance d'espaces sonores, qu'elle nomme « *tutti spatial* », peut annihiler la sensation d'espace.

La temporalité du mouvement est également un élément à considérer. Une scène sonore mise en mouvement perpétuel pendant un temps long peut fatiguer l'auditeur qui aura tendance à avoir une perception de la durée du mouvement comme étant plus longue que sa véritable temporalité. Si les mouvements sont à l'inverse amenés progressivement, l'auditeur aura la sensation d'une séquence plus courte que sa véritable durée. (Guiot, 2017)

Plus largement, le mouvement est un jeu de localisation. Il s'agit de faire varier la position d'un objet au cours du temps, et la précision de localisation ne semble pas avoir de rapport avec la qualité d'un rendu immersif. Rumsey (2002) se réfère à deux études différentes qui affirment qu'il y a peu de corrélations entre précision de localisation et préférence de l'auditeur pour un système en particulier. On a d'ailleurs vu précédemment avec Guastavino et Katz (2004) que, si les oreilles expertes s'intéressent particulièrement à la précision de localisation, ce n'est pas le cas du grand public. Jullien (2021) va dans ce sens, dans un contexte musical :

« Jean Marc Duchêne affirme que le simple contrôle du site, c'est-à-dire la localisation des objets sonores, ne présente qu'un intérêt musical limité. C'est la prise en compte des autres attributs de la masse spatiale, dont l'aire et la densité, qui permet de construire des objets sonores intéressants du point de vue musical. »

Jullien, 2021

Appréhender la localisation et le mouvement dans un contexte de mixage immersif soulève donc certaines questions. D'après ce que nous venons d'évoquer, il n'apparaît pas pertinent de s'attarder sur la création de mouvement. Ce dernier n'est pas primordial dans la sensation d'enveloppement, il peut saturer l'espace, et la localisation de la source en déplacement ne sera pas efficacement perçue. Il apparaît même futile de s'attarder à une localisation précise des sources en général, étant donné que le grand public n'y accordera pas d'importance particulière.

4. Retours d'expérience ; les pratiques immersives actuelles des ingénieurs du son

« Est-ce que finalement tu connais un son qui ne soit pas spatial ? Bien sûr que non. »³⁴

Sharma, dans le cadre de notre entretien.

4.1 La pratique empirique de chacun, un accès aux connaissances

4.1.1 Le recours aux entretiens individuels

Nous avons développé jusqu'ici différents sujets autour de l'audio spatialisé afin d'appréhender de manière éclairée ce qui nous intéresse tout particulièrement : sa grammaire. Afin d'en établir les bases, nous souhaitons confronter les notions évoquées en partie 1, 2 et 3 avec les véritables pratiques des ingénieurs du son et des artistes qui travaillent l'audio spatialisé. Au regard de la définition de la grammaire que nous avons donnée en 3.2.3, dans le but de connaître tous les mécanismes mis en œuvre dans le processus de l'écriture d'une œuvre en audio spatialisé, nous avons choisi de nous entretenir individuellement avec des acteurs du son immersif, qu'ils soient ingénieurs du son ou artistes. Pour ce travail, nous avons échangé avec dix sujets. L'objectif de chaque discussion est le suivant : obtenir une visualisation la plus claire possible de la manière dont la personne conçoit une image sonore en 3 dimensions.

La pratique de l'écriture sonore, immersive ou non, passe souvent par des actions inconscientes, complètement acquises par le praticien et difficilement exprimables. Si l'on souhaite découvrir ces processus inconscients, il est nécessaire de s'entretenir longuement avec la personne concernée, et sur plusieurs entretiens, afin de libérer la parole et d'entrer petit à petit dans les détails des mécanismes mis en œuvre afin d'obtenir un résultat souhaité. Dans notre cas, nous n'avons pas suffisamment de temps pour nous entretenir plusieurs fois avec chaque personne, nous avons cependant la chance d'avoir 10 entretiens à analyser, et un guide d'entretien qui nous permet d'orienter la discussion.

Si la grammaire de l'écriture stéréophonique existe, celle-ci découle de plusieurs dizaines d'années de pratique, d'écrits et de consensus. En ce sens, il nous sera difficile d'atteindre, dans ce travail, un niveau de précision et d'exhaustivité similaire. Nous

³⁴ « Do you know actually a sound that is not spatial? Of course not. » Notre traduction.

pouvons également admettre que nos sujets, aussi professionnels soient-ils, ne disposent d'une connaissance que partielle et que celle-ci peut parfois être altérée par une longue expérience stéréophonique. L'objectif reste tout de même d'avoir le plus d'éléments possibles capables de nourrir notre grammaire. Dans l'objectif d'une exhaustivité maximale, l'entretien doit rester largement ouvert afin que les personnes interviewées évoquent ce qui semble être le plus important pour elles. Nous creuserons ensuite en conséquence les sujets qu'elles aborderont grâce à des relances spontanées (voir 4.2.2). Au regard de notre développement précédent, nous considérons que certaines notions sont particulièrement pertinentes à analyser et nous avons mis l'accent sur les problématiques suivantes (voir guide d'entretien en Annexe 10 p.123) :

- Quelles sont les difficultés rencontrées lors du mixage immersif ?
- Les critères de perception sont-ils utilisés comme base de travail ?
- L'acculturation des artistes au son spatialisé a-t-elle un impact sur le processus créatif ?
- Quelle est l'importance du mouvement dans l'écriture ? La précision de localisation est-elle une préoccupation ?
- Comment sont utilisés les *tops* et les arrières dans la narration ? Un système 2D est-il plus immersif qu'un système 3D ? Comment la frontalité est-elle gérée ?
- Quel est le rapport à la stéréophonie dans l'écriture immersive ?

Avec le développement de la technologie *Dolby Atmos* et la volonté d'*Apple Music* de mettre en valeur des contenus immersifs sur sa plateforme de streaming, il semble que deux manières de travailler se confrontent dans la manière d'approcher l'écriture sonore immersive.

Une partie du catalogue stéréophonique est spatialisée dans l'objectif de répondre au besoin de l'industrie du streaming. Ces contenus ont été à la base pensés, écrits et mixés en stéréo. La première approche consiste donc à écrire une nouvelle forme d'immersion à partir d'une matière prédéfinie et limitée à deux canaux. L'autre approche consiste à penser l'immersion dès la création du contenu.

Il apparaît d'ores et déjà que les réponses apportées à nos problématiques ne seront probablement pas les mêmes en fonction de ces deux approches.

4.1.2 Comprendre les visions individuelles de l'audio spatial

La grammaire stéréophonique est « limitée » par la nature et les contraintes même du dispositif : un champ sonore frontal à 60° défini par un couple d'enceintes. L'écriture immersive s'ouvre quant à elle à un univers de systèmes de diffusion sans véritables limites. Cela laisse penser que sa grammaire existe de manière subjective à travers l'approche personnelle de chaque individu pratiquant l'audio spatial sur un dispositif donné. Il en existerait donc une multitude. Bien que notre objectif soit de trouver une grammaire commune à toutes les pratiques de l'audio spatialisé, il nous semble tout de même pertinent de comprendre les différentes visions personnelles existantes, de manière agnostique au système de diffusion utilisé, car ce sont celles-ci qui définissent la manière d'appréhender l'espace. Elles participent ainsi intimement à la grammaire.

Nous avons souhaité mettre en avant ces individualités. Nous avons ainsi tenu, après chaque entretien, à réaliser un schéma simplifié des principaux éléments considérés par l'individu lorsque celui-ci travaille une scène sonore à 360°.

Nous avons remarqué, de manière empirique et en échangeant, en amont de ce travail, avec Frédéric Changenet, directeur externe de ce mémoire, et Xavier Gibert, responsable du Rfi Lab (profil détaillé en Annexe 8 p.118) qu'il était parfois compliqué de communiquer entre professionnels de l'audio spatial. Au-delà d'un manque de grammaire commune, il s'avère que la compréhension que l'on a d'une scène sonore et la manière dont on veut la traiter ne sont pas simples à verbaliser. Nous avons ainsi confronté chaque professionnel au schéma que nous avons élaboré et qui définit, selon ce que nous en avons compris, la manière dont la personne conçoit l'espace sonore immersif. Cela nous a permis de questionner notre compréhension et d'avoir les corrections éventuelles de chacun³⁵.

4.1.3 Les personnes contactées

Pour ce mémoire, nous ne voulions pas nous restreindre à un seul média. Il était pertinent pour nous de nous intéresser à la fois à la musique, au cinéma et à la radio. Il était également intéressant de ne pas avoir uniquement des ingénieurs du son et d'aller outre la technique en échangeant avec des artistes. Nous avons aussi tenu à échanger avec des étrangers et avons eu l'opportunité de nous entretenir avec 3 professionnels internationaux basés à Berlin. Les profils des 10 personnes contactées

³⁵ Les réponses à cette confrontation, obtenues par mail, ne sont pas considérées dans l'analyse thématique détaillée en Partie 5 p.61.

sont disponibles en Annexe 8 p.118. Avant chaque entretien, les personnes interviewées ont signé un formulaire de consentement disponible en Annexe 9 p.121.

Si nous avons réussi, dans ce groupe de professionnels, à avoir un panel large de médias, de manières de travailler, et un équilibre convenable entre ingénieurs du son et artistes, nous convenons que, mise à part Alessandra Denegri, qui n'était initialement pas prévue, nous n'avons pas eu l'opportunité de nous entretenir avec des femmes. Nous nous sommes orientés vers nos relations et vers les profils conseillés par nos directeurs. Nous convenons qu'un effort supplémentaire aurait pu être effectué afin d'aller vers une parité.

4.2 Analyse des entretiens

4.2.1 Des entretiens qualitatifs et semi-directifs

Pour les entretiens réalisés dans le cadre de ce mémoire, nous utilisons une méthode qualitative, qui a notamment été définie par Paillé & Mucchielli (2021). Il est expliqué que, dans ce contexte de méthode qualitative, les entretiens servent à extraire le sens des informations exprimées plutôt que d'en déduire des statistiques. Les auteurs affirment qu'avec cette méthode, c'est l'expérience individuelle qui est valorisée – ce qui nous intéresse tout particulièrement, comme nous l'avons expliqué en 4.1.2 – à travers un processus « *mené d'une manière naturelle, sans appareils sophistiqués ou mises en situation artificielles, selon une logique proche des personnes, de leurs actions et de leurs témoignages* ». (Paillé & Mucchielli, 2021, p.15).

Au regard de la citation précédente et de notre développement en 4.1.1, il était pertinent pour nous de favoriser une discussion ouverte, sans direction préalable, et de laisser libre cours à un échange naturel avec le participant. Il était important pour nous de laisser la personne s'exprimer largement sur le sujet, tout en gardant une idée précise de ce que nous voulions obtenir : une conception de l'image sonore immersive et la manière dont la personne allait l'obtenir, tout en questionnant les notions que nous avons évoquées en 4.1.1.

Pour ce faire, nous avons choisi de mener des entretiens semi-directifs. Cette technique se situe à mi-chemin entre des entretiens non-directifs, qui tendent à laisser la personne s'exprimer librement et qui donneraient dans le meilleur des cas un monologue, et des entretiens directifs, qui, poussés à l'extrême, prendraient la forme d'un sondage. Les entretiens semi-directifs conviennent très bien à notre contexte puisqu'ils sont constitués d'un mélange entre questions ouvertes, qui laissent le sujet

libre de s'étendre sur des notions qu'il juge personnellement pertinentes, et questions plus fermées, qui permettent de recadrer l'entretien. (Blanchet & Gotman, 2010)

4.2.2 Guide d'entretien et difficultés

Afin de mener correctement un entretien semi-directif, il convient de se baser sur un guide d'entretien préalablement défini. Pour la construction de ce guide d'entretien, nous sommes partis sur l'idée d'un entretien en entonnoir. Nous commençons dans un premier temps par des questions larges, qui nous permettaient de comprendre la vision globale de la personne interviewée concernant le son spatialisé. Nous allions ensuite plus en détail vers les pratiques de mixage et tentions de questionner les notions que nous considérons importantes, en nous basant sur notre guide d'entretien, accessible en Annexe 10 p.123.

Dans leur méthodologie, Blanchet & Gotman (2010), définissent également un système de relance qui permet l'approfondissement de certaines notions. Ces relances peuvent prendre la forme de contradictions, en nous opposant à une affirmation, de questions qui abordent une nouvelle thématique ou encore de relances spontanées, qui consistent à demander le développement d'un sujet abordé par le sujet.

Nous souhaitons souligner ici la difficulté que nous avons rencontré dans la mise en pratique du suivi du guide d'entretien. Notre première question sur la notion d'immersion, la diversité des profils rencontrés et l'étendue du sujet abordé nous ont rapidement conduit vers des territoires vastes et variés. Dans l'idée d'un entretien semi-directif en entonnoir et au regard de notre développement en 4.1.1, c'était aussi l'objectif. Avant de rencontrer les sujets, il nous était impossible de savoir sur quelles notions ils s'attarderaient spécifiquement lors de l'évocation de leurs conceptions d'une image sonore en 3 dimensions. La variété des résultats obtenus – le mouvement, l'énergie, l'acoustique, ou encore le lieu d'écoute – le prouve. Dans ce contexte-là, la relance spontanée était privilégiée. Cela entre en cohérence avec les écrits de Paillé et Mucchielli (2021) que nous avons cités en 4.2.1.

Nous insistons sur le fait que le guide d'entretien a tout de même été consulté systématiquement et nous avons attaché une importance toute particulière à bien valider que tous les sujets souhaités avaient été abordés.

4.2.3 Le recours à une analyse thématique

Après avoir retranscrits nos entretiens, nous avons procédé à une analyse thématique en suivant la méthode définie par Paillé & Mucchielli (2021). Cette méthode consiste à repérer et à documenter. Pour cela, nous devons dans un premier temps extraire des textes les thèmes qui représentent les propos de la personne interviewée et qui correspondent au sujet d'étude. Il s'agit d'identifier l'idée fondamentale d'une prise de parole. Ensuite, il convient de mettre ces différents thèmes en lien les uns avec les autres et de hiérarchiser les notions établies, ce qui donne lieu à la création de sous-thèmes. Nous avons également utilisé la notion de concept, qui définirait une potentielle règle de grammaire, et qui est pour nous l'élément le plus précis de notre hiérarchisation.

« Il s'agit en somme de construire un panorama au sein duquel les grandes tendances du phénomène à l'étude vont se matérialiser dans un schéma (l'arbre thématique). »

Paillé & Mucchielli, 2021, p.271

Paillé et Mucchielli indiquent également qu'il est important de considérer 3 éléments dans notre méthodologie.

1. La nature du support matériel

Dans notre cas, nous avons utilisé le logiciel de traitement de texte *Microsoft Word*.

2. Le type de démarche de thématisation

Pour cette analyse, nous avons utilisé une démarche de thématisation en continu. Cela consiste à identifier les thèmes de manière continue au fur et à mesure de la lecture de nos retranscriptions.

3. Le mode d'inscription des thèmes

Pour ce mémoire, nous avons dans un premier temps relu tous nos entretiens puis surligné petit à petit les thèmes que nous avons identifiés avec des couleurs et utilisé des traitements de texte reconnaissables (soulignage, italique, gras). Une fois la totalité des retranscriptions relues et annotées, nous avons reporté toutes les citations sur un autre document *Word* en les triant par thèmes et en effectuant des premiers regroupements par sous-thèmes. Face au grand nombre de citations et de thèmes retenus, nous avons rapidement utilisé un tableau blanc pour clarifier nos idées et les ordonner facilement, puis nous avons utilisé le logiciel en ligne *Coggle* pour les visualiser

autrement. Nous avons alors obtenu notre arbre thématique. Nous avons ensuite construit deux types de tableaux *Excel* : le premier regroupant toutes les citations par concept, le second renseignant le nombre d'occurrences de chaque concept.

4.3 Résultats d'analyse

En préambule de cette analyse, l'auteur souhaite remercier et relever l'enthousiasme des interviewés face à mon sujet. Le besoin d'une grammaire commune a été manifesté par la majorité d'entre eux et cela nous a conforté dans la pertinence du sujet abordé.

4.3.1 Les résultats de l'analyse thématique

4.3.1.1 L'arbre thématique

Notre première visualisation est l'arbre thématique, disponible ci-dessous en Figure 2. Celui-ci permet de voir sur une seule page l'entièreté de la hiérarchisation de toutes les règles issues de notre analyse.

De la visualisation de cet arbre thématique, nous pouvons relever que nous avons organisé nos concepts en 3 catégories principales :

1. Le mixage, qui factuellement se rapproche le plus de ce qui concerne la grammaire de l'écriture immersive.
2. La technique, qui intègre les contraintes techniques qui peuvent avoir un impact sur l'écriture sonore immersive à partir du moment où elles sont prises en compte.
3. La philosophie, sous-entendu la vision des professionnels et artistes concernant l'audio immersif, leurs manières d'appréhender le medium, de penser l'espace. Nous intégrons à cette partie les éléments historiques et culturels.

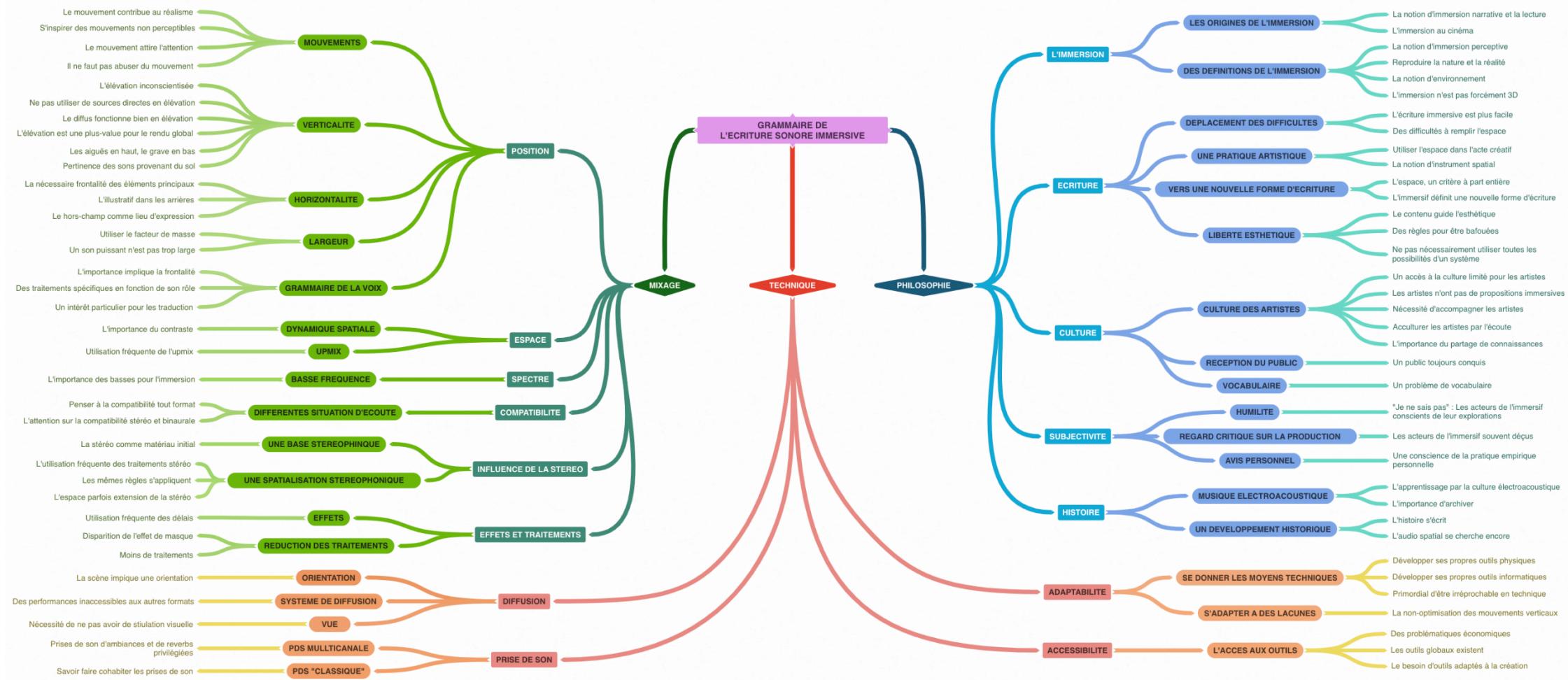


Figure 2 : Arbre thématique de la grammaire de l'écriture sonore immersive

Tableau 1 : Liste des concepts de la catégorie 1 – Mixage – et leurs occurrences

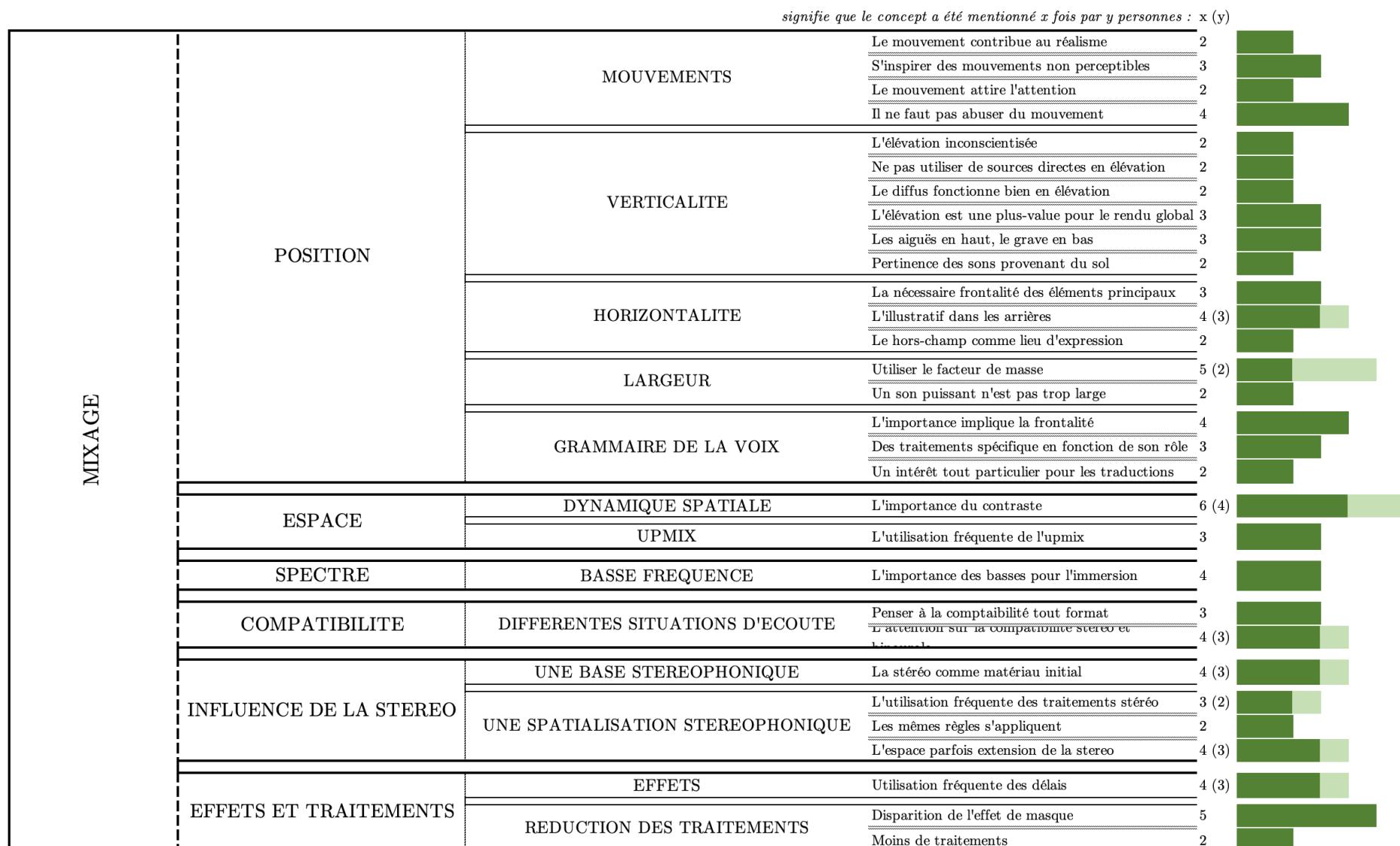


Tableau 2 : Liste des concepts de la catégorie 2 – Technique et Ingénierie – et leurs occurrences

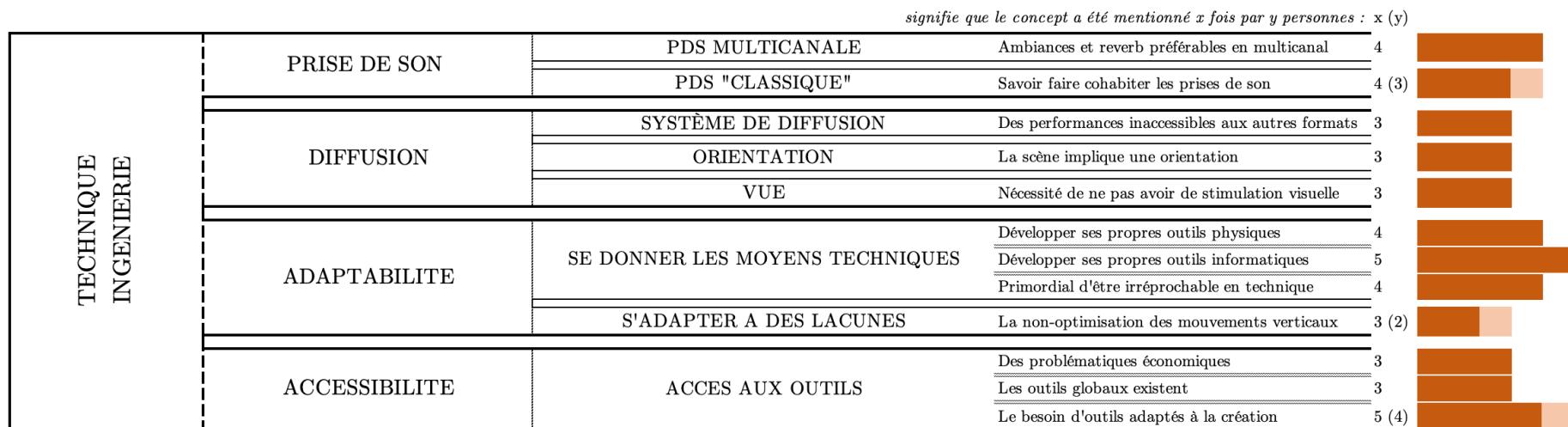
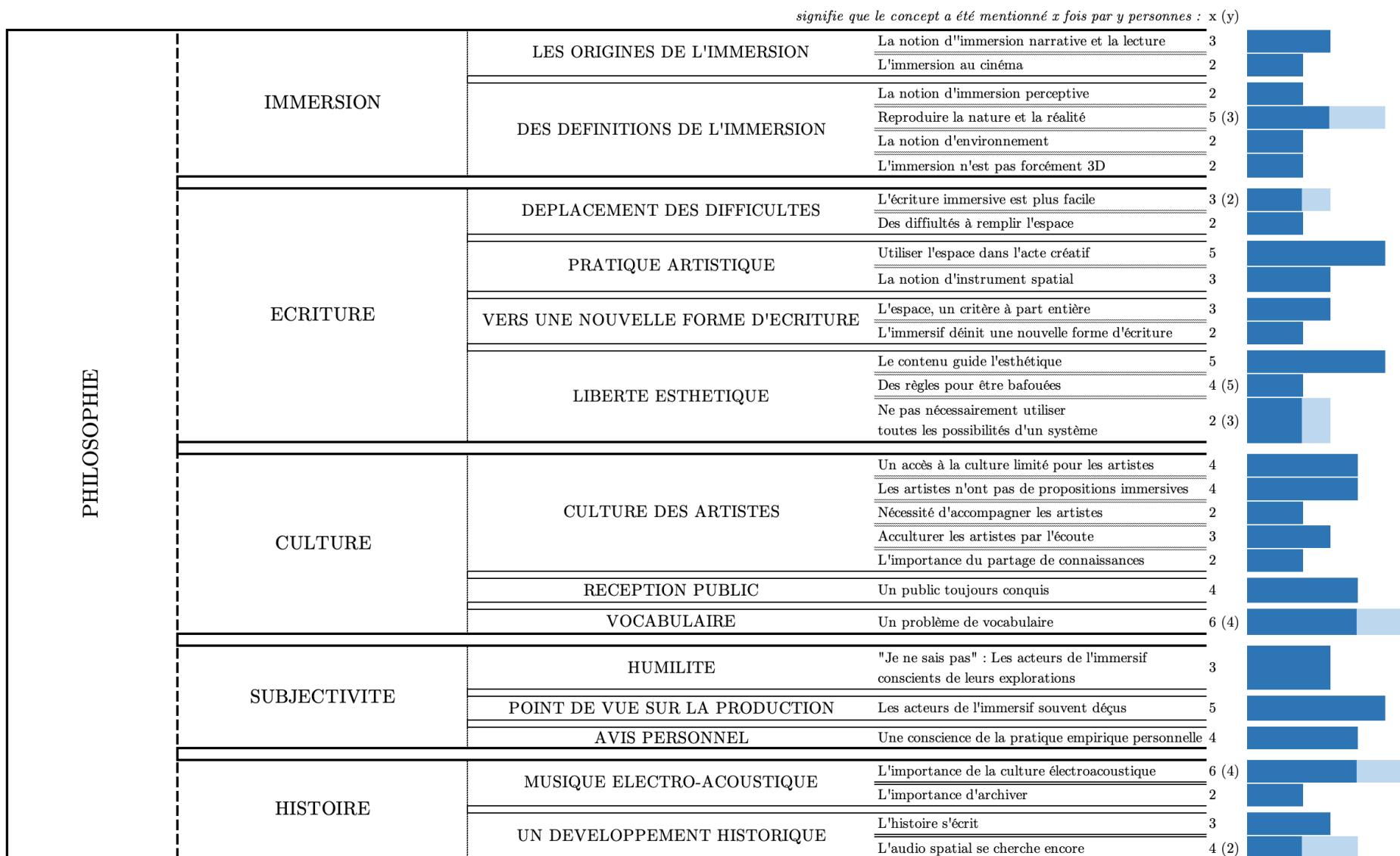


Tableau 3 : Liste des concepts de la 3 – Philosophie – et leurs occurrences



4.3.1.2 La liste des concepts et la grammaire associée

Les Tableau 1, Tableau 2 et Tableau 3 présentent la liste des concepts que nous avons relevés dans le cadre de notre analyse thématique et le nombre d'occurrences pertinentes de chacun d'eux.

Dans notre décompte des citations associées à chaque concept, nous avons remarqué que certaines personnes mentionnaient plusieurs fois le même concept à différents moments de la discussion. Nous considérons important de visualiser ces occurrences particulières. C'est ce que le lecteur pourra comprendre lorsqu'il est écrit « x (y) », avec x le nombre d'occurrence du concept et y le nombre de personne l'ayant abordé.

Si nous prenons l'exemple concret du concept *Mixage > Position > Largeur > Utiliser le facteur de masse* (voir 5.1.1), nous pouvons lire l'occurrence 5(2). Cela signifie ici qu'il a été mentionné par 2 personnes différentes, mais qu'il est apparu 5 fois au total. Antoine Petroff l'a mentionné 3 fois et Hervé Déjardin 2 fois.

Chaque concept peut être compris comme une « règle de grammaire » de l'écriture sonore immersive, du moins en ce qui concerne la Catégorie 1 (p.61) et la Catégorie 2 (p.73). Les concepts de la Catégorie 3 (p.79) ne sont pas véritablement associés à des « règles de grammaire », mais à des réflexions qui ont un impact sur la manière de concevoir l'image sonore à 360°. Il nous semblait donc pertinent de les mentionner.

La liste détaillée des concepts et les citations qui les ont définis font l'objet de la Partie 5 p.61.

4.3.2 Des visions de l'audio spatialisé très personnelles

« Comment [le système de diffusion, ndlr] est utilisé esthétiquement, c'est ça la question qu'il faut se poser. Et pas de quoi il est capable. »³⁶

Sharma, dans le cadre de notre entretien.

En plus de l'établissement d'une partie de la grammaire de l'écriture sonore immersive, dont la pertinence sera discutée lors de notre Discussion, ces entretiens nous ont permis de nous rendre compte des différences marquées entre les visions de chacun concernant l'audio spatialisé. Les manières de concevoir ce médium sont très variées et il s'agit selon nous de philosophies propres à chacun qui peuvent parfois se contredire de manière brutale. Il nous a semblé particulièrement pertinent de mettre en avant certaines conceptualisations de l'audio spatialisé trouvées parmi les personnes interviewées, comme nous l'avons expliqué en 4.1.2.

Le lecteur est vivement invité à consulter l'Annexe 11 p.124 qui comporte des textes résumant la philosophie, ainsi que des graphiques, de différentes personnes avec qui nous nous sommes entretenus. Cette annexe nous permet également de mettre en avant de plus longues citations qui traitent de processus très spécifiques concernant certains traitements associés à une volonté de mise en scène du champ sonore spatialisé.

Ces regards personnels sur la pratique sont, selon nous, un complément non négligeable à l'ébauche de la grammaire, qui constitue l'objet central de notre mémoire, tant les manières de travailler diffèrent encore entre praticiens.

³⁶ « How it is esthetically used, that's the question, not what it is capable of doing. » Notre traduction.

5. Le détail des concepts, l'établissement d'une grammaire non exhaustive de l'écriture sonore immersive

5.1 Catégorie 1 : Le mixage

5.1.1 Thème 1 : La position

--- SOUS THÈME 1 : LES MOUVEMENTS ---

Le travail du mouvement est apparu régulièrement dans nos entretiens. Il n'est cependant pas appréhendé de la même manière et avec la même importance en fonction des visions de chacun. Si certains le considèrent comme un élément essentiel de leur écriture, d'autres considèrent qu'il peut être utilisé comme excuse de certaines productions immersives sans réel enjeu esthétique.

>>> Concept 1 : Le mouvement contribue au réalisme

La volonté de reproduire un environnement naturel est très présente dans les productions immersive (voir 5.3.1). Il s'avère que le mouvement participe au réalisme de la scène sonore.

AP « *Avoir en permanence des trucs qui bougent autour de toi ça participe au réalisme.* »

HD « *Dans la vie de tous les jours nous sommes en permanence sollicités par des sons en mouvement, [...] ça doit être la base de notre travail parce que c'est ce qui construit l'espace naturel.* »

>>> Concept 2 : S'inspirer de mouvements non perceptibles

Les mouvements non perceptibles et lents sont également utilisés. Ils rappellent le comportement de sources sonores naturelles et permettent de simuler les mouvements de tête que nous réalisons en permanence.

HD « *Il y a plein de trucs dans la nature qui bougent que t'entends pas vraiment, tu les conscientises, néanmoins ces choses relatives créent des mouvements spectraux permanents parce qu'on en fait on analyse l'espace. Faut jamais perde ça de vue, c'est ce qui est important.* »

MA « *Des fois ce qui est intéressant c'est les choses qui bougent mais tout doucement.* »

AP « *Alors je joue pas mal avec les mouvements qui ne sont pas forcément perceptibles.* »

>>> Concept 3 : Le mouvement attire l'attention

D'un point de vue psychoacoustique, le mouvement attire l'attention de l'auditeur.

XG « *Créer des mouvements qui vont attirer l'attention sur un moment donné pour faire passer autre chose. »*

MA « *Ce que je veux dire c'est que perceptivement, le mouvement, ça attire quand même l'attention. »*

>>> Concept 4 : Il ne faut pas abuser du mouvement

Au regard du dernier concept, il est assez largement reconnu qu'attirer l'attention de manière trop appuyée n'est pas souhaitable. Le mouvement doit donc être utilisé avec parcimonie.

MA « *Si tu fais bouger 3 ou 4 sons en même temps, le cerveau il sait plus dire qui bouge ou quoi. »*

AP « *Après ça devient vite casse couille [...] au bout d'un moment avoir le même truc qui tourne ... »*

XG « *Ouai vas y on va faire tourner le charley » ouai mais tu vas voir ça va être un peu relou donc non. »*

LW « *Je trouve que les mouvements sont vraiment très intéressants. Il faut bien sûr faire attention à ne pas trop en utiliser. »*³⁷

--- SOUS THÈME 2 : LA VERTICALITÉ ---

La question de l'élévation était un des thèmes que nous souhaitions aborder précisément, notamment au regard de notre développement en 3.3.5. Nous avons choisi d'intégrer à ce sujet la notion de source provenant du sol. Nous avons donc nommé ce sous-thème « verticalité ». Il ressort de ce sous-thème que les personnes interviewées sont conscientes que l'être humain perçoit mal l'élévation.

>>> Concept 1 : L'élévation inconscientisée

Nous pouvons ici faire un parallèle avec les mouvements non perceptibles. Il est intéressant d'utiliser la hauteur pour créer des éléments perçus inconsciemment.

OG « *Dès que c'est vertical on est pas très bon en tant qu'individu et donc ça permet de mettre des choses sans qu'on s'en rende compte véritablement. »*

AP « *Même si c'est inconscient parce que c'est le bordel complet [...] ça marche très bien en hauteur. »*

³⁷ « *I think movement is far more interesting, of course you need to be careful to not overdo it. »* Notre traduction.

>>> Concept 2 : Ne pas utiliser de sources directes en élévation

De la même manière que le mouvement doit être utilisé avec précaution, le positionnement de sources en hauteur pose question.

AP « *Après, honnêtement, de mettre des objets particuliers vraiment en hauteur, je n'ai jamais utilisé comme ça. »*

MA « *Alors après effectivement tu peux mettre ton soliste hein, mais bon ça je trouve que ça marche pas. »*

>>> Concept 3 : Le champ sonore diffus fonctionne bien en élévation

Contrairement aux sources directes, l'utilisation du diffus en hauteur semble bien fonctionner.

MA « *Pareil tu vois j'ai fait des mixs de classique ou je mettais juste un des couples un peu lointains au plafond pas fort. »*

LW « *Pour la musique acoustique tu as généralement des informations diffuses provenant des arrières et de l'élévation. »*³⁸

>>> Concept 4 : L'élévation est une plus-value pour le rendu global

L'élévation est perçue comme une plus-value non négligeable dans le rendu global de la scène sonore. La notion d'air évoquée ici peut très certainement être mise en lien avec le concept précédent.

MA « *Des fois il suffit de mettre un peu d'effet, pas fort, là-haut et tout à coup ça fait de l'air, ça respire. »*

AP « *Oui parce que ça rajoute... ça te permet d'élargir le truc. C'est vrai que l'élévation te permet ça. »*

OG « *Les tops, ça donne une dimension verticale au son [...] une manière de remplir la salle qui est vraiment extraordinaire. »*

>>> Concept 5 : Les aigus en haut, le grave en bas

Notre perception de l'élévation en fonction des fréquences est connue par les personnes interviewées. Le travail de mixage est donc effectué en conséquence.

AM « *Spectralement faire bouger des graves au-dessus de ta tête c'est quand même dur, j'ai pas encore réussi à entendre ça *rire* [...] donc forcément ce serait quand même plutôt des éléments avec un centre spectral quand même plutôt haut. »*

AP « *Je vais plutôt mettre des sons plutôt aigus en hauteur parce que c'est là que tu vas percevoir mieux l'élévation. »*

GS « *Les hautes fréquences sont en hauteur, les basses fréquences en bas. »*³⁹

³⁸ « *Usually for acoustic music, you have diffuse sound information coming from the rear and the top. »* Notre traduction.

³⁹ « *Higher frequencies are high and lower frequencies are down. »* Notre traduction

>>> Concept 6 : Pertinence des sons provenant du sol

Il est également apparu, lors de nos entretiens, que le positionnement de sources sonores provenant du sol peut être un élément particulièrement pertinent dans la création de scènes sonores nouvelles pour l'auditeur.

LW « *Il y a une perspective très intéressante en audio immersif, qui est de ne pas uniquement positionner des sources en hauteur, mais précisément d'en placer en dessous de l'auditeur.* »⁴⁰

AP « *Quand t'as du son qui vient d'en bas ça marche vachement bien aussi.* »

--- SOUS THÈME 3 : HORIZONTALITE ---

>>> Concept 1 : La nécessaire frontalité des éléments principaux

Les éléments principaux sont très largement placés devant l'auditeur (voir également le sous-thème 5 : Grammaire de la voix), dans le cas où, évidemment, la scène sonore est orientée.

AP « *Si tu veux de la puissance je pense qu'il faut que ce soit frontal quoi.* »

XG « *Tout ce qui est important va quand même être dans une forme de 180° frontal.* »

HD « *Si tu as un éléments lead, qui doit être le centre de l'intérêt il va forcément être là devant.* »

La notion de frontalité est cependant remise en question par Gerriet Sharma lorsqu'il dit : « *Tous les trucs ennuyeux qui se passent aujourd'hui en audio spatial assument encore que la scène est devant toi. Il n'y aura pas de changements en musique, artistiquement, esthétiquement, si tu penses toujours que la scène est devant toi.* »⁴¹

>>> Concept 2 : L'illustratif dans les arrières

En conséquence, les haut-parleurs arrières sont utilisés pour l'aspect illustratif de la narration, en complément seulement de ce qui est positionné en frontal.

OG « *Après un ensemble assez diffus sur tout ce qui est l'espace électroacoustique qui est en dehors de l'écran.* »

LW « *Pour la musique acoustique tu as généralement des informations diffuses provenant des arrières et de l'élévation.* »⁴²

XG « *D'où le fait qu'au cinéma [...] derrière c'est que des effets.* »

⁴⁰ « *So, the very interesting perspective in immersive audio content is to not only put sound sources on top but even under the listener perspective.* » Notre traduction.

⁴¹ « *The boring spatial audio stuffs that is happening now is still assuming that stage is in front of you. There's no changing in music, artistically, esthetically, nothing is going to happen if you still think the stage is in front of you.* » Notre traduction

⁴² « *Usually for acoustic music, you have diffuse sound information coming from the rear and the top.* » Notre traduction.

XG « *Quand elle va être en illustration [...], la musique va [...] être derrière.* »

Concept 3 : L'hors-champ comme lieu d'expression

Si les éléments illustratifs sont placés dans les arrières, il ne faut pas pour autant négliger l'importance de l'hors-champ dans la scène sonore immersive. Il est considéré comme un véritable lieu d'expression.

AM « *T'as moins d'expression parce que c'est comme si au cinéma il y a plus de hors champ.* »

OG « *Un moment donné là tu t'éclates [...] tout ce que n'a pas pu filmer la caméra mais que tout d'un coup la spatialisation aide énormément la mise en scène.* »

--- SOUS THÈME 4 : LARGEUR ---

>>> Concept 1 : Utiliser le facteur de masse

Le terme de « masse » est revenu chez deux de nos intervenants de manière particulièrement appuyé.

AP « *Si tu prends un HP que tu le dirige contre le mur t'as un spread de fait qui te donne l'impression que la masse est plus grande.* »

AP « *Quand tu mets des HP aléatoire [aléatoirement dans l'espace, ndlr], t'es vraiment dans un environnement sonore, dans une masse sonore, t'es incapable de dire d'où viennent les sons.* »

AP « *Souvent j'enregistre avec plein d'amplis différents pour avoir des couleurs, [...] t'as une masse sonore démentielle.* »

HD « *Des fois ça se joue là-dessus, la largeur d'une source sonore [...] va aussi jouer de l'importance du son dans le mix à un instant donné et ça c'est les phénomènes de masse.* »

HD « *Si je veux faire une masse je vais dupliquer la piste et cet équilibre entre le haut et le bas je le fait avec les deux pistes et leur niveau relatif. [...] En faisant ça là tu donnes des effets de masse, t'as des trucs monstrueux qui passent en fait. Là c'est aussi pour travailler les masses.* »

>>> Concept 2 : Un son puissant n'est pas trop large

Si nous avons vu précédemment qu'un élément important est souvent placé dans le frontal, notamment pour son impact, il est préférable qu'il garde une largeur limitée.

AP « *Typiquement un son qui doit avoir un impact, il sera très souvent frontal mais surtout il aura une largeur qui sera limité.* »

MA « *Ils ont mis les guitares hyper larges donc il n'y a plus rien.* »

--- SOUS THÈME 5 : GRAMMAIRE DE LA VOIX ---

Il nous est apparu important de dédier un sous-thème à la grammaire de la voix. L'évocations de traitements associés à la voix sont en effet apparus de manière récurrente et c'est une source sonore suffisamment particulière et largement employée pour que nous puissions détailler sa grammaire.

>>> Concept 1 : L'importance implique la frontalité

Nous avons vu que les éléments principaux sont placés dans le frontal. Cela est également valable pour les voix.

- | | |
|-----------|--|
| OG | <i>« Il y a ce qu'on appelle le <i>voco-centrisme</i> parce que la voix elle est très agréable quand elle vient de l'écran. »</i> |
| AP | <i>« Typiquement voix [ou] trucs précis [implique un] HP frontal qui vient vers toi. »</i> |
| HD | <i>« À partir du moment où t'as des voix lead, t'as forcément une frontalité. »</i> |
| XG | <i>« Donc je suis auditeur quelqu'un va me parler, à moi, je vais forcément me le mettre [...] dans une zone qui va être dans mon frontal. »</i> |

>>> Concept 2 : Des traitements spécifiques en fonction de son rôle

En fonction de la sensation que l'on veut donner, ou du rôle de la voix dans la narration, des multitudes de traitements différents peuvent être utilisés, chacun ayant un impact particulier sur la perception de celle-ci.

- | | |
|-----------|--|
| AP | <i>« Tu mets une voix, juste une voix, à un endroit, bah tu vas donner une impression de fragilité [...] et à l'inverse la même voix, si t'as une reverb longue partout autour, tu peux donner une impression de puissance [...] A l'inverse tu mets une reverb courte, tu vas avoir l'impression d'avoir une personne un peu fragile, perdue dans un espace clos. »</i> |
| XG | <i>« Je l'ai placé un peu avant gauche et le podcast est structuré avec le personnage principal centre droit, centre un peu haut mais genre 10° à droite, tous les intervenant sachant sur le 180 devant, tous les personnages illustratifs etc. sur le 180 arrière. »</i> |
| LW | <i>« Il est grand et imposant. [...] Et ça a très bien marché de monter sa voix. Elle était dans le frontal mais pas dans la couronne principale, elle était entre la couronne principale et les tops. »⁴³</i> |

⁴³ « He's tall and big [...] and it worked very well to lift the voice up top. It was on front but between the mid layer and the upper layer. » Notre traduction

>>> Concept 3 : Un intérêt particulier pour les traductions

Les avantages de l'audio immersif, notamment en termes d'effet de masque (voir 5.1.6), permettent de se concentrer sur une source en particulier sans être dérangé par une autre. Cette idée trouve une application très précise dans les contextes de traductions en documentaires.

- XG** « *Une traduction, on va aller la mettre derrière et assez proche. »*
- LW** « *Dans les contenus immersifs, tu peux les avoir côte-à-côte au même niveau, et c'est vraiment intéressant ce qu'il se passe, si tu as par exemple la voix originale à l'avant gauche et la voix traduite au même niveau à l'arrière droit. »⁴⁴*

5.1.2 Thème 2 : L'espace

-- SOUS THÈME 1 : DYNAMIQUE SPATIALE -->>> Concept : L'importance du contraste

Dans la grammaire du mixage stéréophonique, l'idée de jouer du contraste sur la largeur de certaines sources est répandue, notamment entre couplet et refrain. On retrouve cette notion de contraste dans l'audio immersif. Il est largement reconnu qu'il est souhaitable de jouer sur la dynamique spatiale.

- OG** « *Le principe [...] de tout format de spatialisation [...] c'est d'avoir justement une dynamique de spatialisation. »*
- OG** « *Tu as beau être dans un film Dolby Atmos tu vas mixer en Mono et ça va être génial. Après tu profiteras d'autant plus d'un effet de contraste sur des scènes qui le méritent. »*
- HD** « *Il y a eu plein de titres ou les deux premières minutes c'était de la mono, avant d'ouvrir, premièrement parce que ça joue du contraste. »*
- MA** « *Par exemple sur ces morceaux là je proposais de mettre la rythmique sur le côté par exemple. [...] je trouve que ça crée un contraste hypersaisissant. »*
- MA** « *C'est super parce que c'est cette espèce de truc aérien qui arrive tout d'un coup ça fait un contraste majestueux. »*
- ER** « *Le Spatial c'est créer des contrastes. »*

⁴⁴ « *In immersive content you can have them side by side at same level and it's really interesting what happens. If you have some sound sources maybe from the front left original voice and from the back right or rear right the translation at same level. »* Notre traduction.

-- SOUS THÈME 2 : L'UPMIX --

>>> Concept : L'utilisation fréquente de l'upmix

Dans un contexte où les contenus immersifs ne sont pas encore largement répandus, notamment en musique ou la stéréo est encore largement dominante, il est fréquent d'utiliser l'*upmix* afin de rendre le contenu plus immersif.

- XG** « *On l'a joué en disant que quand la musique est in elle est upmixée, pleine balle. »*
- OG** « *Donc à partir de là on fait des upmix divers et variés ça n'a pas empêché de s'amuser avec et puis voilà. »*
- LW** « *Tu peux aussi utiliser Halo Upmix de Nugen ou ce que tu veux pour le rendre plus immersif. »⁴⁵*

5.1.3 Thème 3 : Le spectre

-- SOUS THÈME : BASSE FREQUENCE --

>>> Concept : L'importance des basses fréquences dans l'immersion

Le bas du spectre, notamment les extrêmes graves, est reconnu par beaucoup comme étant un élément essentiel de l'immersion.

- AP** « *Très souvent les subs, les infras etc. participent à te donner cette immersion. »*
- HD** « *Pour avoir une parfaite perception de l'espace [...], ton bas du spectre [...] entre 90 et 200Hz, cette zone-là tu la maitrise totalement. »*
- GS** « *L'ancreage de la pièce est toujours dans les subs. [...] Quelque chose qui n'est pas ancré par les subs t'échappe toujours, tu ne peux pas entrer en connexion avec le contenu. Et pour l'image spatiale, pour la composition de l'espace, tu as besoin d'être ancré au sol. [...] Nous ne devrions pas oublier que l'immersion vient en premier lieu des subs. »⁴⁶*
- LW** « *Utiliser beaucoup d'LFE⁴⁷ ! *rires* [...] Je dirais que le canal LFE est ton ami. »⁴⁸*

⁴⁵ « You can also use Halo upmix from Nugen or whatever to make it more spatial. »

⁴⁶ « The grounding of the piece is always in the sub. [...] Something that is not grounded by subs is really flying away, you cannot connect to it. And for the Spatial Image, for the composition of Space, you need the ground. [...] We shouldn't forget that immersiveness first of all starts with subs. » Notre traduction

⁴⁷ LFE : Low Frequency Effect, fait référence au signal basse fréquence qui est envoyé dans les *subwoofers*

⁴⁸ « Use a lot of LFE! *laughs* [...] I would say that LFE channel is your friend. » Notre traduction

5.1.4 Thème 4 : La compatibilité

-- SOUS THÈME : DIFFERENTES SITUATIONS D'ECOUTE --

>>> Concept 1 : Penser à la compatibilité tout format

Comme nous l'avons souligné en 4.1.2, il existe une multitude de systèmes de diffusion immersive. Une pièce doit pouvoir être jouée sur n'importe lesquels de ces systèmes de la manière la plus fidèle qui soit. Il est donc important de prendre ces diversités en compte dans l'objectif de contrôler l'écoute sur les systèmes dont le rendu sera différent de celui utilisé lors du mixage.

- OG** « *C'est aussi le problème du mix, c'est que ton travail [...] puisse résister à toute sorte de configuration quoi.* »
- XG** « *On ne peut quand même pas totalement s'affranchir du compromis entre les différents formats.* »
- LW** « *Ils sont tellement différents que tu as besoin d'adapter ton mix pour qu'il soit adapté au plus grand nombre de systèmes de diffusion.* »⁴⁹

>>> Concept 2 : L'attention sur la compatibilité stéréo et binaurale

Les contenus audio sont encore majoritairement produits en stéréo (voir 5.1.5). De plus, le contexte d'écoute actuel est tel que la grande majorité des contenus immersifs sont écoutés au casque, que le binaural soit disponible ou non. Il est donc important d'avoir une cohérence entre les contenus immersifs sur enceintes et ceux sur casque, écoutés en binaural ou en stéréo.

- HD** « *Je demande à avoir la stéréo masterisée, premièrement, comme ça on va essayer de faire le lien, d'avoir une cohérence entre la stéréo et le mixage immersif.* »
- MA** « *Le problème c'est de pas se retrouver [...] à avoir, entre ta version stéréo et ta version Atmos, juste un truc qui sonne avec plus d'espace et de reverb [...] puis t'écoutes la stéréo et t'as pas le même disque, ça, ça va pas.* »
- MA** « *Peut-être qu'au binaural ça marche bien, mais bon voilà les mecs ils ont écouté qu'en binaural, bah c'est bien, du coup demain le moteur de binaural change, il change, donc faut écouter ce qu'il y a sur les haut-parleurs c'est ce qu'on met sur la bande.* »
- XG** « *Donc si on place cette traduction [...] dans le binaural de synthèse [...] on va vraiment faire gaffe que dans le down mix stéréo elle revienne bien se poser à sa place.* »

⁴⁹ « *It's just so different so you need to adapt your mix to fit the most reproduction environments.* »
Notre traduction

5.1.5 Thème 5 : L'influence de la stéréo

Les discussions avec mon directeur externe Frédéric Changenet ont rapidement indiqué que, dans l'objectif d'obtenir une grammaire de l'écriture sonore immersive large, il était important de s'éloigner des pratiques qui ont pour principe de se référer à la stéréo pour la construction d'un nouveau champ sonore à 360°. Toujours est-il que la stéréo a encore une influence majeure sur la production de contenus immersifs.

--- SOUS THÈME 1 : UNE BASE STEREOPHONIQUE ---

>>> Concept : La stéréo comme matériau initial

Comme nous allons le voir en 5.2.4, les dispositifs permettant de produire en audio spatialisé ne sont pas encore très répandus. La praticité de la création stéréophonique rend encore cette manière de travailler très largement majoritaire. Les ingénieurs du son doivent ainsi gérer des contenus stéréophoniques dès la genèse de leur travail.

- MA « *Le truc c'est que t'as toujours un stéréo à la base.* »
- MA « *Toujours parti d'un stem d'un mix stéréo [...]. Je pars d'un truc qui se tient.* »
- XG « *Donnez-nous vos stems et on va proposer et on va mettre en place une spatialisation.* »
- OG « *A la base ça reste, et dans énormément de cas, juste de la stéréo.* »

--- SOUS THÈME 2 : SPATIALISATION STEREOPHONIQUE ---

>>> Concept 1 : L'utilisation fréquente des traitements stéréo

En conséquence du concept précédent ou dans une démarche esthétique, les traitements habituellement utilisés en stéréo sont également utilisés dans le contexte de mixage immersif.

- MA « *Donc on a récupéré tous ses traitements de dynamique tous ces traitements tels quels.* »
- AP « *Après ce qui est intéressant avec l'immersif ça pareil [...] moi ça m'arrive souvent que [...] certains éléments, les mixer en stéréo.* »
- AP « *Parce que justement, une basse batterie passé dans des compresseurs de mastering en stéréo ça a une couleur qui est intéressante.* »

>>> Concept 2 : Les mêmes règles s'appliquent

Au-delà des traitements, il est également ressorti chez deux personnes que les règles appliquées dans un contexte stéréophonique sont valables dans un contexte spatialisé.

MA « Je pense que les règles sont exactement les mêmes qu'en stéréo, parce qu'en réalité faire de la stéréo c'est faire de la spatialisation entre deux enceintes. »

HD « Tous les critères qui sont valables dans l'analyse de l'image stéréo tu peux les appliquer à l'image immersive tu vois. »

>>> Concept 3 : L'espace parfois extension de la stéréo

Certainement conséquence, en partie, des deux concepts que nous venons d'évoquer, le travail de l'espace est souvent considéré comme une extension de la latéralisation stéréophonique, autrement dit à un élargissement du *panning*.

MA « On peut refaire les mêmes propositions ou le même type de démarche quand on mixe en spatialisé et on va transporter simplement le paradigme gauche-droite à 360, mais c'est le même paradigme. »

LW « Tout ce que j'ai fait c'était en fait du panning et de la spatialisation »⁵⁰

LW « Donc tu présentes une scène, mais la scène est plus large que le champ stéréophonique. C'est plutôt comme être juste devant un demi-cercle de musiciens. »⁵¹

ER « En gros c'est une extension d'un domaine stéréo et puis point barre, c'est l'ingé son qui va mixer tout ça et ça va être super. »

Nous soulignons le fait que cette dernière citation d'Éric Raynaud doit être comprise comme une critique. Raynaud considère que la plupart des travaux consistant à produire de la « stéréo immersive » ne sont absolument pas pertinent. Nous le rappelons lors du détail de sa philosophie en A11.5 p.136.

5.1.6 Thème 6 : Les effets et traitements

--- SOUS THÈME 1 : LES EFFETS ---

>>> Concept : Utilisation fréquente des délais

La pratique du mixage passe par l'utilisation d'un certain nombre de traitements et d'effets. Dans le cadre de nos entretiens, l'utilisation des délais a été plusieurs fois mentionnée. Par sa nature temporelle, ce traitement est particulièrement adapté à l'audio spatialisé car il permet de créer de la décorrélation et de jouer sur différentes localisations entre signal brut et signal traité.

AP « Un truc que j'aime bien faire avec les délais, [...] j'en ai 5 ou 6 différents et je les paramètre de manière aléatoire. Ça te crée une espèce de masse sonore autour. »

⁵⁰ « All I did was like panning and spatialization actually. » Notre traduction.

⁵¹ « So, you present a stage, but the stage is wider than the stereo field. It's pretty more like you're almost right up front to a semi-circle of musician. » Notre traduction.

- AP** « Si tu veux rajouter un délai qui ne perturbe pas le reste là tu peux utiliser l'élévation ça marche vachement bien. »
- MA** « T'as tout le mix dans le LR, et t'as des breaks et des accents et de temps en temps juste un délai ou quoi qui se balade etc. »
- LW** « Tu as ton signal dry quelque part [...], et là tu as peut-être une reverb artificielle ou un délai [...] et tu le places dans une autre direction. »⁵²

--- SOUS THÈME 2 : UNE REDUCTION DES TRAITEMENTS ---

>>> Concept 1 : La quasi-disparition de l'effet de masque

Concept largement reconnu par une grande partie des interviewés, l'audio immersif implique la quasi-disparition de l'effet de masque, qui est sans doute le paramètre le plus important à appréhender en mixage stéréo. Cette caractéristique de l'audio spatial lui apporte un avantage non négligeable et permet de ne pas avoir recours à certains traitements.

- MA** « Ces effets de masquage qui n'arrivent plus font qu'on peut faire un très joli son avec beaucoup de Headroom beaucoup plus facilement. »
- AP** « Et en live c'est pareil, [...] juste pour séparer les sources tu te fais carrément moins chier pour mixer quoi, ta voix, elle est là et t'as pas besoin de la traiter. »
- XG** « Alors oui dans le sens où, déjà musicalement, on peut en séparant les sources on s'affranchit des effets de masques, donc on a une plus grande liberté là-dessus. »
- OG** « C'est con mais comme les choses se dégagent bien s'identifient bien t'as quelque part moins de problèmes aussi quoi. »
- LW** « La majorité du temps, le problème qui consiste à voir une source en masquer une autre n'arrive plus. »⁵³

>>> Concept 2 : Moins de traitements utilisés

La meilleure discrimination des sources induite par l'audio spatial permet d'utiliser moins fréquemment les traitements fréquentiels ou la suppression de bruit. Le meilleur rendu dynamique permet également le moindre recours aux compresseurs.

- AP** « Déjà le truc je pense que dans les mixs immersif j'utilise quasiment plus de compresseurs sauf pour un son particulier, j'utilise 10 fois moins d'EQ. »
- XG** « Le multicanal et le son 3D en général permet de ne plus avoir à se battre avec le lieu de prise de son. »

⁵² « You have a dry signal somewhere [...] and then you have maybe an artificial reverb or a delay, and put it [...] in another direction. » Notre traduction.

⁵³ « The effect of one sound overshadowing another isn't a problem anymore most of the time. » Notre traduction.

5.2 Catégorie 2 : La technique

5.2.1 Thème 1 : La prise de son

--- SOUS THÈME 1 : LA PRISE DE SON MULTICANALE ---

>>> Concept : Les prises de son d'ambiances et de réverbération privilégiées

Les prises de son multicanales, utilisées fréquemment pour les ambiances et la réverbération, sont largement favorisées dans les productions pour leur réalisme.

AM « *Il y a une différence de perception assez forte entre des champs acoustiques reconstitués de manière artificielle [...] et des champs pris avec un système cohérent. [...] C'est plus une qualité globale de profondeur qui marche hyper bien, qui est impressionnante et qui donne plein d'autres trucs.* »

XG « *C'est là que nous on a intégré que des prises de son en HF, pour pouvoir nous éloigner au maximum des journalistes pour pouvoir avoir toutes nos ambiances en audio 3D, en multicanal, ou en ambisonique, clean par rapport aux micros des journalistes.* »

MA « *Typiquement en classique on va essayer de faire des prises de son d'ambiance multicanale.* »

LW « *Quand il y a ne serait-ce qu'une petite chance qu'il y ait un mix immersif, je pose toujours des micros supplémentaires. [...] Il y a une installation fixe de microphones 3D donc tu n'as qu'à les patcher et les enregistrer, c'est toujours là.* »⁵⁴

--- SOUS THÈME 2 : LA PRISE DE SON « CLASSIQUE » ---

>>> Concept : Savoir faire cohabiter les prises de son

Nous admettons facilement qu'il n'est pas forcément fréquent d'avoir accès à des systèmes de captation multicanale. Il est donc nécessaire de savoir gérer des prises de son mono ou stéréo et de les faire cohabiter avec des prises de son multicanales. Cette cohabitation est un procédé facile à mettre en œuvre et dont les résultats fonctionnent visiblement très bien.

XG « *On a nos deux pistes monos, notre système qui va choper la room ou la pièce ou l'extérieur, les ambiances, tout ça est synchrone et on peut faire un montage clean des parties mono, les repositionner dans l'espace et les baigner dans notre ambiance multicanale.* »

⁵⁴ « *When there's even a small chance that there will be an immersive mix, I always put up some mikes. [...] There's a fix setup of 3D audio microphone array so you just need to patch them and record them, there's always there.* » Notre traduction

- XG** « *On mélange allègrement du binaural natif et du binaural de synthèse. Dans l'Atmos j'ai deux objets qui vont être en binaural natif.* »
- AM** « *Tu mets ta voix anéchoïque dans la reverb [...] et tu mets la petite ambiance que t'as prise à l'Eigenmike⁵⁵, l'illusion elle est instantanée.* »
- LW** « *Tu as bien sûr besoin du signal des microphones additionnels, mais je n'ai jamais effectué un remixage d'un enregistrement stéréo qui avait été enregistré avec uniquement de la stéréo comme dispositif principal de prise de son.* »⁵⁶

5.2.2 Thème 2 : La diffusion

--- SOUS THÈME 1 : LE SYSTÈME DE DIFFUSION ---

>>> Concept : Des performances inaccessibles aux autres formats

En plus de la diminution de l'effet de masque, les dispositifs immersifs, grâce à leur nombre plus élevé de haut-parleurs disposés tout autour de l'auditeur, offrent des performances que ne permettent pas un système stéréophonique classique.

- MA** « *Et clairement ce qu'on peut faire avec le multicanal c'est des expériences plus poussées qu'en stéréo.* »
- XG** « *Ça j'ai pu le faire parce que j'étais en audio 3D et j'aurais pas pu faire si j'avais été en stéréo.* »
- OG** « *C'est vrai que c'est des sensations que tu ne peux pas reproduire dans les autres formats sonores.* »

--- SOUS THÈME 2 : L'ORIENTATION ---

>>> Concept : La scène implique une orientation

Dans un contexte de diffusion, il existe encore, dans la grande majorité des cas, une scène qui implique qu'on la regarde et induit ainsi une orientation de l'audience. Cette directionnalité a été soulignée et critiquée lors de 3 entretiens.

- AM** « *Il y a quand même une directionnalité de l'espace puisque t'as quand même une sorte de scène. Alors on peut discuter de comment est-ce qu'il faudrait la mettre et tout mais en vrai cette directionnalité elle fait que tu as un devant et que tu as un arrière.* »
- XG** « *On ne sortirait pas tout de suite du théâtre à l'italienne et spontanément je ne me suis pas assis en tournant le dos.* »
- GS** « *Dans notre studio, tout doit être mobile pour que tu n'aies pas de sensation d'un avant ou d'un arrière. Tous les contenus immersifs*

⁵⁵ Microphone ambisonique.

⁵⁶ « *You need to have those additional microphone signal for sure, but I never did a remix of a stereo recording which was recorded with just stereo as the main microphone array.* »

ennuyants qui sont produits en ce moments considèrent toujours qu'il y a une scène devant toi. Il n'y aura pas de changements dans la musique, artistiquement, esthétiquement, si tu penses toujours que la scène est devant toi. »⁵⁷

--- SOUS THÈME 3 : LA VUE ---

>>> Concept : La nécessité de ne pas avoir de stimulations visuelles

À l'instar d'une scène que le public a envie de regarder, l'idée d'avoir des stimulations visuelles lors d'une situation d'écoute n'est pour certains pas souhaitable. Il s'agit justement de se concentrer uniquement sur le contenu sonore.

AP « *En fait le principal truc qui résume tout c'est faire disparaître le haut-parleur, parce qu'à partir où la technique disparaît, c'est les émotions qui priment. »*

GS « *Tu peux te rendre compte que quelle que soit la situation où tu places des stimuli visuels contre ou en parallèle à de l'audio spatial, les yeux vont focaliser, les oreilles vont se focaliser. C'est le défi le plus important.* »⁵⁸

WR « *J'ai la sensation que quand les lumières sont éteintes, [...] tu ressens véritablement avec ton corps, et tu es transporté d'une toute autre manière.* »⁵⁹

5.2.3 Thème 3 : L'adaptabilité

--- SOUS THÈME 1 : SE DONNER LES MOYENS TECHNIQUES ---

Le recours à des technologies développées soi-même a été mentionné par la quasi-totalité des personnes interviewées, qu'elles soient physiques ou informatiques. C'est une constante qui semble être particulièrement importante dans un contexte où les outils ne sont visiblement pas encore suffisamment répandus ou adaptés au contexte de production et de création (voir le thème suivant 5.2.4). Cela va de pair avec la nécessité d'avoir une très bonne connaissance des avancées technologiques du milieu de l'audio spatial.

⁵⁷ « *In our studio everything has to me moveable, so that you don't get the feeling there's a front, there's a back. The boring spatial audio stuff that is happening now is still assuming that stage is in front of you. There's no changing in music, artistically, esthetically is going to happen if you still think the stage is in front of you.* » Notre traduction.

⁵⁸ « *You can learn wherever you set visual stimulus against or parallel Spatial Sound, eyes will focus, ears will focus. That's the highest Challenge.* » Notre traduction.

⁵⁹ « *I felt like when you don't have the lights on [...] you're really feeling with your body, and you're completely transported in a different way.* » Notre traduction.

>>> Concept 1 : Développer ses propres outils physiques

Nous avons choisi de faire entrer dans ce concept, qui regroupe enceintes et instruments, l'idée de *workflow*, car il n'entrait pas véritablement pour nous dans la case du logiciel et restait important à mentionner dans ce sous-thème.

- AP** « *Je concevais des systèmes de live pour des projets, le système peut être complètement disparate et pas du tout homogène.* »
- MA** « *Typiquement Xenakis il avait fait des instruments enfin il avait plutôt fait des outils de spatialisation tout ça.* »
- HD** « *[Jean Michel Jarre] a un truc que beaucoup n'ont pas dans la musique classique c'est que lui il fait sa lutherie lui-même quoi c'est que lui il invente ses instruments.* »
- LW** « *J'ai développé un workflow pour le mixage immersif orienté canal en live.* »⁶⁰

>>> Concept 2 : Développer ses propres outils informatiques

Une grande partie de la production audio est aujourd'hui réalisée sur logiciel. Il est donc tout aussi important d'être capable de se fabriquer ses propres outils informatiques pour compléter les performances des outils classiques.

- AP** « *Je concevais des systèmes de live pour des projets, le système peut être complètement disparate et pas du tout homogène.* »
- MA** « *En réalité j'ai pas utilisé les panneaux de Pro Tools j'ai installé le Spat pour pouvoir faire mes rotations de scène.* »
- ER** « *J'ai compris très vite qu'il fallait construire ses propres outils.* »
- AP** « *Je m'étais créé pas mal d'automatismes pour faire bouger les sons [...] qui se déplacent de manière complètement aléatoire.* »
- HD** « *Associé à l'époque au patch Usine qui aurai pu être un patch Max ça marche d'enfer.* »
- WR** « *Nous avons ce device Max4Live qui le meilleur moyen de se plonger dans le dispositif.* »⁶¹

>>> Concept 3 : Primordial d'être irréprochable en technique

L'évolution des technologies est telle qu'il est nécessaire de se tenir au courant à l'aide d'une veille technologique. Cela permet d'une part de produire de manière innovante, et d'autre part, dans le processus créatif, de savoir vers quelle technologie s'orienter en fonction de son besoin et de ne plus se poser la question de la technique.

⁶⁰ « *I developed a workflow for live mixing channel based immersive mix.* » Notre traduction.

⁶¹ « *We have this Max4Live device which is the best easiest accessible way to kind of jump in.* » Notre traduction.

- AP** « *C'est de maîtriser suffisamment la technique pour que quand c'est dans un processus créatif je l'oublie complètement. Dans un processus créatif la technique m'emmerde.* »
- ER** « *C'est sûr que l'apprentissage des outils il doit être global et en fait il faut avoir une connaissance presque holistique quoi, total des possibilités des outils.* »
- XG** « *Dès qu'il y a un outil nouveau qui rentre sur le marché on l'intègre, enfin on l'apprend et on fabrique quelque chose avec, donc on rentre dans la chaîne de production tout le temps.* »
- LW** « *Ça change constamment donc tu dois d'adapter quotidiennement.* »⁶²

--- SOUS THÈME 2 : S'ADAPTER À DES LACUNES ---

>>> Concept : La non-optimisation des mouvements verticaux

Si les technologies immersives sont de plus en plus nombreuses et performantes, il existe encore certains défauts auxquels il faut s'adapter, notamment la gestion des mouvements vers le haut.

- AP** « *Les mouvements vers le haut, c'est mal maîtrisé aujourd'hui.* »
- HD** « *Et si j'ai besoin de faire des mouvements de hauteur, [...] je vais pas utiliser les panoramiques, parce que ça marche pas toujours ça.* »
- HD** « *Je reste dans le haut et je ne fais pas de passage bas-haut parce que à cause de cette triangulation il y a trop de danger.* »

5.2.4 Thème 4 : L'accessibilité

--- SOUS THÈME : L'ACCÈS AUX Outils ---

>>> Concept 1 : Des problématiques économiques

L'audio spatial demande un temps de travail supplémentaire qui est coûteux, si ce n'est pas la technologie elle-même qui impose des frais supplémentaires. Cela freine ainsi son adoption par certaines productions.

- LW** « *Bien sûr c'est encore plus cher de faire un mix immersif. Tu as toujours besoin d'un mix stéréo, le mix immersif est toujours additionnel.* »⁶³
- OG** « *Il n'y a pas encore ce dynamisme des productions d'investir dans le Dolby Atmos parce que quand même ça représente un surcoût qui est pas tout à fait négligeable.* »

⁶² « *It's always changing so you need to adapt in the daily bases.* » Notre traduction.

⁶³ « *Of course it's more expensive to do an immersive mix. You always need a stereo mix, immersive mix is always additional to that.* » Notre traduction.

HD « Écrire sur le plan spatial ça coute du pognon. Et faire des œuvres qui soient vraiment immersives [...], ça va couter beaucoup de pognon. »

>>> Concept 2 : Les outils globaux existent

Utilisé dans la musique contemporaine depuis le milieu du vingtième siècle, l'audio spatial fait preuve de dizaines d'années de développement qui ont donné naissance à un grand nombre d'outils.

MA « C'est que les formats ils sont en train de se mettre en place. [...] Ce que je veux dire c'est que le format et les outils sont là. »

HD « Et aujourd'hui, grâce à la puissance de l'ordinateur, il y a juste à bosser en fait, il y a juste à travailler. Là on a des outils. »

GS « Il y a tellement d'outils dans la nature. »⁶⁴

>>> Concept 4 : Le besoin d'outils adaptés à la création

Pour certains, les outils à disposition ne sont pas adaptés à un processus artistique. Ils sont trop complexes et s'adressent à des ingénieurs plutôt qu'à des créateurs. Pour ces personnes-là, c'est une problématique majeure qui doit être résolue.

MA « Un artiste aujourd'hui quand il mixe en stéréo il sait ce que c'est la stéréo. [...] Ça va finir par être pareil en son immersif. [...] Simplement il leur [...] faut déjà les outils. »

AM « Les créateurs ils ont pas les outils pour et il manque un canal vraiment simple pour diffuser ça. »

AP « Les outils simples n'existent pas et c'est vrai aujourd'hui, hormis le binaural, t'as aucun moyen pour un musicien dans sa chambre à couper de mixer en immersif. »

ER « C'est pas réaliste par rapport à la façon de travailler d'un gars qui peut être dans son train, qui va travailler tiens là j'ai 30 min je vais me générer une petite séquence, explorer aussi ce qu'on peut faire au niveau sonore, créer des trucs, avoir un outil flexible qui permet d'explorer ce qu'on peut faire. »

ER « Effectivement il y a des outils qui existent [...] mais dans la communauté artistique il y avait personne. »

⁶⁴ « There's so many tools out there. » Notre traduction.

5.3 Catégorie 3 : La philosophie

5.3.1 Thème 1 : L'immersion

--- SOUS THÈME 1 : LES ORIGINES DE L'IMMERSION ---

>>> Concept 1 : La notion d'immersion narrative et la lecture

Nos entretiens nous ont permis de distinguer deux types d'immersions. La première est l'immersion narrative, qui peut arriver dans différentes situations, notamment la lecture. La seconde est l'immersion perceptive, que nous verrons dans le prochain sous-thème.

- | | |
|-----------|---|
| HD | <i>« Ce même phénomène [d'immersion narrative] explose avec le livre. [...] Et le livre, l'explosion, du livre elle est là quoi. »</i> |
| GS | <i>« L'immersion était quelque chose qui opérait à la lecture d'un livre. »⁶⁵</i> |
| AD | <i>« Il y a des livres qui peuvent être immersifs, qui sont créés pour te faire perdre la notion de temps et d'espace et qui font que tu es complètement perdu dans l'œuvre. »⁶⁶</i> |

>>> Concept 2 : L'immersion au cinéma

C'est par le cinéma que l'immersion devient accessible au grand public. Deux innovations différentes sont évoquées ici, l'une relative à la vue, l'autre à l'ouïe. Hervé Déjardin parle déjà d'immersion perceptive.

- | | |
|-----------|--|
| XG | <i>« Parce que le multicanal est arrivé par le cinéma. [...] C'est le cinéma qui a commencé à faire venir des sons qui sont à un autre endroit d'émission que l'écran. »</i> |
| HD | <i>« Dans l'histoire des médias le premier qui va inventer l'immersion perceptive c'est le cinéma par le noir salle. »</i> |

--- SOUS THÈME 2 : DES DEFINITIONS DE L'IMMERSION ---

>>> Concept 1 : La notion d'immersion perceptive

La notion d'immersion perceptive est ici appréhendée à travers le prisme de la stéréo, considéré comme le premier dispositif spatialisé. On notera cependant l'affirmation d'Hervé Déjardin qui va plus loin.

- | | |
|-----------|--|
| HD | <i>« L'immersion perceptive au niveau du son c'est vraiment à partir du moment où la stéréo... *hésitation* Pour moi la réelle immersion c'est</i> |
|-----------|--|

⁶⁵ « Immersiveness used to be something that happened when you read a book. » Notre traduction

⁶⁶ « There are books that can be immersive, that are created to make you lose this sense of space and time and you are completely lost in the piece. » Notre traduction

à partir du moment où on commence à mettre des HP tout autour de l'auditeur [...] Tu es au centre des choses, et là ça y est tu es immergé. »

OG « *L'immersion perceptive au niveau du son c'est vraiment à partir du moment où la stéréo, oui ça commence à l'être, parce qu'on a une scène sonore à 60° sur le plan frontal. »*

>>> Concept 2 : Reproduire la nature et la réalité

Les définitions de l'immersion et de l'audio spatial font souvent référence à la nature et font appel à la notion de « réalité », qui doit être reproduite. Plus on arrive à reproduire les sensations que l'on a dans une situation d'écoute naturelle, plus le système est considéré comme immersif.

ER « *L'audio spatial c'est dehors dans la rue en fait, [...] c'est tenter de reproduire des phénomènes naturels de perceptions sonores en environnement. »*

AP « *Le principe du son immersif c'est que tu recrées les conditions réelles d'écoute. [...] C'est en fait d'avoir un système qui peut recréer des conditions physiques d'une réalité. »*

AP « *Justement ce qui m'a intéressé d'être ingénieur pour pouvoir faire de l'ingénierie du son, c'est pour avoir des systèmes qui se comporte acoustiquement comme dans la réalité. »*

XG « *Le multicanal c'est pas pour aller chercher du réalisme, c'est pour amener du réel. Pour moi aller chercher le réalisme c'est inventer quelque chose qui aurait dû être, non le multicanal c'est ramener ce qui a été. Le son 3D c'est ramener ce qui a été. »*

XG « *L'audio 3D, ça nous permet de rentrer le réalisme de l'endroit où les gens parlent. Et comme on parle du principe nous que les gens ne parlent pas de la même façon en fonction de l'endroit où ils se trouvent ça permet de rentrer dans ce réalisme-là. »*

>>> Concept 3 : La notion d'environnement

Nous notons également que le terme « environnement » est apparu à plusieurs reprise, comme un objectif de reproduction à atteindre.

ER « *L'audio spatial c'est être capable de créer un environnement. »*

AP « *L'immersif c'est donner l'impression d'être immergé dans un environnement. »*

>>> Concept 4 : L'immersion n'est pas forcément 3D

Pour conclure ce thème sur l'immersion, il convient de souligner les remarques d'Éric Raynaud et d'Antoine Petroff qui affirment que l'immersion n'est pas nécessairement liée à un système de reproduction 3D. Éric Raynaud évoque d'ailleurs une problématique de vocabulaire, qui nous ramène à ce que nous verrons en 5.3.3.

ER « *L'audio spatial il n'y a pas ... ça peut être du 3D, ça peut être quelqu'un avec 2 haut-parleurs qui joue avec les réflexions de l'espace, il y a pas une définition claire.* »

AP « *Immersif n'est pas forcément lié à la 3D. En fait la 3D participe à l'immersion mais il y a plein de facteurs qui participent à l'immersion.* »

5.3.2 Thème 2 : L'écriture

-- SOUS THÈME 1 : DEPLACEMENT DES DIFFICULTES --

>>> Concept 1 : L'écriture immersive est plus facile

Il apparaît que l'écriture sonore immersive peut être plus facile que l'écriture stéréophonique, notamment grâce à la discrimination plus précise des différentes sources sonores.

AP « *Ouai c'est incomparable, vraiment incomparable.* »

OG « *Par rapport à d'autres système c'est que ça devient presque facile* »

OG « *Quelque part c'est plus facile dans un format dit immersif parce que tu peux plus répartir les choses dans l'espace en fait.* »

>>> Concept 2 : Des difficultés à remplir l'espace

Si le fait de se confronter à un espace à 360° semble ajouter de la facilité à certains, deux personnes ont clairement identifié des difficultés à remplir l'espace, ce qui rend le mixage immersif plus compliqué que le mixage stéréophonique selon eux. Cela serait le cas dans un contexte où l'arrangement ne comporterait pas beaucoup de pistes.

MA « *J'ai envie de te dire oui oui bien sûr mais en fait je pense pas que ce soit plus facile. Je pense que c'est même plus dur. [...] Il faut tout faire en mix parce qu'il y a pas de pistes pour le faire.* »

HD « *Si tu prends quelqu'un qui écrit de manière simple en stéréo, faire de l'espace ça peut pas être plus simple ça peut être plus compliqué parce que t'as pas assez de matière.* »

-- SOUS THÈME 2 : UNE PRATIQUE ARTISTIQUE --

>>> Concept 1 : Utiliser l'espace dans l'acte créatif

Il apparaît très largement que l'espace doit être considéré comme un lieu d'expression et de création.

ER « *L'audio spatial c'est un endroit où il faut laisser exprimer son imagination, filtrer des sons, créer des objets [...] Donc on se rend bien compte que quand on fait ça bah ça doit être une pratique artistique. C'est du geste artistique en fait.* »

- AM** « Je pense que le travail de l'espace sonore c'est un espace expressif en soi et ce n'est pas une extension de la stéréo c'est un autre turc. »
- OG** « Tout d'un coup t'as un potentiel plus grand de création technique et artistique avec le son. »
- HD** « Voilà à un moment tu rentres dans une cathédrale, et après tu rentres dans une bagnole hyper cossue. [...] Ça aussi c'est un acte de création. »
- WR** « On explore les possibilités qu'offre l'audio spatial en termes de composition et d'expression comme un medium artistique. »⁶⁷

>>> Concept 2 : La notion d'instrument spatial

Afin de permettre la création et de libérer l'expression, la notion d' « instrument » a été plusieurs fois employée. Il est reproché de ne pas avoir suffisamment d'outils qui se rapprochent véritablement de l'instrument et qui permettraient de composer facilement, comme nous avons pu le voir en 5.2.4 dans le sous-thème « Accessibilité ».

- GS** « Toutes ces machines que nous avons entre les mains aujourd'hui sont paramétrés comme des machines et non comme des instruments. »⁶⁸
- WR** « J'ai recherché pendant des années quelque chose qui puisse être, en plus d'être un système de diffusion, un instrument, qui te permette de véritablement contrôler les sons dans l'espace. [...] Parce qu'il ne s'agit pas seulement du système, il s'agit du fait que ce soit un outil. [...] C'est comme un instrument spatial. »⁶⁹
- MA** « Je fais une prise de son, puis je la spatialise. Depuis la nuit des temps on fait comme ça. Et c'était tout l'objet de mon mémoire, c'était de créer un instrument qui fait les deux en même temps. »

-- SOUS THÈME 3 : VERS UNE NOUVELLE FORME D'ECRITURE --

>>> Concept 1 : L'espace, un critère à part entière

Il apparaît important de considérer l'espace comme un attribut à part entière et de ne pas le placer au pied de la hiérarchisation des critères à appréhender lors d'une création.

⁶⁷ « We're exploring the compositional and the expressive possibilities of spatial sound like an artistic medium. » Notre traduction.

⁶⁸ « All these machines that we have at hand now are addressed like machines not like instruments. » Notre traduction.

⁶⁹ « I was looking for years for something that could not just be like a sound system but being an instrument, you know when you're really able to control sounds in space. [...] Because it's not just the system it's the fact that it's a tool, like an instrument. [...] It's like a spatial synthesizer. » Notre traduction

- HD** « *Va falloir que [les artistes] sachent travailler parfaitement l'harmonie, la polyphonie, la rythmique, le matériau sonore, l'esprit Schaefferien, plus la spatialisation.* »
- MA** « *C'est ce que Stockhausen disait, c'est une donnée aussi importante que la hauteur ou le rythme.* »
- WR** « *En Dolby, les gens ne mixent pas en pensant l'espace comme un paramètre aussi important que le timbre ou le pitch. C'est justement ce que nous faisons ici, on le pense comme faisant partie du processus de composition.* »⁷⁰

>>> Concept 2 : L'immersif définit une nouvelle forme d'écriture

Certains pensent que c'est à travers l'immersion que se fera l'écriture, et qu'elle se fera d'une nouvelle manière qui intégrera complètement la pensée de l'espace.

- HD** « *Sur les nouvelles formes d'écriture, on ne va pas parler de l'immersion en fait on s'en fout, c'est un tout, l'immersion fait partie, va faire partie intégrante de l'écriture musicale.* »
- XG** « *Bah oui non non ça peut être très bien ce n'est pas dévalorisant [de travailler en stéréo, ndlr] mais j'ai tellement moins de place pour écrire [qu'en audio spatial, ndlr].* »

-- SOUS THÈME 4 : LIBERTE ESTHETIQUE --

>>> Concept 1 : Le contenu guide l'esthétique

En stéréo, un contenu de musique classique ne sera pas appréhendé de la même manière qu'un disque de pop. Similairement, en son immersif, la manière de spatialiser – ou de ne pas le faire – dépend finalement et en premier lieu du contenu.

- MA** « *En fait je pense surtout que ça va dépendre du répertoire qu'ils vont choisir.* »
- AP** « *C'est "format en fonction du projet".* »
- ER** « *Il y a aujourd'hui une forme de musique que j'ai toujours envie d'apprécier en stéréo.* »
- AM** « *Et finalement après la position moi je m'en fou, enfin pour pas mal de choses je m'en fou. Après ça dépend de ce qu'on fait.* »
- OG** « *Il y a des scènes qui se prête très bien à ça dont la spatialisation devient un vrai enjeu et sert la scène et d'autres ou ça sert vraiment à rien parce que vis-à-vis du rapport à l'objet cinématographique il y a pas d'enjeu quelque part.* »

⁷⁰ « *They're not mixing it, in Dolby, thinking about space as something as important as timbre and pitch. And these are those things that we do here, is think about it as part of the compositional process.* » Notre traduction

>>> Concept 2 : Des règles pour être bafouées

Bien que notre dernier concept laisse sous-entendre que la liberté dans l'expression sonore serait bridée par le contenu, il est plusieurs fois rappelé que les règles, si elles existent, sont là pour être bafouées, et qu'il faut laisser libre cours à ses envies.

- ER** « *Mais designer un environnement sonore j'ai envie de dire qu'il y a pas de règles.* »
- AP** « *Après toutes les règles sont... *rires* ça c'est l'autre truc que le contrebassiste m'a dit toutes les règles sont là pour être...* »
- AP** « *Moi je ne mets jamais de règles, de nouveau rien à foutre.* »
- MA** « *Il y a pas de règles.* »
- OG** « *Disons qu'après moi je suis prêt à tout si tu veux ça dépend je n'ai pas forcément de limite mais il faut vraiment que ça s'y prête.* »

>>> Concept 3 : Ne pas nécessairement utiliser toutes les possibilités d'un système.

Selon Gerriet Sharma et Martin Antiphon, être libre de son esthétique spatiale, c'est aussi être conscient qu'il n'est pas forcément nécessaire d'utiliser toutes les capacités d'un système.

- GS** « *Tu leur donnes 24 haut-parleurs et au bout de 2 semaines ils n'en utilisent que 2. Voilà le résultat. Et ce n'est pas "ne pas avoir compris", c'est précisément avoir compris.* »
- MA** « *À la limite ça pourrait être ta liberté, à termes, d'utiliser ce format spatial pour construire un mixage qui serait que frontal.* »
- MA** « *Ils ne mettent rien dans le plafond, justement ils font un mix 7.1. C'est leur choix.* »

5.3.3 Thème 3 : La culture

--- SOUS THÈME 1 : LA CULTURE DES ARTISTES ---

>>> Concept 1 : Un accès à la culture limité pour les artistes

Comme le souligne parfaitement Augustin Muller, si la culture de l'audio spatial existe dans un monde de niche qu'est celui de la musique électroacoustique, elle est encore largement méconnue des artistes.

- GS** « *On a un énorme problème de transfert de connaissances.* »⁷¹
- AM** « *Comme il n'y a pas ça, toute cette culture là pour le son immersif, du coup les gens tâtonnent un peu plus.* »

⁷¹ « *We have a huge knowledge transfer problem.* » Notre traduction.

- MA** « *Mais c'est en train d'arriver, tout ça ça avance donc en fait les artistes vont commencer à se familiariser avec l'espace. »*
- XG** « *En fait ça c'est parce que les gens ont besoin de s'acculturer avec l'objet. »*

>>> Concept 2 : Un manque de propositions immersives de la part des artistes

Certainement en conséquence directe du concept précédent, le manque de culture des artistes implique un manque de propositions spatiales de leur part.

- MA** « *Des artistes qui sont arrivés avec des choses très précises moi en tout cas pour du mixage musique Dolby Atmos bah je n'en ai pas eu. [...] Ça m'est arrivé qu'une fois et c'était avant le Dolby Atmos. *rires* »*
- AM** « *Il n'a pas de conception de l'espace au départ de sa performance, donc c'est hyper lent et hyper faible le niveau ou on arrive. »*
- XG** « *Ils sont arrivés en disant "faites-nous des propositions". »*
- HD** « *- [La spatialisation] n'est pas encore suffisamment appréhendée par les artistes, si ? - Non elle l'est pas du tout. »*

>>> Concept 3 : Nécessité d'accompagner les artistes

Afin de pallier ces manques, certaines structures organisent des résidences et invitent des artistes à questionner leurs relations à l'espace, comme le font le *Spæs Lab* et *Monom* (voir Annexe 8 p.118) Nous aimeraisons souligner ici que Gerriett Sharma et William Russell emploient tous les deux et spontanément le verbe « to care ».

- GS** « *A ce moment-là on en parle, et ce que je souhaite dire c'est qu'on essaye d'être très attentifs aux artistes. »⁷²*
- WR** « *Comparé à d'autres, on a un objectif assez différent ici, qui est véritablement d'aider les gens à comprendre qu'il y a une manière différente de vivre l'art, la musique, les arts sonores particulièrement. Et ça demande juste un petit peu plus d'attention. »⁷³*

>>> Concept 4 : Acculturer les artistes par l'écoute

Ces trois prochaines citations sont à comprendre dans un cadre d'accompagnement d'artistes par des spécialistes du son immersif. Il est important, dans le processus créatif et afin d'outrepasser le problème de culture évoqué par le concept précédent, de faire écouter des contenus aux personnes qui vont travailler les outils de spatialisation.

- AM** « *Il y a aussi une question de culture et d'écoute de ça. Tu vois par exemple un truc que je n'ai pas fait c'est que je ne lui ai pas fait écouter*

⁷² « *But then we talked about it and what I want to say is we take a lot of care. »* Notre traduction.

⁷³ « *So we have a much different goal here than others which is really to help people understand there's a different kind of way you can experience art, music, sound art especially. And that just requires a little bit more care. »* Notre traduction.

- de truc. Et en fait en vrai les gens ils n'écoutent jamais. [...] Donc je pense qu'il y a une question de manque de connaissance. »*
- WR** « *On commence par écouter donc on a écouté différentes pièces, pour commencer à ouvrir les oreilles. »⁷⁴*
- GS** « *Écoute simplement Chowning, Xenakis, Stockhausen. On peut apprendre. »⁷⁵*

>>> Concept 5 : L'importance du partage de connaissances

Dans le même contexte d'accompagnement que celui évoqué précédemment, l'idée de travail collectif est fortement mise en valeur. Il est défendu que chacun peut apprendre de l'autre, ce qui fera avancer la grammaire. Nous pouvons rapprocher ce concept de celui d'archivage (voir 5.3.5).

- WR** « *A chaque fois qu'un artiste vient, l'interface deviens de plus en plus... Elle intègre de nouveaux processus, des nouvelles idées, des nouvelles idées de compositions. »⁷⁶*
- GS** « *Finalement, la plupart des résidents laissent quelque chose. Ça peut être un patch ou une pièce. Ou un texte. Ce qu'on veut vraiment c'est... On veut une sorte de communauté et ça fait maintenant plusieurs années qu'on est dans le processus et ça a l'air de marcher. »⁷⁷*

--- SOUS THÈME 2 : LA RECEPTION DU PUBLIC ---

>>> Concept : Un public toujours conquis

D'une manière générale, il apparait que le grand public est toujours particulièrement réceptif au son immersif.

- AP** « *Je n'ai jamais eu d'expérience du public qui disait "mince ce n'est pas comme d'habitude, c'était moins bien". Jamais. Alors vraiment au contraire, toujours c'est "waouh". »*
- AP** « *Elle avait environ 10 ans, elle m'a dit "écoute ça on écoute tous ça à l'école" et tout, c'était le fameux truc, les 8D machin. [...] Elle a entendu tout de suite la différence. »*
- XG** « *Et c'est là que pour moi il y a quelque chose. Et cet élément se rencontre avec le public. [...] On a eu des réflexions d'auditeurs classiques*

⁷⁴ « *We first start with listening so, we kind of listened to different pieces, different styles, start to kind of open up the ears. »* Notre traduction.

⁷⁵ « *Only listen to Chowning, Xenakis, Stockhausen. We can learn. »*

⁷⁶ « *And every time an artist comes in, the interface becomes more and more... it integrates new workflows, and new ideas, and this new compositional idea. »* Notre traduction.

⁷⁷ « *Most of the residents actually leave something. Like a patch, or a piece. Or a written text. So, what we really want is, we want a kind of community and that's now for years into the process and it seems to work. »* Notre traduction.

qui ont déjà eu cette notion de confort et qui ont trouvé le moyen de l'exprimer. »

- OG** « *Les spectateurs qui ont eu la chance de le voir en Atmos ils en ont quand même pris plein les oreilles. »*

--- SOUS THÈME 3 : VOCABULAIRE ---

>>> Concept : Un problème de vocabulaire

Comme nous l'avons évoqué précédemment (voir le sous-thème « Des définitions de l'immersion » en 5.3.1) à travers le témoignage d'Eric Raynaud, les définitions du terme « immersif » ne semblent pas claires. Un problème de vocabulaire et d'emplois inappropriés a été souligné par beaucoup.

- XG** « *On était tous avec un vocabulaire extrêmement pauvre. »*
- XG** « *Je n'ai pas le vocabulaire, je n'ai pas la méthode d'ordonnancement des notions que je dois appliquer à ce que je ressens. »*
- GS** « *Encore une fois, il y a tellement de mots-valises qui circulent. »⁷⁸*
- AD** « *Je pense que c'est encore un terme nouveau, et on n'a pas beaucoup de vocabulaire pour tout ça, on essaye toujours de définir ce que c'est exactement [que l'immersion, ndlr]. »⁷⁹*
- AM** « *Alors sur le mot immersif j'ai des réserves sur ce mot là c'est le mot un peu valise. »*
- AM** « *En fait ce mot-là [immersif, ndlr] il est un peu... [...] Mais je suis d'accord qu'on en a pas vraiment d'autres hein c'est pour ça. »*

5.3.4 Thème 4 : La subjectivité

--- SOUS THÈME 1 : HUMILITE ---

>>> Concept : « Je ne sais pas » Les acteurs de l'immersif conscients de leurs explorations

Nous soulignons à travers ce sous-thème que plusieurs personnes avec qui nous nous sommes entretenues ont eu une réelle conscience de leurs explorations et qu'elles n'associent pas de vérité absolue à leur pratique. Cela s'est notamment manifesté à travers l'utilisation fréquente du « je ne sais pas ».

- MA** « *Je sais pas » - 5 fois*
- MA** « *Mais est-ce que c'est bien j'en sais rien. »*

⁷⁸ « *Then, again there's so much bus words now around. »* Notre traduction.

⁷⁹ « *I think it's also kind of like a new term, and we don't have a lot of, like the wording for it we're still trying to define what it is exactly to be able to break it down and then catalogize things better. »* Notre traduction

- MA** « Peut-être que je me trompe mais... » 2 fois,
MA « Je dis pas que c'est ce qu'il faut faire... »
AP « Je sais pas » - 2 fois
AM « Je sais pas » - 2 fois
LW « Mais tu peux complètement me prouver que j'ai tort. »⁸⁰

--- SOUS THÈME 2 : UN REGARD CRITIQUE SUR LA PRODUCTION ---

>>> Concept : Les acteurs de l'immersif souvent déçus

Il nous semble pertinent de relever les critiques des sujets vis-à-vis de certains contenus immersifs. Cela remet en question certains éléments de grammaire qui ont été employés par certains et non appréciés par d'autres.

- MA** « Je dirais qu'il y a que 50% des productions qui sont ratées. *rires* et il y en a qui sont vraiment ratées. [...] L'album *Killing The Name* en *Atmos* c'est une catastrophe, mais vraiment c'est une catastrophe. »
AP « T'as plein de merdes aussi, enfin tu vois Sony qui avait ressorti *Space Odity* en spatialisé, je ne sais pas qui a fait ce truc là mais ce n'est pas possible. »
ER « Il faut y aller quoi, il faut tenter des trucs quoi moi je suis super déçu moi de voir des trucs... »
AM « Il y a un aspect naïf d'utilisation de la spatialisation [...] c'est pas spatialement plus intéressant que s'ils l'avaient fait en stéréo. »
GS « Beaucoup de choses qui se vantent d'être nouvelles aujourd'hui sont en réalité absolument incapables de s'approcher de plus de 10% de la qualité de ce qu'ils ont fait à l'époque avec de la bande, des haut-parleurs et des consoles de mixage. »⁸¹

--- SOUS THÈME 3 : EXPERIENCE PERSONNELLE ---

>>> Concept : Une conscience de la pratique empirique personnelle

Au-delà de la conscience d'une certaine humilité, nous constatons également que les sujets ont conscience que leurs expériences personnelles jouent intimement sur leurs pratiques et qu'elles sont très personnelles. Cela a nécessairement un impact sur la grammaire employée.

- MA** « Mais moi j'ai constaté quelque chose à titre personnel qui est je pense n'est pas anodin et je l'ai constaté il n'y a pas si longtemps que ça. [...] Mais ça c'est quelque chose que j'ai constaté personnellement. »

⁸⁰ « But you can definitely prove me wrong. » Notre traduction.

⁸¹ « Lots of things which are praised to be new at the moment are actually absolutely not capable of being, becoming 10% of the quality what they did with analog tape, and loudspeakers and mixing desks back then. » Notre traduction.

- AM** « *Mais de mon point de vue, c'est très personnel.* »
- AP** « *Ça je pense, à force par l'expérience.* »
- LW** « *Encore une fois ce sont mes habitudes empiriques. Il s'avère que je préfère ces paramètres, mais tu peux complètement me prouver que j'ai tort.* »⁸²

5.3.5 Thème 5 : L'histoire

-- SOUS THÈME 1 : LA MUSIQUE ELECTROACOUSTIQUE --

>>> Concept 1 : L'importance de la culture électroacoustique

Lors de nos entretiens, l'importance de la culture électroacoustique a été largement soulignée. Il est précisé par les sujets que la grammaire de l'écriture spatialisée a encore beaucoup à apprendre de ce catalogue.

- MA** « *Tu es allé voir tout ce qui était travaux de musique contemporaine, de Stockhausen et de ces gens-là ?* »
- AM** « *Si tu regardes Pierre Henry tu vois il faisait pas du surround spatialisé il jouait énormément avec l'espace et les différentes façon d'exciter l'espace.* »
- AM** « *Je pense que les compositeurs traditionnels de la musique électronique ont quand même répondu à pas mal de questions en fait.* »
- AM** « *Donc si tu veux, par exemple cette réflexion sur le sens d'un lieu, l'acoustique, en lien avec la position, c'est un truc qui a été traité, qui existe, donc je pense qu'il y a des stratégies qui sont là mais c'est vrai qu'elles ne sont pas explicitées nulle part.* »
- HD** « *Je lui ai fait découvrir la musique électroacoustique et l'influence que pouvait avoir des sons sur le rendu global d'une œuvre.* »
- GS** « *Ça commence avec Stockhausen et Xenakis. Bien sûr nous avons Chowning. [...] Parce que nous voulons que les gens soient capables d'apprendre.* »⁸³

⁸² « *But again, that's my empirical tips. It turned out that I like those settings the most, but you can definitely prove me wrong.* » Notre traduction.

⁸³ « *It starts with Stockhausen, and Xenakis. Of course we have Chowning, [...] Because we want people to be able to learn.* » Notre traduction.

>>> Concept 2 : L'importance d'archiver

Pour ne pas oublier ce qui a été expérimenté et en lien direct avec le concept précédent, nous soulignons la forte pertinence de l'archivage qui permet un transfert de connaissances aux générations ultérieures. C'est également une large source d'inspiration qui ne doit pas être négligée.

- | | |
|-----------|---|
| GS | <i>« Nous avons cette archive de pièces multicanales des 50 dernières années [...] et elles sont parfaitement prêtes à être jouées. »</i> ⁸⁴ |
| WR | <i>« Nous avons cette espèce d'énorme archive de travaux qui sont prêts à être joués. »</i> ⁸⁵ |

--- SOUS THÈME 2 : UN DEVELOPPEMENT HISTORIQUE --->>> Concept 1 : L'histoire s'écrit

Il apparaît que la situation actuelle de l'audio spatial est similaire à celle de la stéréo à ses débuts. Nous serions en train de vivre un tournant dans l'histoire de l'audio.

- | | |
|-----------|--|
| HD | <i>« Donc c'est une étape de l'histoire de la distribution du son. [...] C'est une page qui est en train de se tourner mais elle n'est pas tournée encore pour l'instant. »</i> |
| MA | <i>« La question de faire du son immersif, l'enjeu de l'écriture spatiale, ce n'est pas [...] que c'est obligatoire ou que c'est dans l'aire du temps, c'est que c'est quelque chose qui est en train d'arriver. [...] C'est le chemin de l'histoire, voilà. »</i> |
| XG | <i>« Le public n'a jamais demandé la stéréo, et pourtant les mecs se faisait chier à se dire comment on pourrait mettre deux canaux alors que tout le système de diffusion était mono et ça nécessitait un truc, bon, on en est pour moi à la même période. »</i> |

>>> Concept 2 : L'audio spatial se cherche encore

C'est dans cette période de transition qu'ont lieu les expérimentations, les propositions. L'audio spatial cherche encore sa voie.

- | | |
|-----------|---|
| OG | <i>« Ça montre le marasme dans lequel on est. »</i> |
| OG | <i>« C'est encore au balbutiement. »</i> |
| MA | <i>« Et il y a des trucs qui sont des propositions, des essais. »</i> |
| AM | <i>« Les gens tâtonnent un peu plus. »</i> |

⁸⁴ « We have this archive of multichannel pieces from the past fifteen years [...] and there's all perfectly fit for being played. » Notre traduction.

⁸⁵ « We have this huge sort of archive of works which are ready to play. » Notre traduction.

Discussion

Notre analyse thématique nous a permis de lister un nombre conséquent de « règles de grammaire » de l'écriture sonore immersive que nous pouvons mettre en relation avec les problématiques que nous avons évoquées en 4.1.1.

Tout d'abord, nous remarquons que les personnes interviewées affichent une bonne connaissance des principes de psychoacoustique présentés en première partie. Ils appliquent ainsi la théorie à leur écriture. Cela se manifeste notamment dans la gestion de l'élévation, dans laquelle il ne sera pas question de diffuser des basses fréquences.

Contrairement aux contradictions que nous avions soulevées en 3.3.5, concernant la perception de la qualité globale d'un système en fonction de la présence ou non de haut-parleurs en élévation, il apparaît dans l'analyse que les *tops* apportent une véritable plus-value à l'image sonore. Avec les arrières, et dans un contexte de frontalité, ils sont principalement utilisés pour diffuser du champ diffus, bien que le hors-champ soit considéré comme un véritable lieu d'expression et que des sources directes peuvent parfois y être placées.

Le mouvement est globalement utilisé avec précaution, en cohérence avec notre développement en 3.3.6. Également en accord avec ce que nous avons relevé dans cette même partie, il nous semble que la précision de localisation n'est pas nécessairement une préoccupation chez les personnes interviewées, en faveur de l'acoustique.

Le mixage immersif apparaît comme étant plus facile que le mixage stéréophonique, notamment grâce à la quasi-disparition des effets de masque. En revanche, le manque de culture immersive de la part des artistes semble freiner l'émergence d'idées nouvelles, au même titre que l'omniprésence de la stéréo dans les processus de création de contenus spatialisés. Au regard de notre analyse, nous pouvons même affirmer que, dans les contenus immersifs qui prennent comme référence la stéréophonie, c'est précisément la grammaire stéréophonique qui est utilisée. Les libertés esthétiques apparaissent limitées, notamment par la nécessaire frontalité des éléments principaux.

D'une manière générale, la globalité de la grammaire détaillée en 4.3 et en Partie 5 (p.61) peut être remise en question par le concept établi en 5.3.2 qui s'intitule « Des règles pour être bafouées ». Si nous avons établi un nombre conséquent de potentielles « règles de grammaire » – l'utilisation des guillemets est précisément justifiée par ce paragraphe – celles-ci sont constamment remises en question par les esthétiques de

chacun et relatives aux contenus travaillés. Si cette idée de passer outre les règles, dans un souci stylistique, se retrouve également et évidemment en stéréophonie, l'espace qui s'ouvre à l'écriture immersive semble encore plus à même de laisser libre cours à son imagination en faisant fi de tout consensus.

Pour finir, après analyse de nos entretiens, c'est l'emploi dans ce mémoire du terme « immersif » qui nous pose le plus question. Le concept intitulé « Un problème de vocabulaire », détaillé en 5.3.3, souligne la fragilité de ce terme. Il est souvent utilisé de manière inappropriée et sa définition n'est pas claire. Confrontons ce qui ressort du sous-thème intitulé « Des définitions de l'immersion », détaillé en 5.3.1, avec notre définition, donnée en 3.2.3⁸⁶. Nous pouvons considérer que le concept intitulé « Reproduire la nature et la réalité », détaillé en 5.3.1, valide l'idée de simuler les principes de perception binaurale. Il est en revanche facile de contredire les autres termes que nous avons utilisés dans notre définition. Tout d'abord, la notion d'illusion n'a été employée qu'une fois, par Augustin Muller. Nous parlions également de sources provenant de toutes les directions, idée qui a été contredite par le concept « Ne pas utiliser nécessairement toutes les possibilités d'un système », détaillé en 5.3.2. Notre définition se concentre enfin sur un système de diffusion, ce qui n'est pas nécessairement le cas dans l'immersion, comme le montre la notion d'immersion narrative, indépendante de tout système technique et que nous n'avions absolument pas prise en compte. L'immersion perceptive n'a finalement pour seul but que de soutenir cette immersion narrative en aidant à réaliser cet état de dissociation cognitive – « *pour te faire perdre la notion de temps et d'espace et qui font que tu es complètement perdu dans l'œuvre.* » (voir 5.3.1.) – qui définit l'immersion.

Au regard du débat que provoque la notion d'immersion, nous pensons qu'il aurait été préférable, durant tout le développement de ce mémoire, d'utiliser le terme d' « audio spatialisé » plutôt que d'immersion sonore. Cela nous apparaît plus approprié, étant donné que nous nous sommes situés dans un contexte de restitution sonore par des haut-parleurs qui n'entourent pas nécessairement le public, dans lequel la sensation d'immersion serait une conséquence des règles appliquées au contenu et non une volonté intrinsèque, étant donné que les définition divergent et que la sensation d'immersion serait propre à chacun.

⁸⁶ Immersion : Possibilité donnée à un système de diffusion de faire entendre à un auditeur des sources provenant de toutes les directions, en simulant les principes de perception binaurales, afin de créer l'illusion de scènes sonores complexes.

Conclusion

Nous avons établi, grâce à 10 entretiens étudiés à l'aide d'une analyse thématique, les bases d'une grammaire de l'écriture sonore immersive. Elle est constituée de pratiques et processus communs à différents professionnels avec lesquels nous nous sommes entretenus, et se veut agnostique d'un média ou d'un système de diffusion. Elle traite des pratiques de mixages en lien avec la narration (mouvement, position, dynamique spatiale, spectre, compatibilité, influence stéréophonique), de l'ingénierie sonore associée (prise de son, système de diffusion, orientation, adaptabilité, accessibilité) et de la philosophie de l'audio immersif (histoire, forme d'écriture, pratique artistique, liberté esthétique, acculturation).

Si nous n'avons évidemment pas obtenu la précision et l'exhaustivité de la grammaire de l'écriture sonore stéréophonique, nous avons obtenu un résultat relativement complet et qui correspond à une attente manifestée des personnes interviewées. Nous espérons que ce travail pourra donner des idées, être utilisé comme support théorique pour apprendre ou même permettre à des acteurs de l'audio spatial de trouver des moyens de s'exprimer.

Notre travail trouve cependant ses limites dans le nombre d'entretiens que nous avons effectué. Les philosophies de l'audio spatial étant diverses et les approches du mixage très personnelles, nous aurions certainement eu de nouvelles contributions pertinentes lors de l'analyse d'entretiens supplémentaires. Nous considérons que cette grammaire est à mettre en perspective avec la pratique artistique que représente l'écriture d'une œuvre sonore. L'approche créative se doit d'être en perpétuelle reconstruction à chaque nouveau projet, ce qui implique qu'une méthode appliquée pour un projet ne soit potentiellement plus adaptée à un autre. Cela donne lieu à des contradictions et des avis divergents qui ont pu être constatés lors de nos entretiens.

Nous estimons également que, si nous avons voulu effectuer le travail le plus exhaustif possible, notre résultat est limité par les thématiques précises sur lesquelles nous nous sommes intéressés et il est évident que certaines règles n'ont pas été abordées lors de nos entretiens. Pour obtenir un résultat plus complet, il aurait été souhaitable d'effectuer plusieurs entretiens avec chaque professionnel afin d'aborder plus en détail chacun de leurs processus d'écriture. Certaines règles ont également pu nous échapper lors de notre analyse.

Nous estimons important de rappeler une nouvelle fois, et cela se retrouve à travers la règle intitulée « L'importance de la culture électroacoustique », détaillée en 5.3.5, que la musique électroacoustique a beaucoup travaillé et exploré l'espace et que les pièces qui ont été créées établissent en elles-mêmes une partie de la grammaire de l'écriture sonore spatialisée. Nous regrettons de ne pas avoir eu suffisamment de temps pour nous plonger dans l'écoute de l'immense catalogue disponible dans certaines archives. Nous suggérons au lecteur passionné ou au chercheur intéressé de compléter ce travail par l'analyse approfondie du catalogue de musique électroacoustique.

Enfin, nous aimerais souligner, en nous référant à l'entretien que nous avons eu avec Xavier Gibert, qui lui-même mentionne Hervé Déjardin, que l'intelligence artificielle se nourrit aujourd'hui facilement de la grammaire de l'écriture sonore spatialisée grâce aux contenus diffusés en streaming. Par la nature du fichier ADM, il est possible pour une intelligence artificielle de comprendre avec précision le travail qui a été réalisé lors du mixage en isolant les objets et en analysant la nature des sources sonores qui s'y trouvent. Il est alors possible d'associer ce traitement à l'ingénieur du son qui l'a réalisé et au style de musique correspondant. Il apparaît probable que, dans un futur proche, si ce n'est déjà le cas, l'équivalent de ce travail soit disponible sur un serveur quelque part dans le monde, à la seule différence près que, cette fois-ci, ce travail sera exhaustif.

« Je suis sûr que l'IA va permettre d'obtenir des choses beaucoup plus intéressantes en termes de spatialisation. »⁸⁷

Sharma, dans le cadre de notre entretien.

⁸⁷ « I'm quite sure AI will come up with sometimes much more interesting results when it comes to spatialization. » Notre traduction.

« *Toute recherche sur les sons doit conclure sur le silence.* »⁸⁸

R. Murray Schaeffer, 1977

⁸⁸ Traduit de l'anglais (Canada) par Sylvette Gleize.

Bibliographie

Adair, S., Alcorn, M., Corrigan, C. (2008) *A Study into the Perception of Envelopment in Electroacoustic Music*. School of Music and Sonic Arts, Queen's University, Belfast

Biocca, F., Delaney, B. (1995). *Immersive virtual reality technology*. in F. Biocca & M. R. Levy (dirs.), *Communication in the age of virtual reality*, Routledge, pp. 57–124

Blauert, Jens. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization (revised edition)*. MIT

Blanchet, A. & Gotman, A. (2010) *L'enquête et ses méthodes : l'entretien (2ème édition)*. Armand Colin

Bradley, J. S. & Soulodre, G. A. (1995) *Objective Measures of Listener Envelopment*. Journal of the Acoustical Society of America, 98 (5), Pt. 1, pp. 2590-2597

Cerles, C. (2015) *Caractérisation objective et subjective d'une chaîne de traitement HOA*. Mémoire de Master ENS Louis Lumière

Deacon, T., Bryan-Kinns, N., Healey, P., Barthet, M. (2019) *Shaping Sounds: The Role of Gesture in Collaborative Spatial Music Composition*. C&C '19: Creativity and Cognition

Féron, F.X. (2022) Composer (avec) l'espace : une pratique musicale polysémique in Vogel, C. & Lambinet, S. (dir.). *La perception du son spatialisé : Un son propre pour un sens figuré*. Cahier de Louis-Lumière N°15, pp. 70-92

Féron, F.X., Frissen, I., Boissinot, J., Guastavino, C. (2010) *Upper limits of auditory rotational motion perception* Acoustical Society of America. PACS number(s): 43.66.Qp, 43.66.Mk, 43.75.Cd [JCM] pp. 3703–3714

Guastavino, C. & Katz, B. (2004) *Perceptual Evaluation of Multi-Dimensional Spatial Audio Reproduction*. The Journal of Acoustical Society of America, PACS numbers: 43.66.Lj, 43.66.Qp, 43.38.Md [MK], pp. 1105–1115

Guiot, V. (2017) *De l'interprétation acousmatique*, Filigrane. Musique, esthétique, sciences, société. [En ligne], n°22, Vers une éthique de l'interprétation musicale, URL : <https://revues.mshparisnord.fr:443/filigrane/index.php?id=844>

Hamasaki, K., Nishiguchi, T., Hiyama, K., Okumura, R. (2006) *Effectiveness of Height Information for Reproducing Presence and Reality in Multichannel Audio System*. Audio Engineering Society, Convention Paper 5579, 120th Convention, Paris, France

Howie, W., Martin, D., Kim, S., Kamekawa, T., King, R. (2019) *Effect of Audio Production Experience, Music Training, and Age on Listener Performance 3D Audio Evaluation*. J. Audio Engineering Society, vol. 67, no. 10, pp. 782-794

Inanaga, K., Yamada, Y., Kozumi, H. (1995) *Headphone System with Out-of-Head Localization Applying Dynamic HRTF (Head Related Transfer Function)*. Audio Engineering Society, 98th Convention, Paris, France

Jullien, J.P. (2022) Enjeux de l'immersion dans une acoustique virtuelle pour les systèmes de spatialisation auditive. in Vogel, C. & Lambinet, S. (dir.). *La perception du son spatialisé : Un son propre pour un sens figuré*. Cahier de Louis-Lumière N°15, pp. 96-125

Kamekawa, T., Marui, A. (2020) *Are full range loudspeakers necessary for the top layer of Three-Dimensional audio?* Audio Engineering Society, 148th Convention, Online

Kim, S., Howie, W. (2021) *Influence of the Listening Environment on Recognition of Immersive Reproduction of Orchestral Music Sound Scenes*. Journal of the Audio Engineering Society, vol. 69, no. 11, pp. 834-848

Kim, S., Lee, Y.W., Pulkki, V. (2010) *New 10.2-channel Vertical Surround System (10.2-VSS); Comparison study of perceived audio quality in various multichannel sound systems with height loudspeakers*. Audio Engineering Society, Convention Paper 8296, San Francisco, CA, USA

King, R., Leonard, B., Kelly, J. (2019) *Height channel signal level in immersive audio – how much is enough?* Audio Engineering Society, 146th Convention, Dublin, Ireland

Lindau, A., Erbes, V., Lepa, S., Maempel, H.J., Brinkman, F., Weinzieri, S. (2014) *A Spatial Audio Quality Inventory (SAQI)*. Acta Acustica united with Acustica, vol. 100, no. 5, pp. 984-994

Moélants, A. (2023) *Immersive Audio Test Signals for Musical Applications*. Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio, Huddersfield, UK

Nilles, A. (2016) *La spatialisation des musiques électroniques au regard du processus de composition*. Mémoire de Master 1 Université Paris 8

Nilles, A. (2018) *La spatialisation des musiques électroniques du point de vue de l'auditeur : perception, immersion, concert spatialisé*. Mémoire de Master 2 Université Paris 8

Olivieri, F., Peters, N., Sen, D. (2019) *Scene-Based Audio and Higher Order Ambisonics: A technology overview and application to Next-Generation Audio, VR and 360° Video* EBU Technical Review, Qualcomm Technologies Inc., San Diego, California, USA

Paillé, P. & Mucchielli, A. (2021) *L'analyse qualitative en sciences humaines et sociales*. Armand Colin

Periaux, B., Ohl, J.L. Thevenot, P. (2015) *Le Son Multicanal. De la Production à la Diffusion du Son 5.A, 3D et Binaural*. DUNOD

Ramsey, B. (2011) *Social spatialisation: Exploring links within contemporary sonic art*. Toronto Electroacoustic Music Symposium

Roffler, S.K. & Butler, R.A. (1968). *Localization of tonal stimuli in the vertical plane*. Journal of the Acoustical Society of America, 43(6), pp. 1260–1266.

Rumsey, F. (2001) *Spatial Audio*. Focal Press, Music Technology Series, 2001

Rumsey, F. (2002) *Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm*. J. Audio Engineering Society, Vol. 50, No. 9, pp. 651-666

Sazdov, R., Paine, G., Stevens, K. (2007) *Perceptual investigation into envelopment, spatial clarity, and engulfment in reproduced multi-channel audio*. AES 31st International Conference, London, UK

Shim, H., Oh, E., Ko, S., Park, S.H. (2010) *Perceptual Evaluation of Spatial Audio Quality*. Audio Engineering Society, Convention Paper 8300

Slater, M. (2003) A Note on Presence Terminology.

https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology

Zanni, A. (2020) *De la relation entre acoustique et composition « Espaces d'errances »*. Mémoire de Master ENS Louis Lumière

ANNEXES

Annexe 1 : La stéréophonie

Les premiers essais d'enregistrement stéréophonique remontent en 1931. Alan Blumlein, ingénieur en téléphonie au Royaume Uni, dépose un brevet qui pose les bases de la stéréophonie : les deux canaux, gauche et droite (Périaux *et al.*, 2015). La stéréophonie se répand petit à petit auprès du grand public grâce à son atout économique et sa qualité de restitution : seules deux enceintes sont nécessaires à un rendu spatial convaincant dans l'espace formé par les deux enceintes.

Après les résultats de nombreuses recherches et expérimentations, il est communément recommandé de placer les deux enceintes aux sommets d'un triangle équilatéral qu'elles formeront avec le centre de la tête de l'auditeur (Rumsey, 2001). Il est alors possible de distinguer une image fantôme entre les enceintes. Cela a pour conséquence de créer l'illusion d'un espace acoustique crédible entre les deux enceintes.

L'image fantôme située entre les deux enceintes peut être déplacée de gauche à droite en utilisant un panoramique d'intensité. Cette technique consiste à envoyer le même signal dans l'enceinte gauche et droite et à faire varier l'intensité de chaque signal. L'image sonore qui en découle se déplacera en conséquence. Les deux signaux ainsi modifiés sont dits corrélés. C'est-à-dire que l'un peut être relié à l'autre par une fonction de transfert. C'est la corrélation qui permet l'illusion auditive du champ stéréophonique.

D'autres modifications peuvent être apportées aux signaux afin de les maintenir corrélés tout en modifiant la perception que l'on en aura. La réverbération, utilisée pour créer une sensation de profondeur et de relief, ou l'égalisation, utilisée pour démasquer les sources, en font partie. Ces techniques, parmi beaucoup d'autres, constituent l'art du mixage stéréophonique.

La stéréophonie peut être considérée comme étant un système de spatialisation sonore, le plus simple à mettre en place. Il permet de restituer un champ sonore frontal de 60° de large tout en créant une illusion de profondeur. Par ses caractéristiques physiques, ce système de diffusion possède deux limites principales (Périaux *et al.*, 2015).

La première concerne sa zone d'écoute idéale, qui est très restreinte. On dit que c'est un système de diffusion « égocentrique », c'est-à-dire conçu pour une écoute individuelle (auditeur unique placé au troisième sommet du triangle formé avec les enceintes). Cette position est appelée le *sweet spot* en anglais. Un léger déplacement trop à gauche ou à droite entraînera une modification notable du champ stéréophonique, et un déplacement en-dehors du *sweet spot* entraînera un effet de précédence (l'auditeur perçoit le signal comme provenant que d'une seule enceinte, celle la plus proche de lui).

La seconde réside dans la qualité de la restitution des sources fantômes due à la diaphonie (signal de l'enceinte droite qui arrive à l'oreille gauche et inversement). Le signal stéréophonique diffusé par les enceintes va subir un filtrage en peigne qui colore la source. On note notamment une atténuation conséquente autour de 2kHz et des accidents dans les aigus.

À la suite de ces constats, remarquons que l'écoute stéréophonique au casque est bien différente puisqu'elle envoie chaque signal dans une seule oreille. La perception sonore d'un auditeur lors de l'écoute d'un même contenu stéréophonique au casque ou sur enceintes ne sera donc pas la même. La principale différence réside dans la restitution de la largeur de l'image sonore. Un mixage effectué sur enceintes aura tendance à être plus latéralisé si écouté au casque tandis qu'un mixage au casque donnera trop d'informations centrales si écouté sur haut-parleurs.

Annexe 2 : L'ambisonie et la WFS

« L'objectif est de recréer le champ acoustique qui aurait été généré par les sources sonores que l'on cherche à simuler. »

Jullien (2020)

A1.1 L'ambisonie

L'ambisonie est une technique de restitution du champ acoustique qui se base sur une division de l'espace en différents éléments spatiaux que l'on appelle harmoniques sphériques. Nous pouvons faire une analogie entre cette séparation de l'espace et la transformée de Fourier. De la même manière qu'il est possible de décomposer tout signal périodique en somme de signaux sinusoïdaux, il est possible de diviser l'espace en somme d'harmoniques sphériques. Contrairement aux formats orientés canal, un fichier son ambisonique ne peut pas être lu tel quel. C'est un format qui nécessite un encodage puis un décodage. L'encodage permet justement de regrouper les informations sonores en fonction des harmoniques sphériques. Le décodage va permettre d'interpréter ces harmoniques et de les restituer en fonction du système d'écoute.

Les premiers résultats ambisoniques ont été établis par Gerzon⁸⁹ dans les années 1970. Ils avaient pour objectif de reproduire le champ sonore émis par un certain nombre de sources ponctuelles dans l'espace. Deux grandeurs doivent être analysées : la pression et le gradient de pression. La première étape de leur recherche consiste à utiliser un réseau de 3 microphones bidirectionnels pour permettre la localisation d'une

⁸⁹ Gerzon, M. (1971) *Experimental Tetrahedral Recording*. Studio Sound, vol. 13, Part 1: pp. 396-398, Part 2 : pp. 472, 473 and 475, Part 3: pp. 510, 511, 513 and 515

Gerzon, M. (1975) *Ambisonics. Part Two: Studio Techniques*. Studio Sound, vol. 17 no. 8, pp. 24-26, 28 and 30

source grâce à ses coordonnées cartésiennes. On combine ensuite ces 3 microphones avec un autre microphone omnidirectionnel, qui lui est en mesure d'analyser la pression.

La combinaison de ces 4 directivités permet d'obtenir une première représentation de l'espace en 3 dimensions. Ces directivités correspondent aux 4 premiers harmoniques sphériques. C'est l'ambisonie d'ordre 1. Aujourd'hui, un microphone ambisonique d'ordre 1 est constitué de 4 capsules cardioïdes disposées sur les faces d'un tétraèdre. Ces capsules permettent de reproduire les 4 directivités évoquées ci-dessus après encodage. On distingue deux formats :

- le format A est le fichier son brut, issus des capsules cardioïdes,
- le format B est le format encodé, qui distingue les 4 directivité qui constituent l'ordre 1.

Il est possible de décomposer l'espace en utilisant d'autres harmoniques sphériques plus complexes. On parle alors d'ambisonie d'ordre supérieur, ou HOA. Plus on utilise un ordre élevé, plus on va décomposer notre champ sonore en un nombre élevé d'harmoniques sphériques qui seront plus complexes, et plus on va gagner en précision spatiale. Pour un ordre n, dans le cas d'un rendu ambisonique en 2 dimensions, l'espace sera divisé en $2n+1$ harmoniques sphériques. Dans le cas d'un rendu ambisonique en 3 dimensions, l'espace sera divisé en $(n+1)^2$ harmoniques sphériques. Autant de capsules sont nécessaire pour obtenir des micros capables de capturer un ordre n (Cerles, 2015).

En ce qui concerne la diffusion et afin d'obtenir un rendu optimal, il est recommandé d'utiliser une répartition régulière de haut-parleurs. Soit à la surface d'un cercle pour l'ambisonique 2D, soit d'une sphère pour la 3D. Pour un ordre n, le nombre d'haut-parleurs à utiliser pour ce rendu optimal doit être supérieur au nombre d'harmoniques sphériques utilisés pour diviser l'espace à cet ordre précis. Il est cependant possible d'utiliser un nombre inférieur de haut-parleurs. Par la nature des harmoniques sphériques, un ordre n comporte tous les ordres inférieurs. Il est donc possible de décoder un format ambisonique d'ordre n en utilisant moins d'harmoniques sphériques, et de descendre vers un ordre inférieur. Cela permet de restituer un ordre n sur un dispositif d'enceintes plus petit, bien qu'il y aura un impact sur la qualité de la restitution de l'espace, dont la précision sera réduite de par le fait que le système de diffusion est réduit. (Cerles, 2015)

Après la décomposition en harmoniques sphériques, nous sommes capables par le calcul de « pointer » dans une direction précise et de définir un point de diffusion, en l'occurrence un haut-parleur. Bien qu'il soit recommandé d'avoir un dispositif constitué

d'enceintes correctement réparties, il est possible, grâce au « pointage », d'utiliser n'importe quel système de haut-parleurs, même si celui-ci n'est pas normé. Il suffit de renseigner au décodage la position exacte des enceintes et le son sera décodé comme venant de cette direction. L'ambisonie est donc particulièrement portable et flexible, ce qui en fait son grand atout.

A1.2 La synthèse de front d'onde (WFS)

La synthèse de front d'onde ou WFS – pour *Wave Field Synthesis* en anglais – est une technique de restitution qui se base sur une reproduction du front d'onde à l'aide d'un ensemble de haut-parleurs. Cette approche se base à la fois sur des différences d'amplitudes et de temps (Δi et Δt).

L'avantage majeur et reconnu de la WFS réside dans la stabilité de l'image de la source et de sa position relative vis-à-vis de la position de l'auditeur. En d'autres termes, la WFS permet de contrepasser la problématique de l'effet de précédence d'un système orienté canal et placera la source acoustique à sa position relative réelle. On parle ici de restitution de point source (*Ircam*, 2004).

La WFS peut également reproduire des ondes planes. Ce phénomène sonore n'existe pas dans la nature. Il permet de créer la sensation d'être suivi par un son, celui-ci venant en permanence de la même direction même si l'auditeur se déplace. Contrairement à n'importe quel autre système de reproduction, un dispositif utilisant la WFS aura également la capacité de positionner une source sonore devant les enceintes. On peut utiliser le terme analogique d'« hologramme sonore » dans le sens où la source est située dans la zone d'audience, là où il n'y a pas de sources physiques (*Ircam*, 2004).

Le nombre de haut-parleurs est déterminant pour la qualité de rendu spatial de la WFS, et il est nécessaire d'en avoir un nombre important à disposition. Comme l'explique Jullien (2021) : « *on ne peut reproduire exactement que les fréquences dont la longueur d'onde est supérieure au double de cet espace*. Pour les fréquences plus grandes, le phénomène d'aliasing spatial créé des interférences parasites. » La WFS est une technologie qui fonctionne bien dans les graves, dans le domaine audible où les longueurs d'ondes sont très élevées. Dès qu'on arrive sur des longueurs d'ondes plus courtes et que notre dispositif ne dispose pas assez d'enceintes – donc d'enceintes qui vont être relativement loin les unes des autres – nous serons limités dans la reproduction des hautes fréquences et la qualité du rendu spatial sera moindre.

Le rendu du mouvement posera des problèmes en WFS dus à l'utilisation du Δt dans son algorithme de restitution. Un mouvement implique une modification de délais et aura pour conséquence un *pitch shifting* (un changement de hauteur que nous pouvons associer, dans cette situation, à une tête de lecture qui se déplace sur son support) ou un filtrage en peigne (un *crossfade* entre 2 délais). Dans les deux cas, le timbre en sera affecté.

A1.3 Les limites de l'audio orienté scène et de la WFS

Aussi bien la WFS que l'ambisonie demandent l'utilisation d'un nombre de haut-parleurs conséquent pour la bonne restitution de l'espace sonore qu'ils tendent à reproduire. Si une production souhaite utiliser un de ces systèmes pour une situation d'écoute en public, il sera souhaitable d'élargir la zone d'écoute, augmenter le nombre de haut-parleurs, et par conséquent alourdir le traitement du signal. Ce sont des dispositifs qui deviennent alors coûteux et difficiles à implémenter.

Nous précisons d'ailleurs que, si la WFS est un dispositif allocentrique par nature car conçu pour avoir une image sonore stable quelle que soit la position de l'auditeur dans l'audience, l'ambisonique est quant à lui un dispositif égocentrique par nature car il s'agit de reproduire le champ sonore en un point unique.

Nous pouvons mentionner l'ESPRO à l'Ircam, cité plus haut, qui a installé une ligne de WFS intégrant 264 haut-parleurs, soit un tous les 30 cm. 264 haut-parleurs pour une salle pouvant accueillir 300 auditeurs. On comprend que ce ne soit pas à la portée de toutes les économies de salles de spectacles à la jauge de quelques centaines de personnes.

Annexe 3 : Exemple d'utilisation de l'audio orienté objets

Pour illustrer l'avantage de la NGA et de l'objet, l'EBU (Olivieri, 2019) utilise l'exemple de diffusion d'un match de foot, que nous reprenons ici.

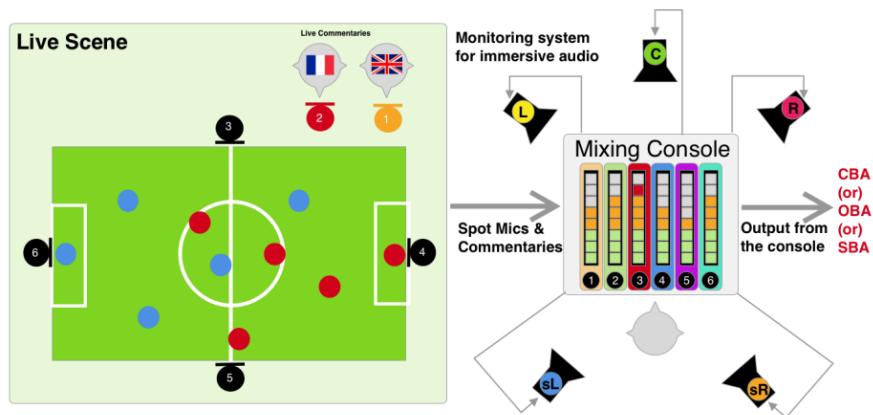


Figure 3 : Événement live avec captation multi-microphonique ; environnement du mixeur (Olivieri, 2019)

Considérons l'installation technique définie par la figure 1. 2 micros sont attribués aux 2 commentateurs, 1 français et 1 anglais. 4 autres sont disposés autour du terrain pour capter les ambiances. Les 6 signaux entrent dans la console, qui a la possibilité d'exporter 3 types d'audio : du *Channel Based*, du *Scene Based*, ou de l'*Object Based*.

Prenons dans un premier temps l'exemple d'une sortie en *Channel Based*. Il conviendrait alors à l'ingénieur du son de livrer différents mixs pour s'adapter à différentes configurations. Cela peut facilement donner lieu à la réalisation de 4 mixs : un stéréo et un 5.1 en anglais, pris en charge par un premier diffuseur, puis un stéréo et un 5.1 en français, pris en charge par un second diffuseur. Ces mixs peuvent cependant ne pas s'adapter à la situation d'écoute de l'auditeur, comme le montre la figure 2.

Dans la première situation d'écoute, l'utilisateur dispose d'un 5.1 mal configuré pour des raisons d'esthétique domestique, et ni la configuration complète, ni la stéréo, peuvent être lue dans des conditions d'écoute normées. La disposition ne correspond pas à la disposition utilisée pour le mixage. Dans la seconde situation d'écoute, la stéréo

est correctement installée et correspond au dispositif utilisé pour le mixage. En revanche, l'auditeur ne parle pas la langue du contenu restitué.

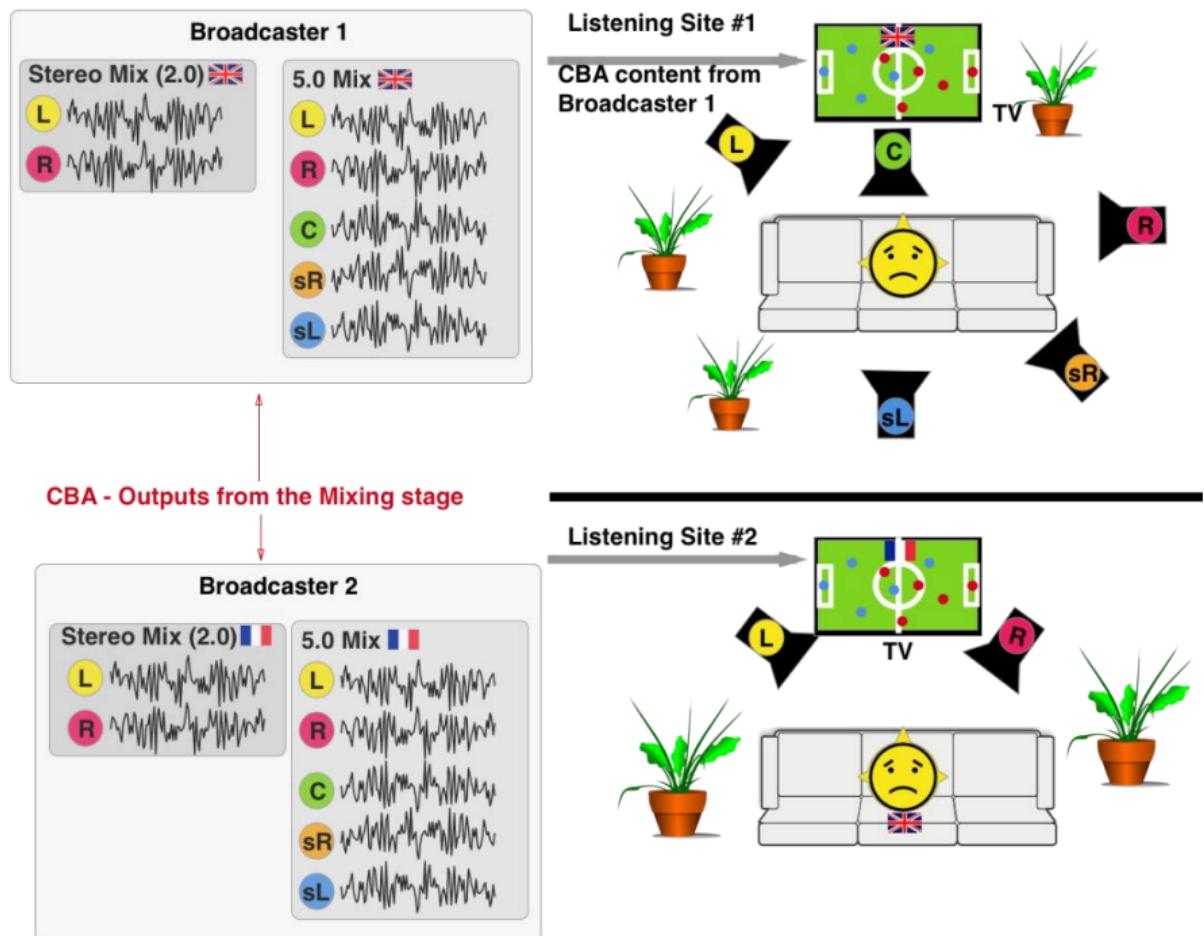


Figure 4 : Export et restitution d'un contenu Channel Based Audio (Olivieri, 2019)

Les raisons qui font qu'un contenu *Channel Based* n'est pas adapté à la NGA sont donc multiples :

- travailler pour différentes nationalités demande d'effectuer différents mixs. Cela peut devenir cher à produire et demander plus de bande passante à l'émission,
- le *Channel Based* ne peut pas être personnalisé. Par exemple, l'auditeur pourrait vouloir baisser le volume du commentateur au profit de l'ambiance du stade, ce qui serait possible grâce au *renderer*,
- le format de rendu n'est pas flexible. Si le système de haut-parleur au point d'écoute ne correspond pas à celui utilisé pour le mixage en *Channel Based*, le rendu sera de moins bonne qualité.

La diffusion en objet semble donc permettre une meilleure adaptabilité, bien qu'elle montre elle aussi ses limites :

- Si le nombre d'objets devient trop important, la bande passante nécessaire pour les diffuser peut devenir grande.
- Nous pouvons distinguer les composantes M&E – pour *Music and Effect* – des composantes vocales. Dans un contexte objet, il est fréquent que les M&E soient mixés en *Channel Based* – typiquement en 7.1.2 comme préconisé par le *bed Atmos*. Nous nous confrontons alors aux mêmes limitations que celles évoquées précédemment dans un contexte de diffusion *Channel Based*.
- Il serait possible de diffuser les M&E en objet, mais cela entraînerait une plus grande bande, ce qui renvoie au premier point.
- Un nombre d'objet élevé demande au *renderer* d'assurer une restitution plus complexe. Le problème ne concerne plus ici la diffusion mais les dispositifs d'écoute. Dans un contexte d'écoute sur dispositifs portables par exemple, la puissance de calcul sera limitée, ce qui aura un impact sur le rendu.

Si nous reprenons nos situations précédentes, l'objet permettrait tout de même au second auditeur d'avoir un rendu optimal. Et on pourrait imaginer des M&E mixée en objet et non en *Channel Based* car le nombre d'objets n'est ici pas trop important, et donc valider une écoute optimale chez le premier auditeur également.

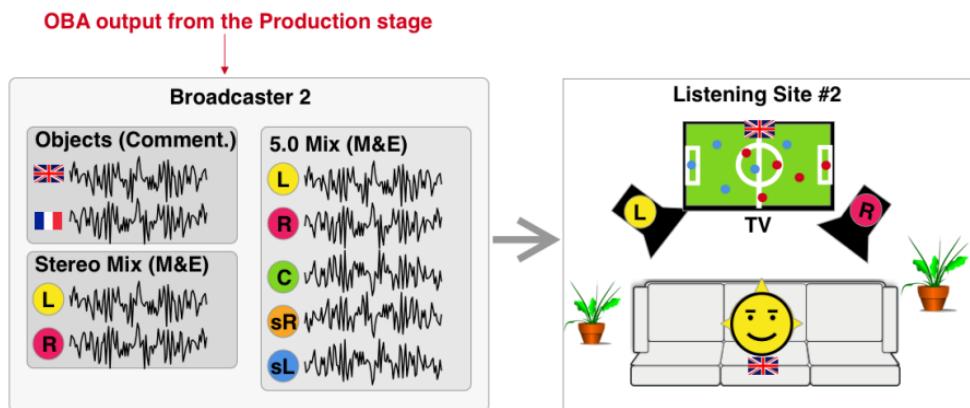


Figure 5 : Contenu Object Based Audio et son environnement d'écoute (Olivieri, 2019)

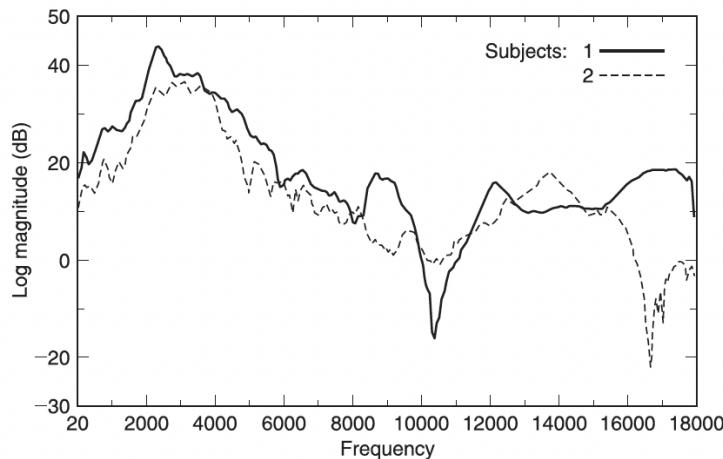
Annexe 4 : Le binaural

Nous pouvons distinguer deux techniques qui permettent de produire du contenu binaural : le binaural natif et le binaural de synthèse.

Le binaural natif consiste à enregistrer du son grâce à deux micros positionnés de telle sorte que le contenu audio pourra être directement restitué dans les écouteurs, avec sensation d'externalisation, sans aucun traitement additionnel. Pour le positionnement des microphones, il faut se servir de la forme de la tête. Il est possible de positionner des microphones au niveau des oreilles d'un individu ou d'utiliser un buste artificiel qui recopiera une forme de tête humaine et de placer des micros à l'emplacement des oreilles fictives.

Le binaural de synthèse consiste à venir appliquer à une source ponctuelle manipulée dans un logiciel les indices de localisation associés à sa position grâce à des filtrages audionumériques. Les filtres ont été initialement constitués en chambre anéchoïque. Pour en construire, on place un sujet (ou un buste) au centre de la chambre anéchoïque avec des capteurs dans ses oreilles. On déplace ensuite une source sonore tout autour du sujet avec une précision donnée afin d'envoyer un contenu fréquentiel à toutes les positions de l'espace. L'analyse des réponses permettra de connaître la fonction de transfert données par les oreilles en fonction des positions. Il sera ensuite possible de filtrer une source sonore virtuelle grâce à la fonction de transfert associée à sa position donnée dans l'espace et l'externalisation aura lieu.

Le binaural a cependant ses limites. En ayant pris connaissance des indices de localisation (voir chapitre 1) et de la manière dont on construit le binaural de synthèse, nous comprenons que le rendu binaural dépend de la HRTF utilisée pour sa production, que ce soit natif ou de synthèse. Les courbes de réponse de l'oreille humaine sont très chaotiques dans les hautes fréquences et très dépendantes de l'individu, comme nous le montre la Figure 6 : HRTFs de 2 sujets pour une source à 0° d'azimut et d'élévation.. Nous arrivons à localiser notamment grâce à notre apprentissage personnel. Une personne qui écoute un binaural produit avec une HRTF différente de la sienne aura ainsi indéniablement un rendu différent et non optimisé.



*Figure 6 : HRTFs de 2 sujets pour une source à 0° d'azimut et d'élévation.
Nous observons des différences considérables dans les hautes fréquences
(Begault, cité par Rumsey, (2001))*

« En fait, obtenir des images frontales à l'aide de n'importe quel système binaural au casque est étonnamment difficile. Cela peut être dû au fait qu'on est habitué à utiliser notre ouïe pour localiser des éléments qui ne peuvent pas être vus, et qu'un élément qui n'est pas visible est probablement localisé à l'arrière. Sans avoir la possibilité de bouger la tête pour résoudre des conflits avant – arrière, le cerveau tend à comprendre une image sonore située à l'arrière. Des scènes de ce type, inversées, sont en conséquence extrêmement fréquentes en audio binaural. »⁹⁰

Rumsey (2001)

Rumsey (2001) précise également que l'absence de visuel corrélé à notre sensation auditive cause d'autres problèmes dans le rendu binaural. L'absence de visuel implique une sensation générale d'écoute où toute la scène sonore proviendrait de l'arrière. La perception de la scène sonore à l'arrière de la tête relève d'une confusion faite par le

⁹⁰ « In fact, obtaining front images from any binaural system using headphones is surprisingly difficult. This may be because one is used to using the hearing sense to localize things where they cannot be seen, and that if something cannot be seen it is likely to be behind. In the absence of the ability to move the head to resolve front-back conflicts the brain tends to assume a rear sound image. So-called ‘reversals’ in binaural audio systems are consequently very common. » Notre traduction

cerveau face à une absence de visuel concordant avec l'image. Nous avons vu précédemment qu'il est possible d'outrepasser des confusions grâce au mouvement de tête.

Répétons Moélant (2023) : « *When a sound source follows our exact head movements (e.g., when we listen through a normal pair of stereo headphones), a sound won't appear to be outside the head.⁹¹* » Comme nous venons de l'expliquer, et contrairement à ce qu'affirme Moélants, le binaural parvient à externaliser le son, tout en reproduisant un signal sonore qui suit exactement nos mouvements de tête. Il est cependant sujet à de fortes distorsions de l'image sonore. Afin de corriger ces distorsions et de ne plus faire de confusions dans la localisation, la solution est de ne plus utiliser un casque stéréophonique classique mais d'utiliser un capteur de mouvement : du *Head Tracking*.

Le *Head Tracking* permet de capter les mouvements de la tête de l'auditeur et d'effectuer une rotation de la scène sonore en conséquence. La position du signal audio n'est alors plus liée aux mouvements de la tête mais en devient relative. Le rendu binaural devient, dans ces conditions, beaucoup plus réaliste (Rumsey, 2001).

⁹¹ « *Lorsque qu'une source sonore suit exactement nos mouvements de tête (autrement dit, lorsque nous écoutons du contenu audio à l'aide d'un casque classique), un son ne se sera pas perçu comme provenant de l'extérieur de la tête.* » Notre traduction.

Annexe 5 : Le travail de l'espace sur logiciel

Pour ce qui concerne le *channel based* audio, la manipulation des sources dans l'espace se fait majoritairement à l'aide d'algorithme de *panning* en Δi . En stéréo ou en 5.1 par exemple, une source sera positionnée dans l'espace grâce à un *panning* qui contrôlera la position en azimut de la source, à distance constante. C'est ce qu'on retrouve sur les consoles ou les DAW – pour *Digital Work Station* en anglais, stations audio numériques. La sensation de distance, de relief et de profondeur doit être ajoutée à la localisation azimutale grâce au mixage et à l'utilisation de la réverbération.

Du côté de l'ambisonique, différents outils ont été développés pour reconstituer des scènes sonores et les encoder en utilisant les harmoniques sphériques. Nous pouvons citer la suite de *plugin* de l'IEM – *Institute of Electronic Music and Acoustics* de Graz en Autriche, les *Spatial Audio Plugins* de Noisemakers, développé par Charles Verron et son équipe en France, ou encore l'*Ambisonic Toolkit* pour Reaper et SuperCollider, projet soutenu par le *Department of Digital Arts and Experimental Media University of Washington* et *The Norwegian Film School Inland Norway University of Applied Science*. D'une manière générale, contrairement à un panoramique stéréo, la représentation spatiale pour le traitement ambisonique est réalisée à l'aide d'un espace en 3 dimensions virtualisé dans lequel l'utilisateur déplace la source sonore. Cette fois-ci, en plus de l'azimut, la distance et l'élévation peuvent être travaillées.

Un autre outil largement utilisé pour le travail de l'espace est le spatialisateur (SPAT) de l'Ircam. Développé initialement comme une réverbération algorithmique permettant de communiquer avec les artistes à l'aide de termes perceptifs, le SPAT est aujourd'hui implanté dans différents logiciels et processeurs de spatialisation. Il a l'avantage d'utiliser sa réverbération dans différents contextes de spatialisation, aussi bien en ambisonique qu'en WFS tout en intégrant d'autres algorithmes de *panning* tels que le KNN ou le VBAP que nous n'allons pas détailler dans ce mémoire. Le SPAT est notamment utilisable dans Max (Cycling74) et via une version interfacé et autonome appelée Panoramix, il est également le moteur du logiciel SPAT Revolution de la société Flux:: (France), et est adapté pour Ableton Live via des *Device Max For Live* comme les SPAT Devices de Music Unit, studio basé à Montreuil, ou le XP d'Éric Raynaud, ingénieur et compositeur français.

En ce qui concerne l'Audio Orienté Objet et le format ADM, nous allons décrire le *Dolby Atmos*, qui nous intéresse particulièrement dans ce mémoire puisque largement utilisé par les ingénieurs du son qui nous ont répondu et actuellement leader sur le marché.

Le *Dolby Atmos* est une technologie propriétaire de reproduction sonore qui intègre *surround* et élévation pour des restitutions sur différents systèmes de diffusion. La société *Dolby* considère le 7.1.2 comme étant la configuration minimale pour profiter au mieux de l'apport de la spatialisation. Le travail de l'espace dans ce contexte passe par le *Dolby Atmos Renderer*. L'*Atmos* – abréviation de *Dolby Atmos* – est aujourd’hui le premier outil utilisé pour exporter un ADM dans l’industrie musicale et cinématographique. Il permet de manipuler jusqu’à 128 canaux audio (*bed* et objets).

Le *Dolby Atmos Renderer* est une interface graphique qui représente notamment une pièce dans laquelle il est possible de positionner des objets associés des sources sonores. Le *Dolby Atmos Renderer* permet d’écouter dans des configurations allant de la stéréo au 9.1.6 en passant par le binaural. Pour une écoute binaurale, il faut noter que le *Dolby Atmos* ajoute automatiquement une réverbération au rendu que l’on nomme BRIR pour *Binaural Room Impulse Response* en anglais.

Le travail en Dolby Atmos comporte deux éléments à appréhender : le *bed* et les objets.

Le *bed* est un format multicanal (*Channel Based*) qui peut comporter jusqu'à 10 canaux. C'est le cas spécifique du 7.1.2, sa configuration la plus élaborée. Les 10 premiers canaux de traitement sont ainsi réservés pour le *bed*. Dans le cas où un *bed* 5.1 est utilisé, par exemple, les canaux numérotés de 7 à 10 seront vides. Les objets commencent ensuite à partir du canal 11. Ils sont constitués, comme nous l'avons expliqué plus haut, d'un canal audio et de métadonnées. Certaines sont dynamiques, comme la position de l'objet ou tout ce qui est relatif aux automations. D'autres sont statiques, comme celles liées à la BRIR associée à l'objet.

Une des différences majeures entre le *bed* et les objets réside dans le rendu binaural. Un objet aura sa propre fonction de transfert au décodage tandis que le *bed* sera restitué à l'aide d'enceintes virtuelles. Une source traitée dans le *bed* peut tout de même avoir des mouvements automatisés via le *panning* du bus associé au *bed*.

Annexe 6 : Quelques définitions de l'immersion et de l'audio spatialisé

« L'expression de “spatialisation auditive” évoque avant tout l'illusion auditive, créée par un système de diffusion, qu'une source sonore peut-être localisée n'importe où dans l'espace. »

Jullien (2020)

« Un “système de spatialisation auditive virtuelle” est un système de sonorisation et de traitement du signal qui permet de créer, d'une part, des sources virtuelles et, d'autre part, une acoustique virtuelle. »

Jullien, (2020)

« L'approche de l'audio spatial tente de créer l'impression de sources sonores qui se déplacent dans l'espace en utilisant des haut-parleurs ou un casque. L'audio spatial utilise certains des indices de localisation que nous utilisons pour décoder la direction d'une source dans l'environnement sonore, nous permettant de savoir d'où le son provient. »⁹²

Deacon *et al.* (2019)

« En utilisant l'audio spatial, les compositeurs tendent à recréer des scènes acoustiques complexes. Pour ce faire, les compositeurs doivent créer des espaces acoustiques virtuels et manipuler les sources au cours du temps. »⁹³

Deacon *et al.*, (2019)

« L'immersion est un terme qui réfère au degré auquel un environnement virtuel submerge le système perceptif d'un utilisateur. »⁹⁴

Biocca & Delaney (1995, p.57)

⁹² « *Spatial audio approaches attempt to create the impression of spatially displaced sound sources by using speakers or headphones. Spatial audio supplies some of the localization cues that we use to decode source direction in environmental sound; allowing a user to pinpoint where sound is coming from.* » Notre traduction.

⁹³ « *Using spatial audio, what composers attempt to recreate are “acoustically complex scenes”. To do this, composers must create virtual acoustic spaces and arrange sources over time.* » Notre traduction.

⁹⁴ « *Immersive is a term that refers to the degree to which a virtual environment submerges the perceptual system of the user.* » Notre traduction.

« Réservons simplement le terme “immersion” à ce que délivre une technologie d'un point de vue objectif. Le plus un système délivre des rendus (dans toutes les modalités sensorielles possibles) et du suivi qui préservent le naturel des sensations ressenties dans la vie de tous les jours, le plus il est immersif. »⁹⁵

Slater, (2003)

« Système de diffusion immersif : un système de diffusion capable de produire du son immersif. Son immersif : inclus des sons qui proviennent de sources provenant du système de base, et au-delà. »⁹⁶

SMPTE (2018)

« Les attributs spatiaux valides sont reliés principalement à la nature tri-dimensionnelle des sources sonores et de leur environnement. »⁹⁷

Rumsey (2002)

Nous partageons également dans cette annexe les conditions requises pour l'obtention d'un « *high-presence audio format* », en l'occurrence un 22.2. Selon la NHK japonaise, un tel système doit :

- permettre à l'auditeur de localiser une source n'importe où sur l'écran,
- pouvoir reproduire une source sonore provenant de toutes les directions entourant l'auditeur,
- doit pouvoir reproduire un espace acoustique en 3 dimensions naturel et de haute qualité,
- doit disposer d'une zone d'écoute optimale élargie,
- doit être compatible avec des formats audios existants,
- doit supporter un enregistrement et une diffusion.

(Hamasaki *et al.*, 2006)

⁹⁵ « Let's reserve the term “immersion” to stand simply for what the technology delivers from an objective point of view. The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is “immersive”. » Notre traduction.

⁹⁶ « Immersive Sound System: A sound system capable of producing immersive sound. Immersive Sound: Includes sound that emanates from sources at and beyond the base layer. » Notre traduction.

⁹⁷ « Valid spatial attributes [are] being primarily concerned with the three-dimensional nature of sound sources and their environments. » Notre traduction.

Annexe 7 : Les processeurs de diffusion spatialisée

- L-Isa, de la société française L-Acoustics

Le L-Isa se veut être une solution qui améliore la restitution sonore en concert. Il permet des installations immersives à 360° sur plusieurs enceintes, contrôlées via le L-Isa Controller. L'algorithme de *panning* et de rendu spatial utilisé par le L-Isa est propriétaire.

- Fletcher Machine de la société canadienne Adamson

La Fletcher Machine est un spatialisateur qui travaille avec un algorithme de *panning* propriétaire et son propre algorithme de réverbération.

- Soundscapes de la société allemande d&b Audiotechnik

Le processeur Soundscapes se base sur deux modules : *En-Scene* et *En-Space*. Si *En-Scene* propose globalement la même offre que ses concurrents L-Isa et Fletcher Machine, *En-Space* permet de modifier la réverbération du site de diffusion à partir de la signature acoustique de 7 lieux renommés mondialement.

- Holophonix, de la société française Amadeus

Le processeur Holophonix, quant à lui, intègre le SPAT de l'Ircam dans son traitement du signal. Grâce à cet outil, et contrairement à ses concurrents, il laisse le choix à ses utilisateurs de l'algorithme de spatialisation à utiliser en fonction du contexte.

Le choix d'Amadeus de laisser l'utilisateur libre reflète une vision bien précise de l'espace de la part de la marque. Elle a l'habitude de construire des solutions sur mesure en fonction du lieu de diffusion. Nous pouvons ici faire un rapprochement avec la vision de Luigi Nono sur la relation entre son et espace. Pour Holophonix, les contraintes du lieu d'écoute sont les premières choses à étudier pour travailler le son et avoir le rendu optimal. Le choix de l'algorithme de spatialisation à utiliser dépend ainsi du contexte, d'où l'idée d'un processeur qui intègre tous les algorithmes.

Annexe 8 : Liste de personnes interviewées et leurs parcours.

Martin Antiphon (FR), Music Unit – Mixeur Dolby Atmos [MA]

Ingénieur du son et mixeur, Martin Antiphon a rejoint le studio Music Unit en tant qu'associé en 20 ?? après avoir été régisseur à l'Ircam. Après avoir effectué un grand nombre de mixages stéréophoniques pour de nombreux artistes renommé.es, il s'intéresse tout particulièrement au son immersif et commence à travailler en Dolby Atmos. En tant que CTO, il accompagne Music Unit dans son positionnement vers l'audio spatial et développe avec le studio différentes technologies qui rendent le son immersif plus accessible.

Hervé Déjardin (FR), Radio France – Ingénieur du son immersif live et studio [HD]

Ingénieur de son à Radio France, Hervé Déjardin travaille dans le cube, un studio équipé d'un système de monitoring à 29 enceintes et de différents systèmes de spatialisation. Il participe activement au développement de nouveaux modes de création et de production de contenus à travers les nouvelles possibilités que propose l'immersion perceptive. Il croit en l'espace comme étant un nouveau paramètre à appréhender pleinement dans l'écriture et travaille en ce sens avec de nombreux artistes comme Molécule, De Laurentis, Irène Drésel ou encore Jean Michel Jarre. Il en accompagne certain sur scène pour spatialiser leur musique à l'aide du système de diffusion spatialisée L-Isa.

Xavier Gibert (FR), Rfi Labo – Responsable innovation [XG]

C'est au moment du rapprochement entre Rfi et France 24 en 2013 que Xavier Gibert persuade la direction de Rfi de passer en multicanal. Le déménagement de la radio dans les locaux de France Média Monde la même année permet d'inaugurer 3 studios 5.1 et le binaural apparaît comme un axe de recherche privilégié : Rfi Labo est né. Les studios sont aujourd'hui labellisés Dolby Atmos et délivrent différents contenus en multicanal, parmi lesquels nous pouvons citer leur émission phare *Session Lab*.

**Olivier Goinard (FR), indépendant – Mixeur cinéma
[OG]**

Mixeur français doublement césarisé (2020, 2021), Olivier Goinard s'intéresse à la spatialisation au cinéma dès ses années de formation à l'ENS Louis Lumière dont il sort diplômé en 1998. Il a mixé depuis plus d'une cinquantaine de long métrage. Le César qu'il a reçu en 2020 pour *Le Chant du loup* récompense son premier, et à ce jour seul, mixage en Dolby Atmos.

**Augustin Muller (FR), Ircam – Réalisateur en informatique musicale
[AM]**

Augustin Muller se forme aux percussions, au jazz et à la musique improvisée avant d'intégrer la Formation Supérieure aux Métiers du Son du CNSMPD dont il sort diplômé en 2010. Il commence alors différentes collaborations dans des ensembles comme *Le Balcon* et travaille avec l'Ircam pour des projets de concerts, de recherche et de création. Il se passionne pour le répertoire électroacoustique et se pose la question de l'interprétation, de la transmission et de la préservation des travaux comportant de l'électronique et un niveau avancé d'informatique musicale.

**Antoine Petroff (CH), indépendant – Entrepreneur, ingénieur, artiste sonore
[AP]**

Formé à Lausanne comme ingénieur du son, Antoine Petroff s'intéresse très jeune au son spatialisé en pensant l'immersion directement dans son espace de diffusion. Pour cela, il n'hésite pas à imaginer et construire des enceintes originales et entièrement dédiées au projet associé. Il intègre la société Devialet à son lancement et est au cœur du développement technologique de leur modèle phare Phantom. Il collabore avec la société Ircam Amplify pour laquelle il développe la technologie « Stereo to Spatial » qui permet de rendre immersif un fichier stéréophonique. Il spatialise dans le même temps un grand nombre de contenus pour différents systèmes de diffusion.

**Éric Raynaud (FR), indépendant – Artiste sonore et numérique, développeur
[ER]**

Passionné de son spatialisé et compositeur, la pratique d'Eric Raynaud évolue entre créations sonores immersives et digitales et développement technologique. Sous le pseudonyme Fraction, il compose plusieurs albums et EPs qu'il sort notamment sur le label Infiné. Ses recherches expérimentales l'amènent à collaborer avec des laboratoires et structures internationales. Il est lauréat de nombreux concours comme celui de l'Institut Français des Arts Digitaux en 2014, de la résidence 4Dsound à Budapest en 2018 ou de la résidence artistique de recherche à l'Ircam la même année. Il développe également le *device Max for Live* et *standalone* intitulé XP, dans lequel il implémente la librairie SPAT de l'Ircam.

William Russell (AU), Monom – Fondateur, musicien, artiste sonore.

[WR]

Accompagné d'Alessandra Denegri (PE), Monom – Actrice, artiste multimédia

[AD]

William Russell, percussionniste de formation, tombe complètement sous le charme du système de diffusion 4DSound lorsqu'il l'écoute pour la première fois il y a une dizaine d'année. Il rêvait alors de manipuler l'espace sonore avec un outil qui ne serait pas seulement un système de diffusion mais qui serait également un instrument. Il intègre la technologie originelle de 4DSound au sein d'un studio, Monom, basé à Funkhaus (Berlin). Il accueille dans cet espace des artistes avec lesquels il tente de placer l'espace au cœur du processus de composition. Au début de notre entretien, William était accompagné de sa compagne Alessandra Denegri, artiste multimédia qui travaille sur la notion d'immersion.

Gerriet K. Sharma (DE), Spæs Lab – Fondateur, artiste sonore, chercheur

[GS]

C'est tout juste après sa formation en art numérique à l'*Academy of Media Arts Cologne* que Gerriet Sharma rejoint l'*University of Music and Performing Arts Graz* et perfectionne sa pratique de la composition, de l'installation sonore et de la réalisation en informatique musicale. Il est spécialiste de l'Audio Spatial depuis 2004 et a co-fondé le Spæs Lab, basé à Funkhaus (Berlin) en se concentrant sur l'esthétique du medium. Il ne s'agit pas de penser la spatialisation en termes de performance mais de mettre en avant les idées novatrices.

Lukas Wilke (DE), Deutschlandradio – Ingénieur du son Dolby Atmos

[LW]

Formé comme *Tonmeister* à la *Filmuniversity Bablesberg* (MFA) en Allemagne, Lukas Wilke officie en tant qu'ingénieur du son à la *Deutschlandradio* à Berlin. Il y pratique le mixage en Dolby Atmos aussi bien pour des fictions radio que pour de la musique classique. Sa pratique musicale personnelle et sa passion pour le son se traduit également par un grand nombre de projets extérieurs.

Annexe 9 : Formulaire de consentement aux entretiens

FORMULAIRE DE CONSENTEMENT A L'ENTRETIEN

Introduction et présentation du projet : Dans le cadre de mes études à l'ENS Louis Lumière, je, soussigné Vincent Meurisse, l'auteur, réalise un mémoire qui s'intitule « Tentative d'établissement de l'écriture sonore immersive. »

Ma recherche s'intéresse aux pratiques des ingénieurs du son et artistes qui travaillent le son spatialisé et tente de mettre en lumière les similitudes et divergences des différentes approches au regard d'un travail bibliographique. Ce travail prend sa source dans le constat qu'un grand nombre de professionnels de l'audio spatial travaillent aujourd'hui de manière empirique et en utilisant leurs bases stéréophoniques, et que la grammaire de l'écriture sonore immersive n'a pas encore été véritablement formalisée. Mes entretiens serviront à la tentative d'en poser les fondements.

Déroulé et de l'entretien : L'entretien que je réalisera avec vous, participant, prendra la forme d'une discussion qui sera enregistrée. Je vous poserai des questions sur votre pratique du son spatialisé, vos habitudes, votre vision.

Droit du participant : La participation à cette étude est sur la base du volontariat, la participant a la possibilité de quitter l'étude à n'importe quel moment sans fournir d'explications. Les enregistrements seront retranscrits mot pour mot en annexe de mon travail, en intégralité ou en partie, et serviront de support à mon analyse. Il est possible que ce mémoire devienne accessible publiquement. Il convient donc au participant de demander à rester anonyme s'il le souhaite.

Merci de cocher les cases correspondantes à vos souhaits :

- J'ai lu la description du projet de recherche et par la présente, accepte d'y participer. J'ai pris connaissance des conditions d'utilisations des résultats de cet entretien.
- Je donne mon consentement à l'enregistrement, la transcription et l'utilisation de certaines citations de cet entretien dans le cadre du mémoire de l'auteur.
- Je souhaite rester anonyme.

Date :

Signature du participant :

Signature de l'auteur :

CONSENT FOR THE USE OF AN INTERVIEW

Project presentation: I, Vincent Meurisse, the author, am writing a MASTER thesis called “Attempt to establish a grammar of spatial audio writing”. This work is being made as I am a student of the Paris based sound engineer school ENS Louis Lumière. For this research, I am interested in the practices and habits of sound engineers and artists working with Spatial Audio and I aim to find similarities or divergences between the different approaches, which can be considered as being as numerous and personal as the number of people working in the field. My work derived from the fact that most of the professionals are working with the only help of their own experience and if the stereo grammar seems to be known and applied, even in the spatial realm in an extended way, the grammar for immersive audio is yet to be formalized.

Form of the interview: The interview will be a discussion between you, the participant, and me, the author. It will be recorded. I will ask you questions about your way of working with spatial audio and your vision.

Participant's rights: You are considered as a volunteer, and you have the possibility to stop this interview any time without notice. The entire or parts of the recording will be written down and will be the base of my analysis. My thesis might be made public. You have the right to be anonymous.

Please, fill the right cases:

- I have read this paper and I accept this interview.
- I give my consent for the author to use the recording and to write it down.
I also give my consent for the author to use my words in his work and to quote some of my sentences.
- I wish to be anonymous.

Date:

Signature from the participant:

Signature from the author:

Annexe 10 : Guide d'entretien

GUIDE D'ENTRETIEN		
THEMATIQUE	QUESTIONS	DEVELOPPEMENT
Introduction sur l'immersion	<p>Qu'est-ce que le son immersif selon toi ?</p> <p>Quelle est ta vision vis-à-vis de l'immersif et de son futur ?</p> <p>Dans quel format immersif travailles-tu et pourquoi ?</p>	<p>Y a-t-il des formats que tu trouves plus pertinents dans leur rendu ?</p>
La méthodologie d'écriture	<p>A quel moment intervient le travail de l'espace dans le processus d'écriture ?</p> <p>Les artistes ont-ils des propositions spatiales ? Si oui lesquelles ?</p>	<p>Quelles sont les différences que tu remarques dans ta manière de travailler entre une situation où tu pars d'un stéréo et une où tu composes l'espace dès le début du processus créatif ?</p> <p>Est-ce que tu considères qu'il y a un manque de démocratisation de l'espace ?</p>
Le contexte de mixage	<p>Quelles compétences le mixage immersif demande en plus d'un mixage stéréo ?</p> <p>Comment tu appréhendes la localisation et le mouvement ?</p> <p>Quelle est ta relation avec la frontalité ?</p>	<p>Quelles difficultés rencontres-tu ?</p> <p>Est-ce que ça demande pas de manier les outils classiques (EQ, delay, compression) différemment ?</p> <p>Qu'est-ce que tu mets dans les arrières ?</p> <p>Qu'est-ce que tu mets dans les tops ?</p>

Annexe 11 : Les philosophies de chacun

A11.0 Avant-propos : la vision de Frédéric Changenet

Il nous apparaît important d'introduire ces « portraits » par celui de Frédéric Changenet, directeur externe de ce mémoire, avec qui nous avons échangé longuement pour l'élaboration de ce sujet et son bon déroulement. Ingénieur du son à Radio France, il est l'une des références du son spatialisé au sein de la Maison de la Radio et travaille notamment sur le mixage immersif de fictions et documentaires. La manière dont on construit l'espace sonore à 360° est une question fondamentale qu'il se pose régulièrement.

Selon Frédéric, le son immersif est sans aucun doute le futur de l'audio. Il intervient lorsqu'un système de diffusion est capable de reproduire des sources sonores provenant de toutes les directions, et il a pour objectif de se rapprocher de la perception naturelle. L'audio immersif est selon lui défini selon 3 grades :

GRADE 1 :

- Scène orientée
- Pas de rôle sémantique de la spatialisation
- Musique Classique, certaines fictions et documentaires radio
- Son “HD”, meilleure transparence

GRADE 2 :

- Scène orientée
- Rôle sémantique de la spatialisation
- Fictions radio, musiques spatialisées
- Donne de nouveaux outils pour raconter des histoires

GRADE 3 :

- Scène non orientée, rôle sémantique de la spatialisation
- VR, Live électronique
- Interaction, écriture sonore très différente

En fonction du projet et du grade qu'il implique, la construction de l'espace sonore ne sera pas appréhendée de la même manière. Certaines réflexions restent constantes dans sa manière de travailler. Il tente notamment d'abolir la hiérarchie entre la scène

frontale et le *surround*. Et si la majorité des contenus semblent aujourd’hui encore soumis à une certaine symétrisation des éléments qui les constituent, il essaye de passer outre cette habitude et se pose systématiquement la question de l’asymétrie. Enfin, la problématique de la corrélation des signaux sonores dans la restitution à plusieurs haut-parleurs est selon lui primordiale.

Sa construction de l'espace passe par différents éléments et préoccupations :

- La scène est constituée d'un fond immersif constitué d'ambiances, de nappes, de certaines sources mobiles et fixes, qui ancrent la scène dans un environnement.
- Viennent s'y ajouter des objets sonores, éléments ponctuels, qui amènent la narration. Ils peuvent eux aussi être fixes ou en mouvement en fonction de leur rôle dans le contenu.
- Ces deux premiers éléments constitutifs du champ sonore baignent dans une réverbération multicanale, associée au système de diffusion, qui vient apporter du liant et former une cohérence dans la scène sonore globale.
- Le système d'écoute cible (celui qui sera le plus utilisé pour la majeure partie des écoutes) et la compatibilité des systèmes va avoir une influence directe sur la manière de mixer.

Nous admettons qu'un entretien, similaire à ceux que nous avons réalisés dans le cadre de ce mémoire, avec Frédéric aurait été particulièrement pertinent pour compléter ce « portrait » trop succinct et nourrir notre grammaire.

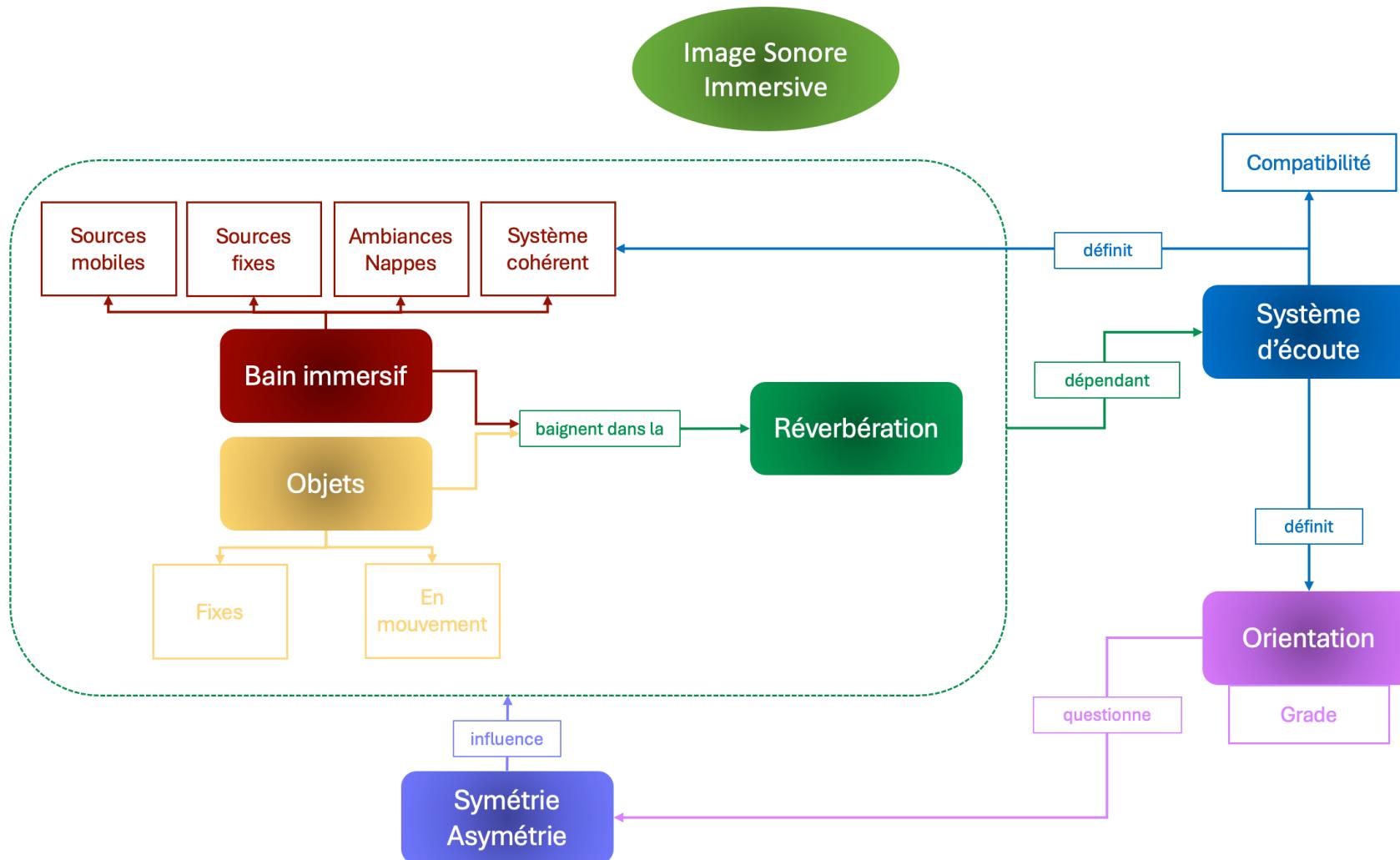


Figure 7 : Conception d'une image sonore immersive selon Frédéric Changenet, d'après la compréhension de l'auteur

A11.1 Martin Antiphon

Interprétation de l'auteur et schéma associé, envoyés à Martin Antiphon :

« Lorsque tu t'apprêtes à travailler sur un mixage immersif, la première question que tu te poses est celle du contenu. De quel style musical s'agit-il ? Est-ce un interlude ou un single ? Comment la pièce est-elle écrite ? Cela va jusqu'à la source. Sa spatialisation ne sera pas la même en fonction de son rôle.

La musique classique a une place à part car c'est un style très codé, et ces codes doivent être respectés. Ton premier objectif est de procurer une impression de réalisme à l'auditeur. Pour cela, tu n'hésites pas à positionner des couples en élévation, pour apporter de l'air et faire respirer le mix. Il est aussi très important de respecter le mixage stéréophonique. Un mix de musique classique immersif peut passer par l'ajout de réverbération latérales. Il faut cependant les utiliser avec modération pour ne pas modifier les valeurs de plan qui existent en stéréo.

D'une manière générale, quand tu travailles un mix spatial, les premières étapes passent par le déploiement des pistes, dans l'idée de faire un mélange spatial qui créera une matière enveloppante. Comme en stéréo, cela passe par dans un premier temps par le volume, puis par le positionnement des sources, cela sans reverbs ou effets particuliers. Tu te réfères ensuite à tes connaissances théoriques et empiriques en termes de perception : positionner un soliste en hauteur ne fonctionnera pas, faire bouger plus de 3 ou 4 sources en même temps ne sera pas perçu par l'auditeur...

Dans ta pratique du mixage immersif, tu pars toujours de quelque chose qui se tient musicalement, issu en général d'un travail stéréophonique effectué en amont. Tu aimerais cependant que la spatialisation soit pensée dès l'écriture et tu défends l'idée que la spatialisation est une donnée aussi importante que la hauteur ou le rythme. La problématique la plus importante aujourd'hui, maintenant que les technologies sont mises à disposition, est de donner les bons outils aux artistes pour qu'ils puissent penser l'espace dans le processus créatif.

Selon toi, le mixage immersif n'est pas plus facile que le mixage stéréophonique. Il est au contraire plus difficile car il y a tout un nouvel espace à remplir. Alors qu'en stéréophonie, le masquage sert justement la fusion, la disparition de l'effet de masque en spatial va rendre toutes les pistes plus lisibles et il sera ainsi plus difficile de les traiter correctement.

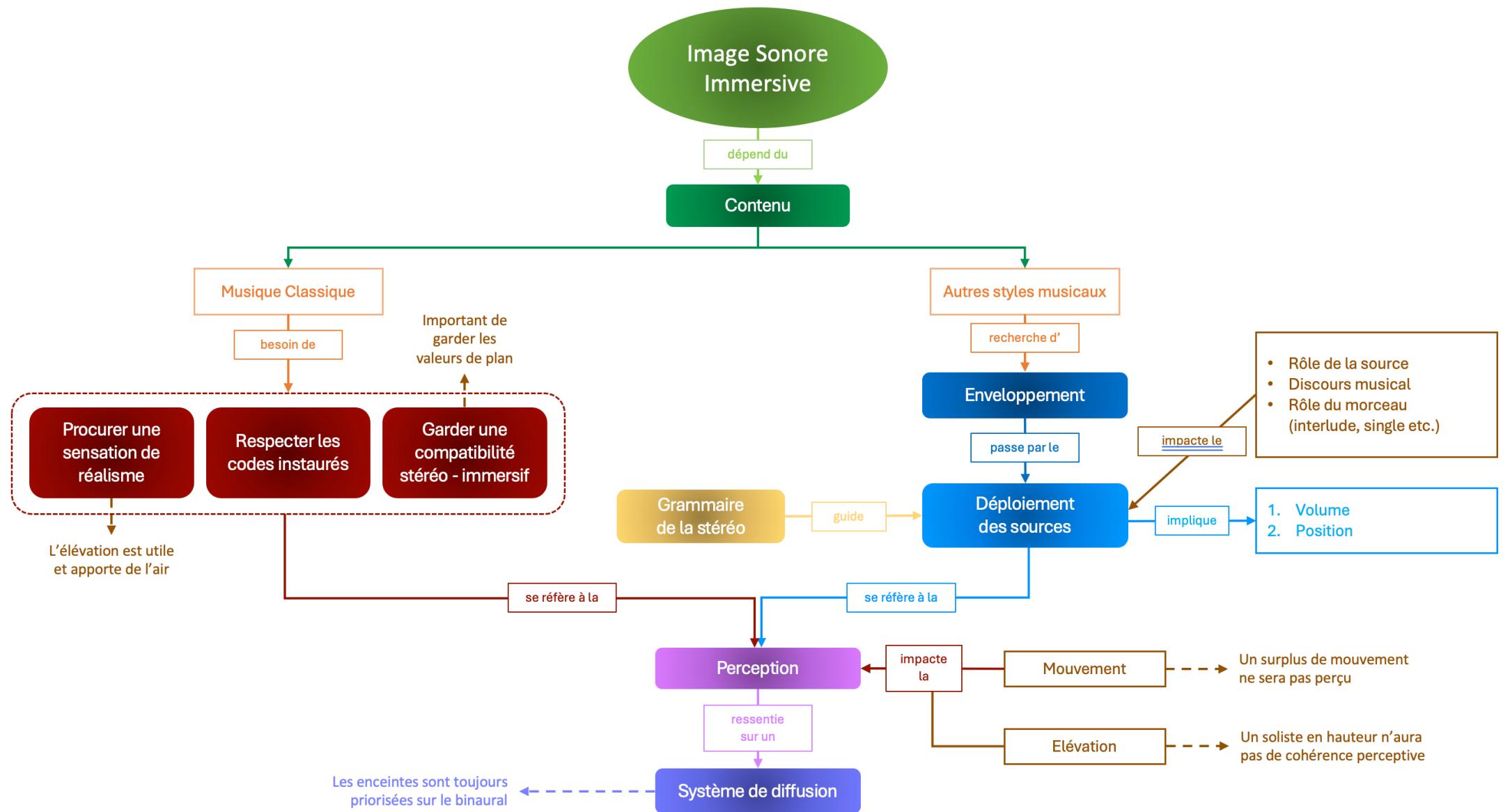


Figure 8 : Conception d'une image sonore immersive selon Martin Antiphon, d'après la compréhension de l'auteur

A11.2 Olivier Goinard

Interprétation de l'auteur et schéma associé, envoyés à Olivier Goinard :

« Avant tout, quel que ce soit le format de diffusion utilisé, le son au cinéma est au service de la dramaturgie. La spatialisation sert ainsi elle aussi à répondre à l'enjeu cinématographique. Le Dolby Atmos, en ce sens, représente pour toi une véritable évolution qualitative grâce à 3 paramètres qui sont beaucoup plus précis : le timbre, la dynamique, et l'espace. La précision de localisation est plus élevée, et le fait de travailler dans cet espace permet une discrimination des sources plus importante. La répartition devient plus intéressante et le mix en est plus facile.

Il est également important pour toi de savoir que tu peux te fier à la diffusion en salle. Si en 5.1 on n'est jamais sûr de la qualité du *surround* dans une salle quelconque, il est pertinent de savoir que le rendu Atmos sera maîtrisé dans une salle labellisé Atmos. Cela permet notamment d'assumer beaucoup plus le hors champ, donc les arrières et l'élévation. En 5.1, cela a pour conséquence de n'envoyer que du très diffus dans les *surround*, en complément de la façade. Pour le Chant du Loup, tu n'as pas hésité à envoyer le son à l'arrière dans les moments d'écoute attentive.

Tu considères que l'Atmos est particulièrement séduisant dans les scènes d'action. Les différents éléments cohabitent ensemble plus facilement et il est aisément de se servir de tout le système pour des mouvements rapides ou des effets impressionnantes. Il est tout aussi performant dans des scènes plus intimes, où les ambiances peuvent considérablement gagner en qualité.

D'une manière générale, il est important d'utiliser la spatialisation avec parcimonie, toujours dans l'idée de servir la narration. Il est primordial pour toi de jouer sur la dynamique de spatialisation. Certains passages peuvent être en mono, d'autres en stéréo, puis d'autres utiliseront la spatialisation du système au plus fort de ses performances. Ce sont ces variations qui vont provoquer les sensations chez l'auditeur. »

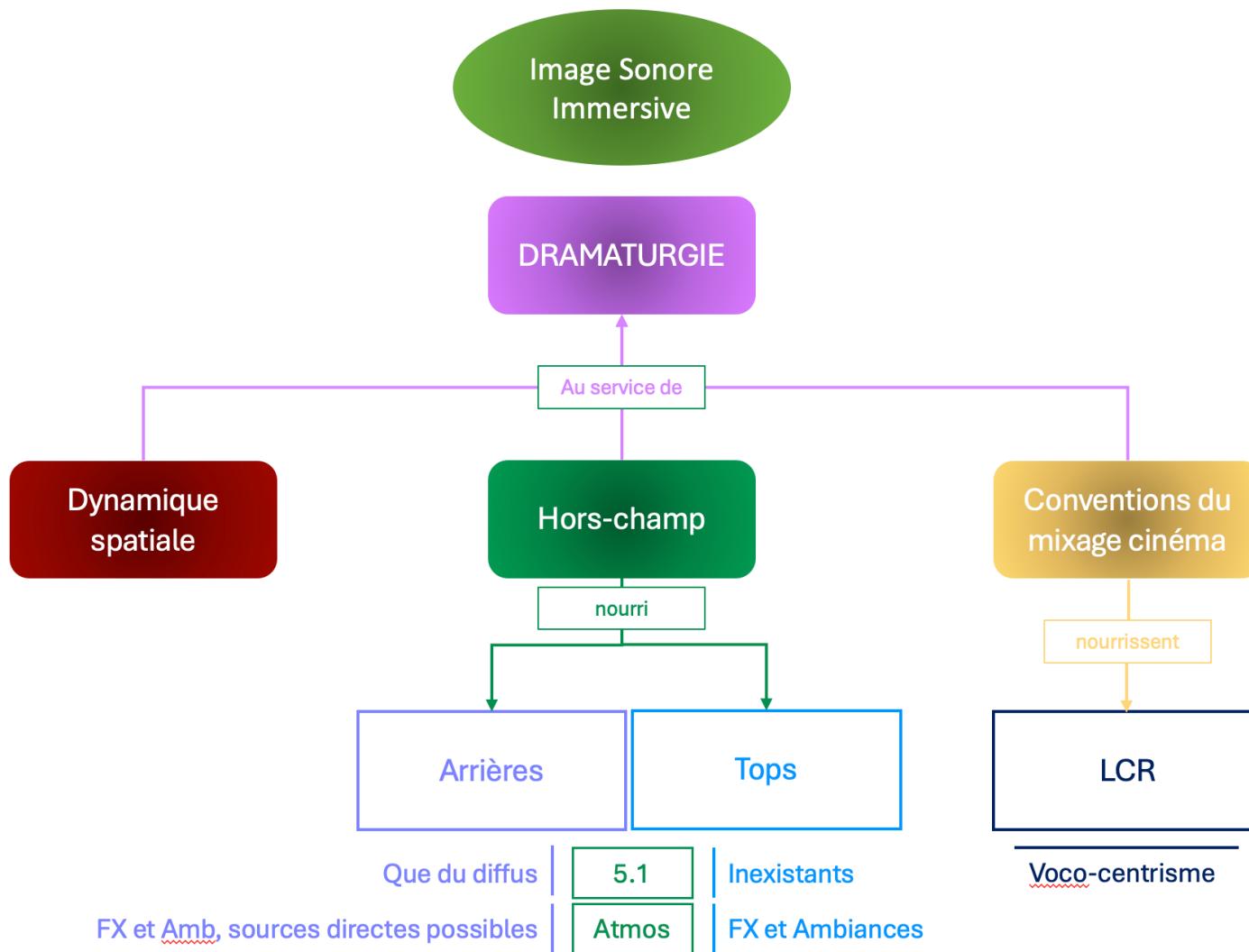


Figure 9 : Conception d'une image sonore immersive selon Olivier Goinard, d'après la compréhension de l'auteur

A11.3 Augustin Muller

Interprétation de l'auteur et schéma associé, envoyés à Augustin Muller :

« Pour toi, le terme immersif pose question car il te ramène à la notion d'illusion. Tu soulignes qu'aujourd'hui, on associe immersif à la diffusion de son spatialisé, mais il ne faut pas oublier qu'un concert acoustique est très immersif. Cette notion de salle est d'ailleurs importante. Selon toi, avoir un objectif d'immersion en faisant fi du lieu d'écoute est problématique car la salle sera toujours présente, avec ce qu'elle implique. Dans un contexte de live, il est nécessaire d'établir une relation entre le lieu où sont les gens physiquement et là où on les amène mentalement mais il ne peut pas y avoir de dissociation entre les deux.

L'espace sonore est pour toi un espace expressif en soi. Il est pertinent de le travailler dès l'étape de création, à l'image de ce que font les compositeurs et compositrices de musiques électro-acoustiques depuis les années 50. En ce qui concerne la grammaire de l'écriture spatialisée, tu considères d'ailleurs que ces personnes-là ont apporté beaucoup d'éléments de réponse, mais que les résultats de leurs recherches ne sont pas explicités. Ce manque de grammaire commune, couplé au fait que les ingénieurs du son utilisent dans la phase de mixage des notions de psychoacoustiques connues, entraîne une homogénéisation des contenus et de la manière dont on travaille l'espace. D'une manière générale, le grand public, et même les professionnels, manquent encore de connaissances et n'ont pas la culture du son spatialisé, à l'image du travail effectué avec Richie Hawtin qui n'est pas allé très loin.

Dans ta pratique de la spatialisation, tu attaches particulièrement d'importance à l'acoustique et à l'équilibre dynamique. La réverbération vient en premier dans la construction de l'image sonore immersive, tu cherches à travailler la profondeur et les valeurs de plans. Tu favorises pour ça des systèmes de prise de son cohérents plutôt que de la spatialisation artificielle. Tu t'en fiches de la position de tes sources ou des trajectoires, sauf si le contexte le demande précisément. Dans ta palette créative, tu aimes aussi jouer avec le hors champ, et il y a chez toi un grand intérêt pour la scène frontale. Remplir l'espace à 360° de manière homogène ne fait pas sens et signifie même la perte d'un moyen d'expression. Cela rappelle ton rapport à la salle et son importance : dès qu'il y a quelque chose à voir, ce n'est plus homogène. Il faut le prendre en compte et s'en servir. »

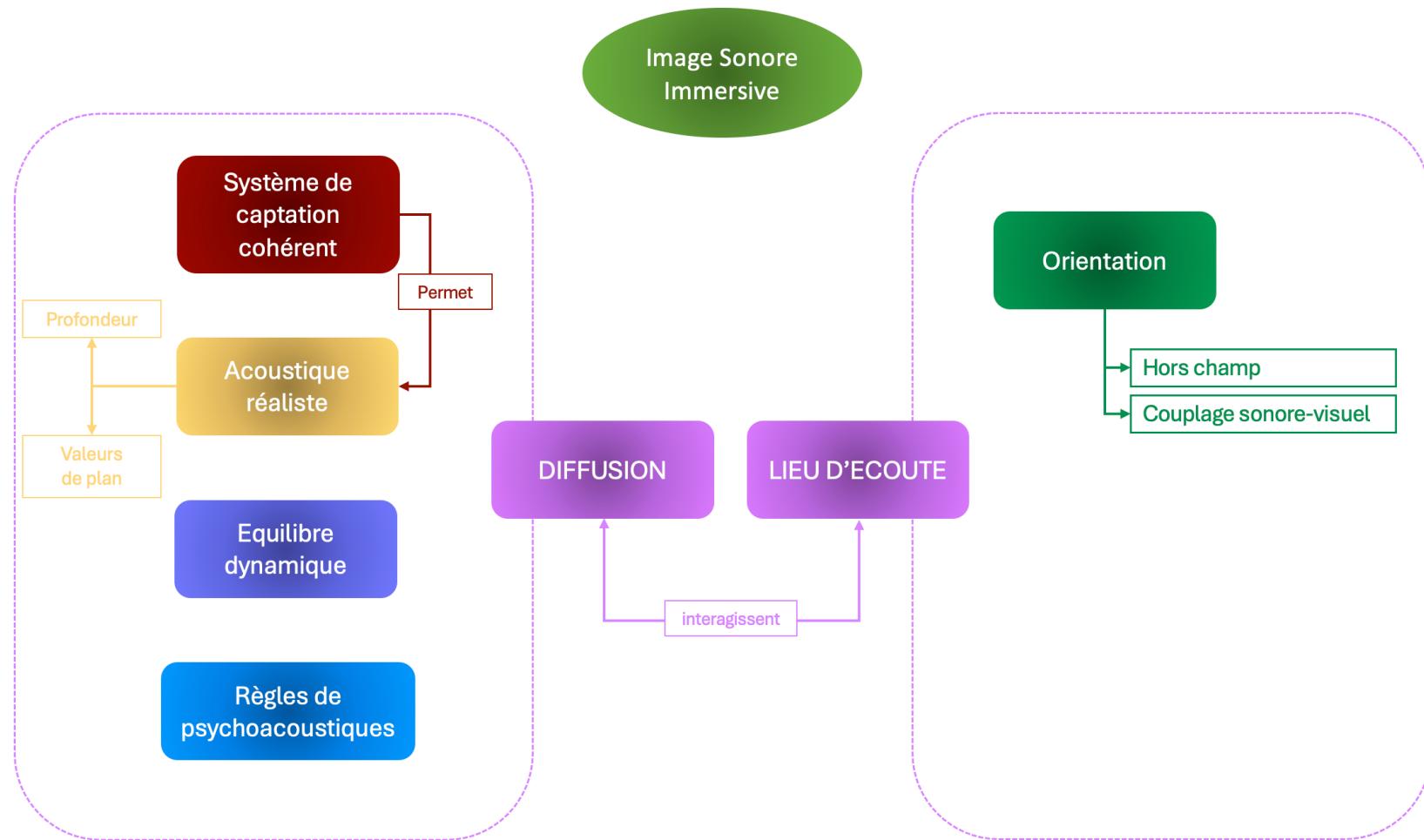


Figure 10: Conception d'une image sonore immersive selon Augustin Muller, d'après la compréhension de l'auteur

A11.4 Antoine Petroff

Interprétation de l'auteur et schéma associé, envoyés à Antoine Petroff :

« Avant d'entrer dans le détail de la conception d'une image sonore immersive, il est important selon toi de définir précisément le terme "immersif". Tu précises que l'immersif n'est pas forcément lié à la 3D et que beaucoup de facteurs jouent sur l'immersion. Un des éléments importants à prendre en compte est notamment le rôle important des infrabasses.

Pour toi, une des choses les plus importantes lorsque tu travailles dans un contexte immersif, particulièrement dans un contexte d'installation, est de faire disparaître la technique. Lorsque celle-ci n'est plus visible, ce sont les émotions qui prennent. Le mot émotion est d'ailleurs un terme qui revient plusieurs fois dans ton discours, et semble diriger la création de l'image sonore. Si l'immersion tend à recréer les conditions physiques de la réalité, ce sont les émotions que l'on fait passer dans le son qui vont avoir un impact sur l'auditeur, et créer l'immersion, aussi bien en stéréo que dans des systèmes "immersifs" d'ailleurs.

D'une manière générale, tu aimes jouer avec l'acoustique des lieux et le positionnement des enceintes. Tu n'aimes pas utiliser les systèmes de diffusion classiques et tu vas souvent même jusqu'à concevoir des enceintes spécifiquement pour un projet et un lieu précis. Tu n'as pas peur de retourner des enceintes contre les murs pour jouer sur le champ diffus. Ajouter des haut-parleurs permet de te rapprocher d'une sensation réelle, que ce soit pour reproduire la véritable largeur d'une source ou pour assumer les réflexions, même si le rendu global sera moins précis en termes de localisation.

Dans ta manière de créer les images sonores immersives, tu parles de certaines astuces acquises avec l'expérience :

- Les sons avec de l'impact, ou au premier plan narratifs, doivent être frontaux, et avec une largeur limitée. Le kick, et la batterie plus généralement, qu'elle soit acoustique ou électronique, en font partie.
- Souvent, tu mixes certains éléments à la manière stéréophonique, pour leur donner une couleur particulière.

- En ce qui concerne les mouvements, tu utiliseras des déplacements peu perceptibles, ou des trajectoires aléatoires, à l'image de ce qui se passe dans la vraie vie où en permanence les choses bougent autour de nous. Cela joue d'ailleurs en faveur du rendu binaural, qui sera plus réaliste
- Tu fais une différence entre les sons ponctuels, que tu localise et fais interagir avec des *rooms*, des délais, des *flanger* etc. et les textures ou nappes, que te n'hésite pas à envoyer dans toutes les directions, dans des zones beaucoup plus larges, et jamais sur un seul HP.

Pour toi, il serait beaucoup plus pertinent, dans nos formats de diffusion, de séparer les sources en fonction de leur rôle. Il faudrait penser en termes de couches. Concrètement ça n'a pas de sens d'encoder une voix, une reverb et un instrument de la même manière. Il faudrait que la notion spatiale soit présente dans les formats. »

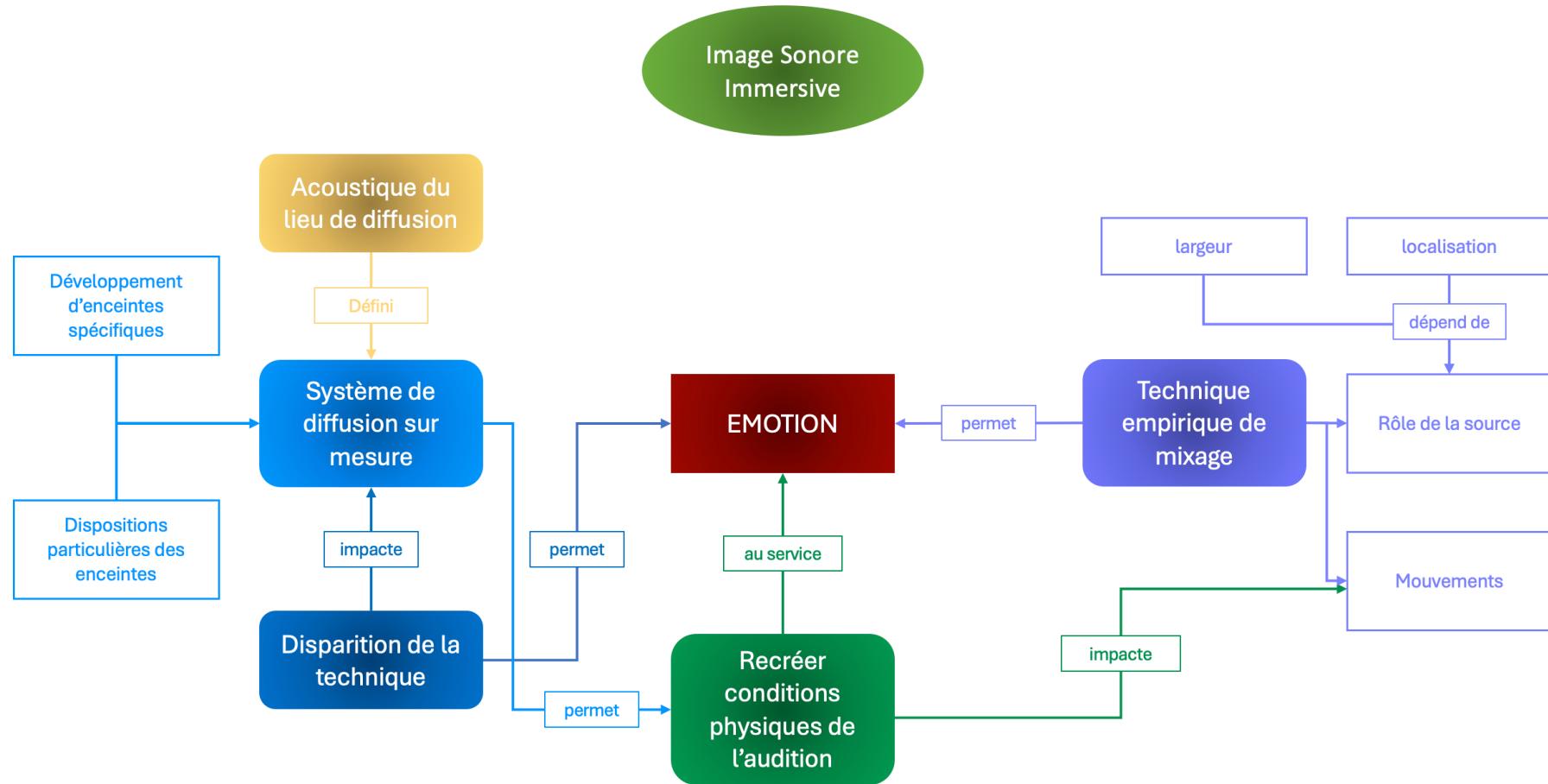


Figure 11 : Conception d'une image sonore immersive selon Antoine Petroff, d'après la compréhension de l'auteur

A11.5 Éric Raynaud

Interprétation de l'auteur et schéma associé, envoyés à Éric Raynaud :

« Le son spatialisé tend à reproduire une sensation réelle d'écoute, et l'artiste intervient dans la déconstruction, reconstruction de la matière pour provoquer de nouvelles sensations chez l'auditeur.

Selon toi, mixer de la musique populaire, actuelle, en immersif n'a pas de sens. Certains contenus doivent rester en stéréo, d'autres même en mono. Il n'y a pas d'intérêt à localiser une source à l'arrière ou sur les côtés, à créer du mouvement, à jouer sur le spatial, si c'est simplement pour construire une extension du domaine stéréophonique.

De ton point de vue, l'intérêt du son spatialisé réside dans l'exploration complète et sans à priori de ce nouveau paramètre – l'espace – qui s'offre à nous. Il n'y a pas de règles à suivre dans ta conception de l'image sonore immersive. Il faut expérimenter. La construction d'une image sonore à 360°, dans ta pratique, résulte justement d'un travail de recherche, et finalement de jeu du hasard. A l'image de ta pratique de la musique électronique qui prend racine dans la musique expérimentale, tu as toujours allié recherche et composition.

Si tu considères qu'il n'y a pas de limites à l'imaginaire spatial, tu admets trouver de la difficulté dans la construction d'une image sonore à 360°. Une des premières questions que tu te poses est relative à l'énergie. Il faut remplir l'espace sonore afin de submerger l'auditeur. Le fait d'avoir quelques sources ponctuelles à des localisations précises ne suffit pas. Il faut avoir de la matière. Dans ce sens, tu t'intéresses particulièrement à un travail de texture, de granularité, de densité. Ton travail peut être défini comme un travail qui place le timbre au sommet de la hiérarchie des paramètres sonores à travailler. Le filtrage, le *stretching*, les délais sont des traitements privilégiés.

Dans ton travail de l'image sonore immersive, tu considères que de séparer le signal en un *bed* et des objets est une approche pertinente. Le *bed* en *channel based* peut justement permettre de répondre à la problématique de l'énergie. Il sert à asseoir la structure globale de la création. Les objets offrent une liberté qui viendra compléter la composition et jouer sur l'espace avec précision.

Les difficultés rencontrées se trouvent dans la technique. Les outils sont encore difficiles à manipuler et trop complexes (*Spat Revolution*, *Panoramix*, *L-Isa*). Ils sont puissants et leur existence est une bonne chose pour la démocratisation de l'immersif mais ils ne sont pas adaptés aux besoins des artistes qui cherchent à explorer, expérimenter dans un temps court, poser une idée rapidement. C'est pour ça que tu ressens le besoin, et que tu considères important, de développer ses propres outils. XP découle de cette réflexion.

En PJ, un schéma des paramètres qui interviennent dans ton processus de création d'une image sonore à 360°. »

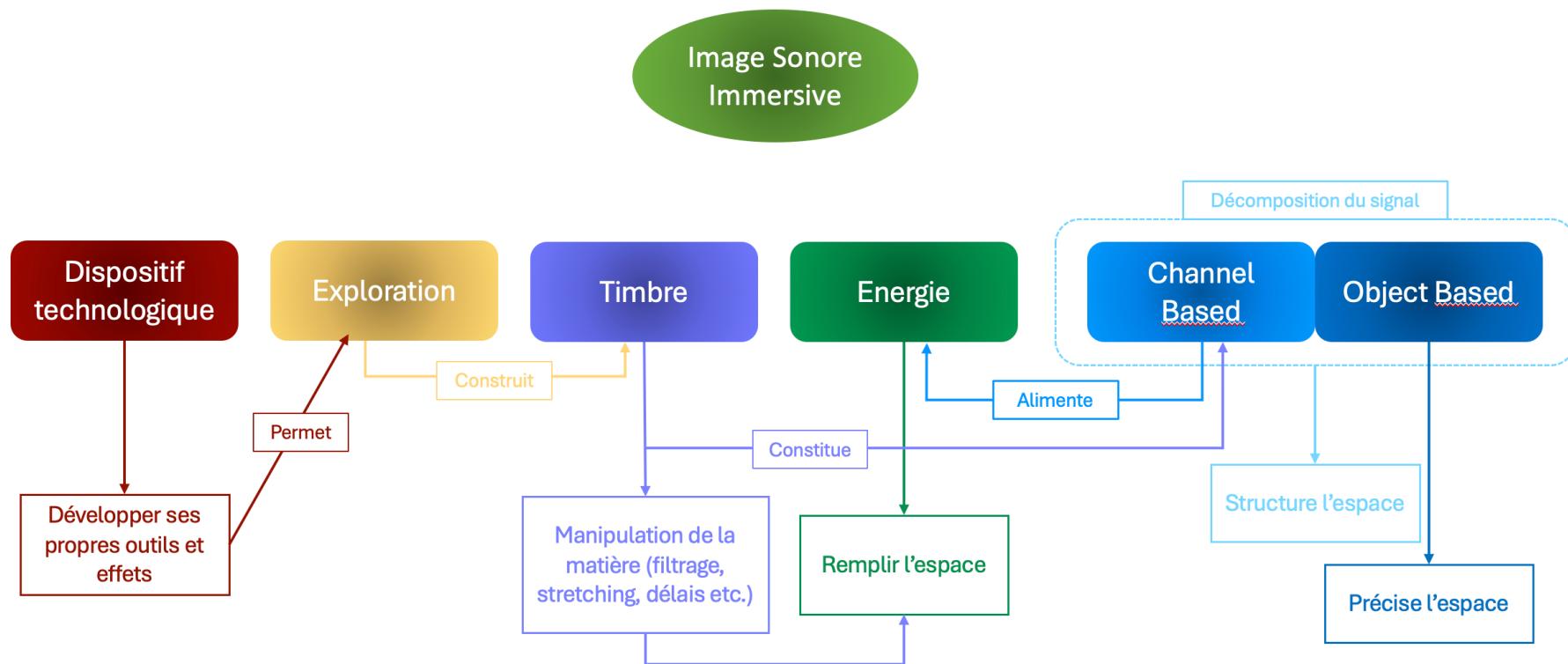


Figure 12: Conception d'une image sonore immersive selon Éric Raynaud, d'après la compréhension de l'auteur

Annexe 12 : Retranscriptions des entretiens

Nous prions Alexandre Garcia et Jérémie Besset de bien vouloir nous excuser de ne pas avoir trouvé le temps nécessaire à la retranscription et l'analyse de leurs entretiens. Nous les remercions une nouvelle fois pour le temps qu'ils ont bien voulu nous accorder.

Nous précisons également qu'un problème technique est arrivé durant l'enregistrement de l'entretien d'Éric Raynaud et que nous n'avons pu retranscrire qu'un tiers de notre discussion. Nous en sommes particulièrement attristés, tant le regard critique d'Éric Raynaud sur la manière actuelle d'appréhender l'espace par le plus grand nombre était pertinent pour notre sujet. Nous avons tenté de retranscrire au mieux sa vision dans notre interprétation disponible en A11.5 p.136. Nous le prions, ainsi que le lecteur, de bien vouloir nous excuser pour cet incident.

A12.1 Entretien avec Martin Antiphon

V : Présentation du mémoire.

M : Tu es allé voir tout ce qui était travaux de musique contemporaine, de Stockhausen et de ces gens-là ?

V : Je n'ai pas eu véritablement l'occasion d'écouter beaucoup d'œuvres de musique contemporaine non. Et puis il y a la problématique du contenu, qui a été pensé pour un espace donné dans beaucoup de cas, qui pose la question de la manière dont on réécoute ces créations et dont on les appréhende dans un contexte différent, notamment au casque.

M : A part en concert, c'est compliqué. Après, cela dit, tu pourrais te rapprocher, pour juste compléter l'état de l'art, de RIM qui jouent ces pièces là pour voir comment ils les approchent. Est-ce que c'est renseigné dans la partition ? Comment est-ce qu'ils se positionnent par rapport à cette question de la re-jouabilité ? Tu vois, tout un tas de trucs comme ça. Par exemple dans mon mémoire, j'avais parlé de *Dialogue De L'ombre Double* (Boulez, 1985) où t'avais sur la partition des flèches qui te disent dans quel haut-parleur il faut envoyer le son et à quel moment, un peu en mode Acousmonium où vraiment il y a quelqu'un qui interprète. Après, ce n'est pas la question que toi tu poses. La question que tu poses c'est « c'est quoi le vocabulaire qu'on utilise ? ».

L'autre chose à laquelle j'ai pensé en lisant ton sujet c'est au travail de Pierre Schaeffer avec le *Traité des objets musicaux* (Schaeffer, 1966) qui a essayé de faire la même chose mais sur les éléments sonores. Voilà je t'expose ce à quoi ça m'a fait penser. Nécessairement il y a des choses qui ont été pensées mais tu le soulèves très justement, c'est peut-être appliqué à des cadres qui sont plus ceux dans lesquels on est, dans les paradigmes dans lesquels on est aujourd'hui quand on parle de son immersif.

V : En tout cas, des RIM de l'Ircam je vais en rencontrer, notamment pour discuter du travail qui a été fait avec Richie Hawtin⁹⁸.

[Présentation de l'entretien et des questions]

M : On peut le faire aussi en deux temps, ce sont des sujets sur lesquels j'ai mon avis et c'est intéressant qu'ils soient challengés parce que je vais donner mon avis. Et si ça se trouve quand tu vas parler autour de moi, tu vas dire « en fait il est le seul à avoir dit ça et les autres ils ont dit ça » et je vais me dire que j'ai une vision tronquée. Je n'en sais rien. Je trouve ça intéressant même que moi tu réagisses, et ce sera intéressant de rediscuter par curiosité et pour bavarder. Il y a plein de choses que tu ne pourras peut-être pas mettre dans ton mémoire, parce que dans un mémoire, on est obligé d'être efficace. Et quelquefois ce n'est pas qu'on emprunte des raccourcis, c'est qu'il y a des sujets et des théories qui sont compliqués à étayer et qu'on ne décrit pas dans le mémoire, et des choix que tu vas faire dans ta tête. Ça m'intéressera.

⁹⁸ Fait référence ici à l'entretien avec Augustin Muller, qui a accompagné Richie Hawtin pour sa performance spatialisée à l'Ircam dans le cadre de la série de concert Électro-Odissée en 2024
<https://www.ircam.fr/agenda/electro-odyssee-fuse/detail>

Ce que je peux dire en préambule, c'est que la question de faire du son immersif, l'enjeu de l'écriture spatiale ce n'est pas que c'est obligatoire ou que c'est dans l'aire du temps, c'est que c'est quelque chose qui est en train d'arriver. C'est que les formats, ils sont en train de se mettre en place, que ce soit dans le broadcast ou dans l'industrie musicale, donc on a cette ouverture de l'espace qui est désormais accessible aux créateurs de contenus. Alors c'est vrai que pour l'instant, si tu veux distribuer un contenu Dolby Atmos avec juste du son dans le LR, chez Apple tu te fais rejeter. Mais à la limite ça pourrait être ta liberté, à termes, d'utiliser ce format spatial pour construire un mixage qui serait que frontal. Ce que je veux dire c'est que le format et les outils sont là. La question c'est qu'est-ce qu'on en fait et quels outils on a pour l'exploiter ? Donc ça je pense que c'est important. C'est à dire que nous, pourquoi on a fait cette démarche à Music Unit ? C'était quelque chose qui était dans la continuité logique de notre démarche, parce qu'on faisait déjà du son immersif en installation sonore et en musique contemporaine et qu'on avait une autre branche d'activité avec des associés. Et notamment moi où j'étais sur les deux branches et je faisais plein de mixage en stéréo et je rêvais que d'une chose c'est de voir ce qu'allait apporter la construction de l'espace sur des musiques qui ne l'avaient jamais questionnée, parce que sur les musiques contemporaines c'est évident. Même en musique de film finalement.

En musique de film on fait de la spatialisation depuis toujours, c'est juste qu'on met un peu des ambiances et qu'on se sert de la spatialisation pour donner une sensation d'immersion ou du réalisme. Ceci-dit c'est exactement ce que je fais quand je fais un mixage de musique classique aujourd'hui en Dolby Atmos. Mon but c'est de donner une sensation de réalisme. Je n'utilise pas l'espace autrement. On pourrait bien, mettre le spectateur au milieu de l'orchestre ... d'ailleurs je travaille avec un quatuor à cordes qui a envie d'interroger l'espace sur de la musique de répertoire et je leur ai dit « Super ! c'est de la recherche musicale, de la recherche de création. »

Ce que je veux dire c'est que tout ça est accessible, ça n'est pas obligatoire mais en même temps c'est un champ des possibles qui est désormais accessible. Voilà. C'est juste pour dire que c'est le chemin de l'histoire. À un moment il y a la stéréo qui est arrivée, les gens ils se sont dit « ça sert » et aujourd'hui tu n'achèterais pas une chaîne hi-fi de qualité s'il n'y a pas deux enceintes. Peut-être que dans vingt ans plus personne n'acceptera un format stéréo. Va savoir.

V : Est-ce qu'on ne peut pas dire que ça avait déjà été essayé à l'époque avec le 5.1 mais que ça n'avait pas marché ?

M : Oui mais en réalité attention. Ce qui a été essayé quand on dit le 5.1 c'est « l'industrie musicale distribue des contenus au format 5.1 au grand public ». C'est ça qui n'a pas marché. En revanche, les musiques mixées en 5.1 ou en 7.1 on fait ça depuis... Je ne sais pas quand c'est vraiment arrivé dans l'industrie du cinéma de faire du multicanal en musique de film, mais ce que je veux dire par là c'est que ça a été un échec parce que le modèle économique et le modèle commercial étaient pas bons. C'est-à-dire que les gens, il fallait qu'ils achètent de nouveaux lecteurs CD, des nouvelles enceintes, qu'ils mettent des enceintes partout chez eux, pour une qualité que les gens ne perçoivent pas forcément. Et aujourd'hui ce qu'il se passe, ce n'est pas que les gens vont migrer ou pas migrer vers le son multicanal, c'est que ça va être fonctionnel dans tous leurs appareils, et ils n'ont rien à faire. Si la télé se met à diffuser et à embarquer de l'AC4 donc le format Atmos, ou d'autres codecs, le contenu il est là. Et en plus aujourd'hui

c'est un contenu qui est capable de s'adapter au système de rendu. Alors tout ça, ça se met en place petit à petit mais ce sont des choses qui vont être transparentes pour l'utilisateur. Alors tu auras toujours les geeks qui auront envie de se faire les super systèmes à la maison mais eux ils avaient déjà acheté des 5.1 pour leur Home Cinema. Pour tous les autres, ils veulent juste écouter de la musique dans de bonnes conditions. Les codecs, les machins, le multicanal, ils s'en fichent. L'important c'est que ce soit agréable, que l'expérience soit là. Et clairement ce qu'on peut faire avec le multicanal ce sont des expériences plus poussées qu'en stéréo. Alors après il y a le débat est-ce que ça sonne mieux, moins bien, mais ça à la limite ce sont des problèmes de contenus, ou de façon de les restituer mais c'est une adéquation qui ne va pas influencer le consommateur. Le consommateur il achète des AirPods, ça fait du spatial, c'est cool !

V : Tu dis que les gens avec un 5.1 chez eux ne voyaient pas forcément la différence avec la stéréo. Aujourd'hui le son spatial est accessible à tout le monde principalement via les AirPods comme tu viens de le mentionner. Est-ce que tu penses qu'ils en ont conscience et que les retours utilisateurs sont bons vis-à-vis du spatial ? Est-ce qu'ils sentent vraiment une différence ?

M : Je ne sais pas je n'ai pas fait d'enquêtes. J'ai lu les enquêtes d'Apple et effectivement ils disent que c'est génial et que tout le monde adore ça. Bon. Effectivement quand on parle avec nos proches ils disent « ah oui c'est un peu bizarre quand on tourne la tête ». C'est vrai. Mais moi j'ai constaté quelque chose à titre personnel qui est je pense n'est pas anodin et je l'ai constaté il n'y a pas si longtemps que ça. C'est vrai qu'au tout début quand Apple Music a sorti les AirPods évidemment je me suis jeté dessus et puis c'est vrai que je me suis dit « c'est chiant on tourne la tête c'est vrai que c'est bizarre », c'est vrai. Et l'autre jour, j'écoutais de la musique dans le train, j'écoutais des trucs en Atmos et voilà je ne bougeais pas beaucoup la tête, je la bougeais un peu pour choper un truc dans mon sac, pour regarder le paysage... Et à un moment donné j'ai désactivé le Head Tracking et en fait ça m'a tout de suite manqué. Je me suis dit « ah il manque un truc », et vraiment il manquait un truc important. C'est-à-dire que c'était beaucoup moins bien. Et pourquoi c'était beaucoup moins bien ? Parce que le Tracking ça reproduit l'écoute qu'on a naturellement. C'est-à-dire que si je suis dans mon salon, je mets le son sur mes enceintes en face et je tourne la tête, la musique elle reste là. Donc autant c'est un peu schizophrénique c'est-à-dire que dans un cas tu tournes la tête le chanteur il change de position ça te refocalise l'attention sur la musique alors que tu es en train de faire autre chose, parce qu'effectivement il se passe quelque chose, sauf qu'en fait, moi mon constat c'est qu'une fois qu'on s'y est habitué, on s'y habitue beaucoup plus vite, les gens vont s'y habituer beaucoup plus vite que ce qu'on pense. Et finalement s'en déshabituier est beaucoup plus compliqué que s'en habituer, je pense. Mais ça c'est quelque chose que j'ai constaté personnellement. Mais je pense que ça fait sens et que ça va faire sens chez les gens. C'est que, comme ça reproduit la manière naturelle d'écouter des sons. Bon là c'est un avis, une théorie, un pari que je fais, et encore une fois là on parle de musique en binaural écouté au Head Tracking, ce qui n'est pas le seul cas d'usage du son multicanal en général.

V : On peut mettre ce que tu dis en lien avec l'idée d'apprentissage de l'oreille au son spatialisé. Par exemple certaines études montrent qu'un Surround 2D serait plus immersif qu'un système 3D, et que la conclusion est que peut-être que c'est parce qu'ils n'y sont pas habitués. Quand on n'est pas habitué on ne se rend pas compte et on n'arrive pas à entendre la plus-value.

M : Parce que c'est une question de contenus. Typiquement moi je vais te répondre sur l'élévation parce que je me suis encore fait avoir hier en mix où il y avait un morceau où vraiment je trouvais ça mou. Et en fait il y avait un truc que j'avais mis au plafond, et je l'ai remis en bas et ça a tout changé, c'est-à-dire que tout d'un coup ça regroovait. Je ne dis pas qu'il n'y avait rien en haut, je dis juste que, perceptivement... Tu vois typiquement on parle de grammaire. Je suis en train de remarquer, et ça ne m'avait pas fait ça jusque-là, peut-être parce que je n'avais pas d'aides drastiques sur les propositions artistiques, justement en utilisant l'élévation, et en plus j'utilise un Bed où c'est un .2 [2 enceintes en élévation, ndlr] donc c'est juste un gauche-droite matrixé sur 4 haut-parleurs donc ça fait beaucoup de trucs qui font que c'est brouillon le rendu. Donc typiquement il y avait un instrument donné, ou l'effet de la caisse claire ou un truc comme ça, quand je le mettais en haut je perdais toute la netteté, j'avais un effet de masquage presque, perceptif, où du coup je sentais moins l'espace. Parce que finalement, je ne sais pas, peut-être que j'avais le même type d'effet de masquage que je peux avoir dans des enceintes mono entre deux éléments, tu vois j'avais un masquage spatial. En tout cas ce truc-là prenait le pas perceptivement sur d'autres éléments plus importants du mix et rendait le truc pâteux. Et j'ai juste pris la piste que j'avais mise au plafond, et je l'ai mise en bas, et j'avais plus de problèmes. Pourtant je n'ai pas changé les niveaux, je n'ai rien changé sauf ça.

Tout ça pour dire que par rapport à ta remarque « ouai c'est plus immersif sans l'élévation » en fait ça dépend du contenu. **Petit rire** Ça dépend de ce qu'on fait. Et ce que je peux te dire c'est qu'avec une bonne reverb multicanale, c'est beaucoup plus immersif avec de l'élévation. **Rires** Une reverb en 2D ou une reverb en 3D, tu as l'impression que tu as allumé les enceintes quoi. Enfin c'est un peu ça que tu fais, tu allumes les enceintes en haut... Après tout dépend comment c'est utilisé. Pareil j'ai fait des mixs de classique où je mettais juste un des couples un peu lointains au plafond pas fort. En fait, si tu le coupes, tu as le son qui redescend dans les cheveux quoi, tu as le son sur un plan, tu entends que c'est 2D et parfois il suffit de mettre un peu d'effet, pas fort, là-haut et tout à coup ça fait de l'air, ça respire. Alors après effectivement tu peux mettre ton soliste hein, mais bon ça je trouve que ça ne marche pas encore très bien, mais c'est logique on n'est pas fait pour... Disons que perceptivement déjà on perçoit mal l'élévation, en plus elle est très filtrée. Donc c'est logique que quand tu mets des sons là-haut une fois sur deux ça ne se passe pas très bien. Mais c'est logique, c'est perceptif.

V : Perceptif et logique, parce que tu es dans une logique de réalisme pour des musiques qui s'apparentent au classique.

M : Alors attention non non non. L'exemple d'hier c'était un truc de pop et j'ai quand même conservé un arpeggio de synthé sur le pont qui est au plafond et il est super au plafond. Et c'est super parce que c'est cette espèce de truc aérien qui arrive tout d'un coup ça fait un contraste majestueux. Mais par contre ma room de caisse claire, mon chorus de caisse claire je sais plus un truc que j'ai fait pour spreader le son que j'ai mis au plafond marchait pas du tout parce que ça écrasait tout le reste. Par contre mon synthé mon arpeggio tout seul au plafond lui il est génial. Alors après pourquoi est-ce qu'un truc est génial et l'autre pas ça c'est ... La magie du mixage et de la création quoi, c'est que on peut pas... Si on avait des recettes toutes faites ça se saurait.

V : Et par exemple sur ce mix de pop comment tu appréhendes ton mix spatialisé ? Qu'est-ce que tu fais en premier ? Et est-ce que t'avais un stéréo à la base ?

M : Alors, oui. Le truc c'est que t'as toujours un stéréo à la base. Après ceci-dit là en l'occurrence c'est un morceau qui n'a pas été mixé en stéréo. Donc j'ai attaqué le mix direct en multicanal. Et en plus c'est un morceau où il y avait beaucoup de pistes, ce qui me permettait des fois... Pour un son il peut y avoir 4 pistes de synthé pour faire un son justement, de créer un mélange spatial pour créer une matière qui soit enveloppante, soit localisée, soit large, soit réduite, mais sans utiliser d'artifices ou d'effets ou de reverbs c'est-à-dire juste en déployant les pistes.

Et j'avoue que ... **attente** Est-ce que j'ai des méthodologies en tête, c'est une bonne question, c'est-à-dire que je ne me la suis pas posée. En tout cas je pense que les règles sont exactement les mêmes qu'en stéréo, parce qu'en réalité faire de la stéréo c'est faire de la spatialisation entre deux enceintes. Et que déjà si tu spatialise mal tes sons, dans un mix stéréo, ça ne se passe pas bien. Si par exemple tu mets tout hyper large tout le temps, ton refrain il arrive et il n'est pas plus large que le couplet et du coup bon la patate elle vient pas trop alors que si tu t'arranges pour que ce soit plus large sur le refrain que sur le couplet, tu accompagnes avec une technique de mix le discours musical. Tu fais pareil avec du son multicanal. Du coup moi en stéréo le plus important en premier tu fais le volume et en deux tu fais le panning. Bah là c'est pareil, en premier c'est quoi le volume de la piste et en deuxième c'est quoi sa position. Et ensuite c'est quoi ses effets. Donc je pense que oui c'est la même logique c'est à dire que tu écoutes ton morceau... Après encore une fois là on parle d'étapes où l'artiste arrive avec une proposition spatiale qui est nulle. C'est-à-dire qu'il n'y en a pas. Typiquement quand tu fais un mixage stéréo, tu arrives avec des propositions de l'artiste ou pas. Par exemple il va avoir mis une guitare à gauche et de l'orgue avec une Leslie hyper large. Après au mixage, bon parfois on conserve ces propositions mais on peut décider de faire un truc complètement différent, l'orgue tout petit filtré à droite, à contre champ dans un couplet, on veut que ça pop, mais on ne veut pas que ça prenne la tête donc on ne le met pas fort dans un coin, ça fait qu'on entend que ça mais ce n'est pas fort, ça ne dérange pas, ça ne bouffe pas les leads, voilà... Et finalement si l'artiste arrive au mixage Atmos et qu'il arrive avec des envies ou des pistes déjà spatialisées ce n'est pas pareil.

J'avais fait, avant qu'Apple arrive, j'avais travaillé avec un artiste qui composait pour faire de l'audio spatial sur de la pop et il avait tout composé en stéréo dans Live. En revanche, le mec est arrivé avec 4 pages PDF, « ça c'est là, ça ça bouge de tant à tant », il avait tout écrit. Donc du coup on a pris son cahier et on a tout reproduit, et c'était super. Parce qu'il avait eu la vision, il avait projeté tout ce qu'il fallait faire spatialement.

V : Mais c'est rare d'avoir des artistes qui viennent avec ce qu'ils veulent. Ça n'arrive jamais.

M : Ça m'est arrivé qu'une fois et c'était avant le Dolby Atmos **rires** C'est un titre qu'on a distribué, qui est sur les plateformes en Dolby Atmos qu'on avait envoyé à Dolby parce que lui il était un artiste labelisé Dolby mais c'était avant que Apple lance le spatial audio donc il n'y avait pas de canaux de distribution donc on avait livré un Atmos mais dans lequel il n'y avait pas de canaux de distribution. Mais typiquement voilà donc après des artistes qui sont arrivés avec des choses très précises moi en tout cas pour du mixage musique Dolby Atmos bah

je n'en ai pas eu. Mais tu vois là typiquement j'ai mon quatuor à corde pour lequel j'ai mixé leur album *Brahms* et eux ils veulent absolument faire du répertoire en spatialisé. Mais ce n'est pas commun et moi non plus parce que mettre un musicien dans chaque coin pour écouter un quatuor je ne sais pas, en fait faut écouter. En fait je pense surtout que ça va dépendre du répertoire qu'ils vont choisir. Soit ça fait sens parce que généralement une écriture pour quatuor c'est fait pour que les sons se mélangent donc si on les dissocie dans l'espace est-ce que la musique marche toujours il y a que le quatuor qui va pouvoir le dire parce que c'est eux les spécialistes du répertoire et de la musique qu'ils jouent, pas moi. Moi je peux faire des propositions et donner mon avis mais c'est eux qui vont dire si ça marche ou si ça ne marche pas, musicalement j'entends.

V : Tu dis que ça doit se mélanger dans l'écriture de la musique et là on est dans un contexte de mix Atmos, orienté, et est-ce qu'il n'y aurait pas une discrimination spatiale à mettre un instrument à l'arrière vis-à-vis de ceux qui vont être devant ?

M : C'est ce qu'on va tenter, c'est-à-dire qu'un moment donné il y aura un lead et on va essayer. Là pareil c'est page blanche, en tout cas les trucs que j'ai entendus où les mecs ils ont testé en classique ça m'a paru intéressant mais pour moi ce sont des expérimentations, ce ne sont pas des œuvres. C'est bien d'expérimenter mais est-ce que musicalement il y a une émulsion ? Est-ce qu'il se passe quelque chose ? C'est une autre question. Tu vois pareil les prises d'orchestre où on a la tête à la place du chef, bon c'est marrant mais en fait c'est fatigant comme écoute... C'est marrant, c'est instructif.

V : Tu dis que le travail spatial c'est encore une page blanche, mais tu utilises ton apprentissage stéréophonique. Il y a des réflexes qui restent et il y a quand même une image frontale qui prend le dessus.

M : Oui. Mais est-ce que c'est bien je n'en sais rien. Je vais te donner un exemple : à chaque fois que j'ai fait des disques de pop, il y a toujours dans un album de 12 titres 1 ou 2 titres qui sont, ou souvent, qui sont des titres pas anecdotiques mais qui sont des amusements, des interludes, des choses comme ça. Par exemple un groupe de Rock, ou sur un groupe d'électro, où d'un coup tu auras un morceau un peu court avec une boîte à rythme et un gros kick etc. par exemple sur ces morceaux-là je proposais de mettre la rythmique sur le côté par exemple. Et c'est rigolo parce que moi j'adore ça, je trouve que ça crée un contraste hyper saisissant, et dans 90% des cas, au moment des retouches de mix, la rythmique elle revenaient au centre, parce que les artistes ont des habitudes et ça leur faisait trop bizarre, ils écoutaient au casque avec la rythmique d'un côté etc. Tout ça pour dire qu'effectivement on peut refaire les mêmes propositions ou le même type de démarches quand on mixe en spatialisé et qu'on va transporter simplement le paradigme gauche-droite à 360, mais c'est le même paradigme. Et pour répondre à ta question de la frontalité, bon si ça se trouve un jour je vais faire un album comme ça, sur un interlude je vais m'amuser à mettre la rythmique derrière et la voix derrière, si ça se trouve ce sera bien ou les artistes vont me dire « ah c'est trop bizarre. » Je ne sais pas. Et typiquement, aujourd'hui ce qui est intéressant c'est qu'il y a quand même beaucoup de sorties en Atmos, en tout cas plus que ce qu'on pensait ou en tout cas que moi je me serais attendu, et que dans ces sorties il y a quand même plus de choses bien que ce à quoi je m'attendais aussi. Je dirais qu'il y a que 50% des productions qui sont ratées. *rires* Et il y en a qui sont vraiment ratées. Et il y a des trucs qui sont des propositions, des essais, et je pense notamment à un album d'un groupe de métal sur lequel je suis tombé par hasard, et les gars ce qu'ils ont fait c'est que t'as

tout le mix dans le LR, et tu as des breaks et des accents, et de temps en temps juste un délai ou quoi qui se balade. Donc en fait ce sont des effets hyper ponctuels et hyper francs, et c'est hyper bien. Tu vois c'est une proposition, c'est pas du tout didactique parce que tu as le groupe qui joue dans le LR et tout d'un coup t'as une grosse reverb de voix sur les côtés, un break de guitare qui passe de gauche à droite, un fill de batterie au plafond. Je n'aurais jamais osé et je trouve que ça marche super bien. Alors après, moi si tu me demandes de faire un mix Atmos je vais essayer de faire en sorte que ce soit immersif parce que j'aime ça. Mais je trouve que cette proposition artistique, c'est du coup vraiment une proposition, je la trouve vachement bien après c'est purement esthétique. Et c'est comme quand Hervé Déjardin fait de l'Atmos avec Jean Michel Jarre où ils ne mettent rien dans le plafond, justement ils font un mix 7.1. C'est leur choix.

V : Dans des contextes très codés de musique classique sur lesquels tu travailles, tout se joue sur la reverb ?

M : Il faut bien dissocier la reproduction fidèle d'une acoustique, même si on l'invente, en tout cas de vouloir mettre un auditeur dans un espace réel, et l'utilisation de l'espace comme un outil de composition, ce qui est quand même deux choses très différentes. C'est important parce que typiquement, en classique on va essayer de faire des prises de son d'ambiance multicanale, là où en pop on va chercher, si on peut mettre des micros de room de batterie, ça dépend comment la musique est produite, mais on va clairement utiliser l'espace comme un champ de création pour mettre un effet, une piste... chose qu'on ne fera pas en classique, ou alors ce que je disais tout à l'heure, quand c'est fait en classique je trouve ça raté. À voir ce que je vais faire avec ce quatuor. Mais du coup pour répondre à ta question, c'est vrai que ce qui est difficile avec le classique, mais je pense que c'est valable dans toute musique, c'est qu'il y a des codes, et typiquement je crois que la musique classique c'est celle où il y a le plus de codes... Non je retire ce que je viens de dire, ce n'est pas celle où il y a le plus de code, c'est qu'en pop il y a beaucoup de liberté, puisque justement on utilise même la stéréophonie comme champ créatif. À la limite on peut faire ce qu'on veut. En classique tu ne fais pas ce que tu veux, alors certes il y a des codes et indépendamment de ça tu ne fais pas ce que tu veux parce que si à la fin ton son de piano il est moche et que ton son de violon il ne sonne pas assez personne sera content alors que sur ton son de variet', si ton son de Wurli il est medium et que ça marche bien avec le morceau et que c'est groovy tout le monde va trouver ça super. Donc sur du classique l'enjeu n'est pas le même. Après ce que je trouve difficile en classique c'est que typiquement pour entendre l'espace, de l'espace autour de soi, il faut de l'espace donc de l'effet de salle, et que le problème c'est de pas se retrouver, quand tu vas faire un Atmos d'un truc de musique classique, avec, entre ta version stéréo et ta version Atmos, juste un truc qui sonne avec plus d'espace et de reverb parce que t'as rajouté de la room sur les côtés tu as rajouté de la reverb. Alors super c'est immersif, puis tu écoutes le stéréo et tu n'as pas le même disque. Ça, ça ne va pas et c'est vrai que quand tu immerges le spectateur dans un effet de salle, ce qui se passe c'est que tu sens que ton plan frontal il recule et il faut arriver à le mettre au même plan que dans la stéréo et ça ce n'est quand même pas facile

V : Avant que le Dolby Atmos arrive, quels outils vous utilisiez ? Et pourquoi avoir pris le pas du Dolby Atmos ? Qu'est-ce que vous en pensez ?

M : C'est le seul qui est distribué donc c'est le seul que les maisons de disques me commandent. *Rires* C'est-à-dire que jusqu'à présent les musiques en spatialisé que je faisais

c'était soit pour du spectacle vivant, soit pour des installations sonores, soit pour de la VR mais en discographie il n'y avait rien, donc évidemment l'Atmos c'est un format... Déjà il faut dissocier la marque Dolby Atmos, qui n'est pas un format, le format c'est l'ADM, le format c'est éventuellement le codec, c'est l'AC4 de Dolby et le Dolby Atmos c'est la marque qui enveloppe tout ça. Et la force de Dolby c'est qu'ils ont créé des outils pour fabriquer des contenus, ce qu'on appelle globalement le Dolby Atmos, qui est clairement l'outil de spatialisation le plus pauvre qui existe à ce jour en termes... Je sais je prends un exemple très con mais je bosse dans Pro Tools parce que je bosse dans le truc qui est le plus répandu aujourd'hui et faudrait que je commence à considérer à migrer parce que, ça va changer, mais pour l'instant le panning dans Pro Tools il est orthogonal. Quand tu mixes en Atmos c'est un enfer, quand tu veux changer une courbe d'automation d'une trajectoire faut que tu aies 4 courbes à la fois, c'est une horreur, alors que je sais dans Nuendo ils ont intégré un panner AED, Azimut, Élévation, Distance, puisqu'en plus l'Atmos n'a pas de distance, enfin si tu peux passer des choses à l'intérieur du cube, ce qui a pour conséquence de distribuer de l'énergie en amplitude dans les différents haut-parleurs donc c'est horrible. Bref tout ça c'est que qu'il y a de pire. Mais ça marche. C'est-à-dire que ok j'aimerais bien migrer sur Nuendo mais faut que j'apprenne tout parce qu'il y a un panner en AED et qu'il y a plein de pistes multicanales et qu'il y a plein de fonctions hyper bien. Ok. Mais pour l'instant je bosse dans Pro Tools parce que pour l'instant c'est que je maîtrise c'est ce que je connais et malgré les complexités de routing et de panning et de machin, ça marche, et je peux bouncer mes fichiers, je fais un *pomme alt bounce* et puis je sors mes ADM quoi, il y a deux ans et demi personne ne pouvait imaginer qu'on en serait là, donc ça c'est génial, clairement.

Et là j'ai produit un album avec inFiné de Bruemaker et en réalité j'ai pas utilisé les panner de Pro Tools, j'ai installé le Spat pour pouvoir faire mes rotations de scène, parce que sinon c'était un enfer tu peux pas vraiment linker des objets dans Pro Tools, alors t'as un outil qui permet de faire ça mais je l'avais essayé c'était un peu beugué... Faudrait que j'insiste parce que ça permet de faire vraiment de l'objet en AED et de coupler ces objets donc apparemment on peut faire tout ça donc c'est vrai que c'est pas mal, mais sinon j'utilise d'autres algorithme des fois donc le Spat, par-dessus l'Atmos, que je printe dans l'ADM. Mais je le fais sur des projets où je sais que j'ai envie de prendre le temps d'expérimenter et je suis avec des artistes qui sont ouverts à qui j'explique ce que je fais qui sont ok qui écoutent qui font « ok ça vaut le coup ».

V : Tu as parlé de rotation. Est-ce que c'est fréquent pour toi d'utiliser tout ce qui se réfère au mouvement et à quel niveau ? Est-ce qu'il y a une manière de travailler le mouvement qui est différente et qu'est-ce que ça apporte ?

M : Alors je dirais deux choses très importantes, et d'ailleurs je l'ai encore lu récemment dans un bouquin très intéressant, faudrait que je retrouve la citation, mais il y a une citation qui dit que si tu montres à l'image un arbre à un spectateur il voit un arbre. Tu lui montre deux arbres il voit deux arbres. Tu lui en montre 3 et il voit une forêt. Le bouquin que j'ai lu récemment c'est un bouquin sur la cartographie parce que je cherche à faire des analyses de données pour créer de la musique générative, j'ai lu un bouquin sur la cartographie. Comment tu crées un graphique qui soit lisible et en fait le mec il disait qu'en fait au-delà de trois variables, le cerveau il ne sait pas saisir une information telle quelle. Au-delà de 3 variables il est obligé d'utiliser sa mémoire. « Qu'est-ce que j'ai entendu ? Qu'est-ce que j'ai lu ? » Et en spatialisation j'ai envie de dire que c'est pareil. Il n'y a pas d'études perceptives mais je pense

que si tu fais bouger 3 ou 4 sons en même temps, le cerveau il sait plus dire qui bouge ou quoi, si mais il est obligé de se concentrer « tiens celui-là il bouge, lui aussi, ils sont combien, faut que je compte faut que je mémorise », alors que s'il y en a 2/3 pas besoin de compter tu le sens perceptivement. Ce que je veux dire par là c'est que le mouvement des sources est quelque chose qui est très bien perceptivement... Mais c'est pareil en stéréo. C'est-à-dire en stéréo, en gros, c'est rare qu'on utilise dans un mix un ping pong delay, un phaser gauche droite vraiment très prononcé ou alors on n'en met pas beaucoup parce que sinon on ne comprend plus rien. Alors après on peut très bien décider de faire que des sons qui bougent et qui tournent tout le temps, c'est un choix artistique mais perceptivement c'est vrai qu'il y a des choses qu'on sait, comme ce que je viens de dire même si ce n'est pas théorisé et que je serai ravi de lire des travaux là-dessus, et moi j'ai fait plein de disques en Dolby Atmos où il n'y a rien qui bouge. C'est pas du tout obligatoire de faire bouger un truc. C'est pareil en musique contemporaine, sur des choses spatialisées parfois il y a des trucs qui ne bougent pas c'est très bien et voilà. Parfois ce qui est intéressant, ce sont les choses qui bougent mais tout doucement. Parce qu'en fait quand ça bouge vite ça va attirer l'oreille donc c'est l'exemple que j'ai utilisé tout à l'heure avec l'orgue tout filtré. C'est-à-dire que si tu as un chanteur, le but c'est de porter l'attention sur le chanteur, si tu mets ton son d'orgue filtré d'un côté, il va beaucoup attirer l'attention. Tu vas faire ça si tu as envie que ton orgue dialogue avec ton chanteur et qu'il n'y ait rien d'autre qui pop et tu ne vas pas le mettre trop fort parce que tu ne veux quand même pas que ça prenne trop le pas. Ça va être pareil si tu fais bouger un son. Mais ça dépend du son que tu fais bouger c'est-à-dire que si tu fais bouger une nappe de synthé avec un chorus bah ça va juste être immersif, ça va être beau, mais si tu as un petit thème, une petite mélodie et qu'elle bouge, peut-être que je me trompe mais... Ce que je veux dire c'est que perceptivement le mouvement ça attire quand même l'attention. Donc c'est un élément à prendre en compte dans le discours concret.

V : Le mouvement qu'on choisit dépend intimement du rôle du son auquel il est associé

M : Tout à fait, c'est pour ça que ce qui est intéressant dans ton travail c'est que clairement la spatialisation, et c'est ce que Stockhausen disait, c'est une donnée aussi importante que la hauteur ou le rythme. Surtout quand elle est utilisée à ce point-là c'est-à-dire que, je ne sais pas tu fais un ping pong delay, qui fait gauche droite gauche droite, et puis il va tourner, ça a le même impact que quand tu l'écoutes en stéréo mais ce que je veux dire c'est que c'est un discours qui est hyper fort. Alors que si tu mets un delay sur un instrument et que le delay se situe derrière l'instrument bah ça prolonge l'instrument alors que si c'est un ping pong ça créé clairement des événements, tu vois, importants. Donc par rapport à ce que tu viens de dire en effet la spatialisation avec du son en mouvement est clairement indissociable du son auquel il est associé et indissociable du morceau et de l'environnement dans lequel il est placé parce que voilà si... Si tu mets une section rythmique et si tu mets ta caisse claire derrière ou même, dans un mix stéréo, à droite, ou alors typiquement je suis en train de penser à un truc très con, à un album que j'adore avec deux batteries, une à gauche une à droite, et les deux mecs ils jouent à l'unisson quasiment tout le temps, sauf de temps en temps. Et du coup c'est hyper beau parce que la batterie est hyper large parce qu'ils sont à l'unisson, mais ça fait des flas, et puis de temps en temps il y en a un qui met un break et ça se répond gauche droite et ça marche hyper bien, et c'est de la spatialisation tu vois.

V : Est-ce qu'en immersif tu travailles tout seul ou les artistes travaillent avec toi ?

M : J'ai envie de dire c'est comme en stéréo. J'ai toujours été quelqu'un qui a aimé travailler en présence des artistes. Et j'ai fait énormément de mixs et j'ai mixé en analogique en présence des artistes parce que je trouvais que l'intérêt il est double, déjà l'artiste prend part à des choix assez tôt dans le processus de mixage et que deuxièmement il valide des choix de mixage que tu ne remets pas en cause une fois que tu as fini ton travail et... Une fois que, toi, perceptivement, tu as fait un traitement sur la guitare en fonction de ce que t'as fait sur la basse et l'artiste il arrive et il dit que sur la basse ça ne va pas et 2 jours plus tard on se rend compte que la guitare en fait c'est une basse. *Rires* Donc le fait que de mixer avec les artistes ça fait que ce genre de situations arrive moins souvent. C'est pareil en Atmos. Cela étant, j'aime mixer en Atmos en présence des artistes parce que je trouve ça intéressant de partager, de challenger, et qu'eux me challengent. Maintenant j'aimerais bien que ça arrive plus souvent. *Rires*

Et puis après il y a aussi le truc c'est que... Mais c'est pareil quand on fait du mixage stéréo, c'est que... On fait de la pédagogie, c'est à dire que tu vois, effectivement tu commences à... En Atmos le premier truc que le mec dit c'est « - Ouai vas-y on va faire tourner le charley. - Ouai mais tu vas voir ça va être un peu relou donc non. » En fait ça c'est parce que les gens ont besoin de s'acculturer avec l'objet. Mais c'est pareil en stéréo. Quand c'est la première que tu mixes avec des artistes en stéréo ils ont des demandes ou des attentes, ils se projettent dans des trucs où ils ne se rendent pas compte ou ce qu'ils demandent est plus en lien avec la composition que le mixage, et ça, ça arrive tout le temps en mixage stéréo. Il donne des références, il veut que ça sonne comme ça mais fallait mettre de la guitare, c'est de la guitare qu'on entend, et ce genre de choses arrive et arrivera de la même manière en spatialisation, de la même manière.

V : Est-ce qu'il y a un ressenti dans la technique de mix pour toi qui est plus facile parce qu'on est en immersif, notamment vis-à-vis du démasquage. Est-ce que tu ressens ça ou pas du tout ?

M : J'ai envie de te dire oui bien sûr mais en fait je ne pense pas que ce soit plus facile. Je pense que c'est même plus dur parce que le fait qu'il n'y ait pas d'effet de masquage fait que si on veut maîtriser justement des synchronicité dans des éléments musicaux ou des mélanges de textures, c'est plus difficile. En fait je vais te prendre un exemple très net, tu vois mixer un album en stéréo où il y a 50 pistes d'arrangements, je vais avoir aucun problème parce que je sais que je vais mettre quelques arrangements forts et quelques arrangements pas forts et puis ça va remplir le truc et puis ça va marcher tout seul. Alors que quand il y a une guitare basse batterie avec 2 rere, pas de choeurs et rien d'autre, si c'est du rock garage ça va marcher parce que c'est de l'énergie brute, alors que si on a envie de créer des contrastes et de créer des textures, il faut tout faire en mix parce qu'il y a pas de pistes pour le faire, donc des fois des trucs avec plein de pistes c'est plus facile à mixer que ce qu'on croit grâce aux effets de masque, alors que du coup tu vas te retrouver en multicanal avec toutes cette armée de pistes à mettre en espace... Et tu vas tout entendre tout le temps et ça va être trop le bordel et c'est exactement ce qui m'est arrivé hier. C'est un titre, ça fait plusieurs jours que je suis dessus et que j'y reviens parce que je n'y arrive pas, parce qu'il y a trop de pistes. Et en stéréo ça se passe beaucoup mieux.

Donc non il n'y a pas de règles. Après il y a un autre truc c'est que ces effets de masquage qui n'arrivent plus font qu'on peut faire un très joli son avec beaucoup de Headroom beaucoup

plus facilement. Mais après je n'ai jamais essayé encore de faire un album où j'ai fait un son de batterie de A à Z en Atmos dès le départ. Toujours parti d'un stem d'un mix stéréo, éventuellement redéployé les rooms, mais je pars d'un truc qui se tient. Et c'est pour ça qu'aussi dans les productions Atmos je pense qu'il y a des choses qui marchent bien aussi c'est que pour l'instant les gens partent de stems stéréo et que même si c'est très contraignant et que parfois ils font n'importe quoi avec, ils partent quand même avec un truc qui se tient... musicalement. Tu as des équilibres entre les pistes, musicalement qui se tiennent parce que faits avec des techniques de mix acquises depuis des décennies.

V : Tu dis que tu pars de stems de batterie stéréo ? Sans le multipiste de la batterie ?

M : Ah bah tout le monde fait ça.

V : Et il ne te manque pas des informations pour travailler l'espace justement avec une simple stéréo ?

M : Tout le monde fait ça, les mecs ont des stems stéréo et ils se démerdent avec ça. Ce qui est con d'ailleurs. Mais les gens c'est ce qu'ils ont aujourd'hui. Après... Tu vois j'ai fait un live de Souchon. Pareil je suis parti du stéréo mais comme c'est un pote qui a mixé et que j'avais vendu à la prod' que je voulais y passer un certain temps, et puis il a réussi à les convaincre que justement on allait repartir de ses sessions et pas des stems, et on a rouvert ses sessions et j'ai passé le bus stéréo en bus multicanal. Donc on a récupéré tous ses traitements de dynamique, tous ces traitements tels quels mais on pouvait doser un peu de centre qui n'était pas juste un repliement gauche et droite. On pouvait élargir un peu des pistes qui n'étaient pas juste une piste stéréo, qui étaient élargies, on pouvait spreader un peu les micros etc. Et ça, ça change tout. Mais tout le monde ne peut pas faire comme ça en revanche. Donc non aujourd'hui ils gens ils ont des stems stéréo en Atmos.

V : Ce serait plus agréable d'avoir le multipiste finalement.

M : Oui et non, parce qu'en fait qu'est-ce que tu vas faire avec le multipiste ? Typiquement les Rage Against The Machine ils ont remixé l'album *Killing In The Name* en ATMOS, c'est une catastrophe, mais vraiment c'est une catastrophe. Je pense en fait que ce qu'il s'est passé, déjà ils ont fait comme Gojira et ces gens-là, ils ont mis les guitares hyper larges donc il n'y a plus rien, plus que du basse batterie au centre, donc il y a plus de liant avec les guitares du coup il y a plus de patate, c'est le premier problème. Le deuxième problème c'est qu'ils ont mal rééquilibré l'énergie et du coup, par exemple chez Gojira on a trop de basse batterie par rapport aux guitares, alors peut-être qu'au binaural ça marche bien, mais bon voilà les mecs ils ont écouté qu'en binaural. Super, du coup demain le moteur de binaural change, ça change tout, donc faut écouter ce qu'il y a sur les haut-parleurs puisque c'est ce qu'on met sur la bande. Et je pense que pour *Killing In The Name* ils n'avaient pas de stems parce que c'était mixé sur bande et du coup ils sont repartis du multipiste et du coup bah je suis désolé mais ils n'ont pas... Le petit grain de saturation sur la voix il n'est pas là, ce n'est pas bien. Je peux me tromper. Autant pour Gojira je pense qu'ils sont partis des stems qu'ils devaient avoir, autant pour Rage Against The Machine je ne vois pas pourquoi ils auraient eu des stems donc je pense qu'ils sont repartis du multipiste et du coup je pense qu'ils auraient mieux fait de s'abstenir.
Rires

V : Est-ce que tu réfléchis à de la théorie de l'enveloppement de l'auditeur, notamment dans les salles l'importance des premières réflexions latérales, quand tu mixes en Atmos ?

M : A bah bien sûr. *Rires* C'est tout le secret. *Rires* Bien sûr.

V : A quel moment ça intervient ?

M : Il n'y a pas de règles, c'est qu'il y a des sons ils sont très bien spin spotté dans un haut-parleur, d'autres très bien partout, d'autres très bien qu'avec de la reverb, tu vois il y a pas de... ça après c'est du sensible c'est de la technique de mix tu vois. Il y a des choses que je commence à reproduire... Pas systématiquement évidemment, mais des trucs que j'aime bien que je refais. Mais je dis pas que c'est ce qu'il faut faire. Tu vois par exemple quand j'utilise une 480 Stéréo dans un mix, il y a deux preset que j'aime bien c'est la *brick wall* et l'*autoparc*, ça ne veut pas dire que c'est les deux meilleurs preset de la 480, c'est juste que c'est ceux que j'aime bien parce que je sais quoi obtenir et que ça va macher avec ce que je veux faire. Quand je vais mixer en Atmos ce sera pareil. J'ai des trucs, voilà, qui font partie de mes ingrédients et que des fois je réutilise et puis des fois ça marche des fois pas, par exemple ce que je disais en tout début de conversation avec ma reverb au plafond qui était catastrophique alors que des fois ça marchait tu vois.

C'est intéressant parce que. Franchement je crois qu'il y a beaucoup de choses que je t'ai dites sur lesquelles je n'avais jamais vraiment verbalisé. Donc je trouve ça intéressant. Et on va pouvoir conclure parce qu'après j'ai un autre coup de fil. On pourra se rappeler si tu veux parce que ça m'intéressera que tu me dises toi ce que t'as retenu de ce que je t'ai dit, parce que peut-être que tu vas mettre le doigt sur des trucs sur lesquelles j'ai moins insisté.

Mais en revanche, le truc le plus important, c'est que depuis 1h on discute on discute de ce que fait le mixeur au moment du mixage d'un contenu en son spatialisé. Mais moi la question c'est : et les artistes ? Qu'est-ce qu'ils vont en faire ? Je vais te répondre très simplement. Un artiste aujourd'hui quand il mixe en stéréo il sait ce que c'est la stéréo. C'est-à-dire il sait quand est-ce qu'il veut un ping-pong delay, ou quand il veut un son à gauche ou un son au centre. Et ça s'entend dans les maquettes que tu reçois. Donc ça va finir par être pareil en son immersif, il n'y a pas de raisons. Simplement il leur faut deux choses pour ça. Il leur faut déjà les outils. Mais c'est en train d'arriver tout ça, ça avance donc en fait les artistes vont commencer à se familiariser avec l'espace. Parce qu'aujourd'hui, si je veux créer un son spatialisé, je fais quoi ? Je fais une prise de son, puis je la spatialise. Depuis la nuit des temps on fait comme ça. Et c'était tout l'objet de mon mémoire, c'était de créer un instrument qui fait les deux en même temps, ou finalement les deux sont associés alors je dis pas que c'est nouveau parce que typiquement Xenakis il avait fait des instruments... Enfin il avait plutôt fait des outils de spatialisation tout ça. Moi j'ai essayé de faire un instrument vraiment qui soit le plus en phase possible avec ce que les artistes ont l'habitude d'utiliser, c'est à dire un synthétiseur avec des oscillateurs des filtres, pas du tout ésotérique. Même s'il y a plein de fonctions ésotériques sous le capot pour faire de la FM et tout ça. Mais pour tout te dire J'ai fait une version alpha, qui était un POC, qui a été clairement concluant, j'ai fini une version standalone avec une interface graphique tout finalisé il y a un mois, que j'ai filé à mes associés et qui fonctionne très bien. Ils font des prises avec là ils sont en train de tester. Et depuis un mois j'ai repris la programmation je suis en train de tout réécrire pour que ce soit plus simple, moins gourmand, et que ça ait une meilleure interface pour en faire un device M4L. Mais du coup, toujours dans l'idée, et

c'est là où j'ai progressé par rapport aux versions que j'ai faite l'année dernière ou même à celle que j'ai faite il y a un mois, c'est que c'est aussi de rendre l'interface de spatialisation digeste, inspirante. Inspirante ça c'est important.

A12.2 Entretien avec Hervé Déjardin

V : [Présentation du projet]

H : Je me pose la question de savoir ce qu'on lâche ou ce qu'on ne lâche pas en infos. Je vais être transparent, il faut que je fasse un bouquin là-dessus, avec Radio France. Parce que Radio France ça fait 20 piges qu'ils nous paient à faire ça donc c'est Radio France qui doit en profiter. Je vais te donner les tucs généraux, que tout le monde va te dire, qui confirmera ce qu'on t'a dit. Ce que je retiens de ce qui se fait à l'extérieur. D'ailleurs il y a Dolby en ce moment qui parle à tout le monde de leur histoire d'équilibre spatial, c'est comme si tu parlais de la stéréo en disant qu'il ne faut pas tout mettre à droite ou à gauche. Bon ok. Une fois qu'on a fait ça ? C'est le discours de Dolby Aujourd'hui, ils ont l'impression d'inventer une grammaire simplement parce qu'ils équilibrivent un espace, super. Mais l'affaire doit aller beaucoup plus loin. Je vais être là-dessus. Je ne vais pas te parler de perspective, de là où je pense qu'il faut qu'on aille.

V : Qu'est-ce que c'est l'immersion ?

H : L'immersion c'est le fonds de commerce de tout média et de toute personne qui veut attirer dans son champ quelqu'un... Je définis 2 types d'immersions :

L'immersion narrative, c'est « raconte-moi une histoire » et je me baigne dans cette histoire et donc mon état mental se décorrèle totalement, en tout cas se détache de mon état physique puisque seul mon subconscient veille au grain, et mon conscient lui est complètement barré dans l'histoire qu'on est en train de me raconter.

L'immersion perceptive, c'est ajouter à l'immersion narrative le fait de se décorrérer du monde réel par la perception. Donc il y a différentes voies, si on prend l'histoire des médias, le premier qui va inventer l'immersion perceptive c'est le cinéma par le noir salle puisqu'il va occulter la vision périphérique et permettre de se concentrer sur la vision frontale. Et à l'époque où le cinéma vante le noir salle, début 20^e ou fin 19^e, les images sont pas géniales, je pense que la luminosité des projecteurs ça devait pas être encore ça, donc à mon avis il devait y avoir peut-être, à regarder dans l'histoire du cinéma, mais que peut-être que les projo n'avaient pas la capacité d'envoyer énormément de lumière sur l'écran et donc faire le noir salle devait améliorer la perception de l'image, t'occultais de fait la vision périphérique et donc on se concentre sur cette image, et donc on est un peu plus immergé dans cette image.

L'immersion perceptive au niveau du son c'est vraiment à partir du moment où la stéréo arrive. Oui ça commence à l'être parce qu'on a une scène sonore à 60° sur le plan frontal, mais pour moi la réelle immersion c'est à partir du moment où on commence à mettre des HP tout autour de l'auditeur. Ou c'est le casque, parce qu'il y a une construction mentale au casque plus ou moins réaliste à partir des indices, même si la stéréo n'a jamais géré parfaitement les indices, sauf sur un peu de Δt et... Comment dire... On a une distorsion angulaire entre un mix fait pour HP et sa perception sur casque. Mais le fait d'avoir un son immersif, c'est à dire à 360°, déjà le 5.1 ou le 4.0... Tu vois le 4.0 il est génial, c'est que ça y est tu n'écoutes plus les choses de manière frontale mais tu es au centre des choses, et là ça y est tu es immergé. Et ça c'est vraiment l'immersion perceptive. C'est là où le binaural est essentiel, enfin même la stéréo au casque. La stéréo au casque c'était déjà une immersion perceptive même si comme je viens

de le dire on était décorrélé d'un réalisme sur le plan des indices parce que les itd et ild n'étaient pas reproduis en similitude avec ces indices dans un monde réel. Mais je me rappelle quand moi j'écoutais dans les années 80 le walkman d'un pote qui avait la chance d'avoir un super casque de chez Sony qui était une tuerie, là tu écoutais ça et tu écoutais ce que t'avais l'habitude d'écouter sur HP et tu avais les kicks et les basses dans la tête et tout un univers autour de toi, c'était une construction mentale. Je ne le percevais pas de manière frontale en fait. Dès que j'ai mis un casque pour la première fois et que j'écoutais de la stéréo ; et à l'époque j'écoutais des trucs comme *Eurythmics*, *Foreigners* des groupes pop rock ; je baignais dans les sons c'était un vrai bonheur ! Mon envie de faire ce que je fais aujourd'hui elle est là ! Quand j'avais 11-12 ans et que j'écoutais ces musiques là à la récré avec ce Walkman Sony à la récré.

Donc l'immersion c'est ça et l'immersion elle qualifie une perception, un mode de perception et sous-entendu un état mental

V : Ça part d'une volonté artistique ou technique ?

H : Déjà l'immersion narrative, ce qui fait le succès de la radio, c'est que si tu prends un peu l'histoire des échanges entre les êtres humains, je ne veux pas dire de bêtises mais en premier tu as les conteurs qui viennent dans les villages et qui racontent des histoires, et on vient les voir sur la place du village, on vient s'immerger dans une histoire, parce que le mec il raconte des histoires qu'on n'a jamais entendues et c'est hyper nouveau, c'est un phénomène quoi.

Ce même phénomène explose avec le livre quand on commence à savoir lire en masse et qu'on distribue des histoires avec des livres. Et l'explosion, du livre elle est là. Et à chaque fois qu'il y a un nouveau truc on dit « ouai c'était mieux avant. » Et après tu as le média radio qui est le premier où là, avec un micro, un émetteur et avec les ondes courtes ou les grandes ondes, d'un seul tout tu peux alimenter un pays. À l'échelle nationale tu as un speaker qui raconte une histoire et tout le monde écoute cette histoire. Et en plus derrière cette histoire on peut commencer à l'illustrer avec des sons, d'où les bruiteurs et toute l'histoire de la narration radiophonique qu'il y a derrière.

Ça c'est l'immersion narrative, c'est le truc le plus puissant au monde. Pourquoi c'est le plus puissant au monde, parce que l'être humain passe son temps à essayer d'échapper à son réel ou essayer de faire des pauses dans son réel, parce que nous sommes qu'imaginaire en permanence quoi, c'est un aller-retour entre le concret, ce que je vois... Et là moi en permanence je suis en train de partir dans mon imaginaire quand je te parle, je vais chercher des trucs, des références, et puis te voir et regarder ma tasse de thé, c'est en permanence ça l'être humain, donc c'est la drogue la plus puissante qui puisse exister. C'est là-dessus que ce sont construit les médias. Les chaînes d'infos elles se sont construites sur un truc très stroboscopique. Alors maintenant c'est le café du commerce, mais c'est toujours une immersion en fait, une immersion narrative. Elle vaut ce qu'elle vaut aussi, on peut la juger en termes de morale ou d'influence sur les populations, mais sur le plan de l'immersion c'est de l'immersion narrative. Comme les clips. Après l'être humain est capable de développer du génie pour essayer de faire avec du vide un truc qui t'attire, je veux dire le nombre de clip vidéo dans les années 80/90 ou c'est 10 0000 images à la seconde, en fait ça change de plan ça ne raconte rien quoi. C'est le phénomène aujourd'hui TikTok, ou sur les réseaux sociaux, je pense que ça t'arrive et ça m'arrive aussi où d'un seul coup tu es là tu ne sais pas quoi foutre et tu regardes des vidéos débiles et tu les fais

défiler et tu es capables de passer un quart d'heure à regarder des vidéos débiles. Enfin quand je dis débile je m'entends tu me comprends mais là j'ai perdu un quart d'heure en fait. *Rires* Tu t'es détendu un quart d'heure à regarder un truc qui ne t'apporte rien quoi.

C'est ça l'immersion narrative. C'est le truc le plus puissant qui soit. Et nous on travaille sur un truc grand plus qui est l'immersion perceptive. Par le son. Après tu as l'immersion perceptive par l'interactivité, l'interaction, ce sont les jeux, c'est ce qui fait que moi quand j'étais gamin, avec le truc noir et blanc les deux raquettes et la balle carrée je jouais des heures avec mes cousins sur un truc ou c'est deux barres blanches et des balles, parce que tu interagis quoi et tu es parti, tu es immergé dans le truc, c'est ce qui fait la puissance du jeu. Le jeu c'est une des plus grosses économies de l'Entertainment. Si ce n'est la première. Je crois que le jeu a une économie bien plus solide que celle du cinéma.

Et c'est quoi ? Le fonds de commerce il est là, et pourquoi Apple ou Meta mettent des milliards, je pense qu'ils mettent beaucoup d'argent, sur l'immersion avec l'Apple visio ou les Quest, parce que c'est là qu'est le pognon, parce que c'est là qu'est l'attraction.

V : Et l'immersion perceptive en musique, elle n'est pas encore suffisamment appréhendée par les artistes, si ?

H : Non elle l'est pas du tout. Après la musique elle a cette capacité. En fait il y a une frontière, elle va passer par l'écriture, et là je pense qu'il y a 10-15 ans de boulot, elle va passer par l'écriture, l'évolution des logiciels, il y a une révolution à faire. Écrire sur le plan spatial ça couture du pognon. Et faire des œuvres qui soient vraiment immersives avec l'image en plus, ça va couter beaucoup de pognon. Donc là il y a une révolution à faire sur le plan de la répartition des richesses, et ça, ça ne va pas être facile. Et aujourd'hui tu es dans un monde où les maisons de disques, elles prennent des habitudes : le mec arrive avec sa prod, c'est lui qui a tout payé. Mais une prod stéréo en Home Studio bien léchée, c'est 10-15 000 bals d'investissement, sans parler des heures de travail du musicien qui ne sont pas payées. Si je rajoute le musicien, admettons que le musicien soit très bon il fasse ingénieur du son et compositeur, un album c'est au bas mot 7-8 mois de boulot quoi. 7-8 mois de boulot avec quelqu'un qui toucherait 2500 bals pour vivre net, je compte vraiment au plus bas. On est à 25 000, un album c'est 30 000 bals, au bas mot, c'est vraiment un gamin qui bouffe des pizzas dans sa chambre, si on tient compte de ces heures de travail.

Aujourd'hui la réalité économique c'est quand je discute avec Cécile⁹⁹, on regardait l'autre jour avec Apple Music, ce sont ceux qui paye le mieux en te disant qu'ils vont te payer chaque Stream, et là j'ai un doute parce que je pense qu'il y a des arnaques là-dessus. À 1 centimes du Stream en Atmos, sinon c'est 0.9 centime de Stream, admettons que chez Apple elle fasse 1 millions de Stream, faut les faire hein, à la fin elle a pris 10 000 bals. 10 000 bals si je prends ces 7 mois de travail acharnés, sans parler de l'argent qu'elle a lâché à payer le mix stéréo. Ici à Radio France elle ne paye pas le studio, on a fait des deals avec elle ce qui fait que le studio pour l'instant il est gratos et que nous on investit en recherche. Rien que sur ces heures de travail à la fin j'arrivais à un 8-9 euros de l'heure charges comprises. Ça ne va pas pouvoir durer quoi. Ou alors ça va durer comme ça, mais rajouter un travail d'écriture réel, et rajouter les médias immersifs c'est à dire faut travailler de l'image et compagnie, moi je vais me lancer

⁹⁹ Mentionne l'artiste Cécile DeLaurentis avec qui Hervé travaillait en avril 2024.

avec France TV, on va voir s'ils signent à essayer de trouver des workflows qui soient très rapide en création image, même très rapide ça va couturer des sous, ce n'est pas rentable, c'est un truc de taré. Un musicien il est aussi peu payé, enfin peut-être un peu mieux payé [que les paysans, ndlr] parce que malheureusement aujourd'hui, les paysans, le principe c'est ils perdent du fric quoi, on est quand même dans un monde de tarés, les producteurs, ceux qui produisent de quoi nous nourrir perdent de l'argent quand ils produisent parce qu'il y a des gros malins qui prennent une valeur ajoutée.

Ce monde-là il est en train d'exploser parce que de toute façon nous sur le plan de la création ça ne va pas être possible. Parce que selon moi aujourd'hui les musiciens qui vont vraiment être des gens extrêmement intéressants sur le plan de la spatialisation, sur le plan de la création immersive, sur les nouvelles formes d'écriture, on ne va pas parler de l'immersion on s'en fou, c'est un tout, l'immersion fait partie, va faire partie intégrante de l'écriture musicale. Faut qu'ils sachent, pour moi va falloir qu'ils sachent travailler parfaitement l'harmonie, la polyphonie, la rythmique, le matériau sonore, l'esprit Schaefferien, plus la spatialisation. Et le mec il touche 8 bals de l'heure ? Même pas en rêve quoi.

V : Les artistes n'ont pas le temps d'écrire l'espace donc à chaque fois tu l'écris avec eux.

H : Ouai, après j'ai de la chance, ça fait 20 ans que je fais ça donc j'ai une vision du truc. Les artistes aujourd'hui viennent chercher ma compétence en premier parce qu'ils savent que ça va le faire, et la démonstration elle s'est faite avec Jean Michel Jarre, il vient me voir parce que je le connais, c'est quelqu'un qui écoute avant tu vois, il vérifie, et après quand il dit « ok ça, ça m'intéresse j'en ai besoin pour ma musique », il vient nous voir quoi. Parce que j'ai 20 ans de ça et que je peux être force de proposition. Et après faut démontrer que tu es force de proposition, tu peux écrire, ou alors tu peux t'adapter. Maintenant moi, personnellement, dans mon deal aujourd'hui, je travaille avec des artistes qui viennent chercher mon écriture. Et qui vont la faire évoluer, et avec qui je vais partager. Je n'ai pas envie de travailler avec quelqu'un qui vient me voir et qui me dit « bon super tu maitrise Nuendo et L-Isa, tu maitrises ta chaîne, donc maintenant je vais bosser avec toi et tu vas faire comme ça comme ça comme ça », non ça m'intéresse pas.

V : Et c'est cette écriture sur laquelle tu ne souhaites pas communiquer et dont Radio France doit profiter ?

H : Là en l'état oui. Mais dans le déroulé, sans rentrer dans les détails, l'artiste m'envoie le multipiste, et je vais écrire à 80-70%, je vais écrire le truc tout seul, et après il vient avec moi en studio, c'est-à-dire que sur un album, sur 10 titres il me faut entre 5 et 7 jours pour faire les 10 titres dégrossis en écriture, et après il va venir et puis on va finaliser, maintenant dans les très grandes lignes de l'écriture, oui c'est de l'équilibre spatial, oui c'est de faire plaisir à l'oreille. Mon travail s'appuie essentiellement, et c'est ce qui fait peut-être mon originalité, et je suis atterré que dans un boîte comme le CNSM on enseigne pas plus ça, c'est l'esprit des objets musicaux de Schaeffer. Si tu veux savoir où je prends ma source d'inspiration, c'est ça quoi. *En montrant le livre de Schaeffer*

C'est ce truc-là. Tout mon travail il est là-dedans. C'est l'objet musical, c'est cette philosophie. Oui c'est une forme de philosophie parce que déjà Schaeffer pose un point philosophique le jour où il dit qu'à partir du moment où on avait enregistré sur un microsillon,

le son on sait le lire à l'envers à l'endroit, le son est plus seulement l'énergie et un moyen de transport d'une modulation, il devient matériau. Après moi l'écriture spatiale elle est basée sur le mouvement, et le mouvement ou pas, mais l'équilibre d'un espace. Et le mouvement il est basé sur l'analyse de la morphologie des sons.

V : Est-ce que cette référence est aussi pertinente en stéréo ?

H : Oui. Oui. Bien sûr. Là aujourd'hui je pense que Schaeffer et Boulez, et plus Schaeffer d'ailleurs que Boulez, mais Schaeffer n'a pas su faire de l'institutionnel, ce que Boulez a su faire. Il a fait l'IRCAM, il a fait la Cité de la Musique. Boulez il a vraiment posé le truc.

V : Après Schaeffer il a fait le GRM.

H : Oui enfin tu vois d'un point de vue institutionnel, l'IRCAM derrière ils ont dépoté. Tu me diras il y a les GRM Tools, d'accord, enfin l'IRCAM eux ils ont des process qui aujourd'hui marchent dans l'industrie, ils se sont faits bien pompés au niveau de l'industrie culturelle, mais l'industrie industrie, tu vois l'automobile et compagnie il y a eu plein de process dans les années 90 c'est l'IRCAM qui les avait pondus et ça a servi. Quelque part Boulez a réussi ce coup-là, après ils ont quand même pas mal théorisé de choses, on peut leur reconnaître ça. Le GRM c'est beaucoup plus light, le seul qui a théorisé ça reste Schaeffer.

Je pense que Schaeffer dans l'esprit, la manière dont il voit le matériau sonore, pour moi la continuité aujourd'hui c'est Live d'Ableton. C'est vraiment là-dedans. Pour moi c'est l'outil Schaefferien. Je suis incapable de dire si c'est vraiment inspiré de ça, après les choses tu sais, une pensée elle finit par commander, par aller dans les esprits et faire son œuvre. Je pense que Schaeffer il y a de ça. Mais pour moi l'outil aujourd'hui sur lequel un musicien doit bosser et travailler leur art c'est Live d'Ableton.

V : Et quand tu parlais d'équilibre de l'espace, on parle d'énergie ? De timbre ?

H : Je ne sais pas si tu as vu les boulot de Guy Laporte sur l'analyse de l'image sonore. C'est un nombre de critères d'analyse qui permettent de qualifier une image sonore en stéréo. Tous les critères qui sont valables dans l'analyse de l'image stéréo, tu peux les appliquer à l'image immersive tu vois. Qu'ils soient sphériques déjà, les équilibres, l'équilibre spectral, les profondeurs, les plans, la largeur. Ça s'applique de fait à l'espace. Je vais te dire un truc, si je devais former, je n'ai pas cette chance de former des élèves, mais si je formais des élèves à l'espace je peux te dire que pendant un mois ils ne touchent pas un 5.1 ou un stéréo. Je les fais bosser sur un mono. Parce que savoir-faire une fuyante, savoir-faire une profondeur en mono, quand tu as compris les concepts avec une mono, après en espace tu vas tout déchirer. Parce que quand tu attaques directement soit par la stéréo ne soit pas l'espace, quand tu n'as pas l'expérience quand tu n'as pas la compétence, de fait tu vas étaler les sources pour t'affranchir des problèmes de masquage, de compréhension de l'image et compagnie. Ça ne va pas te permettre de comprendre les effets de masque et surtout la manière dont le cerveau interprète une image sonore en fonction des contraintes. A partir du moment où tu mets de l'itd et l'ild, voilà. Alors que construire une image sonore par les codes, notamment, de la balance spectrale, par le spectre, par les jeux de réverbérations, de modulation des spectres, alors là après tu vas déchirer, là tu vas envoyer du lourd. Même les musiciens faudrait les former à ça.

Après tu as des gens qui ont ça naturellement. Cécile avec qui je travaille, Jean Michel Jarre je n'en parle pas, lui il l'a développé, comme il a beaucoup bossé, je crois 1 an dans les ateliers de Schaeffer, et puis après Jean Michel Jarre il a un cerveau qui va super vite donc il acquiert très vite parce que c'est vraiment un passionné de musique et de construction. Cécile, qui a un bon niveau en technique, sur cette construction-là elle est beaucoup plus empirique et elle sort des trucs, tu te dis « c'est magique ». Parfois les couleurs, c'est un peu agressif, mais quand tu mélanges le tout tu fais « bah oui la construction elle est belle, il y a un ensemble. » Et puis c'est son univers. Tout ça pour te dire qu'il y a des musiciens qui l'ont naturellement ce que j'appelle la construction. C'est-à-dire je mets dans ma construction musicale, dans ma composition, il y a un moment je vais aller chercher tel spectre ou tel spectre. Je pense que dans la tête de Cécile elle ne parle pas en termes de spectre, tiens je vais rajouter telle voix ou telle voix, tiens là j'ai un espace, je pense que c'est plus ou moins conscient qu'elle va remplir et ça donne de la musique qui a un certain niveau de complexité et qui joue bien de la perception. Et aujourd'hui un ingénieur du son, quoi que le terme m'embête parce que je crois qu'aujourd'hui c'est plus vraiment de ça dont on a besoin, mais je crois que quelqu'un qui va participer à la création dans le domaine de la musique, il faut que ça il l'ai parfaitement en tête. Et donc je reviens à ce que je disais, je pense que le meilleur, le mieux, à une époque j'étais en électroacoustique, il y a eu plein de titres ou les deux premières minutes c'était de la mono, avant d'ouvrir, premièrement parce que ça joue du contraste, et le jeu c'était de garder l'attention sur une mono. Donc travailler au mieux cette diffusion.

V : Si je te pose la question de la grammaire du mixage immersif, tu me parlerais de critères du même type que ceux de Guy Laporte ?

H : Oui, comme je te disais dans un mix je vais être en permanence à essayer, c'est pour ça que je demande en général pour avoir une référence, je demande à avoir la stéréo masterisée, premièrement, comme ça on va essayer de faire le lien, d'avoir une cohérence entre la stéréo et le mixage immersif qui sera sur la distribution. Tu n'es pas à une balance tonale totalement différente faut pas que ce soit non plus... Après on s'autorise avec l'artiste ou le producteur des variantes, je ne vais pas coller, surtout parfois la version master elle est cracra, c'est un peu boucherie. Après ouai si je fais ça j'ai plus d'espace quoi, j'aurais 5 haut-parleurs qui vont gueuler pareil, mais au moins ça m'oriente vers la balance tonale.

V : Et comment se passe la communication avec les artistes ?

H : En général ça se passe bien. Ce qui est bien compris s'explique bien. A partir du moment où toi tu sais pourquoi tu fais ça, tu l'as analysé, que t'as mis des mots dessus, tu les transmets et celui qui t'écoutes, s'il n'est pas con, tu vois s'il dit une connerie ou si ça fait sens quoi. Maintenant dans la vie tout est relatif, une vérité peut avoir sa contre vérité le lendemain. Donc il faut se méfier. Je me méfie de la vérité en fait parce que je pense qu'elle n'existe pas, elle nous sert juste à nous rassurer.

V : Quelles sont les difficultés que tu rencontres quand tu travailles une image immersive à 360° ?

H : Je n'ai pas de difficultés particulières, la seule difficulté que je peux rencontrer c'est quand je me plante sur la personnalité d'un son. C'est à dire que je colle quelque chose qui ne correspond pas, un mouvement par exemple, qui ne correspond pas à ce son, ou une position

qui ne correspond pas, parce que je l'ai mal compris dans le mixage, je l'ai mal compris dans la composition. La difficulté en musique, mais de toute façon en fiction tu auras le même problème... Si tu n'as pas compris l'intention du compositeur, tu vas te planter. Jean Michel Jarre, c'est une écriture très établie et parfois très subtile, il m'a fallu 2 albums, et même sur Oxymore il y a des trucs que j'ai laissé de côté et puis dès qu'il arrive c'est la première question que je vais poser : « Pourquoi tu l'as fait ? Tu veux où ? C'est quoi son rôle ? » Et au bout d'un moment j'ai compris, parce que tu veux une fusion, parce que si un musicien utilise un élément sonore pour donner une fusion, c'est à dire donner une nouvelle couleur avec un autre élément, tu ne peux pas les séparer, le truc il fait flop, ce sont des trucs comme ça. C'est là où il ne faut pas te planter, mais après c'est là où tu dégrossis et c'est là où il y a tout le travail ensemble où il dit « ah non là il faut le mettre là tu ne peux pas faire ça », ça, ça arrive, et après c'est un travail... Parfois le musiciens il n'entend pas toujours. Je fais des trucs et le musicien il dit « ah c'est super », une fois qu'il a dit ça tu ne peux pas trop revenir dessus alors que tu sais que tu t'es planté sur le mouvement ou sa position. Et quand je parle de sa position ce n'est pas forcément sa position azimutale, c'est sa profondeur, sa position globale, sa grosseur, son ampleur, sa masse, sa valeur de masse. Chez Schaeffer c'est très important les valeurs de masses. Est-ce que t'es pointilliste ? Est-ce que c'est une masse que tu veux ? Et parfois ça se joue là-dessus, la largeur d'une source sonore va faire premièrement qu'elle va rentrer ou sortir du mix en fonction de ce que tu veux faire mais elle va aussi jouer de l'importance du son dans le mix à un instant donné et ça ce sont les phénomènes de masse, c'est le critère de masse.

Et la balance tonale bien évidemment, largeur, profondeur, est-ce que j'ai vraiment l'impression d'être dans un espace, ça part loin ou est-ce que je suis dans un espace hyper constraint. Parfois on cherche l'espace hyper constraint, et moi d'ailleurs parfois j'ai du mal avec ça, putain ça manque d'espace, alors que Cécile elle est là « non non, le truc il est là en permanence voilà il n'y a pas d'espace je ne veux pas d'espace ». Voilà à un moment tu rentres dans une cathédrale, et après tu rentres dans une bagnole hyper cossue, voilà c'est ça, c'est ni plus ni moins que ça, et ça aussi c'est un acte de création.

V : Est-ce que c'est plus facile que le mixage stereo ?

H : Ça dépend. Prends quelqu'un qui écrit de manière simple en stéréo. Faire de l'espace ça peut ne pas être plus simple, ça peut être plus compliqué parce que t'as pas assez de matière. Tu vois le truc. Si t'as une stéréo qui est écrit avec un minimum d'élément et qu'en plus tout est basé sur de la fusion alors là ça devient complexe parce que tu sais pas quoi foutre tu vois, Les 1er mixs qu'on a eu avec Frederic Changenet on en rigole encore, le truc tu dis « le mec il dit "j'ai tout fait pour de la mono". Oui je confirme parce que là si j'étais, ça fait flop quoi c'est nul », à partir du moment où ça sonne dans un seul haut-parleur, déjà en stéréo ça doit être limite alors dans un système 5.1 j'ai cassé toute sa musique tu vois, c'est aussi ça la difficulté. Le mec il maîtrisait, il était bon, il savait de quoi il parlait, et j'étais d'autant plus mal, il me dit « mais ouai la prod elle ne m'avait pas prévenu que c'était pour faire du mix immersif pour séquences », il me dit « ah ouai fallait pas que je joue ça », et ce qui était d'autant plus compliqué c'est que le mec savait de quoi il parlait, il savait pourquoi il faisait ça quoi. Voilà ça peut être aussi une difficulté. Et à l'inverse, c'est de pas avoir assez d'élément pour spatialiser alors que le musicien te le demande.

Si la musique est bien écrite en stéréo, tu n'as pas plus de difficultés, les arrangements sont bien faits, l'orchestration est bien faite, donc sur le plan spectral tu n'auras pas de recouvrements, d'effets de masque. Après l'ingénieur du son qui a suffisamment d'expérience, sur les quelques galères d'effet de masque, je sais pas s'il va jouer d'autres comportements avec soit des filtres dynamiques, soit du MS, des trucs tricky pour donner de l'espace à son mix et faire ressortir des éléments à un instant donné en faisant revenir les autres à un autre moment, après c'est de la technique de mix en fait, c'est comme de cuisiner quoi à un moment faut aussi avoir des astuces, de l'expérience pour ressortir des trucs. En spatialisation t'auras les mêmes choses quoi.

Maintenant la spatialisation elle va permettre de plus en plus d'exprimer des œuvres orchestrales et je pense qu'à termes il y aura des trucs bien plus complexes que l'écriture symphonique telle qu'on la connaît. C'est pour ça qu'ils me font tous rires avec de la musique savante avec des arpeggiator, des synthés et des samplers, ils commencent déjà à se la prendre dans la gueule, elle existe déjà. Tu prends le travail de Max Cooper c'est de la musique savante c'est hyper chiadé c'est hyper pensé, c'est de la putain de composition. Mais il n'y a pas que lui, tu prends Phillip Glass, tu prends tous ces gens-là même Jean Michel Jarre tu as un niveau d'écriture... les mecs ils peuvent s'accrocher. L'harmonie, la mélodie et la rythmie c'est bon il connaît ça, mais en plus de ça il a un truc que beaucoup n'ont pas dans la musique classique c'est que lui il fait sa lutherie lui-même, c'est que lui il invente ses instruments. Et ça c'est la piste à suivre. Avec Cécile on s'est rencontrés et je lui ai fait découvrir la musique électroacoustique et l'influence que pouvait avoir des sons sur le rendu global d'une œuvre et je sais que là ça y est, elle a compris elle est tombée accro du truc. Ça change complètement. Un son que t'as bidouillé à partir d'un enregistrement de n'importe quoi, d'un crayon qui roule sur une table, sur le coup tu bidouille un peu... Tu vas chercher le moment où t'as un truc bizarroïde, tu le mets dans la compo et d'un coup t'as un truc, c'est comme les aromes en cuisine. Tu as mis un truc et là « c'est quoi ce goût ?! ». C'est exactement ça, c'est la même chose.

Ce travail-là il est fondamental. Et l'œuvre de Schaeffer est basée là-dessus. C'est écouter les sons, écouter leur musicalité, et recréer des musiques à partir de ça. Composer tes propres univers à partir d'un univers réel déformé et là on rentre dans cette chimie. C'est un truc de dingue. Bach était un passionné de la facture d'orgue parce que je suis sûr que pour lui c'était les synthés. Tu vois je peux délicher avec plein de couleurs, si je me rappelle bien quand il voyait des facteurs d'orgue, il leur prenait la tête pour avoir des jeux supplémentaires à chaque fois au même titre qu'un musicien aujourd'hui qui voit un synthé et qui voit plein de trucs et qui se pose plein de questions. C'est la même chose, même raisonnement, même démarche, même gourmandise, même curiosité. Je crois que c'est le point commun entre Einstein et Bach, quand ils s'adressaient à leurs élèves et qu'ils étaient un peu agacés du regard émerveillé de leurs élèves. Ils disaient « mais ne nous prenez pas pour des génies, on est simplement un peu plus curieux que vous. » Je pense que c'est ça, c'est la curiosité qui génère la créativité. Et aujourd'hui faire des orchestres de synthèse c'est à la portée d'un musicien, c'est une réalité.

V : Avec Cécile, dans votre travail, vous avez aussi bossé avec l'image pour le live que vous préparez, est-ce que ça change quelque chose pour toi ?

H : Pour l'instant l'image va suivre le son et la musique. Dans le process de création des images 360, c'est le créateur vidéo qui part d'un univers graphique, simple, c'est ce que travaille

Cécile pour ses pochettes, une pochette par titre, et le créateur graphique a deux choses, il a cette pochette-là, et la stéréo. Et à partir de là il fait sa trame il raconte son histoire. Pour l'instant ça marche bien mais fallait qu'on trouve un workflow. Pour l'instant il sort une image à partir de la stéréo, et tu te rends compte que c'est génial parce que ses mouvements de l'image sont basés sur les mouvements musicaux et ça c'est essentiel. Et après on va associer des mouvements d'image avec des mouvements de son dans l'espace. C'est notre workflow actuellement.

V : Les méthodes de travail n'ont pas changé.

H : Pour l'instant c'est la musique et le son qui génère l'image. Peut-être que la prochaine fois on va travailler sur des process inverses ou en tout cas qui vont dans les deux sens, c'est à dire que l'image influe sur le son et la musique, et la musique et le son influent sur la matière de l'image, mais là-aussi c'est une grammaire... On y va, on y va vraiment pas à pas.

V : Quand tu dis « c'est une grammaire », c'est à dire ?

H : De l'écriture de l'image et de l'écriture du concert. Parce que dans les émotions du concert on veut que déjà c'est la musique, le son et la musique, qui mergent, il y a l'image, il y a les gens sur scène, et donc là il y a une grammaire à établir d'attention, c'est-à-dire qu'il y a des moments où il y aura plus d'image, on va remettre dans le noir. On veut que l'imaginaire reprenne le pas. Et puis Hop, on détourne l'imaginaire, on réimpose quelque chose. Et puis après c'est ce qui se passe sur scène qui attire l'attention. On va être dans ces travers-là.

V : Donc finalement l'écriture spatiale va être très différente dans un contexte live ou dans un contexte studio.

H : Dans les grandes lignes elle va être identique, mais adaptée au live. Déjà je vais passer dans la tuyauterie de la console de mixage etc. Après faut que ce soit facilement reproductible, et puis l'axe de travail c'est aussi d'avoir quelque chose de super écrit mais sur lequel on prend la main où on veut et quand on veut. Ce qui sous-entend que, comme je le dis avec humour, c'est que si je fais une crise cardiaque ça marche tout seul, Cécile elle peut démarrer le show et le truc il va rouler tout seul. Ou alors tous les deux on prend des sons à n'importe quel instant et on joue avec, d'une manière ou d'une autre, à partir de capteurs, de son gant, de ses contrôleurs de l'iPad etc...

Donc moi mon travail sur le live, premièrement parce que je pense que ça correspond à un besoin de ce que fait Cécile, ça répond aussi au besoin de travail en recherche en écriture pour le live avec ce que fait L-Acoustic et ça correspond de toute façon à ce qui a toujours été fait en musique c'est à dire que tu as des musiciens qui jouent à la partouche et tu as des musiciens qui jouent totalement à l'inspis et tu as des musiciens qui jouent entre les deux. Et je pense que ce qui est intéressant c'est de pouvoir répondre aux deux. Enfin au trois, soit totalement écrit, improvisé ou entre les deux. En fonction du mood ce qui permet aussi de faire évoluer le live au fil des dates.

V : Tu utilises le même setup qu'avec Molécule¹⁰⁰ ?

¹⁰⁰ Fait référence ici à la tournée *Acousmatic 360* de Molécule en 2019.

H : Oui

V : Parce que c'est le L-Isa est le système le plus abouti ?

H : Il y a plusieurs aspects, le premier c'est que premièrement ça permet d'écrire un album et d'écrire un live en toute fiabilité. J'ai reçu le L-Isa pour la première en 2018 et je n'ai jamais eu un seul beug, pas un seul. Il ne m'a jamais planté. Comparé au Spat autant te dire que c'est le jour et la nuit. Ensuite associé à Nuendo en Studio ça marche d'enfer et associé à l'époque au patch Usine qui aurai pu être un patch Max ça marche d'enfer. Après avec L-acoustics on est dans une démarche d'innovation de dev en innovation et de développement d'un outil de référence dans le modèle, dans l'exploration de l'écriture et d'un show live ou d'album. Et à chaque fois que je discute avec eux, la première fois que je les ai vu à Graz ou je faisais écouter la série Séquences, là on s'est dit qu'il fallait qu'on bosse ensemble. Je n'étais pas trop sono je m'en foutais un peu. On s'est dit qu'il fallait qu'on fasse un truc et puis avec Molécule on a essayé de le faire et en 2018 on a commencé à signer des NDA, des accords en innovation et ce travail-là continue. Moi ça fait 6 ans qu'à chaque fois qu'on s'appelle on doit s'appeler avec L-Acoustics 5/6 fois par ans, à chaque fois on parle de la même chose et on voit l'avenir de la même manière, voilà pourquoi je travaille avec eux. Quand je parle avec D&B et d'autres je n'ai pas cette vision-là, je ne perçois pas de vision, je perçois des gens qui suivent dans ce domaine-là ou qui ne sont pas convaincus pour certains. Ils suivent en fait, ils en font parce que tout le monde est en train d'y aller parce que c'est un mouvement. En revanche quand je suis avec L-Acoustics je vois des gens qu'ils ont une vision. Quand en plus cette vision correspond à celle qu'on partage à Radio France je vais aller on bosse avec eux. C'est ça l'approche.

Et puis après leur système fait qu'aujourd'hui tu achètes un processeur L-Isa ou tu prends un L-Isa gratos, tu as un outil super quoi. Et puis tu vois l'avenir, il y a ce que je te dis et ce qui est en train de se préparer qui va sortir là, un gros truc avant la fin de l'année, d'autres qui arrivent l'année suivante, enfin c'est monstrueux, ce sont des visionnaires, au même titre qu'Apple est visionnaire dans son domaine, là c'est vraiment de ce niveau-là, et avec une ambition qui me colle à la peau. c'est une entreprise française, ça pour moi c'est très important, et donc européenne, et je pense qu'il faut qu'on pense européen.

V : Comment le L-Isa gère l'élévation ?

H : Le L-Isa c'est le processeur, avec le contrôleur, et là-dedans tu as de l'élévation. En revanche tu as de l'élévation qui a le même défaut qu'avec le Spat Révolution, il y a que Dolby qui n'a pas ce problème-là, c'est la fameuse triangulation que tu as dans le Spat qui fait que tu as des positions où le son il peut partir à l'opposé, dans un dôme, parce qu'à un moment il va chercher à un troisième haut-parleur le plus proche et il y a un passage où il est là-bas et ça te donne des incongruités. C'est dû à des panoramiques de type VBAP et autres, ils sont en train de bosser là-dessus. C'est un point faible mais que t'as aussi sur le spat.

V : Sur un set up live tu n'utilises pas d'élévation si ?

H : Si sur le live nous par exemple on va en avoir. Si on le fait à la Jam Capsule, il y a 8 haut-parleurs en hauteur et je vais les exploiter, mais de manière assez simple. C'est-à-dire que

premièrement je reste dans le haut et je ne fais pas de passage bas-haut parce que à cause de cette triangulation il y a trop de danger. Mais ça me va très bien pour l'instant. Et si j'ai besoin de faire des mouvements de hauteur, sur un truc qui monte, je vais plus le faire avec des pistes qui, en énergie, vont être... Je vais construire ça vraiment au mix, je ne vais pas utiliser les panoramiques, parce que ça ne marche pas toujours.

V : En studio, c'est la même manière de travailler sur la hauteur ? Tu te limites en termes de mouvements ?

H : Non j'ai fait des mouvements de masses, mais là je duplique les pistes entre le haut et le bas, c'est à dire que si je veux faire une masse je vais dupliquer la piste et cet équilibre entre le haut et le bas je le fait avec les deux pistes et leur niveau relatif. Parce que les panoramiques ça ne marche pas très bien et puis en plus en faisant ça là tu donnes des effets de masse, t'as des trucs monstrueux qui passent en fait. Là c'est aussi pour travailler les masses.

V : En studio, à quel point tu considères une image frontale et à quel point tu gardes une frontalité dans le mixage ?

H : Ça dépend. Là, à partir du moment où t'as des voix, enfin plus exactement à partir du moment où tu as des voix lead, tu as forcément une frontalité. Parce que c'est ton sujet qui va déterminer ça ou pas. Avec Jean Michel Jarre on n'est pas toujours d'accord là-dessus, parce qu'il veut de la frontalité de fait parce qu'il veut être celui qui est le centre de l'intérêt, c'est comme ça que ça s'écrit aussi le show. Moi ça va dépendre, par exemple Arthur H il y a forcément un plan frontal. Sur les musiques électro, si tu as un éléments lead, qui doit être le centre de l'intérêt il va forcément être là devant. Après est-ce que c'est dans le central ou dans la stéréo, ou équilibre entre central et stéréo. C'est à la fois la décision qui se fait en fonction des stratégies de mix et en fonction de si je veux quelque chose de pointilleux ou de plus large et puis si je me dis que ce master là il y a peu de chance qu'on l'écoute dans une salle de cinéma donc je peux me permettre de faire un stéréo fantôme ou une image fantôme je joue avec ça tu vois, ça va être des critères d'esthétique et des critères d'équilibre et de construction d'image.

V : Pourquoi avoir choisi ce set up dans la boite avec un dôme qui virtualise des enceintes plutôt que d'utiliser une enceinte par point de diffusion ?¹⁰¹

H : Tout simplement parce que le studio, quand on l'a pensé avec Frédéric... Premièrement à l'époque on bossait beaucoup avec l'Ircam, on voyait les limites du 5.1 et de toute la clique, et on était sûr que de toute façon c'était la piste de l'Ircam qui était la bonne, et avoir un studio agnostique c'était une bonne idée. Associé à Nuendo avec sa barre de monitoring c'était la deuxième bonne idée. Le spatialiseur qu'il y a dans le cube c'est le vieux Waves, qui marche très bien, dedans en permanence quand tu le démarre tu vas virtualiser un 5.1, un 8.0 et un 7.1.4. tu as un certain nombre de canaux, les 8 premiers canaux, les 12 canaux suivants etc... Et dans Nuendo, barre de monitoring, tu appelles ces canaux là en fonction de ce que tu veux monitorer. Déjà gros gain. Nous on était persuadé et on est encore plus persuadé aujourd'hui que les formats industriels ils vont disparaître, ça va plus vouloir dire grand-chose. C'est le pipe

¹⁰¹ Fait référence au studio d'Hervé Déjardin à Radio France, « la boite », qui est équipé d'un dôme de haut-parleurs qui virtualise d'autres systèmes d'écoutes.

et les métadonnées qui arriveront qui donneront une restitution en fonction de ce qu'il y a, et là tout va devenir possible, par des réseaux de HP, etc.

Donc aujourd'hui le Dolby Atmos faut vraiment en avoir rien à foutre et juste être un suiveur, c'est l'essentiel de la prod mais je ne leur en veux pas c'est comme ça, ce n'est pas quelque chose de négatif, les gens ont besoin de faire de l'argent ils gagnent leur vie avec ça ils n'ont pas forcément envie de se poser des questions existentielles par rapport à ça, ils suivent un mouvement et Dolby vit de ce mouvement. Tant mieux pour eux, mais ce mouvement là il va s'arrêter dans les 5 ans qui viennent. Ce truc là il va tomber parce que de toute façon, premièrement, tu vois que les pipes qui sont en train de sortir, ce qui préparent notamment chez les Chinois, n'est plus basé sur du canal, c'est vraiment basé sur de l'objet. Et il y a plus aucune philosophie canal mis à part si tu veux transporter un stéréo et un 5.1 là-dedans.

Ici c'est agnostique et la prochaine révolution dans l'audio, s'il y a une révolution, elle va se faire sur la restitution. Parce que je suis désolé mais un 5.1 ou un 7.1.4 on va essayer de passer à autre chose au bout d'un moment. C'est à dire que les images fantômes c'est bien gentil mais il va falloir arrêter le délire. On va arrêter de contraindre le binaural parce qu'on est sur du speaker, le pipe Dolby c'est un pipe qui n'est pas pensé pour ça d'ailleurs. Il n'a pas été pensé pour ça c'est un pipe optimisé pour un truc qui n'étais pas pensé en natif, il est pensé pour des salles de cinéma et pour être véhiculé sur du DCP, pas par de l'AAC sur 16 canaux. Alors Dolby ils peuvent être gentils de débouler en disant « ah ouai nous on fait de l'AC-4. » Super mais personne ne le diffuse à part Tidal, très content pour eux, mais les mecs dites-vous bien que vous ayez quand même raté un truc : Spotify, le plus gros distributeur de musique, a fait un peu d'immersif mais ils ont plus ou moins abandonné, mais il va falloir revenir là-dessus, ça c'est un problème, Deezer je ne sais pas, en France on a Qobuz, mais je ne sais pas, j'ai l'impression qu'ils sont encore au 5.1. Les seuls qui fassent de la diff à grande échelle : Tidal mais aucune audience, ou en tout cas peanuts, l'audience de masse c'est Apple.

Sony a fait le Sony 360. Quand tu discutes avec des artistes avec qui je travaille ils sont tous chez Sony, Et pour eux t'as aucune recommandations Sony 360, c'est la recommandation Apple Music. Tout est dit. Dans l'Apple Music, c'est le pire de l'Atmos, vaut mieux écouter de l'Atmos chez Tidal que de l'Atmos chez Apple sur le plan de la restitution et du respect du travail de la production. Mais aujourd'hui l'essentiel de l'audience elle est chez Apple, tu te dis « je vais essayer de faire un master qui fait que, dans 10 ans, je n'aurais pas l'air d'un con » mais pour l'instant faut qu'il soit optimisé et sur la balance tonale et sur le plan de spatialisation sur la galère d'Apple, parce qu'Apple c'est une boucherie. Donc ça c'est la réalité d'un marché. Aujourd'hui c'est évident que sur les 5 à 10 ans à venir il y en a forcément un qui va tirer sur des choses beaucoup plus précises, sur des choses qui seront beaucoup plus adaptées parce de toute façon c'est le sens du vent de l'objet, l'objet est adapté pour ça. Ce n'est qu'une première étape. L'objet, en officiel, il sort en 2012. Il y a 12 ans, c'est tout jeune, c'est vraiment tout jeune. 12 ans il y a moi et Fred et des gens comme nous qui connaissions, des aficionados. Aujourd'hui la plupart des gens que je croise même chez Sony tu ne peux pas parler objet avec eux ils ne comprennent rien.

Donc c'est une étape de l'histoire de la distribution du son. Une nouvelle étape. Et c'est un intermédiaire, c'est une page qui est en train de se tourner mais elle n'est pas tournée encore pour l'instant.

V : En termes de grammaire immersive, est-ce que tu as des choses à ajouter ?

H : Mis à part de manière générale, et je suis toujours étonné que ça n'apparaîsse pas comme une évidence pour beaucoup de gens qui font le même métier que moi, c'est que le cerveau lui il aime la différence, le cerveau il discerne les choses par le contraste entre les choses, et que dans la vie de tous les jours nous sommes en permanence sollicités par des sons en mouvement, des mouvements de fusion, de séparation de sources, ce n'est que ça. Ça doit être la base de notre travail parce que c'est ce qui construit l'espace naturel en fait. Et la plus belle, le plus beau des commentaires que j'ai eu il y a une semaine, on m'a dit ce qui est vachement agréable, quand on écoutait avec des gens de scènes nationales, notamment Montbéliard, ils me disaient « mais ce qui est vachement bien », ou c'était les gens de la Seine Musicale, eux m'ont dit un truc étonnant et c'est vachement agréable c'est que « oui il y a du mouvement ça bouge beaucoup mais ça sonne extrêmement naturel », il y a une fusion dans ces mouvements simplement parce que je m'inspire de la nature de ce que j'entends dans la nature tu vois, il y a plein de trucs dans la nature qui bougent que t'entends pas vraiment, tu les conscientises, néanmoins ces choses relatives qui créent des mouvements spectraux permanents parce qu'en fait on analyse l'espace, faut jamais perde ça de vue, c'est ce qui est important.

[...]

Je peux passer beaucoup de temps à préparer un projet Nuendo mais après les plats je te les sers à la vitesse éclair. Parce que c'est ce qu'on m'a enseigné en lycée technique. D'abord, on apprend vraiment la grammaire de base. Energie efficace, énergie moyennée. Tu comprends très vite pourquoi un boomer il ne donne pas ce qu'il doit donner. Et je me suis aperçu que parfois je nettoyais mon image sonore sur des concepts basés sur la loi d'Ohm et sur des concepts d'électricité. Au même titre que ça cette conception je la transpose dans l'espace, je m'aperçois que mes profs de BEP m'ont éduqué un truc de base qui me dit « concentre toi sur les trucs de base ». Sur la construction spatiale je m'inspire de la nature en permanence.

Aujourd'hui tout le monde veut travailler avec des traitements dynamiques d'il y a 40 ans, ça n'a aucun intérêt pour moi quoi dans mon travail. Ce sont des trucs dont tu as du mal à maîtriser la temporalité parce que ce sont de vieux systèmes, tu n'es pas à la micro seconde, tu es au feeling. Alors effectivement tu as des ingénieries qui travaillent au feeling, et donc oui ils vont trouver le swing, léger moment de pompage, légèrement retarder ma pulsation grâce à ce moment de dynamique. Tu peux aussi le faire avec un compresseur d'aujourd'hui.

Aujourd'hui je travaille beaucoup dans Nuendo avec des outils de dynamique qui travaillent à la centaine de micro secondes. Par exemple pour avoir une parfaite perception de l'espace que ce soit sur haut-parleur chez toi, dans une sono ou sur casque, ça veut dire que ton bas du spectre qui est une zone de masquage phénoménale, notamment entre 90 et 200Hz, cette zone-là tu la maîtrise totalement. Sachant qu'elle a une exigence c'est qu'elle a besoin d'envoyer beaucoup d'énergie parce que c'est celle qui va créer le phénomène de trame, le petit coup de poing dans le ventre qui fait du bien, qui donne une pulsation avec de la proprioception, ça veut dire qu'il faut envoyer beaucoup d'énergie. Mais si cette énergie-là elle n'est pas maîtrisée dans le temps, elle va venir masquer le reste, parce qu'elle va se comporter comme si je te mettais un flash dans la gueule, et si dans la seconde qui suit je te dis « tu vois la petite bougie derrière ? - Ah non je ne vois rien du tout là. » Les basses fréquences c'est exactement ce phénomène-là, donc c'est comment envoyer des raies d'énergie qui donnent la pulsation,

donnent la couleur qu'on aime avoir aujourd'hui parce que c'est un rapport esthétique au son, tout en ayant une clarté parfaite de l'espace, et ça tu ne peux pas l'avoir avec des vieux compresseurs, ni des vieux filtres, là il faut des outils qui bossent à la centaine de micro seconde près.

La démarche elle doit être là. Aujourd'hui on est à une nouvelle grammaire mais on est aussi à une nouvelle manière d'utiliser les outils. Et les outils qu'il y a dans Live Ableton et les outils qu'il y a dans un Nuendo après tu auras les mêmes dans Pro Tools ce n'est pas le sujet, mais en tout cas les outils que tu amènes dans Nuendo de base... Moi aujourd'hui je sais que je préfère travailler avec des outils qui sont des outils qui, sur le plan temporel, sur la gestion des énergies, me permettent une réelle prédiction du résultat et donc viennent confirmer ce que j'obtiendrais, ce que je vais visualiser avec un oscilloscope. C'est à dire que je vais avoir quelque chose d'extrêmement précis et qui va être à l'échelle de la milliseconde minium. Parce que dans l'espace, de toute façon à termes c'est ce qui va permettre de placer le maximum de choses en ayant le maximum de visibilité, et entre le travail de l'harmonie, de la mélodie, de la polyphonie, le travail de la matière et le travail du mix c'est à dire de la construction au mixage des pulsations, des mouvements, on va obtenir dans les 10/15 ans qui viennent pour les plus doués d'entre nous des mosaïques qu'on aura jamais entendues.

On est dans le domaine du vitrail en fait. De la mosaïque, plutôt de la mosaïque. C'est à dire que tu vas amener énormément de petites énergies, ou de grosses, et qu'à la fin ça va te donner un truc tu vas faire « mais je suis où ? » C'est ça la piste que je suis, c'est la mosaïque, la mosaïque parfaitement établie, ou il y a des centaines d'éléments, ou ils donnent une perception spatiale et une composition qui n'existe pas dans la nature mais qui s'en inspire totalement par sa diversité et sa richesse. Et donc mon cerveau va kiffer. Et aujourd'hui quand je travaille avec des musiciens j'interviens de plus en plus, et c'est pour ça que j'ai tendance à avoir des exigences de gamins gâtés, parce que j'ai besoin de bosser avec des musiciens sur qui je peux exercer une influence sur le choix du synthé, sur l'écriture d'un arpéggiator, sur l'écriture d'un kick, sur l'écriture d'une caisse claire, ou de plusieurs caisses claires. J'ai besoin de ça pour avancer dans mon écriture spatiale.

Après Cécile on a d'autre projet sur une œuvre de Bach, je vais commencer la mosaïque avec eux. Ce sont des centaines et des centaines de son qui interagissent les uns avec les autres. Ça c'est vraiment une piste extrêmement intéressante et très prometteuse pour développer de nouveaux univers à la fois musicaux et à la fois sur le plan perceptif. Parce que moi je ne suis pas dans la reproduction d'un monde, je ne suis pas en train d'essayer de reproduire un orchestre, ce n'est pas un travail qui m'intéresse. Il faut le faire et il y a des gens qui sont beaucoup plus doués que moi pour le faire. Mais mon travail il consiste à savoir comment je donne le plaisir de l'immersion dans un cerveau au même titre que mon cerveau prend plaisir. Comme toi t'as la même chose au bord de la mer ou dans la forêt qui font que tu aimes bien être là, parce que tu as les oiseaux le vent, des millions de feuilles qui oscillent autour de toi qui génèrent une harmonie, qui génèrent un univers. Et aujourd'hui, grâce à la puissance de l'ordinateur, il y a juste à bosser en fait, il y a juste à travailler.

Là on a des outils. Ce qui n'est pas encore le cas pour l'image, ils galèrent encore, il n'y a pas assez de puissance, et tant mieux. C'est ce qui fait qu'on va avoir des ordinateurs encore plus puissants qu'avant. C'est ce domaine d'application qui pousse à donner plus de puissance. Mais dans le domaine du son maintenant il va falloir organiser, rationnaliser, parce que quand

tu commences à avoir 120 pistes il y a tout un travail à faire... C'est pour ça que là Cécile va me brancher avec les développeurs qui sont derrière Ableton, je vais les rencontrer dans peu de temps, et je vais mettre le pied dans un truc ou je ne vais pas les lâcher les mecs. A la fois sur la spatialisation, et à la fois sur la relation entre la spatialisation et l'écriture. Et Ableton on peut peut-être même développer un truc où il n'y a pas besoin de tant de panoramique que ça, parce que le panoramique pour moi à termes j'en veux plus. L'écriture ultime c'est sans le panoramique. Il y en aura forcément mais l'écriture ultime c'est de poser les choses, et elles interagissent les unes avec les autres et là la couleur de la mosaïque évolue, et dans le temps et dans l'espace, et en énergie. Tu vois c'est ça l'écriture ultime. Et tu l'as compris il y a peut-être 20 ou 30 ans de boulot. Mais des mecs comme Max Cooper sont déjà là-dedans. Et ça c'est Schaeffer, tu es chez Schaeffer là-dedans.

Ce que je trouve intéressant, j'adore l'harmonie, et je suis un passionné de Bach et de la polyphonie, je suis intimement convaincu qu'il faut qu'on retrouve dans la musique électronique, qui est, par rapport à la musique urbaine, hip hop et compagnie... les mecs bossent beaucoup plus la musique. Réellement. C'est à dire que le plus fainéant des mecs de l'électro bosse plus la musique que les autres. Dans le Hip Hop le plus important c'est la voix. Donc la musique ils s'en foutent et ça fait chier parce que la voix associée à des univers c'est un truc de dingue. Je suis sûr que le hip hop va y venir un jour mais ça met du temps. Les seuls qui ont vraiment bossé là-dessus c'est de l'autre côté de l'Atlantique. Où les mecs bossent vraiment des univers, tu écoutes plein de musicien de la scène même new yorkaise des années 80, même Prince par exemple ce sont des univers à chaque fois, il y a 4 sons le mec pose un univers. Ce que tu n'as pas aujourd'hui dans le hip hop français. Il n'y a pas d'univers. Tu mets un kick et tu gueules dessus. Ce qui est dommage parce que le hip hop est extrêmement intéressant à développer. En revanche dans la musique électro, si en plus tu n'as pas la voix, soit tu ne bosses pas la musique tu fais boucher ou boulanger, donc même le plus fainéant des fainéants, et j'en croise dans la musique, le mec est obligé de bosser un minimum. Pour moi la musique électro est la musique la plus prometteuse dans le domaine.

Il faut qu'on soit stratège parce que sinon on va se faire avoir quoi et je ne veux pas qu'on disperse de la connaissance comme t'as pu le comprendre. Mais je pense que le propos que je viens de te tenir est bien plus intéressant que de te donner des détails sur l'écriture. Parce que ce que je viens de te dire là à la fin, entre ce que je t'ai dit sur le traitement du bas et ce que je viens de te dire sur mon approche globale, je viens de te dire là où je vais. En revanche je suis paisible parce que les escrocs ils ne vont pas y arriver.

Ce que je viens de te donner à la fin c'est plein de jus de cerveau qui va faire gratter, il va falloir développer des outils, développer l'écriture, reprendre des vieux écrits. Cécile je lui ai dit d'écouter un max de polyphonie, « écoute le moyen âge, écoute la renaissance, tout est là. Parce que si tu maitrise ça, tu maitrise la variété du matériau sonore, tu maitrise l'assemblage dans l'espace, alors là, tu vas claquer des trucs. Ça va faire de la musique qu'on n'a jamais entendu, dans sa globalité, mais ça sort d'où ? » Déjà Cécile c'est plus ou moins le cas.

Ce qui ressort de *White Opening*¹⁰², c'est la base de la base du travail que j'ai envie de mener. Ce sont des sons enregistrés au titre Schaefferien. On a enregistré des sons de peintures, des sons de pinceaux, des sons de bâches qu'on secoue dans l'espace, et ce qui ressort c'est que

¹⁰² Ouverture de l'album *Musicalism* (2025), mixé par Hervé Déjardin à la Maison de la Radio

tout le monde bloque sur *White Opening*. Oui c'est normal c'est un truc qu'ils n'ont pas l'habitude d'entendre. Mais qui ramène aux codes de ce qu'on a l'habitude d'entendre avec à la fin un gros Kick, un apreggiator qui tourne autour, ce genre de trucs.

Faut aller là où les autres ne vont pas. Et pour aller là où les autres ne vont pas, il ne faut pas dire « je veux faire différemment des autres », faut juste faire ce qu'on a envie de faire.

A12.3 Entretien avec Xavier Gibert

X : [Introduction sur l'histoire de RFI Lab] On était à la maison de la radio à l'époque, parce que finalement, entre la restauration sonore et le mastering, il y a beaucoup de points communs. Ensuite on a commencé à faire tous les produits de communication qui ne rentraient pas dans la chaîne de production d'RFI : CD de promo musique etc. Après on est arrivé ici et, là, je leur ai dit : « on se rapproche de France 24 il faut absolument qu'on soit multicanal ». Donc on a commencé à se procurer, avant le déménagement, 3 studios qui étaient 5.1 et, de là, j'avais fait connaissance entre temps avec Hervé Déjardin, et je connaissais Edwige de l'époque où j'étais à Radio France. Elle a créé avec Hervé cette partie-là, elle était responsable de la couleur d'antenne et était une ingénieur CNAM qui est partie en retraite assez récemment. À partir de là je me suis formé au multicanal, au binaural, et on est devenu pote avec Hervé et Frédéric Changenet. RFI labo est né comme ça.

Nous on a le titre de labo mais on est en production réelle. C'est à dire que, nous, on ne peut pas faire d'expérimentation si ça n'aboutit pas à un programme. On sert à faire quelque chose.

C'est pour ça qu'on gère toute la partie prise de son musique ; c'est pour ça qu'on doit, dès qu'il y a un outil nouveau qui rentre sur le marché, l'intégrer. On l'apprend et on fabrique quelque chose avec, donc on rentre dans la chaîne de production tout le temps. Donc *Session Lab*¹⁰³ c'est un POC¹⁰⁴ qui a commencé, au départ, en disant que quand, nous, on était formé en multicanal, on nous disait droit dans les yeux : « la musique ne peut pas être en audio 3D si elle n'est pas écrite pour ça ». Nous qui tournions à environ 300 titres sur les émissions de musique par an, on disait : « c'est ça, tu vas apprendre à tous les *home studio* et tous les créateurs que, désolé pour eux, ils n'auront jamais accès à l'audio 3D parce que ça n'a pas été écrit comme du Varèse. » Donc on a commencé à travailler ce qui n'était pas du remix à l'époque, mais qui était une idée de dire aux maisons de disque : « donnez-nous vos stems et on va proposer. On va mettre en place une spatialisation. » Et donc ça a été *Session Lab* qui a été juste un truc, mais tout de suite, quand on l'a fait, on faisait valider par les ingénieurs du son stéréo, les maisons de disques, les directeurs artistiques, et on le diffusait en audio 3D sur les réseaux. Mais on faisait aussi une interview des artistes et on faisait également une vidéo et un podcast, pour faire en sorte que ce bloc puisse être, pour les maisons de disques, un outil de communication. Et nous permettre d'exister, donc on travaille que comme ça.

V : En quelle année avez-vous commencé ?

X : C'est devenu, après, une émission complète. On a appliqué ça à nos systèmes de reportage. C'est-à-dire qu'on a fait travailler nos journalistes en disant : « qu'est-ce que ça veut dire d'écrire en audio 3D quand tu es en reportage ? » C'est là que nous on a intégré que des prises de son en HF¹⁰⁵, pour pouvoir nous éloigner au maximum des journalistes et pour pouvoir avoir toutes nos ambiances en audio 3D – en multicanal ou en ambisonique – propre par rapport aux micros des journalistes. Donc on s'est aperçu que, comme nous on avait des journalistes

¹⁰³ Podcast original de RFI et programme phare de RFI Labo. Mixé en *Dolby Atmos* pour une écoute immersive au casque.

¹⁰⁴ Proof Of Concept

¹⁰⁵ Hautes fréquences

qui travaillaient seuls depuis des années à RFI, si on leur retirait l'outil de médiation, qui était le micro, en disant : « on va venir, nous, en tant que preneur de son, interagir dans l'interview », ça gênait les journalistes. Il fallait qu'on les laisse quasiment seuls parce qu'ils avaient pris, pendant des années, l'habitude de faire ça. Donc on est parti du principe que, ok, ça on leur laisse. Mais nous, ça ne nous intéresse pas ce côté monophonique, c'est presque gênant. Donc on va les mettre en HF, ce qui fait que nous on va pouvoir travailler en Double MS, en IRT, en ambisonie, suffisamment loin, pour que cet interview-là ne rentre pas dans notre système. Donc on travaille comme ça. C'est-à-dire qu'on a entre 4, 5 ou 6 pistes selon nos besoins. On a nos deux pistes monos, notre système qui va choper la *room*, la pièce ou l'extérieur, les ambiances, tout ça est synchrone et on peut faire un montage clean des parties mono, les repositionner dans l'espace et les baigner dans notre ambiance multicanale. Enfin 3D.

V : Vous avez donc commencé par faire de la musique multicanale, montrer aux maisons de disques que ça fonctionne, puis vous êtes allé sur du reportage 3D.

X : C'était parallèle, on a fait les choses de front.

V : Et en termes de technologies, comment ça se passe ?

X : On était en mode canal vers du mode binaural, après on a intégré la technologie *Dolby Atmos* parce que finalement, quand on était en mode canal vers le binaural, on avait une première phase de mix qui était pour le binaural de synthèse et qu'on était obligé de reprendre pour faire une stéréo. Donc en fait on avait quand même nos 3 formats.

V : Vous ne faites pas comme à Radio France où vous envoyez le binaural à l'antenne.

X : Et non !

V : Vous faites quand même attention à avoir un mixage stéréo classique.

X : Exactement. Et donc avant on le faisait en 2 passes pour *Session Lab*. Avec la technologie Atmos, on arrive maintenant à quelque chose qui est satisfaisant pour nous, avec l'ADM, qui est de pouvoir sortir une stéréo qui se tient en parallèle du binaural de synthèse et en parallèle du 7.1.4, multicanal, sachant que nous on garde le 7.1.4 comme format d'archivage, pour le moment. On en est à, on fait quand même un ADM mais comme l'ADM est incapable d'être repris par l'INA...

V : L'INA n'archive pas les fichiers ADM ?

X : Je les tanne depuis un an et demi. On a organisé une séance de travail avec Hervé [Déjardin] et Frédéric [Changenet], j'ai secoué l'INA en disant : « je vous signale que vous avez déjà raté le passage de l'ambisonique et de tout ce qu'on a fait en VR. »

V : Ils n'ont rien archivé de tout ça ?

X : Non. Ils ne peuvent archiver que de la stéréo. La seule chose qu'ils ont archivé en multicanal, c'est Le Singe Soleil. On est en 1995 ou 1990.

V : C'est un choix non ? Ça dépend d'eux.

X : Ce n'est pas qu'ils ne veulent pas, c'est qu'ils n'ont aucune conscience de l'évolution du métier. Comme j'ai enseigné à l'INA, en ayant géré les archives de nos contrats d'archives avec Rfi, je les connais bien. Je leur ai dit : « mais vous vous rendez compte que tout ce qui est en train de se passer ? C'est que dans 10 ans, quand vous voudrez récupérer une interview propre d'une artiste que l'on fait actuellement, ça sera une plateforme qui aura récupéré l'ADM qui pourra faire un solo sur la voix. Vous, vous serez encore en train d'essayer de faire de la séparation de source alors que le programme aura été fait en ADM, et vous ne serez pas capable de le faire [le solo sur la voix]. » C'est les a un peu dérangés, ça a bougé. On leur a livré un ADM pour voir comment ils pourraient faire. Et depuis... C'est l'INA si tu veux. Si tu ne remets pas de la drogue dedans ils ne se réveillent pas toujours.

Donc je reboucle. On produit quand même, nous, un 714. Toujours. Parce que je me dis que dans 35 ans, ils seront quand même un peu près capables de dire : « bon la piste 1 on va plutôt la mettre à gauche, la 2 sur la droite, peut-être que la 3^{ème} c'est le centre » et ils vont retrouver le truc. Mais la stéréo et le binaural de synthèse ne sont pas pour nous des formats d'archivage.

V : Vous faites *Session Lab* en multicanal. Y a-t-il d'autres contenus en audio 3D ?

X : D'autres émissions ? Toutes ! Nous on ne fait rien qui n'est pas à partir de multicanal, toutes les séries d'apprentissage du français sont faites en binaural de synthèse à partir de multicanal. Tout ce qu'on fait. Là on a fait le printemps des poètes, c'est pareil. Là on a fini un album de Chinese Man, c'est le dernier qu'on a fait, en Atmos.

V : Ce sont des contrats privés.

X : C'est une prestation. Là on travaille pour le printemps des poètes, c'est une prestation aussi. Là on travaille pour *Alto*, qui a remporté le marché du panthéon. On a présenté le podcast *Femmes de Ring* en février qui est en Atmos, c'est toutes nos productions. Nous, on ne va pas dessus si ce n'est pas multicanal.

V : Ça a été facile de faire bouger la direction à Rfi ? Je sais qu'à Radio France ce n'est pas évident. Est-ce qu'il y a une conscience de l'évolution du métier ?

X : Personne ne voulait être au 6^{ème} étage quand on est arrivé ici parce que c'est l'étage de la direction. J'étais le premier à dire que je veux y être. Donc en fait je suis en face du directeur technique groupe. On a cette facilité, par rapport à Radio France, de pouvoir aller exposer ce qu'on fait et oui, on est suivi. C'est pour ça qu'on est 7 à temps plein sur le sujet, c'est pour ça qu'on est... J'ai zéro contrainte financière, j'ai ce qu'on veut quand on veut, ça passe sans problème. Et en plus on n'injecte pas beaucoup, on ne couture pas très cher par rapport à d'autre. Donc oui, on est complètement...

Au niveau de la direction de l'information, en gros, c'est différent parce qu'un journaliste est sclérosé dans sa façon de penser depuis l'école donc... Il ne faut pas me lancer sur la sclérose des médias depuis que ce sont des journalistes qui les dirigent, mais ça commence à venir. C'est à dire que là, ils ont été obligés de reconnaître qu'un programme qu'ils n'ont pas demandé, qui

a été financé entièrement par un autre puisqu'on a payé la journaliste pendant 2 ans, *Session Lab*, Jean Marc Four arrive de Radio France il dit : « j'ai besoin d'une émission de musique », bim : elle est toute faite. « J'ai plus qu'à en remettre une deuxième pour faire les deux autres dimanches » et il les a confiées à la même journaliste. Donc oui, il y a une reconnaissance du fait que c'est nous qui gérons la communication en gros. Quand la communication a besoin de briller, si on va à Longueur d'ondes¹⁰⁶ et qu'on fait un salon d'écoute pendant tout le salon en *Dolby Atmos*, qu'on est capable de ramener *Dolby*, *Neumann* et *Trinnov*, c'est bien parce que on travaille là-dessus depuis un certain temps et qu'on est reconnu sur le secteur. Donc la direction nous tolère comme ça, parce qu'on est complètement dans la production. Tout ce qui est fait ici est valorisé en termes de communication.

V : Tu as parlé de Longueur d'ondes. C'est ce genre de salon d'écoute qui vous permet de tester les contenus immersifs avec du public. Sinon c'est que de la diffusion binaurale.

X : Oui, et avant on faisait des trucs en binaural dans ce genre de salon. Et depuis qu'on est en Atmos on peut faire des écoutes en mode canal.

V : Qu'est-ce que vous en ressortez ?

X : Pour moi, l'évolution de l'écriture de podcast va passer par l'audio 3D, c'est évident. Même s'il y a toutes les notions de marché du podcast etc. Il y aura une vraie différence, il y aura l'aristocratie du podcast qui sera en audio 3D et il y aura le podcast qui sera l'équivalent du transistor dans la salle de bain pour la radio. Il va y avoir cette différence-là. Franck Bessière¹⁰⁷, qui est venu la semaine dernière pour présenter son futur projet qu'on aimeraient monter ici, a eu cette réflexion drôle. Il disait « oui, je suis en train de produire un truc pour la RTBF, mais bon c'est en stéréo ». Et ça a fait rigoler les mecs : « pourquoi tu dis ça ? » Et il dit « oui non non ça peut être très bien, ce n'est pas dévalorisant, mais j'ai tellement moins de place pour écrire. » Et cette réflexion-là, qui n'émanait de rien en fait, qui était vraiment spontanée, qui n'était pas du tout stratégique, il est là. C'est-à-dire que c'est cette place qu'on a pour pouvoir augmenter le nombre d'informations simultanées et avoir la possibilité de déplacer l'attention dans cet espace sonore. C'est ce qu'on a toujours recherché pour moi en stéréo, et finalement, cette attention, elle était que gauche, droite, avant, arrière, en fonction d'une réverbération, mais il n'y avait pas cette possibilité de dire qu'à ce point-là, j'ai trois informations qui ne m'intéressent pas mais qui sont présentes. Il y en a une qui m'intéresse et sur laquelle je vais pouvoir m'accrocher, tout en mettant les autres en traitement de fond de ma pensée. Ce qui est une écoute naturelle qu'on fait au quotidien. Et c'est là que pour moi il y a quelque chose, et cet élément se rencontre avec le public. Ça s'est vu à Longueur d'onde sur les 3 jours. C'est à dire qu'on a eu des réflexions d'auditeurs classiques qui ont déjà eu cette notion de confort et qui ont trouvé le moyen de l'exprimer. D'ailleurs, j'ai un rendez-vous la semaine prochaine pour voir comment on pourrait remonter ça, parce que ça a été monté avec des efforts de nos partenaires et avec la confiance qu'ils ont eue. Comment le reproduire, c'est la question.

¹⁰⁶ Festival de radio qui se déroule chaque année à Brest.

¹⁰⁷ Réalisateur de contenus immersif, dont le podcast Visiter les vivants pour Radio France en 2024.

V : Est-ce que tu as senti que les retours que vous avez eu ont été différents que quand vous faisiez des salons d'écoute binaurale ? Est-ce qu'il y a une plus-value d'avoir des enceintes vis-à-vis du casque ?

X : Le côté collectif. Moi je trouve que les gens s'expriment plus quand ils sont dans une salle et qu'il n'y a pas de... Comment je vais le dire.

Quand tu accueilles quelqu'un et que tu lui mets un casque, il va avoir tendance à être plus content parce qu'il a eu un rapport de politesse avec toi. Parce qu'il y a eu une proximité dans la relation. Donc il y a eu ce côté *waouh effect*, « ah ouai c'est dingue » mais ce « waouh c'est dingue » ne va pas forcément au-delà de la perception spatiale. Ça ne rentre pas dans le sujet. Moi je pense que les gens, quand tu les mets au casque pour écouter un contenu en binaural, ils vont plus s'attendre à l'effet que ça va donner plutôt qu'au contenu du programme. Quand tu les mets dans une salle, ils sont sur des critères qui sont beaucoup plus cinématographiques d'écoute collective, et donc finalement je trouve qu'ils sont moins... Moins intimidés par le côté sonore. Je trouve. Ce n'est peut-être pas très clair mais pour moi, personne n'irait écouter un programme parce qu'il est en audio 3D, ça ne changera rien. Ou alors tu fais le coiffeur quoi. Et ils vont dire « oh c'est génial, tu as entendu ? Il y a des ciseaux qui ont coupé ! ». Oui d'accord, mais ils iront parce que l'histoire est bien écrite, parce que le sujet est bien. Et ce qu'ils en percevront va être du rapport du non-dit. Dans une expression de « c'est vachement bien, je me suis senti bien, le son est bon etc. » mais ça ne va pas se jouer sur un effet, ça va se jouer sur un contenu.

Autre exemple, mais je ne sais pas si ça a un rapport. Élise Roche a critiqué, selon moi, très bien Femmes de ring, le podcast qu'on a présenté à Brest [à Longueur d'ondes] en avant-première, qui est un podcast en 4 épisodes sur la box féminine. Dans sa critique, elle trouve que le producteur [un personnage narratif, ndlr] est trop présent et elle dit qu'heureusement, il n'est pas au centre. Elle critique un peu ça. Elle dit que le mec parle de la boxe féminine mais il en parle beaucoup, il parle de sa mère finalement, il tire la couverture à lui. Je traduis dans mes mots. « Heureusement il n'est pas au centre, il est placé sur la droite », et elle fait tout un truc sur l'écriture 3D là-dessus. Pour moi, là, on est sur quelque chose d'extrêmement pertinent, parce qu'effectivement, moi, je l'ai placé un peu avant gauche et le podcast est structuré avec le personnage principal centre droit, centre un peu haut, mais genre 10° à droite. Tous les intervenants sachants sont sur le 180 devant, tous les personnages illustratifs etc. sur le 180 arrière. Et donc elle n'a pas perçu tout ça, mais le fait de pouvoir dire que le programme, lui, a convenu quand même, malgré la présence narrative trop importante du producteur parce qu'il était à un autre endroit de l'espace et qu'il avait justement ce côté annexe et strapontin, on est là dans l'intérêt de l'audio 3D.

V : C'est marrant parce que dans le podcast Visiter les vivants de Frack Bessière, les intervenants scientifiques sont aussi placés à l'arrière. Pourquoi tu as fait ce placement de voix autour de l'auditeur ? Qu'est-ce que tu t'es dit quand tu as dit « ok pourquoi pas le narrateur à l'arrière et les intervenants annexes devant » ? Pourquoi garder un frontal et une répartition comme celle là ?

X : Parce que qu'on le veuille ou non, on ne sortirait pas tout de suite du théâtre à l'italienne et, spontanément, je ne me suis pas assis en tournant le dos. Donc je suis auditeur, quelqu'un

va me parler à moi, je vais forcément me le mettre dans un périmètre... enfin dans une zone qui va être dans mon frontal. Il y a quelqu'un qui va être comme moi, qui est... donc là si on reste sur Femmes de ring, une des femmes qui témoignent des violences féminines, qui témoignent de « oh putain je me suis bien défoulée etc. », elle est finalement étrangère au discours parce qu'elle n'a pas été invitée, on l'observe, donc je vais la mettre ailleurs que dans le frontal. Quelqu'un qui va me parler, me transmettre une information, mais finalement qui répond au producteur, j'assiste à cette discussion-là. En revanche, si j'assiste à cette discussion-là, je vais la mettre devant mais je vais avoir un côté équilibre à gérer parce que si je mets tout à gauche, au bout d'un moment il y a une vraie fatigue auditive dans le discours. En revanche, quand ce producteur-là, je vais devoir garder une question qu'il a posé lors de l'interview, là il va être déplacé et je vais le détimbrer. Enfin je vais changer sa couleur pour qu'il rentre dans un autre endroit où je suis bien sûr que là, il va poser la question non pas à moi, auditeur, mais à l'expert qui parle. Et dès qu'il va être avec les micros qui vont concerner la structure du podcast, il va reprendre sa place et donc je peux, comme ça, dire à l'auditeur « là il te parle, là il parle à son invité », et voilà un exemple de la façon dont j'aborde les choses. Maintenant pourquoi est-ce que je n'ai pas mis... Ça m'est déjà arrivé de mettre [des sources directes, ndlr] à l'arrière. Maintenant comme on a quand même une idée de diffusion stéréo, quand je vais passer en Dolby Atmos et quand j'ai ma stéréo qui va être générée, je n'oublie pas que finalement, elle est générée à partir d'un 5.1. Donc finalement il y a une première forme de *downmix* de mon 7.1.4 qui va devenir un 5.1 et à partir de là, dans mon Atmos Renderer, je vais appliquer mes valeurs de *downmix* que j'aurais pu paramétrier en disant « est-ce que j'atténue de -3 dB mes arrières etc.

Et donc si je mets des personnages importants un peu trop à l'arrière dans mon *down mix*, je vais un petit peu les perdre par rapport à ce qui est devant. Donc en fait on ne joue pas beaucoup avec ça nous. C'est à dire que tout ce qui est important va quand même être dans une forme de 180 frontal. Et on a vraiment expérimenté ça sur les traductions. C'est à dire qu'une traduction, on va aller la mettre derrière et assez proche mais, on va avoir ce côté chuchoté en fait, parce que la précision de perception à l'arrière est bien plus grande qu'à l'avant. Parce qu'on n'a pas de centre en biaural, vraiment.

C'est notre sens de la protection, donc on a une précision de localisation qui est bien plus importante. D'où le fait qu'au cinéma leur obsession c'est de ne pas faire sortir de la narration des spectateurs et que finalement derrière c'est que des effets.

Donc si on place cette traduction parce que c'est utile pour nous, dans le binaural de synthèse, un peu à l'arrière, on va vraiment faire gaffe que dans le *downmix* stéréo elle revienne bien se poser à sa place. Et donc si la voix principale est à plus 15 degrés, là, à l'avant, et que la voix arrière on va la mettre finalement dans un rapport qui va être à 25 pour qu'elle vienne se recaler pas au même endroit, qu'elle ne vienne pas dessus entrer en conflit.

V : Il y a quand même une démarche de penser toujours à la stéréo, malgré le fait qu'on travaille en multicanal. Est-ce que ça ne peut pas brider la création ?

X : Brider non mais il faut quand même penser à tout. Et on l'a expérimenté aussi à Brest. C'est à dire que nous on travaille avec des studios centriques, avec un centre, et des studios de positionnement en rond. Là on s'est retrouvé certes avec un ADM, mais avec une disposition en rectangle. Et donc c'est l'avantage de l'ADM : ça passe, mais la salle étant un peu longue et

le fait qu'on n'était qu'en 7.1.4, ça aurait nécessité un 9. Les choses qui étaient placées derrière, elles étaient vraiment derrière tu vois, et on aurait eu envie, nous, finalement si on avait pu, refaire baver un peu les arrières pour les ramener entre les *side*. C'est à dire essayer de recompresser un peu cette pièce-là. Donc oui on ne peut quand même pas totalement s'affranchir du compromis entre les différents formats. On est obligé de... Mais ça s'est intégré au cours de l'expérience de façon assez intuitive tu vois on ne l'a pas modélisé. Ça maintenant c'est un truc que tu ressens, c'est la magie de nos métiers, c'est que finalement tu fais des trucs tu t'en rends plus compte c'est en te le racontant que tu te... tu vois.

V : Et tu as utilisé tout à l'heure le terme « cinématographique » quand tu parlais des gens qui viennent écouter vos contenus. Pourquoi l'analogie du cinéma ?

X : Parce que le multicanal est arrivé par le cinéma qui est lui-même une représentation du théâtre à l'italienne avec un évènement visuel qui est frontal. Mais dans le théâtre à l'italienne, le son venait du même endroit que l'écran et que c'est le cinéma qui a commencé à faire venir des sons qui sont à un autre endroit d'émission que l'écran. Donc on va mettre du temps à se séparer de ça. On a beaucoup été sensibles à ça, nous, quand on a fait beaucoup de vidéos 360. Le marché s'est complètement écroulé, donc on en fait plus du tout, la XR ne tiens que sous assistance respiratoire de ceux qui y croient encore. Donc en fait ça veut dire qu'il y a que les studios qui arrivent à faire ça. Mais c'était une vraie question de savoir, parce que finalement le derrière disparaissait, parce que finalement il pouvait devenir frontal à partir du moment où on tournait la tête, donc c'était assez particulier de mixer comme ça. Parce que, nous, on a tout de suite eu le masque et on l'a intégré au *Pyramix*, on a tout de suite pris un *AmbiEye* qu'on avait financé d'ailleurs à l'époque pour pouvoir avoir une écoute binaurale dynamique instantanée au moment du mix, mais on travaillait déjà quand même en mode canal tu vois, toute la préprod elle se faisant en mode canal, c'était une vraie question de, finalement, ce qui est derrière va devenir devant à un moment.

V : Quels types de contenus vous avez fait en non-frontal ?

X : Non, nous on a fait que de la VR pour la cité des sciences, du documentaire en VR. On n'a pas fait d'ambisonique, de binaural de synthèse dynamique à partir de l'ambisonique en casque seul. On n'a jamais eu de demandes de cet ordre-là.

Tout ça pour dire que cette idée de devant derrière, nous c'est parce qu'elle a disparu... en fait on s'en est... toute la question de pourquoi le cinéma, on a vraiment pris conscience de ça quand on a travaillé dans la VR, du fait de dire « bah ouai putain en fait, la stéréophonie a été mise... les deux haut-parleurs de la stéréo ont été mis devant parce que, finalement... pourquoi est-ce qu'on a eu besoin de regarder nos haut-parleurs, pourquoi est-ce qu'à la racine, Schaeffer etc. Pourquoi est-ce qu'on a construit des studios avec les hautparleurs devant ? » Personne ne s'est posé la question. Pourquoi est-ce qu'au départ, on n'a pas construit des studios en mettant les haut-parleurs derrière ? En disant qu'en fait on va se baigner dans le son et qu'en fait le son va venir dans le sens de notre regard et pas contre notre regard, c'est bien parce qu'on a été, parce qu'on gère le son comme des prédateurs, et on ne l'a pas géré comme des proies. Et ce rapport-là du frontal, ça aurait pu très bien se faire autrement et on aurait écrit différemment si on avait placé la stéréophonie à l'arrière. Et le 360 m'a permis de prendre conscience de ça. Ce serait... je ne l'ai jamais fait mais ce serait rigolo de proposer aux gens une écoute stéréophonique et de leur mettre à l'arrière, en leur disant « tu vas écouter

avec moi, et finalement tu ne vas pas regarder tes haut-parleurs ». Et moi quand on fait des écoutes avec les gens je suis assez... surtout maintenant, dans notre culture, si je ne leur mets pas un *vumetre* à regarder... Si je ne mettais pas les écrans, tu vois ce que je veux dire, personne ne regarde l'enceinte centrale, tout le monde regarde l'Atmos Renderer. Et ils regardent les objets bouger, et donc je suis sûr que les mêmes personnes, je leur mettrais une stéréo en mettant les enceintes derrière, ils se retourneront en disant mais pourquoi c'est derrière ?

Alors que finalement ils décalent l'intégralité de la scène dans leur écoute mais comme ils ont la proposition, la perception nécessaire, ils recalent tout. Mais cette idée de cinéma on est lié à ça et c'est pour ça qu'on s'est fait chier pendant des années, on s'est emmerdé à essayer de trouver la HRTF qui permettrait d'avoir le devant de la scène stéréo, je leur disais « mais putain faite le simplement vous prenez un multicanal vous le foutez dans un binaural de synthèse, bous déplacer votre bruit blanc, il monte », c'est un effet casquette classique parce que de toute façon on en avait pas besoin, on a le déplacement du bruit blanc en face il est pile au bord du champ visuel parce que on avait pas besoin d'identifier le son de la source que l'on voyait arriver en tant que prédateur, donc on l'a pas développé, donc tout ce sentiment de devant... enfin je dis pas qu'en fermant pas les yeux je sens pas le devant mais ma perception du son de face elle est acquise par puiseurs sens et pas uniquement l'audition.

V : Est-ce que dans la pratique de mixage multicanal vous avez-vous avez un souci de réalisme ?

X : On a le souci de réalisme à partir du moment où on est en reportage. C'est-à-dire que le multicanal et le son 3D, en général, permettent de ne plus avoir à se battre avec le lieu de prise de son. Donc moi, là, je te prends un exemple d'un truc. On faisait un documentaire sur l'histoire de l'électro, et en parallèle on a un des anciens qui est interviewé. Il donne un rdv dans un bar pour l'en province, avec un énorme bruit de frigo. Mais le bar avec le congélateur des glaces d'à côté dans lequel normalement tu dis « non en fait on ne va pas faire l'interview. » Mais on est avec notre système audio 3D. Et finalement, ce son-là et cette interview du mec qui était un mec qui avait été au top, qui avait été complètement exclu du milieu, qui avait changé de vie, qui était en province et qui parlait de ces années de gloire du début de l'électro, ce son-là devenait signifiant dans sa façon de parler. Et il n'était pas gênant parce que comme on était en binaural, il était extrêmement localisé donc on pouvait s'intéresser à lui. Le réalisateur fait le truc et après il interview un des mecs de l'électro. Le mec dit « on va faire l'interview dans je sais plus quelle boîte de nuit-là où je mixe etc. » Là, le mec propose de faire l'interview et tu as tous les ventilos des amplis etc. et on garde les deux. Et finalement, les montages de ces deux mecs, avec un qui était dans un bruit extrêmement moderne de système de refroidissement de chambre de local technique d'une boîte de nuit, et l'autre qui était de la même génération que lui, qui avait foiré et qui était dans le frigo... Il se passait quelque chose.

L'audio 3D, ça nous permet de rentrer le réalisme de l'endroit où les gens parlent. Et comme on parle du principe, nous, que les gens ne parlent pas de la même façon en fonction de l'endroit où ils se trouvent, ça permet de rentrer dans ce réalisme-là. Et maintenant, nous, on ne se bat absolument plus, on intègre tous les défauts du lieu de la prise de son dans la prise de son, parce qu'on a plus cet effet de masque, elle est plus gênante. Après, pour ça on a des techniques. Tous nos invités ils sont pris avec du Shoeps Omni, donc on a un rapport avec le champ proche et le champ lointain qui est hyper bien, maintenant dès que tu as un petit problème tu sors un

Ceddar là qui te nettoies, tu peux descendre de 10dB le bruit de fond ça n'a absolument pas touché à ta voix.

V : L'audio 3D simule une écoute réelle, ce qui le rend plus agréable.

X : Le réalisme il est cherché là-dessus [sur le documentaire, ndlr]. Après quand on est en production et en création, le réalisme on s'en fout complètement, ce n'est pas le sujet. En tout cas le multicanal, c'est pour dissocier. Le multicanal, ce n'est pas pour aller chercher du réalisme, c'est pour amener du réel. Pour moi aller chercher le réalisme c'est inventer quelque chose qui aurait dû être. Non, le multicanal c'est ramener ce qui a été. Le son 3D c'est ramener ce qui a été. Quand on est dans le mode reportage. Quand on est dans le mixage musique, ce n'est absolument pas pour amener le réalisme parce que de toute façon, la musique, ça se crée à l'italienne avec des musiciens qui sont la plupart du temps rassemblés au même endroit, et quand c'est de la musique électronique ça peut être des haut-parleurs qui sont disposés dans l'espace. Mais de toute façon, l'intention de création, elle est quand même à un endroit donné parce que de toute façon, même quand tu écoutes un concert de musique électronique, personne ne tourne le dos au DJ. Enfin on est toujours... Alors qu'il ne se passe absolument rien.

Et tout ce qu'on a essayé de faire, ce qu'on faisait aussi, tous les mixs pour Mille Grammes, pour Culture Box, France TV, c'est nous qui les faisons, on a essayé de travailler sur des cohérences de point de vue de caméra, de champs de caméra avec un point de d'écoute. Effectivement cette déstabilisation elle est insupportable. Pourquoi ? Parce qu'on casse notre habitude d'écoute là-dessus et dès qu'il n'y a pas de... Tu ne peux pas faire tourner la scène, ce n'est pas vrai. C'est ce qu'on a du mal à apprendre aux gens qui débutent en prise de son en binaural natif. Qu'il ne faut pas bouger la tête. On leur fait faire mais parfois en fait c'est compliqué, parce que parfois tu t'aperçois qu'ils ne comprennent pas et en fait tu te dis « non c'est lui qui a enregistré donc en fait il revit sa scène où il a tourné la tête donc il entend... »

V : Dans les moments où tu fais le positionnement de voix dans l'espace, au moment du mixage, que tu évoquais tout à l'heure, tu étais le mixeur ? Combien y avait-il de personnes dans le studio ? Comment ça se passe dans ces cas-là ?

X : Nous on est des techniciens chargés de réalisation à Rfi. On a fusionné les métiers de la technique et de la réalisation.

V : C'est intéressant de combiner les deux.

X : Tout le monde ne dirait pas ça. Nous on fait ça, après ça ne veut pas dire qu'on est tout seul. Nous on est en équipe, en écoute multiple, et tout ça. Mais oui, donc tout ce dont je te parlais oui, c'est de l'expérimentation, on est à la prise de son, on fait le montage avec le producteur et le journaliste, et on fait le mix.

V : Et comment se passent les échanges avec les personnes qui ne sont pas techniciens ? Est-ce qu'il y a une conscience de ce qui se passe ? Comment se passe le dialogue ?

X : Je n'ai jamais rencontré de producteur ou de journaliste qui ne soit pas d'accord. Parce que du moment que t'es dans la cohérence, je ne sais pas ce que c'est que la cohérence mais en tout cas quand tu as une opinion commune sur le but du programme, il n'y a pas de problèmes.

Alors peut-être que ça vient du fait qu'on a intégré tout ça et que si j'avais décidé de mettre le commentateur ou le producteur à l'arrière, ça générerait peut-être des questions, mais non. Quand je te dis « non », je ne sais pas si c'est parce qu'il y a une forme d'intimidation, c'est à dire qu'on a encore quand même plein de journalistes pour lesquels c'est très nouveau. Par exemple, Hortense qui bosse pour *Session Lab* en tant que journaliste, elle dit « ah oui ça c'est chouette j'aime bien » mais elle n'ira jamais dire à Jérémie « euh non je préférerais que ». Parce qu'à aucun moment elle est gênée par ce qui est proposé parce que c'est intégré quoi.

V : Et est-ce qu'elle le ferait en stéréo ?

X : Non plus !

V : Donc c'est une confiance aveugle envers la technique.

X : Pas forcément aveugle c'est une confiance dans la personne avec laquelle elle travaille. Je pense qu'en fait je ne crois pas que ça pose un problème, on se pose... C'est une fausse question ça, parce que toute façon, un journaliste ou un producteur actuellement dans cette génération-là du moment qu'il comprend bien, finalement il est satisfait. Quelle que soit la forme. Dans notre expérience qui n'est pas universelle du tout.

V : Est-ce qu'à ce moment-là, finalement, on ne s'en fout pas que ce soit en multicanal ?

X : Fondamentalement, un journaliste, il s'en fout totalement. Il est gratifié quand c'est en audio 3D, il peut dire « ah ouai c'est super bien », mais il est pas du tout capable, comme nous on peut se le dire en disant « ça j'ai pu le faire parce que j'étais en audio 3D et j'aurais pas pu faire si j'avais été en stéréo, j'aurais été obligé de faire des choix qui aurait découlé de mes contraintes d'effets de masque, de rapport dynamique de mes contraintes de compréhension etc. » et ça nous on leur dit « ce truc-là impossible de le faire si on avait pas été en binaural ». Ça aurait été incompréhensible, mais il y a que nous qui le savons et puis on ne passe pas le temps à reprendre le journaliste en disant « regarde comment on aurait pu échouer si on avait été en stéréo ». Non, on passe au suivant et ça roule. Mais c'est peut-être un handicap, c'est peut-être pour ça que les directions des grands médias tolèrent mais ne poussent pas. Je dis des grands médias, c'est Radiotélévision Suisse Romande, où leurs deux preneurs de son et producteurs qui sont absolument géniaux, avec lesquels on a bossé pour des prises de son de concert, sont obligés d'aller à Genève parce qu'ils n'ont pas un studio multicanal à Lausanne. Ils refont entièrement leur immeuble et il n'y a pas de salle Dolby Atmos de prévue. La RTBF s'est refaite entièrement en restant entièrement stéréo. Parce qu'ils sont dirigés par des journalistes et des gens qui ne voient pas du tout l'intérêt des choses.

V : Et le grand public ?

X : En fait le public se dirigera là où il est le plus confort. Mais ce n'est pas lui qui sera leader. Le public n'a jamais demandé la stéréo, et pourtant les mecs se faisaient chier à se dire « comment on pourrait mettre deux canaux alors que tout le système de diffusion était mono » et ça nécessitait un truc... Bon, on en est pour moi à la même période. Le problème qu'on va avoir ça c'est plutôt politique. Mais les seuls qui ont la capacité de sentir ça ce sont les GAFAM et c'est pour ça que le vrai risque actuellement est qu'on est en train d'offrir à Apple l'intégralité de nos ADM, avec toutes les maisons de disque. Il y a que Hervé Dejardin qui alerte tout le

monde correctement là-dessus en disant « vous savez qu'en fait ils sont forcément en train de faire bosser des IA en fonction du nom des ingés son et que comme ils peuvent faire des solos sur l'ensemble des objets... » Enfin je pense que, tu vois, dans les années 90, s'il y en avait un qui aurait pu faire un solo sur la guitare de Prince il aurait bien aimé, se l'échantillonner et pouvoir foutre ça dans une IA qui compose la partoche pour voir comment il a fait son *rif*... C'est ce qui est en train de se passer. Et donc c'est la peine tous les grands médias comme nous.

V : Si je repars sur le mixage multicanal, on a pas mal parlé des positions des voix. Pour ce qui est du reste des habillages sonores, est-ce qu'il y a plus de libertés ?

X : Alors oui dans le sens où, déjà musicalement, on peut, en séparant les sources, on s'affranchit des effets de masques, donc on a une plus grande liberté là-dessus. Tout ce qui est... Nous on mélange allègrement du binaural natif et du binaural de synthèse. Dans l'Atmos j'ai deux objets qui vont être en binaural natif, ma divergence dans le mode canal n'affecte pas mes deux canaux quand je les mets en OFF au niveau des indices de binauralisation.

V : Mais sur Apple Music ils envoient tout en mid.

X : Alors oui, mais je ne vais pas sur Apple Music, je suis dans les programmes de reportage dans les environnements de RFI. À chaque fois je suis en mode canal ou binaural de synthèse, déjà on peut le faire, on fait ça comme ça. Dans Femmes de ring il y a du binaural natif qui rentre là-dedans et qui est beaucoup plus travaillé dans les *side* ; recorrigé un petit peu avec de la divergence et ça me fait un fond de sauce qui, quand il va être en binaural de synthèse, va revenir et va retrouver sa place sans problème.

V : Tu ne peux pas passer le binaural natif en synthèse.

X : Non, je l'enlève au niveau des indices. Quand il va revenir dans mon binaural de synthèse de toute façon il va redevenir natif parce qu'il n'aura pas subi de HRTF. On est d'accord qu'on est sûr de l'artisanat on prend un marteau tordu et on obtient ce qu'on veut, on se fout de la méthode. Et donc oui, le positionnement des sources sonores va permettre de faire des contrepoints avec ce qui est dit, ou aller accentuer ce qui est dit, ou créer des mouvements qui vont attirer l'attention sur un moment donné pour faire passer autre chose. Oui on a tout ça cette liberté-là. Et nous, on a un truc qui est pas du tout orthodoxe mais qu'on utilise allègrement, c'est qu'on fait de l'*upmix* en musique et ça marche hyper bien.

V : Quel outil vous utilisez ?

X : On est sur du Nugen 7.1.4., Stéréo vers 7.1.4. Et une fois qu'il est en 7.1.4 tu gères ta largeur stéréo, tu peux virer ton centre, tu peux travailler ton espace sonore, ton *upmix* pour réserver des places et pour pouvoir placer tes voix. Du coup, finalement, ta musique elle est enveloppante mais elle te laisse la possibilité de placer tout tes autres éléments, mais nous on *upmix* à fond. Par exemple *Session Lab* on a mis en place une petite tech, parce *Session Lab* on en sort un tous les 15 jours donc les artistes qu'on a, on ne va pas récupérer les stems. Donc on fait de l'habillage traditionnel radio, c'est que quelquefois on va avoir de la musique qui va

être *in* donc plein pot, et puis après on va avoir de la musique qui va être à l'arrière. Un autre de *Session Lab* demande de valider le positionnement des voix d'un programme.

Par exemple, habillage traditionnel en radio, le *jingle*, la musique, comme c'est une émission de radio, on l'a joué en disant que quand la musique est *in* elle est *upmixée*, pleine balle, quand elle va être en illustration, c'est à dire que sur le pont la journaliste va revenir pour continuer son interview, la musique va passer de 7.1.4 en stéréo en deux autres objets, passer à l'arrière et rentrer dans une autre room, ce qui fait qu'elle va être derrière, nous mais cette fois dans une room qui est pas forcément une room 7.1.4 qui peut être une *Altiverb* ou une *room* stéréo classique et on va pouvoir comme elle est placée derrière continuer à apprécier et à écouter cette musique-là, à un niveau qu'on aurait jamais pu mettre si on avait été en stéréo. Et quand la phrase va être terminé et qu'on va avoir besoin de ramener cette musique, les deux objets vont repasser dessus, vont retrouver leur place stéréophonique et simultanément les pistes vont passer en *upmix* et on va de nouveau être embarqué dans la musique, donc plutôt que d'avoir : ce côté j'avais mon micro de voix, j'ai balancé mon intru j'ai mis mon instru sous la voix et après j'ai cassé avec la musique stéréo ; on fait en gros la même grammaire, mais on fait du positionnement en plus avec de l'audio 3D ce qui nous permet d'avoir des niveaux beaucoup plus important.

V : Tu m'avais parlé au téléphone de ta discussion avec les techniciens qui ont installé le studio d'écoute à Brest, qu'est-ce qu'il s'est passé ?

X : On était à Brest pour monter le salon d'écoute à Longueur d'ondes, on a fait venir les haut-parleurs de Neumann, qui sont arrivés de toute l'Europe parce que c'est leur truc de démo ; on a fait venir l'équipe de Dolby parce que c'est eux qui nous prêtaient l'Atmos Renderer et qui venait aussi avec la voiture Dolby ça faisait partie de la création ; et on a amené Trinnov, la boite d'optimisation, avec leur igné son pour étonner tout ça parce que nous on ne travaille pas dans un studio qui n'est pas étalonné. Nous, si notre Trinnov il tombe en panne on arrête de mixer, il est hors de question qu'on fasse une diffusion en écoutant la combinaison d'un haut-parleur et d'une salle avec tout ce que ça implique.

Et donc on a installé l'ensemble du système. On était avec des gens qui se connaissaient sans vouloir se connaître. Ce que je veux dire c'est que le mec de chez Neumann était un peu embêté qu'on fasse venir Trinnov parce qu'il vivait ça comme un manque de capacité de ses haut-parleurs à s'auto-satisfaire d'une salle. Le mec de Trinnov lui était beaucoup plus à l'aise, c'est quelqu'un qui est plus « rentré » etc. Les mecs de Dolby voulaient que ça pète et qu'on entende les objets tourner à fond la caisse pour montrer que le Dolby ce sont vraiment des sources très précises. Et donc on installe le système, on envoie les premiers bruits blancs, roses, avec la sonde Trinnov en 5 points parce que le Sweet spot il était hors de question qu'il soit sur une ligne là-dessus, donc comme ils font dans une salle de cinéma, et on fait travailler le processeur de façon classique. Et là on met notre premier mix qui était le Chinese Man, et là on se retrouve avec un son où il y avait eu des aberrations dans le traitement du Trinnov. C'est-à-dire que les mesures qu'il avait prise en différents endroits avaient généré une surexposition aux basses fréquences qui faisait qu'il y avait un biais dans l'algorithme et on s'est retrouvé avec un son avec des basses hyper sèches mais qui devaient pas descendre en dessous de 150/100 Hz, il y avait pas de *bass management* donc malgré tout on voyait les *boomers* des grosses 410 devant qui se déplaçaient comme ça on se disait « merde », et un truc très haut medium mais pas saturé du tout, un truc très sec t'avais l'impression de voir 2 filtres

comme ça. Et sidération de tout le monde. Le mec de Neumann était en train de dire « il est hors de question que mes gamelles elles sonnent comme ça », les mecs de Dolby se disaient « mais qu'est-ce que ça va donner come image » etc. nous avec Raphaël on commence à mettre des mots c'est-à-dire « ouai de toute façon il y a un problème à 2 kHz ça s'est clair, il manque de basse, on est pas bien », et on était tous avec un vocabulaire extrêmement pauvre parce qu'émotionnellement il y avait une peur de tout le monde qui voulait que la salle sonne bien et il y a une déception collective et personne n'était cabale de mettre les mots sur la notion de profondeur, sur la notion de couleur, sur la notion d'équilibre, sur la notion distance etc... J'enchaîne les termes comme ça de façon un peu empirique.

Et voilà, résultat des courses, le garçon de Trinnov dit « non », lui qui a l'habitude de ce genre de chose n'a même pas le vocabulaire pour dire « non mais attendez ce que vous n'aimez pas je l'aime pas non plus mais je sais d'où ça vient » et finalement personne ne s'est parlé et on est tous allé se coucher parce qu'on était cramé, il était minuit et demi et Alexandre de Trinnov a bossé, lui, tout seul sur son algo dans sa chambre, et sur ses différents points de mesure il était capable de voir à quel endroit l'algorithme s'était trompé et on est arrivés le lendemain et hop super beau. Et finalement il n'y a pas eu de re discussion après d'avoir une vraie lecture, un vrai vocabulaire en lien avec nos réactions.

X : Pourtant tu m'as cité plein de vocabulaire qui existe. Pourquoi ça n'a pas suffi ?

X : Parce qu'on n'est pas en mesure de le convoquer pour dire ce qu'on ressent quand on est en multicanal. Qu'est-ce qui vient de se passer¹⁰⁸ : tu as des mecs qui ont 35 ans de boite, qui sont Benoit, c'est lui qui vient de finir le Chinese Man, qui a été validé avec un retour où ils ont demandé un dB de moins à 2k sur un truc Atmos, qui mixe toutes les Séquences en musique électro etc. et qui me dit « tu vois j'ai mis un eu la vois derrière ». Mais finalement l'ensemble des termes qu'il a expliqué il a pas pu dire « j'aimerais bien que tu valides parce que – je dis n'importe quoi hein mais – je trouve qu'en termes de timbre et de narration je suis pas sûr que ça vaille le coup de mettre la narration là j'ai fait ça mais est-ce que ça pourrait être autrement » et finalement on est tous un petit peu dans une... alors ça vient peut-être de notre formation de base, on est tous un peu handicapé dans la façon de dire les choses et de les exprimer à autrui. Moi je suis aussi dans cette situation là quand tu travailles avec des journalistes, pour expliquer, pour sortir si j'aime - je n'aime pas. Je n'ai pas le vocabulaire, je n'ai pas la méthode d'ordonnancement des notions que je dois appliquer à ce que je ressens. Alors peut-être que je ne l'aurais pas plus en stéréo, en tout cas...

Par exemple Abigail Wincott qui est une Anglaise avec laquelle il faut que je te mette en contact, elle veut travailler avec l'université de Montpellier qui est une université de linguistique, où eux ils sont capables de se faire des études sur la place de l'interjection dans l'écriture journalistique etc... Elle veut travailler là-dessus parce qu'elle a noté ce biais-là et elle pense que l'audio 3D évoluera à partir du moment où il y aura une cohésion dans la terminologie. La discussion qu'on avait eu sur le parfum et sur ce que fait l'IRCAM¹⁰⁹ est toujours un petit peu... relève pour moi du même problème. C'est-à-dire que tout ce que les ingénieurs du son de l'IRCAM peuvent mettre dans l'intention du mixage qui est censé

¹⁰⁸ Xavier fait référence à un moment dans notre conversation où Benoit est entré dans notre salle pour demander validation d'un choix de mixage auprès de Xavier.

¹⁰⁹ Xavier fait référence à la technologie SpeaK, développée par l'équipe PDS de l'Ircam et utilisée notamment par Ircam Amplify dans ses Forums pour les pouvoirs du son.

représenter un autre sens que l'audition, on ne peut pas trouver un point commun verbal qui vous permettent de dire « ah ouai je comprends ce qu'ils ont voulu faire à ce moment là parce que je suis dans le j'aime - je n'aime pas, je sens - je ne sens pas ». Peut-être qu'il n'y a pas de solution mais franchement, si moi on pouvait me dire « quand tu abordes un programme en multicanal, la première chose auquel tu dois penser pour l'exprimer c'est ça. Après c'est cette partie-là, ensuite c'est cette partie-là. En tout cas tu ressens la totalité des émotions simultanément mais la façon d'en parler c'est ça, ça et ça. » Mais moi mon rêve ce serait de savoir-faire ça. Je ne sais pas le faire.

V : Ce que tu dis reviendrait à hiérarchiser les paramètres sonores d'une scène à 360°.

X : Exactement. Donc peut-être que c'est la même terminologie que la stéréo, mais il y a quand même quelque chose d'autre qui est de l'ordre de l'espace. Et ça on ne l'a pas. C'est pour ça que le sujet m'intéressait et que quand Frédéric m'en parlait en disant qu'on n'a pas ce truc-là, on ne sait pas en parler, on ne sait pas l'exprimer, en tout cas on peut le faire qu'entre gens qui se comprennent, et lui il est beaucoup plus confronté à ça quand il fait des formations avec des gens qui ont une grosse compétence stéréophonique, il est plus gêné que ça. Nous on est un peu dans notre cercle fermé, on se connaît tous et on est dans le non verbal presque, mais voilà je suis assez à court, et c'est pour ça que ton sujet est intéressant et impressionnant parce qu'en gros c'est de l'aide au handicap quoi.

V : Et pour le Chinese Man, comment ça s'est passé avec l'artiste ? Il avait des volontés spatiales ?

X : Non, ils sont arrivés avec leur album, c'est leurs 20 ans, ils sont venus parce qu'on avait déjà bossé avec eux plusieurs fois, on leur avait fait leur premier clip en 360, ils sont arrivés en disant « faites-nous des propositions ». Donc c'est leur ingé son stéréo qui a d'abord fait l'interface, ou nous on a fait une proposition de spatialisation, il est revenu lui en orientant, en validant, en disant oh non ça aurait voulu que ça vogue plus, ça plus derrière etc. ça a duré que dalle quoi 4h, après les artistes sont revenus, ils ont écouté tout, ils ont validé, et ont vérifié qu'ils retrouvaient leur intention artistique, après c'est reparti ils l'ont écouté au calme chez eux, on a bien vérifié, on leur a envoyé des mp4 pour qu'ils puissent être dans les conditions d'écoute Apple, et après il y a eu un retour, et finalement leur retour principal était de dire « ça bouge pas assez sur certains titres ». Et là Benjamin, que t'as vu, qui est musicien et qui fait un peu la direction artistique musique avec Benoit, là on a dit « on veut bien aller plus loin mais on n'ira pas au-delà de ça parce que sinon on sort du mix ». Et c'est nous qui finalement les avons un peu freinés par rapport à leurs intentions en leur disant « gaffe, si on s'aperçoit que waouh ça bouge, c'est que déjà on écoute plus ». Et du coup hyper facile, hyper fluide, très agréable, mais ils étaient vraiment dans les débats, proposez-nous, ouverts à ce qui se passait. Et ça correspondait à peu près à ce qu'ils attendent de nous.

Moi je craignais qu'on ait l'inverse. Qu'ils soient « non on ne retrouve pas notre mix qu'est-ce que ça fait là » et que finalement on reconstruise une stéréo classique. Et pas du tout, ils étaient tellement en confiance qu'ils s'attendaient à ce qu'on fasse plus. Et on a dit « non on ne peut pas, on peut faire plus, mais dès qu'on pense au fait que « ah elle chante là », c'est fini tu es sorti du mix, ce n'est pas le but, tu es dans la Foire du Trône quoi.

V : Parlons du mouvement.

X : Il y a des moments dans la musique ou ça permet de servir le discours et d'autres pas. Je suis plus dans le docu donc j'ai plus de libertés de faire ça mais même dans le documentaire, quand tu commences à faire bouger tu sais pourquoi tu le fais quoi.

Mais est-ce que le mouvement serait... même la définition du mouvement et le vocabulaire autour du mouvement n'existent pas. Je me suis fait critiquer par mes collègues, tu écouteras sur Femmes de ring, j'ai décidé de mettre tout un discours qui était issu d'un reportage Radio donc j'ai vraiment traité dans une idée de transistor etc. Et qui tourne, parce que la femme en lien avec la musique nous raconte autour de nous la façon dont elle se faisait frapper par son mec et comment elle s'en est sorti et ses rapports à l'hôpital, c'est une petite scène qui dure 2 min 10 à peine, mais pour moi elle était un discours quasi général, représentatif de tout ce qui peut se passer toutes les semaines dans les services des urgences de France et de Navarre et du coup je la voulais pas quelque part, parce que je voulais qu'elle soit personne, et mes copains ont dit « ouai mais bons pourquoi est-ce que tu l'a mis dans un transistor, pourquoi est-ce que tu rajoutes cette idée de produit radiophonique et qui en plus bouge ? » Justement, pour moi elle est issue d'un média, donc c'est le média qu'on entend, donc c'est elle qui parle et ce n'est pas son histoire, c'est l'histoire de toutes donc pour moi ça bouge. Je n'ai pas cédé. Mais l'argument était juste mais dans ces cas-là pourquoi est-ce que tu lui ajoute ce côté transistor ?

Et donc pour certains de mes collègues c'était ou l'un ou l'autre. Mais j'étais dans un... Ils m'ont dit « en fait on entend qu'elle bouge on n'écoute pas ce qu'elle dit. » Alors peut-être que c'est parce qu'aussi, quand tu fais un truc, tu le connais tellement par cœur que bon. Il y a qu'eux qui m'ont critiqué mais voilà nos échanges ils sont de cet ordre-là.

V : On aborde le côté subjectif du jugement.

X : Oui mais cette subjectivité-là pourrait être entre guillemets canalisée par un vocabulaire commun qu'on n'a pas parce qu'on est tous dans le subjectif. Et peut-être que c'est un idéal auquel on aspire et qui n'existe pas hein mais je pense que ça nous permettrait de mieux nous rencontrer. Et par exemple, ça permettrait dans une note d'intention d'avoir des trucs. Franck Bessière était dans cette difficulté-là devant les mecs de la cellule podcast. « - Mais pourquoi le faire en audio 3D ? - Euh c'est tellement évident que je ne sais même pas comment te le dire man. » Tu vois, il avait un peu envie de répondre ça donc il s'en est bien sorti en parlant de la place, en parlant du fait qu'il voulait mettre l'auditeur dans une situation de ressenti de ce qui était à cette époque-là l'espace sonore. Le thème c'est la migration. Et que finalement les experts devaient être en contrepoint, c'était eux qui devaient être les bases solides et tout le reste devait être quelque chose d'enveloppant pour transmettre l'impression. L'auditeur on raconte l'histoire de l'humanité et son thème c'est tous migrant et raconter les différentes phases de la migration depuis les 3 derniers milliers d'années. Mais si ça pouvait être de dire « parce que en fait l'audio 3D s'il est abordé dans le Field Recording, d'une certaine façon, amène cette couleur-là etc. » J'invente mais c'est ça.

V : Les contenus *Session Lab* dont tu m'as parlé ils sont en binaural sur le site ?

X : Oui. Binaural de synthèse. Là c'est émission de radio, c'est tourné avec 2 micros d'appoints, ambisonique au niveau de l'ambiance ou en double MS. Ambisonique FOA (First Order Ambisonic) qui lui-même après va vers un 7.1.4 donc ça le Pyramix fait ça très bien, on

repasse en fait en mode canal donc l'ambisonique retourne dans le *Bed*, et puis tout l'habillage, lui est à partir d'*upmix 7.1.4* de version stéréo. Enfin du positionnement stéréo dans l'espace. Après t'as Femmes de ring que tu vas pouvoir trouver dans les podcasts de RFI aussi.

A12.4 Entretien avec Olivier Goinard

V : Tu as mixé plusieurs films en Atmos, comment ça s'est passé ?

O : Le pluriel est un petit peu tiré par la corde parce que j'en ai vraiment mixé un seul qui est *Le Chant du Loup*, d'Antonin Baudry. Après j'ai fait une série en Home Atmos d'Olivier Assayas qui s'appelle *Irma Vep* pour HBO. Donc je ne suis clairement pas le mixeur le plus expérimenté en France vis-à-vis de l'Atmos, si ce n'est de l'avoir fait une fois, et en l'ayant fait une fois je sais quels sont les enjeux, quelles sont les possibilités et *Le Chant du Loup* c'est un film qui s'y prêtait bien donc c'est intéressant.

Après, objectivement c'est quand même un format qui a du mal à décoller malgré son implantation qui n'est pas négligeable en termes de nombre d'écrans en France. La réalité est qu'on en a pris conscience très vite, et *Le Chant du Loup* est un des seuls films qui fait exception à la règle. C'est que quand un film sort en France et occupe des salles Dolby Atmos et est exploité dans ce format de son, ce sont des films américains à 99.9% donc il n'y a pas la place pour l'instant.

Je n'ai pas dit que ça n'allait pas pour autant évoluer mais par rapport au cinéma qu'on fait en France et comme il est exploité il y a une sorte de dichotomie. Il y a un truc qui ne fonctionne pas. *Le Chant du Loup* ça s'y prêtait donc ça a quand même pris quelques écrans, il y a quand même eu une exploitation réelle mais ce n'est pas loin d'être le seul. Il n'y a pas encore ce dynamisme des productions d'investir dans le Dolby Atmos parce que ça représente un surcout qui n'est pas tout à fait négligeable. Parce que qui dit vrai Dolby Atmos dit éventuellement une prise de son adaptée au Dolby Atmos, un montage son dans ce format là si on veut bien faire la chose et que ce ne soit pas juste un upmix d'une version 5.1 ou 7.1 du mix ciné pour faire un faux Atmos. Ce qui est intéressant c'est de faire du vrai Atmos sur toute la chaîne pour vraiment creuser des idées, explorer tout le potentiel de ce format-là qui est très intéressant. Donc tu imagines le surcout que ça représente dans une économie qui n'est pas du tout préparé à ça pour aboutir à une exploitation inexiste à l'heure actuelle, au moins en France, et même ailleurs. C'est-à-dire qu'un film Français n'a aucune chance d'avoir une exploitation en Atmos dans le monde.

Ce qui aurait pu tout changer mais ce qui est en train de se faire enterrer une fois de plus par les producteurs, c'est que Canal a commencé à demander des delivery, c'est à dire les formats pour la diffusion du film sur ces plateformes, en Home Atmos. Il y a eu ce cas avec *Le Règne Animal* qui est mixé en 7.1, là ils ont pris 3 jours, et c'est Canal qui a financé la chose, ils le voulaient vraiment comme ça, ils voulaient en faire un petit événement pour avoir le film vraiment tamponné en Home Atmos et que ce ne soit pas juste une répartition ou on met du 7.1 dans du Home Atmos. Ils voulaient vraiment utiliser la technologie Home Atmos. Là pour le coup le film a pas du tout été prévu pour ça à la base et donc a été upmixé. C'est une sorte de faux Dolby Atmos mais ça existe comme ça. Si jamais tu as tous les diffuseurs qui s'y mettent, d'exiger un format Home Atmos... Même du Home Atmos à la fin ça veut dire qu'en amont il faut qu'on parte aussi d'un vrai Atmos...

Et il y a cette idée-là, que j'ai eu avec HBO quand je faisais *Irma Vep*, c'est à dire qu'eux aux US et même en France, n'ont pas les moyens de diffuser en Home Atmos et pour autant ils m'ont demandé de faire une version en Home Atmos, que personne n'écouterait jamais mais

parce que c'est dans l'idée qu'un moment donné, à termes, c'est ce qu'ils doivent avoir en tête, c'est qu'il n'y aura plus que ce format. Ce qui n'empêchera pas des gens, si tu veux, de recevoir le single Home Atmos et il y aura des décodeurs qui mettront ça dans différents formats, ce sera juste un downmix si tu veux et ça, ça faciliterait tout un truc de flux, de deliveries etc.

J'ai pas mal divergé mais ça montre le marasme dans lequel on est. Aussi ça te montre quelque chose vis-à-vis de la réalité de l'exploitation du son immersif pour les films français qui est quand même relativement médiocre. En tout cas inexistante. Et pour l'instant elle ne peut pas exister. Et il y a une autre raison à mon sens, c'est le fait que le festival de Cannes n'est pas équipé en Atmos. Le jour où les films d'auteurs qui sont présentés vont être potentiellement diffusés en Atmos, ça va générer un effet boule de neige, on peut l'imaginer. Et à la fois je mettrais un petit bémol là-dessus parce que Berlin est équipé et sur 20 ou 30 films en compétition il y en a que 2/3 qui sont en Atmos. Et à Venise ils ont une salle, pas la grande, mais une salle aussi en Atmos, et personne ne présente des films en Atmos. Donc à la fois sur festival ça ne prend pas. Après je pense que Cannes... Et pour peu que le film qui aura été mixé en Atmos remporte la Palme, peut-être que ce sera là le grand coup de départ. Mais on sera peut-être à un point où il y aura un nouveau format sonore qui ne sera pas de l'Atmos.

V : Quand tu parles de Home Atmos, tu entends un format ADM livré aux diffuseurs ?

O : Oui c'est ça.

V : Je n'avais jamais entendu parler du terme Home Atmos.

O : Home Atmos et Atmos ce n'est pas la même chose, c'est une déclinaison. Tu as quand même des tops et des objets, mais c'est beaucoup moins... Si tu veux, le signal va se déplacer entre 2 tops, 2 rails de tops, tandis que dans une salle de mixage ou une salle de cinéma tu peux en avoir entre 10 et 15 par rail. Donc tout d'un coup tu as la précision qui est augmentée. Après tout ça c'est des boîtes noires, tu ne sais pas trop ce qu'y s'y passe, ça n'empêche pas que ton mouvement est traduit entre 15 enceintes sur le top gauche, et en Home Atmos il va être traduit entre 2 enceintes. Mais le Home Atmos c'est quand même un format, c'est lui qui a le plus de chance de se développer parce que c'est lui qui a le plus d'utilisateurs, comme le Home Cinema. Mais à mon avis c'est aussi minoré parce que tous les Parisiens et habitants de la région Ile de France devraient être les grands acheteurs de Home Cinema, sauf que vu le contexte immobilier on a des petites pièces et le Home Cinéma ne se développe pas du tout. Ou il se développe mais pas autant qu'on pourrait l'imaginer

V : Ou alors il y a des barres de son et ce genre de choses, du transaural.

O : Oui après il y a plein de truc un peu bizarre. Tu mets le stickers Atmos partout. Du moment où tu mets un haut-parleur qui va vers le plafond ça devient un Home Atmos. Ça c'est la foire commerciale.

V : Pour toi quelle est la différence fondamentale entre un stéréo et de l'immersif ?

O : Même un mono. Ce qui est pas très clair c'est le pouvoir de discrimination des différentes couches et la coexistence des différentes couches de son qui sont bien plus puissantes, plus fortes, plus enthousiasmantes dans un format Atmos que dans un format plus réduit, Parce

que tu as une manière de répartir les sources sonores qui est vachement intéressante. En tout cas elle permet à ton oreille d'avoir une meilleure discrimination, donc une meilleure cohabitation de l'ensemble des sources sonores. Donc le mix devient plus facile, plus intéressant, c'est comme si quelque chose se libère. Tu le vois quand tu mixes et que tu commences à surcharger, parce que maintenant on empile des dizaines de sons ensemble donc pour faire tout entendre. C'est tout un art de mixage et tout un travail. Et quelque part c'est plus facile dans un format dit immersif parce que tu peux plus répartir les choses dans l'espace.

C'est très basique finalement. Et la précision de spatialisation elle est aussi plus forte par exemple avec les objets parce que ça permet vraiment, par une combinaison d'ondes, par l'ensemble du maillage et des haut-parleurs, d'avoir un positionnement de la source sonore que tu veux diffuser à un endroit beaucoup plus précis que si jamais tu joues dans un espace spatial plus réduit type 5.1 ou ça va être un équilibre avec l'enceinte avant, les enceintes arrière etc. C'est beaucoup plus précis. C'est comme un maillage, ça reste très simple finalement. C'est juste que la définition est meilleure tout simplement.

V : Tu disais que *Le Chant du Loup* s'y prête bien. Pourquoi ?

O : Parce que c'est dans la mise en scène même du film. Il y a tous ces moments d'écoute où tout d'un coup on pouvait vraiment créer des situations très complices avec le spectateur. Quand l'oreille d'or joué par François Civil se met à écouter par exemple, on renversait l'écoute vers l'arrière de la salle et dans la définition sonore qui est propre au Dolby Atmos, tu as presque aucune déperdition de finitions sonores entre l'avant et l'arrière, avec toutes les histoires de Bass Management. Quand tu déplaces un son dans un espace 5.1, tu vas te servir des arrières uniquement pour des choses très diffuses, pour des sons qui vont être globalement des compléments de la façade et pour mettre de l'espace, créer ces environnements-là. L'Atmos quant à lui te permet vraiment de placer des sons qui deviennent d'autant plus des acteurs, qui sont beaucoup plus prédominants et que tu acceptes mieux parce qu'ils ont une très bonne définition dynamique et spectrale. Il y a beaucoup de salles où les arrières... Tu as quand même une bande passante et un état des enceintes qui est souvent assez douteux, donc tu ne peux pas t'appuyer vraiment sur ces enceintes-là, tandis que tu sais que si jamais tu vas dans une salle Dolby Atmos elle est forcément moderne, elle a aussi un cahier des charges pour avoir la validation de Dolby d'être exploitée en tant que Dolby Atmos qui fait que tu peux vraiment faire confiance à ce système. Il y a une qualité de restitution de ton travail. Tu peux aller mettre des voix à l'arrière et les voix sont vraiment diffusées. Si tu vas dans une vieille salle pas révisée depuis longtemps et qui est toujours en exploitation 5.1 depuis 20 ans, tu sens bien dans ton mix que tu n'as pas forcément intérêt à jouer avec. Le fait d'être en Atmos ça te libère de ça. Après ça reste un risque parce que ton film sera aussi exploité en 5.1 et en 7.1. Donc ce sera aussi diffusé dans des salles où tu auras joué avec la spatialisation maximale et il y aura une sorte de reflet à tes risques et périls dans ces salles de moins bonne qualité.

C'est aussi le problème du mix, c'est que ton travail puisse résister à toute sorte de configuration. Si vraiment tu commences à t'amuser à des choses de spatialisation, alors après, dans des formats inférieurs, qui sont quand même majoritaires en exploitation, tu peux te retrouver avec des choses qui ne sont pas très heureuses. Ce n'est pas pour autant qu'il faut mixer en mono mais faut savoir ce qu'on fait, tu n'es pas forcément dans une liberté totale. Les salles ne sont pas pareilles, elles ne sont pas toutes dans un même état et il faut surtout tenir compte de la compatibilité de tous les formats sonores.

V : Et le fait de faire le choix de mettre beaucoup d'éléments à l'arrière quand l'oreille d'or écoute comme tu disais, est-ce que ça ne posait pas des problèmes en 5.1 ?

O : En tout cas, la matrice qui fait ça crée des downmix de manière un peu automatique que tu peux contrôler. Par exemple, tu as quelques paramètres, assez obscurs certes, mais que tu peux quand même maîtriser un petit peu. Pour le downmix que tu fais de ton format Atmos vers 7.1 ou vers 5.1, il y a des petits paramètres sur lesquels tu peux agir mais ça reste globalement la boîte noire qui fait le downmix à ta place. Donc tu peux te rendre compte de l'effet que ça fait mais globalement c'est quand même un downmix assez magique, ça marche plutôt très bien. Cela dit j'entendais bien que le mouvement se répétait à l'arrière et je ne l'aurais peut-être pas osé en 5.1. En tout cas, quand ça se passe bien en Atmos, tu sais globalement que la transposition est assez heureuse tout en insistant sur le fait que ça fait malgré tout prendre un risque sur l'exploitation 5.1. Mais on avait choisi cette audace là et on avait de toute manière envie d'avoir un mix Atmos qui soit un peu clinquant. Sachant aussi que dans un auditorium, tu n'as pas le temps physiquement de faire un mix total Atmos, un mix total 5.1 un mix total 7.1. C'est impossible en termes de timing. Donc à un moment donné tu fais des petites écoutes pour aller vérifier et tu vois bien les moments qui sont un peu problématiques. Et finalement c'est 10-15 min du film. Et ceux-là tu vas vraiment les écouter. Pour le reste il y a des moments de compatibilité qui sont très forts.

Le principe de l'Atmos et de tout format de spatialisation ce n'est pas le fait de faire de l'Atmos tout le temps. Ce qui est vachement intéressant c'est à mon sens d'avoir justement une dynamique de spatialisation. Vraiment je n'aime pas les montages son ou les mixages où c'est en permanence un truc qui ne respire pas, qui a tout le temps la même identité sonore. Ce que je trouve intéressant c'est de faire des moments mono, des moments stéréo, des moments de mono dirigée, et des moments où là tu peux t'éclater à faire exploser le système Atmos. Parce que le film le mérite, parce que les scènes le méritent, il faut juste adapter ton train de spatialisation à la dramaturgies du film.

V : Et quels sont les moments qui méritent de pousser l'Atmos à fond ? À quel moment tu te dis « là il faut y aller » et pourquoi ?

O : De fait ça s'impose un peu à soi-même par... C'est vrai que là où l'Atmos a des grandes réussites c'est sur des scènes d'action. On peut le constater assez aisément. Par exemple il y a un moment où un hélicoptère attaque un sous-marin et c'est quelqu'un qui a un bazooka qui monte sur le sous-marin qui détruit l'hélicoptère. Autant te dire que là tu t'éclates parce que tu as tout ce que n'a pas pu filmer la caméra qui tout d'un coup est énormément aidé par la spatialisation. Ce genre de scène comme ça, tous les passages, les tirs, la mer, et surtout aussi la musique, là il y a tout le côté spatialisation qui entre en jeu. Les trajets des différents éléments. Mais il y a aussi la coexistence des sources sonores dont je te parlais au début. Tu mets la musique à un endroit, tu mets le fond sonore à un autre, l'ambiance de la mer, des vents, et après tout le côté combat. Et en fait c'est là où il y a une énorme différence par rapport à d'autres systèmes, c'est que ça devient presque facile. Et puis c'est super agréable pour les oreilles, c'est moins fatigant, c'est plus enthousiasmant, c'est fascinant, ça fait peur, ça fait des sensations. Évidemment là tu sens vraiment les différences. Après pour autant tu peux aussi avoir des super utilisations de l'Atmos juste dans les moments d'ambiances, dans les moments beaucoup plus intimes. Particulièrement les tops donnent une dimension verticale au

son qu'évidemment tu n'as pas du tout en 5.1 et en 7.1. Ça rend une manière de remplir la salle qui est vraiment extraordinaire, surtout qu'on a une très mauvaise discrimination verticale en tant qu'être humain. Donc ça permet de mettre des choses sans qu'on s'en rende compte véritablement et ça donne vraiment une autre dimension. Pour le coup ça, ce n'est pas volé. Rien que pour les ambiances. On n'a pas besoin d'être dans des choses spectaculaires ou des choses d'effets. Tout ça devient vraiment très joli en Atmos.

V : Dans ce que tu biens de dire, tu as commencé à séparer les éléments de ton mix : effets, musique, voix... Comment tu te représentes l'espace sonore en immersif ? Si tu devais décomposer ton image sonore, comment tu la construis dans ces moments-là ? Quels éléments on y trouverait ?

O : Comme ça, intuitivement je dirais que tu as plus l'impression d'être dans une potentielle sphère de sons. C'est le fait de pouvoir mettre les sons dans cette sphère. Ça donne cette sensation-là, que tu retrouves au cinéma quand tu vas voir le film. Tout d'un coup on peut dire que le son devient 3D. Après tout ça est un peu tiré par les cheveux mais c'est vrai que ce sont des sensations que tu ne peux pas reproduire dans les autres formats sonores. Ça reste le format sonore aujourd'hui le plus abouti. Ça marche plutôt très bien dans ce que nous on veut faire et c'est une vraie évolution qualitative.

Ensuite, où je place les sons, ça dépend de chaque scène. Il n'y a rien de stable en soi. Globalement on reste sur des conventions qui marchent bien et qui sont fortes au cinéma parce qu'il y a ce qu'on appelle le *voco-centrisme* par exemple. La voix elle est très agréable quand elle vient de l'écran et un film ça reste très globalement des dialogues. Ça, ça ne change pas énormément. Après ce qui change par exemple ce sont les acoustiques, beaucoup, parce que tu peux te servir de tout le système Atmos. Mais c'est encore au balbutiement parce que les reverbs sont pas encore développées pour ça. Elles commencent à l'être mais tout ça prend vraiment son temps. La techno est arrivée et rien n'était prêt pour l'Atmos. Il n'y avait aucun système de prise de son adapté quasiment, c'est en pleine recherche encore et c'est pas du tout abouti donc ça prend vraiment son temps. Pour que ce soit à maturité il faudra encore entre 5 et 10 ans facile. Même Pro Tools, dans Pro Tools n'es pas au max mais ils restent un peu en avance par rapport à d'autres, tout ne converge pas au même moment. De toute manière c'est toujours un jeu de chat et de chien parce que personne ne bouge tant qu'un autre n'a pas bougé et une fois qu'un a bougé, ça force les devs à se bouger à leur tour, et pour le moment les demandes des utilisateurs sont quasiment inexistantes. Donc pour l'instant on se débrouille avec des choses qui ne sont toujours pas vraiment adaptées.

V : Tu disais que ça permettait de prendre des risques. Est-ce que ça te permet de sortir du fait qu'on met peu de choses dans les arrières en cinéma, de sortir de l'idée que l'image reste reine et de s'amuser plus ?

O : Oui, mais ça c'est vraiment une question de choix de mise en scène. Après un film comme *Le Chant du Loup*, ni le tournage, ni le montage image a au fond véritablement pensé la spatialisation du son, c'est un truc qui s'est construit à posteriori du montage du film. Il faudrait avoir une démarche qui soit vraiment intimement liée à la mise en scène dès le début de l'écriture, même du scenario. Ce serait vraiment très intéressant et ça pour l'instant ça n'a pas été fait.

V : Ça s'est pensé après ? Le réalisateur s'est dit plus tard qu'il pourrait le faire en Atmos ?

O : Ça s'est vu assez vite parce que le film avait des moyens. Ça s'est peut-être négocié de base, qu'il serait fait en Dolby Atmos, surtout qu'Antonin est un passionné de son. Mais après, un truc qui est central, c'est la question de l'exploitation, tant qu'on ne savait pas vraiment quel film on avait et ce que ça allait donner... Il y avait un gros budget et tout allait converger pour que ce soit quand même en Atmos mais ce n'était pas central de la part des producteurs qui n'en savent rien quelque part. Ils savaient à peine ce que c'était que l'Atmos. Ça pourrait être central, ça pourrait être un argument, c'est une très bonne chose qu'il l'ait été, et puis le son a été salué, c'était une très bonne expérience pour tout le monde. Je pense que les spectateurs qui ont eu la chance de le voir en Atmos en ont quand même pris plein les oreilles, c'était quelque chose et c'était super pour le film. Mais la mise en scène en soi elle n'est pas réfléchie en Atmos. On s'adapte, on met en valeur, on effectue le travail qu'on fait aussi en 5.1 ou en 7.1 ou autre chose. C'est à dire qu'on compense des faiblesses de mise en scène puis après on bénéficie aussi. Parfois c'est vachement bien pensé ou vachement bien foutu, ou le montage il est déjà très en place et il n'y a plus qu'à aller plus loin. On est toujours dans ce même rôle, c'est-à-dire qu'avec le son on compense des faiblesses ou on met en valeur des choses qui sont déjà là mais ça ne diffère pas spécialement si ce n'est que tout d'un coup tu as un potentiel plus grand de création technique et artistique avec le son.

V : Est-ce que les réalisateurs sentent vraiment la plus-value de l'Atmos ?

O : Antonin lui c'est un passionné de son et il l'a vraiment sentie lui la plus-value. Tu cèdes, tu commences à écouter et après tu peux plus t'en passer, c'est addictif. D'ailleurs ce serait intéressant de faire un sondage : quels sont les réalisateurs qui ont déjà vu un film en Atmos dans leur vie ? Je pense que le résultat serait effarant. Après il faut voir que s'ils ne s'y intéressent pas, c'est que c'est impossible pour 98% des films français d'avoir un mixage en Atmos. C'est aussi ça la réalité. Sur les 250 films produits, combien sont mixés en Atmos ? J'exagère peut-être avec le nombre mais il y en a peut-être 10%, je n'en sais rien. Mais je sais qu'il n'y en a pas énormément. En tout cas je sais qu'il y a beaucoup d'auditoriums Atmos à Paris parce qu'ils se sont équipés des subventions du CNC. Il y en a un dynamisme. Mais il suffit juste d'appeler des auditoriums où il y a de l'Atmos, Polisson, Creative Sound, Yellow Cab, Boulogne, Transperfect et tu peux leur demander : dans vos auditoriums Dolby Atmos combien de films se mixent par an en proportion ? Il doit y avoir 12 films qui se mixent dans chaque auditorium Atmos, ce serait intéressant de voir la proportion. À mon avis elle est quand même relativement faible.

Malheureusement c'est comme ça parce que la post prod est un parent pauvre. Ça ne coutera pas cher, c'est ça qui est bizarre. Une post prod son et image en France ça coutera dans les 300 000 euros à peu près, par rapport au budget d'un film ce n'est pas énorme, mais on arrive à un point où ils ont tout dépensé, les prods, avec les surplus de tournage, les imprévus etc.. Donc on arrive avec juste ce qu'ils avaient prévu au maximum. Il n'y a pas un investissement très fort sur la post prod son, ou la post prod en général, c'est le tournage qui prend tout.

V : Et il n'y a pas eu de prises de son en Atmos ?

O : Pour ce film il y en a eu aucune. Après c'étaient des belles prises de son, avec un tournage dans des sous-marins etc. Mais il n'y a pas eu de prises de son dédiées à l'Atmos. Ce

n'est pas forcément ultra indispensable, c'est juste que là, pour l'instant, on est en train de faire des essais et on arrive sur des trucs qui sont concluants parce qu'on a enfin des relations de phase entre les différents canaux du Dolby Atmos. Au lieu de faire de la mono ou de la stéréo dirigée, on pourra appeler ça comme ça, mais qui marche aussi, on va essayer de trouver un système le plus intéressant possible. De toute manière, tous ces systèmes de couple ils ont vraiment leurs limites parce qu'ils marchent fondamentalement à un endroit dans la salle précis, exactement comme une écoute stéréo où il y a un Sweet spot et pas deux. Après ça n'empêche pas que c'est quand même bien d'écouter même si tu es à deux mètres de l'enceinte droite, mais tu n'écoutes plus que l'enceinte droite. Donc tout ça guide aussi beaucoup tes gestes en mixage. Tu ne fais pas un mix pour un point précis dans la salle de cinéma, faut que tout le monde en profite et que ce soit génial pour tout le monde. Ça affecte tes gestes.

V : J'aimerais revenir sur la scène d'action dont tu m'as parlé, comment tu l'as construite précisément ? Tu places les voix au centre ?

O : Tu es libre. Il y a aucune norme qui t'interdirait de faire autrement mais ça n'empêche pas que le désir que je peux ressentir à l'heure actuel et le désir des producteurs, réalisateurs, équipes autour de moi quand je fais le mix ne posent pas question : on ne va pas mettre la voix à l'arrière alors qu'on voit l'acteur sur un écran, c'est tout. Après il y a des choses très intuitives qui se sont construites depuis des dizaines d'années et ils ont quand même réussi à établir une sorte de langage inconscient de la spatialisation qui existe. On ne repart pas de zéro. Après chacun a son goût, ses styles etc.

Moi je suis pas du tout un spécialiste des films d'action. J'ai plutôt un créneau dans les films d'auteurs, dans lesquels ça parle beaucoup, où il y a une sensibilité qui passe énormément dans les ambiances. Je suis spécialiste de ça, j'ai mixé très peu de films d'action grand public. Donc il y a des personnes qui peuvent avoir un tout autre rapport avec ce que je dis là. Je suis plus dans une école où, en effet, il y a des voix à la façade, il y a des belles acoustiques, il y a des sons d'ambiance plutôt très travaillés avec des mouvements à l'intérieur, de niveau etc. Il y a vraiment une démarche d'avoir des sons plutôt diffus à l'arrière, parce que toute ponctualité est souvent gênante. Ça fait tourner la tête comme on dit. Ce n'est pas forcément une sensation très agréable. Après dans des films d'action, tu es de toute manière submergé par tous ces sons et c'est là que tu peux déborder et t'amuser avec un espace sonore qui est beaucoup plus englobant. Et la grosse différence se fait du côté transitoire des sons qui va envahir l'espace sonore, chose qui n'arrive pas dans des scènes beaucoup plus classiques où t'as une scène à la campagne etc. Tu ne vas pas mettre des choses très ponctuelles et dérangeantes parce qu'elles vont te faire tourner la tête et elles te détournent l'attention de la dramaturgie. Moi je vois un intérêt dans l'Atmos et je pourrais très bien mixer un film simple que j'ai mixé en 5.1 en Atmos, pour que ce soit encore plus intéressant et plus sensible. Parce que même dans les scènes du *Chant du Loup* il y en a pas mal où c'est beaucoup plus simple et on pourrait dire que ça ressemble à un film d'auteur classique. On retrouve ce côté avec des sons de transitoires à l'avant parce que tu les associes vraiment à l'écran, et après un ensemble assez diffus sur tout ce qui est l'espace électroacoustique qui est en dehors de l'écran.

V : Donc là typiquement, ce sont des bruits d'impacts dans l'eau, ou sur le sous-marin, que tu t'es permis de mettre sur les côtés, dans les arrières, en élévation.

O : Tu as un hélicoptère qui passe qui est capable de traverser la salle et qui tire en même temps, donc tu n'hésites pas, et c'est un grand plaisir de se prendre les balles du côté inférieur gauche de la salle et tu finis à l'arrière droit et tu traverses plus de 100m. C'est dément quoi. Ça apporte énormément de plaisir à la scène.

V : Et les musiques, est-ce que tu les repositionnes différemment dans l'espace ?

O : Totalement parce que tout ça est toujours lié à cette histoire de discrimination qui est bien meilleure. C'est à dire que tu peux avoir un partitionnement des sons. Avant, en tout cas dans un format 5.1, tu étais un peu obligé d'empiler des sons au même endroit. Et tu n'as pas beaucoup d'alternatives. Si jamais tu veux faire une spatialisation sur la façade, tu restes sur ton LCR, et tu as quelque chose d'un peu diffus à l'arrière. C'est déjà pas mal. Là tu peux profiter notamment des tops et puis des objets parce que si jamais tu veux positionner des instruments de manière très précise dans l'espace, ça devient possible. Donc le fait que chacun prenne sa place, admettons que tu aies en simultanée 100 sources sonores, plutôt qu'elles soient réparties entre 5 enceintes principales, là elles sont réparties entre des dizaines d'enceintes. Ce n'est pas la même chose. Tout d'un coup tout devient plus lisible, plus fascinant.

V : Tu avais finalement des musiques qui étaient livrée en multipiste.

O : Disons que oui mais tout ça venait de musiciens qui bossent en stéréo chez eux dans leur Home Studio. Donc c'était beaucoup de couches de leur stéréo. Après tu joues avec ces différentes couches mais en stéréo. Et ce qui va durer très longtemps comme ça. Après tu peux avoir des enregistrements de vrais orchestres, ce qui n'a pas été le cas pour *Le Chant du Loup*. Donc tu peux avoir un enregistrement qui est plus dédié au Dolby Atmos. Mais la majorité des musiciens aujourd'hui restent dans une écoute stéréo chez eux donc ce qu'ils te livrent, quand c'est eux qui livrent directement les choses, ça reste de la stéréo. Donc à partir de là on fait des upmix divers et variés. Ça n'a pas empêché de s'amuser avec. À la base ça reste, et dans énormément de cas, juste de la stéréo.

V : Et tu parlais des reverbs qui visiblement représentent un des autres éléments qui se prête bien à un système immersif comme l'Atmos. Comment tu les travaillais ? Tu disais qu'il n'y avait pas forcément beaucoup de reverbs multicanales.

O : A l'époque ça n'existant pas encore la Cinematic Room ou l'Altiverb qui vient de passer en 7.1.2. Donc après tu te débrouilles avec 2/3 reverbs. Ça n'empêche pas que les canaux existent, et tu as tes reverbs de façade, tes reverbs d'arrière et ta reverb de tops, et tu te débrouilles avec ça. Ça n'empêche pas que tu peux mettre des valeurs de réverbération, des temps de réverbération différents selon les canaux donc tout d'un coup tu as une propagation du son, des pré-délais, enfin tu changes la manière dont tu travailles tes reverbs. Ça ne change pas forcément grand-chose le fait qu'à l'époque, comme ça n'existant pas, je les faisais avec 3 différentes et maintenant c'est plus ergonomique et je le fais avec une seule. Ça n'empêche que les paramètres sur lequel je joue sont les mêmes, c'est à dire que tout d'un coup tu peux faire des déplacements de sons qui sont vachement plus enthousiasmants et qui donnent des sensations d'espace qui sont assez magiques. Tout ça rejoue toujours la même chose, c'est par rapport au système qui est meilleur à tous les niveaux. Pas seulement en termes spatial, c'est aussi fréquentiel et dynamique. Ce sont les trois grands atouts de ce système.

V : Est-ce que la question se pose, en cinéma, de la symétrie et de l'asymétrie.

O : Disons que je suis prêt à tout, ça dépend, je n'ai pas forcément de limites mais il faut vraiment que ça s'y prête parce qu'on reste dépendant de l'image. Sur *Le Chant du Loup* ça ouvrait des espaces, on trouve une sorte de langage inconscient qui fait qu'il y a des choses, des manières de faire qui appartiennent ou qui n'appartiennent pas au film. C'est très difficile à exprimer ou à sentir mais il y avait une sorte d'évidence, parce qu'il y avait des effets de contraste à faire du fait d'être enfermé dans un sous-marin et tout d'à coup il y avait des scènes où on sortait parce qu'il y avait des scènes de combat à la surface. Ce qui était vachement intéressant c'était de faire des contrastes de spatialisation. Et un peu dans tout : des contrastes de dynamique, et fréquentiellement aussi ça se faisait aussi naturellement, dans l'ambiance un peu ronflante d'un sous-marin.

Malheureusement je n'ai pas encore essayé des choses assez variées, mais je ne demande que ça, c'est juste que pour l'instant je ne suis pas sur ces projets. Ça ne risque pas forcément de m'arriver. Je ne sais pas ce qui va m'arriver, je serais potentiellement prêt à tenter toutes sortes de choses.

V : Est-ce que c'est plus facile de mixer en immersif ? Est-ce que tu as des souvenirs de difficultés que tu as rencontrées ?

O : Pas tant que ça, si ce n'est que quelque part tu sens bien que ton cerveau, en tant que mixeur, il est un peu plus mobilisé par tout ce dont tu dois te rappeler. Tout d'un coup tu te mets à envoyer des sources sonores à des endroits divers et variés et qui dépendent de l'interaction de beaucoup de haut-parleurs, donc forcément tu es dans quelque chose qui te sollicite plus intellectuellement, même si c'est inconscient. Ça n'empêche pas que le fait de mixer c'est aussi mémoriser toutes les intentions passées après passées. C'est être capable de se souvenir « tiens j'ai fait ci, j'ai fait ça, les choses sont là, les choses s'emboitent là », parce que tu as beau faire quelque chose, une semaine plus tard on réécoute et puis il y a des choses qui se remettent en question et il faut avoir conscience de ce qui a été fait pour pouvoir replonger les mains dedans. Je me souviens de ça. Cognitivement c'est plus encombrant. À la fois ça libère au niveau de l'écoute mais d'un point de vue de ton travail tu es obligé de mémoriser pas mal de choses en plus que tu n'as pas forcément l'habitude de retenir quand tu mixes dans un format plus réduit. Ça c'est notable.

En termes d'autres difficultés non, puisque comme je t'ai dit ça rend les choses quand même plus faciles pour l'intelligibilité. Par exemple c'est bête mais comme les choses se dégagent bien, s'identifient bien, tu as quelque part moins de problèmes. Tout s'emboite mieux, il y a des choses qui sont un peu plus dures à faire, un peu plus dans la lutte au niveau de l'exercice de mix, qui deviennent plus facile. À la fois faut bien les faire et faut bien travailler les choses pour qu'elles existent aussi en 5.1 et 7.1 mais il y a quelque chose qui t'aide à bien travailler quelque part parce que tu sens après l'impact sur les autres formats, où peut-être que tu aurais pataugé du fait de cet encombrement de niveau, spatial etc. et là tu l'as moins et donc il y a quelque chose qui fait que ça t'aide à faire les bons gestes sans doute.

V : Tu n'en as pas fait d'autres parce que tu n'as pas eu l'occasion ou tu t'es dit qu'un ça t'avait suffi ?

O : Non ça ne m'a pas suffi du tout, je serai très heureux d'en refaire. On a parfois envisagé la chose mais l'argent n'était pas là tout simplement. Ça ne s'est pas présenté mille fois non plus. Ça dépend du type de film que tu fais, je suis sur des films qui sont entre 2 millions et 5/6 millions c'est à peu près mon créneau et après il y a 4/6 mixeurs en France qui font des budgets au-delà. Ce sont eux qui les font, c'est mécanique. Ce n'est pas un créneau qui me concerne, donc les films en Atmos m'échappent pour l'instant, malheureusement. Mais je serais bien plus heureux de faire ça.

Et se pose la question, si on revient au fait qu'il y a quelque chose qui se passe au niveau de la matrice qui est super chouette dans les down mix, de mettre en place le fait de mixer tous les films en Dolby Atmos mais sans les écouter en Dolby Atmos. C'est-à-dire en laissant la matrice faire pour ramener en 5.1. Ce qui après, si jamais il y a besoin de faire un format 7.1 ou Atmos comme ça quelque part, d'avoir une configuration prête. Cette matrice a une manière d'agglomérer les sons qui est assez magique, et qui est différente que de la faire de manière discrète, en répartissant les sons tels que je le fais avec un panner 5.1. Je ne saurais pas te l'analyser, je n'ai pas la réponse forcément, mais il y a une agglomération des choses qui est vraiment très chouette. Il y a très bon downmix. Elle doit sûrement s'analyser scientifiquement mais je n'ai pas la réponse.

V : Je ne vais plus pouvoir voir *Le Chant du Loup* en salle de cinéma en profitant du son Atmos, mais en location en streaming ?

O : Ce ne sera pas l'Atmos que tu écoutes ce sera le Home Atmos. Sur le Blue Ray HD c'est en Home Atmos. Il a été particulièrement soigné parce qu'il y a très peu de réalisateurs qui suivent les mastering et qui permettent de faire les différentes versions vidéo du film et lui il était là jusqu'au bout donc ce sont vraiment des choses qu'il a signées. Et puis le mix Home Atmos est assez différent du mix Atmos original. D'ailleurs même Antonin le préférait. Parce que c'est encore des jours de travail et on a fait une version dont on était très content.

V : Alors là je ne comprends pas. Le mix est différent ? Même dans les mouvements ? Dans les positionnements des sources ?

O : Oui, si tu veux ton mastering tu le fais avec des éléments séparés, ce n'est pas l'Atmos qui devient du Home Atmos, ce n'est pas un truc où tu appuis sur un bouton et hop. Tu as encore une action possible.

V : Pour moi c'était une sorte de Downmix en fait

O : Ça pourrait être aussi simple que ça mais je fais toujours les mastering de tous mes films, ce n'est pas quelque chose que je délègue.

[...]

Je sors des éléments beaucoup plus éclatés qui me permettent après de refaire pour un environnement de vidéo, de perception vidéo, un mix qui est dédié si tu veux. Ce n'est pas pour autant que je repars dans des choses comme changer une post synchro ou des machins, tous les choix sont faits mais je travaille avec quelque chose qui est encore très évolutif.

V : Jusqu'à parfois penser l'image sonore différemment.

O : Oui parce que ce n'est pas tant différemment, c'est juste que par exemple, et notamment sur un film en Atmos justement, l'écart de perception du film et le fait de faire de l'Atmos, tout d'un coup tu fais quelque chose de très spectaculaire, tout d'un coup tu te retrouves dans une salle, tu as vu ça au Pathé Beaugrenelle avec 400 places en HDX. Donc tu as un souvenir dingue du film qui est concret, et après tu te retrouves dans une pièce de 8m2 avec le réalisateur, tu as beau faire du Home Atmos ça ne sonne pas pareil. Les sensations ne sont pas les mêmes donc pour recréer des sensations qui se rapprochent que ce que t'avais fait faut resculpter les sons, les faire passer un peu autrement, ça coexiste différemment. Le mix était plutôt bien, on n'a pas souffert on a juste voulu travailler encore mais ce n'est pas une transcription toute droite.

V : Mais les leviers de spatialisation sont les mêmes

O : Oui le travail a été fait et ça reste une transcription, une transduction, mais ça n'empêche qu'on a quand même beaucoup retravaillé. De toute façon c'est quelqu'un qui ne s'arrête jamais. Tu l'appelles demain tu lui dis « on refait 3 jours de mix » il y va. Il ne s'est jamais arrêté. Quand il est parti il n'était pas content parce que ça ne s'arrête jamais, ce n'est pas grave.

Moi comme je te le dis je ne suis pas, et de loin, le grand spécialiste de la spatialisation du son au cinéma. D'ailleurs j'avais fait un mémoire sur la spatialisation au cinéma tu peux retrouver à Louis Lumière dans leur archive, ce n'était pas intérressant.

[...]

Et en fait tu voyais finalement qu'il y avait une variance et ça rejoint ce que je te dis aujourd'hui, finalement ça n'a pas tant bougé en 29 ans, en fait il y a des scènes qui se prêtent très bien à ça dont la spatialisation devient un vrai enjeu et sert la scène et d'autres ou ça ne sert vraiment à rien parce que vis-à-vis du rapport à l'objet cinématographique il n'y a pas d'enjeux quelque part, et un bon mono il y aussi ça. La spatialisation ça veut dire aussi mobilisation de l'attention qui te détourne des enjeux d'un film par exemple, en tout cas de la substantifique moelle dramaturgique, cinématographique de l'objet de cinéma. Le fait d'être en mono parfois ça t'apaise les sens et ça te permet de te concentrer et c'est peut-être ça la priorité. Tu as beau être dans un film Dolby Atmos tu vas mixer en mono et ça va être génial. Après tu profiteras d'autant plus d'un effet de contraste sur des scènes qui le méritent. Tu peux d'ailleurs avoir tout un film en mono mixé en Dolby Atmos et tout à coup toute une scène de dingue pendant 5 min mixée en Dolby Atmos et c'est dément. Il ne faut vraiment pas être dans ce truc-là de « je suis en Dolby Atmos donc il faut en mettre partout », ce qui est intéressant c'est tout le potentiel dynamique que ça offre. Et c'est en gros les conclusions que j'avais eu il y a 20 ans dans ce mémoire-là.

A12.5 Entretien avec Augustin Muller

V : [Présentation du mémoire]

A : Est-ce que tu as une délimitation esthétique dans ton travail?

V : Non, je ne me limite pas à une esthétique. L'idée de mon mémoire et de travailler sur tous les milieux qui pratique l'audio spatialisé. Jusque-là je distingue tout de même deux visions qui ont l'air de se confronter. La première semble utiliser le spatial dans un contexte de musique actuelle, dans la création de contenus à destination du streaming, et utilise l'immersif comme une extension de la stéréophonie. La seconde rejette le paradigme stéréophonique et considère l'espace comme un nouvel espace d'expression à part entière, qui se doit de donner lieu à de nouvelles créations. Quelle est ta pratique de l'immersif et ta vision de tout ça ?

A : Dans les choses que tu décris je me rapproche plus de la deuxième situation. Je pense que le travail de l'espace sonore est un espace expressif en soi et que ce n'est pas une extension de la stéréo, c'est autre chose. Ça vient aussi du type d'esthétique dans laquelle je travaille. Et j'ai des réserves sur le mot « immersif ». C'est un mot valise pour faire certaines choses mais il y a quelque chose d'un peu totalisant avec ça. Il y a différentes façons de travailler la spatialisation sonore qui soient justement immersives ou pas. Ce n'est pas justement l'immersion qu'on recherche avec ça. Mais c'est peut-être un autre débat.

Je travaille comme réalisateur pour des pièces de musique électroacoustique et électronique et pour un répertoire qui traite de l'espace sonore depuis très longtemps. J'ai passé beaucoup de temps à faire du Stockhausen, à travailler du Boulez, à travailler du Xenakis, et ce depuis très longtemps. C'est quelque chose que je travaille en tant que musicien électronique et en tant que réalisateur et il y a dans ces musiques là une conception de l'espace à l'origine de la création et je pense que ça marche mieux quand c'est comme ça. Et moi, en tant que créateur sonore, c'est une chose qui m'intéresse évidemment et que j'essaye de travailler. Alors pour le coup on pourrait dire que c'est immersif mais plus en termes d'espaces acoustiques ressentis, en termes de lieux par exemple, ou de ressentis internes de l'espace acoustique plus qu'en termes de spatialisation : « à tiens j'ai telle ou telle source et je vais la mettre à tel ou tel endroit » ; même si bien sûr, de temps en temps, on fait ça. Voilà, je dirais que je ferais plutôt partie la deuxième catégorie.

V : Tu poses la question du terme immersif, qu'une création spatialisée peut être immersive ou non. Qu'est-ce que tu entendes par là ?

A : Sur l'aspect de son immersif il y a un truc qui à la fois m'intéresse et me pose question qui est toute la question de la déréalisation, « on va donner l'impression », de l'aspect simulation. Un bon exemple ce sont les musiques fictives. J'en ai fait une avec Aurélien Dumont sur *L'autre fille* de Annie Ernaux, on est parti sur un truc pas très immersif... Mais quand même un peu, très recréation d'acoustique et d'espaces mentaux liés à ces acoustiques, donc un truc que t'as plutôt envie d'entendre dans un studio très calibré ou en tout cas très silencieux parce que c'est des micro-détails hyper minimalistes. Mais d'un autre côté, je me rends bien compte que quand tu fais un spectacle vivant, que t'es dans une salle avec des performeurs qui sont là et un public qui est là, que tu vas écouter un solo et un récital d'András Schiff de piano à la Philharmonie, ça va être super immersif alors qu'il est tout seul sur scène. En fait ce mot-là il

est synonyme de diffusion du son spatialisé. Mais je suis d'accord qu'on n'en a pas vraiment d'autres donc c'est pour ça.

Moi je fais du spectacle vivant, dans un endroit vivant, où on est, où les gens sont là et forcément l'aspect d'immersion au sens réalité virtuelle, en fait ça ne marche pas. Pour moi c'est une fausse piste de se dire « ah on va essayer de mater le plus possible les salles, avoir un système de haut-parleurs les plus réguliers possibles ». Mais en fait il y a toujours un sweet spot, il y a toujours autant de perception que de public, il y a toujours une salle qui vit, il y a toujours l'intentionnalité du regard, tu as toujours ça. Donc c'est un truc que j'essaye de prendre en compte aussi, que ce ne soit pas une immersion abstraite. C'est ce qui vient un peu de l'IRCAM et du modèle du spatialisateur. Quand tu ouvres le spat tu as le *listener* au milieu et tu crées tout ce paysage pour lui. Mais en fait la situation d'écoute ce n'est pas ça, c'est un autre truc. C'est pour ça que j'ai des réserves sur le mot mais ça vient aussi du fait que je fais du spectacle vivant.

V : Tu parlais des musiques fictions où vous avez pris le temps de faire des acoustiques, tu parles de l'immersion du piano dans une salle de concert. Pour toi l'immersion c'est une acoustique ?

A : Oui ça peut venir par l'acoustique mais pour moi le mot d'immersion signifie que tu es dans un environnement dans lequel tu n'es pas réellement. Par « immersion » j'entends « illusion », et je pense que c'est un terme problématique dans le sens où : est-ce qu'on a besoin d'être dans un autre endroit que là où on est ? Mais si on met de côté cet aspect-là, oui je pense que pour reconstruire une image... enfin justement c'est pas une image ; pour reconstruire un champ ou une acoustique ou une impression de composition spatiale qui soit réaliste, l'aspect de l'acoustique à mon avis c'est super important, et ça on est pas les premiers à l'avoir fait.

Si tu regardes Pierre Henry, il ne faisait pas du surround spatialisé, il jouait énormément avec l'espace et les différentes façons d'exciter l'espace. C'est vrai pour la musique électroacoustique, de créer un univers acoustique qui soit autour de toi. Et je pense qu'effectivement la diffusion spatialisée a servi historiquement à faire naître ces images mentales, et ce depuis hyper longtemps. Ce qui vient à l'esprit c'est du Stockhausen où il y a quasiment tout le temps une multi diffusion, de plusieurs sortes, mais en gros l'idée c'est que tu as un système de haut-parleurs loin du public mais tout autour de lui, qui joue fort et qui fait résonner le lieu. Déjà de la musique d'orgue, si tu vas écouter un orgue dans une cathédrale ou une église, tu as un instrument spatial déjà

Oui à mon avis tu as quand même un truc important dans la relation entre le son, l'acoustique, le point d'écoute, les choses que tu vois, les choses que tu ne vois pas et le fait d'être dans quelque chose. Et du coup le mot immersif pourquoi pas de ce côté-là.

V : Si je comprends bien, finalement, dans une salle à quoi bon faire de l'immersif parce que ça vit mais...

A : Alors non ça ne veut pas forcément dire ça, ce n'est pas « à quoi bon », on peut faire de la diffusion dans l'espace spatialisé, mais pour moi je pense qu'il y a une tendance – mais je sais pas encore où on en est – une tendance à la virtualisation, à se dire « ah ouai je vais recréer ça là ou transformer ça » et je me rends compte que, plus ça va, plus je suis sensible aux

interactions entre lieu réel et... Dans l'ESPRO¹¹⁰ on a fait pas mal de concerts. On a pu avoir l'espro tout mat et des reverbs artificielles. C'était cool, mais c'est quand même une salle très particulière, tu ne peux pas faire ça partout. En tout cas pour de la musique vivante, ou pour une installation c'est peut-être un peu différent, ou pour une pièce électroacoustique, mais pour la musique live il faut établir une relation entre la présence des gens là où ils sont et là où ils sont mentalement. On ne peut pas complètement dissocier les deux, c'était ça que je voulais souligner.

V : En revanche dans un contexte d'écoute très égocentré avec sweet spot pour une seule personne, là ça peut mieux marcher.

A : Oui, ça marche même plutôt bien. J'ai fait des choses avec Le Balcon, des spectacles au casque, on a fait des choses ambisoniques... Je trouve ça génial et effectivement le fait que tu aies un système bien calibré dans un espace très maîtrisé te permet de reproduire des champs. Ça vient peut-être de mon éducation CNSM avec la prise de son au couple, mais je suis assez convaincu qu'il y a une différence de perception assez forte entre des champs acoustiques reconstitués de manière artificielle, en replaçant des points, et des champs pris avec un système cohérent, que ce soit un couple, une tête artificielle ou un micro ambisonique. Pour moi la différence, en général, j'arrive à reconnaître. Tu vois c'est surtout en termes de profondeur, même pas en termes de discrimination des positions en degré, c'est plus une qualité globale de profondeur qui marche hyper bien, qui est impressionnante et qui donne plein d'autres trucs.

Je ne sais pas si tu sais, j'ai pas mal bossé sur des reverbs impulsionales 3D, j'ai fait une banque avec le compositeur Pedro Garcia Velasquez. On a fait une résidence tous les deux de recherche artistique à l'IRCAM. Tu peux taper sur internet « théâtre acoustique convolué »¹¹¹. En fait on a échantillonné plein d'endroits qui ne sont pas des salles de concert justement, qui sont des lieux un peu particuliers, soit pour leur acoustique, soit pour leur caractère. On a fait un château en ruine, ce genre de truc. On a pris des réponses stéréo, binaurales, et à l'Eigenmike. Et on a fait une installation qui s'appelle *Étude pour Théâtre Acoustique*. L'idée c'est qu'on a fait des petits tableaux qu'on a présenté ça à l'IRCAM et au ZKM. Ça a un petit peu tourné, on a fait des petits tableaux électroacoustiques à partir de sons enregistrés parfois en chambre anéchoïque, ou d'autres trucs très sec qu'on a mis en situation avec des acoustiques dans plein d'endroits. Et en fait on s'est aperçu que si tu avais le Field Recording correspondant à ton empreinte et que tu la mettais ça marche hyper bien. Tu mets ta voix anéchoïque dans la reverb de je ne sais pas quoi, de la gare de Karlsow et tu mets la petite ambiance que t'as prise à l'Eigenmike, l'illusion elle est instantanée. C'est impressionnant.

Donc on a fait pas mal de choses comme ça. Après j'ai beaucoup travaillé là-dessus justement, sur *L'autre fille* avec Daniel et Aurélien, où on a mis la voix d'Annie Ernaux en situation dans des lieux qu'elle décrit. Donc de la mettre dans une chambre, c'était chez moi en plus pendant le confinement. J'ai fait des réponses chez moi, et oui effectivement l'impact de l'acoustique ou de la non acoustique, c'est-à-dire un truc très sec etc. par rapport à ce qu'elle est en train de dire fait que t'en as une perception qui est... En termes perceptifs c'est très fort, en termes de sens, la contextualisation ou au contraire l'assèchement ça marche très bien. Et d'ailleurs il y a une version binaire qui est horrible. Je trouve que ça marche pas du tout et

¹¹⁰ Espace de PROjetction, salle de concert et studio à acoustique variable à l'IRCAM

¹¹¹ <https://www.lieuxperdus.com/convolueur/>

je trouve que ça marche que dans une acoustique très contrôlée. Franchement la version binaurale n'est pas bien. Je ne suis pas sûr qu'on aurait réussi à faire la même chose en stéréo. On ne comprendrait pas ces variations. Le truc c'est que tu analyses trop ces variations de réverbération comme des variations de timbre quand t'as pas l'espace autour de toi. Surtout sur une petite reverb. Et du coup tu ne comprends pas, au casque, pourquoi ta voix est un peu masquée comme ça. Alors que quand tu es dans le dôme tu comprends parce que tu as cet effet acoustique d'holophonie. Pour le coup ça fonctionne vraiment sur l'idée d'illusion immersive.

V : Donc dans les paramètres qui interviennent en premier dans ta conception d'une image sonore il y aura la réverbération en premier lieu, ou l'acoustique de salle.

A : La réverbération en général. Et finalement, après, la position je m'en fiche. Pour pas mal de choses je m'en fiche. Après ça dépend de ce qu'on fait, mais pour moi, pour les contenus un peu immersifs je pense plus en termes d'acoustiques et de plans, c'est à dire loin proche. À mon avis c'est un nouveau truc donné par l'ambisonique que je n'arrivais pas à faire aussi bien avant, c'est à dire proche de moi ou loin de moi. Je pense même que si tu utilises des field recording avec des gens, il y a une question de proxémie. Je ne sais pas si tu connais ce concept-là. Si tu regardes les choses qu'a faites Isabelle Viaud-Delmon à l'IRCAM par exemple. Elle bosse plus sur les aspects sociaux. Tu as la sphère publique, loin de toi, et puis tu as les choses interpersonnels et puis tu as carrément l'aspect intime. Entendre un son très proche de toi ça ne t'arrive pas tous les jours.

Je pense que la diffusion ambisonique est le seul endroit, avec la diffusion binaurale au casque, où tu peux arriver à avoir un mode d'expression qui fait ça, qui fait qu'il y a des sons qui font partie du contexte. Par exemple même une voix elle peut être très contextualisée, pas très proche de toi, faire partie d'un paysage, vraiment très loin ou alors tout d'un coup elle peut devenir très proche de toi. Pour moi c'est intéressant quand je travaille en mode « immersif » entre guillemet. Et c'est pour moi plus intéressant que les questions de trajectoires azimut élévation... C'est cool en fait mais pour ça tu n'as pas besoin d'un système ultra immersif. Tu fais comme Stockhausen, tu fais un son qui fait une spirale les gens ils deviennent fous parce que c'est cool mais pour moi ce n'est pas forcément immersif justement.

V : Tu as la considération de remplir l'espace à 360° ?

A : Pas forcément, ça c'est un autre truc qui est pour moi hyper important et qui est justement véritablement lié à la situation d'écoute. Le truc de se dire « on va avoir 360° en azimut homogène » c'est vrai uniquement si tu es dans le cadre d'une installation dans laquelle tu ne contrôle pas où sont les gens, mais de mon point de vue, c'est très personnel. Quand tu fais ça, tu n'as pas plus d'expression, tu as moins d'expression parce que c'est comme si au cinéma il y a plus de hors champ. Pour moi, quand j'entends une voix qui vient de derrière moi on sait que psycho-acoustiquement ce n'est pas la même chose, et on sait que ce n'est pas la même perception par rapport au sens, par rapport au type de reflexe que tu vas avoir. Et quand j'ai fait des projets de musique mixte, il y a quand même une directionnalité de l'espace puisque tu as quand même une sorte de scène. Alors on peut discuter de comment est-ce qu'il faudrait la mettre cette scène, mais en réalité cette directionnalité fait que tu as un devant et que tu as un arrière. Il y a des choses que tu vois et des choses que tu ne vois pas, du coup ça donne un autre statut. Quand tu vas déplacer un son pour le mettre dans un endroit que tu vois, du coup

il a une qualité en plus, et quand on fait un truc très immersif au sens un peu commercial du termes où il n'y a plus de directionnalité, c'est cool mais en fait tu neutralise un peu tout.

J'ai l'impression de perdre un effet de palette finalement. Et pour *L'autre fille*, au début, j'ai été un peu nazi, je n'en avais rien à taper je mets tout le temps la voix au centre, elle ne bouge jamais sauf à des moments où elle est derrière. Et avec Daniel on s'est rendu compte que la configuration scénique – c'est pour ça que c'est hyper lié à la scénographie – fait que le spectateur est assis un peu dans tous les sens, donc on a un peu cassé cette directionnalité-là. Mais selon moi c'est un petit peu à regret parce que je pense qu'il y a des choses qui ont un peu perdu de leur force. Après c'est vrai que dans ce cadre-là, où des gens peuvent être dans n'importe quel sens, c'est cool de ne pas privilégier un sens. Après sur des questions plus de mix musique où on va équilibrer des choses pour qu'il y en ait partout, oui je pense qu'on peut le faire bien.

Après je me rends compte que dans l'espace à plat, compte tenu de ce que j'ai dit il n'y a pas vraiment de hiérarchie. Si on parle d'un truc complètement différent qui a été l'accompagnement pour Richie Hawtin, qui n'était d'ailleurs pas merveilleux... Le seul truc qu'on a fait c'est de se dire qu'il y a des zones qui sont à plat, qui sont égales et riches en timbres et qui peuvent être perçues de manière claire, et qu'il y a peut-être des zones plus particulières qui sont le haut et le très bas. Bon. Mais je n'ai pas plus de règles que ça. Je pense que c'est quand même vraiment lié à des questions de scénographie. Je pense que si tu as quelque chose à voir à un endroit, ton espace n'est pas homogène. Il n'y a pas d'homogénéité de l'espace parce que tu as forcément un point qui est attiré et à ce moment-là Front Left ce n'est pas pareil que Back Left. Et est-ce que c'est immersif ça ? Est-ce que c'est 360 ? Je ne sais pas, t'as du son partout mais il n'a pas le même statut.

V : Avec ton ensemble Le Balcon vous êtes toujours dans un contexte de musique frontale ?

A : Pas tout le temps, ça dépend

V : Et de diffusion multipoint ?

A : Ça dépend, mais oui. C'est un ensemble qui s'est créé il y a 15 ans et qui a tout de suite été conçu comme sonorisé. C'est un ensemble sonorisé donc il y a un ingé son, qui est le fondateur de l'ensemble, et très vite on a commencé à faire de la multidiff, très vite on a commencé à faire du mix objet sans savoir qu'on faisait du mix objet. Et on travaillait un répertoire qui nous amène à travailler sur des idées de spatialisation, en tout cas ça fait partie de notre langage, mais c'est un ensemble de performeurs. Il y a toujours des musiciens. C'est rarissime qu'on ait fait des trucs où on ne voit pas les musiciens, donc on fait quand même le rapport entre le geste et le son. C'est vraiment important.

V : Donc là l'homogénéité de l'espace sur scène est liée au fait que les musiciens soient visibles. Comme ça on les regarde.

A : Il y a un statut de l'espace qui est différent. Il y a des compositeurs qui ont eu des délires comme ça. J'ai fait des pièces de Nunes à l'Ircam où lui, s'il avait pu, il aurait mis en live les musiciens dans une chambre anéchoïque derrière le plateau, on les voit juste et on les entend. Genre t'as le violon devant toi mais il sort de l'enceinte de derrière. Stockhausen il fait

ça aussi dans *Michelstreise*, il replace tous les instruments dans l'espace par rapport à un point fictif et il fait un truc assez marrant avec un arbre à micro. Au milieu de *Michelstreise*, le trompettiste il joue dans des micros qui sont placés dans les coins de la salle. Du coup quand il fait des mouvements de trompette t'as des mouvements géants.

Alors c'est vraiment cool au sweet spot, mais si tu n'es pas au sweet spot tu as du tuba ultra fort dans une enceinte pendant 40 min. Donc il y a un côté de praticité qui n'est pas très bien géré. Ce sont des trucs qu'on a faits mais je n'ai pas plus de stratégie, enfin j'ai des stratégies artistiques sûr qu'est-ce que je veux donner comme sens à la spatialisation. Et je me rends compte qu'elles sont plus en termes de profondeur et d'équilibre dynamique que vraiment de position.

Ensuite, en ce qui concerne le mixage pur et dur qui serait peut-être extension de la stéréophonie, c'est ce qui s'est passé avec Hawtin. En fait il s'est passé deux choses. Déjà c'est très étanche, il est venu une semaine et c'est très difficile de rentrer dans son workflow, on ne peut pas vraiment. Il n'a pas de conception de l'espace au départ de sa performance, donc c'est hyper lent et le niveau où on arrive est hyper faible, même si on a des bons outils. Quand tu n'as pas cette conception-là, tu te retrouves à faire le même truc que tout le monde, c'est-à-dire que t'as des sources qui tournent dans un sens et dans l'autre et des points fixes. C'est sympa mais bon.

En termes de mixage, les choses que je retiens c'est que tu as des évidemment des éléments structurels que tu veux pouvoir avoir dans de bonnes conditions de timbre, de niveau, de phase. Donc il y a des trucs que tu pourras faire ou ne pas faire. Forcément dès que tu vas avoir du mouvement, et notamment du mouvement vers le haut, un peu rapide, spectralement faire bouger des graves au-dessus de ta tête c'est quand même dur. Je n'ai pas encore réussi à entendre ça. *Rire* Je ne sais pas si c'est physiquement possible, donc forcément ce serait quand même plutôt des éléments avec un centre spectral plutôt haut. Il y a des règles de psychoacoustique qui font qu'on fait tous à peu près le même genre de trucs au bout d'un moment.

Mais parfois je ne sais pas si ça apporte grand-chose. L'autre jour à l'IRCAM il y a eu plusieurs trucs [en immersif, ndlr] et à la fin, Ryoji Ikeda était en stéréo. Franchement c'était hyper bien, et c'était hyper bien spatialement. Je me suis dit « bon ». On l'a d'ailleurs fait jouer en salle sèche et il trouvait ça trop sec et il a demandé qu'on remette pas mal d'acoustique. Pour donner cette immersion en fait alors qu'il jouait en stéréo.

Je sais que Frédéric Changenet a pas mal travaillé ça. Bon ils n'appellent pas ça musique fiction mais c'est pareil pour la radio. Il a fait des choses comme ça et je sais qu'il se pose vraiment ces questions de comment on écrit l'espace pour un mixage. Je pense que les compositeurs traditionnels de la musique électronique ont quand même répondu à pas mal de questions. Il y a pas mal de stratégies à l'œuvre là-dessus, il y a pas mal de gens qui ont réfléchi à ça. On n'a pas parlé d'Olga Neuwirth. Elle a fait une pièce assez importante à l'IRCM qui s'appelle *Encantadas* dans laquelle ils ont utilisés justement beaucoup de reverb 3D en live, c'était une des premières fois où on faisait ça et c'était une pièce qui faisait référence au *Prométéo* de Luigi Nono. Elle est allée sampler l'acoustique de San Lorenzo, c'est cool.

Donc par exemple cette réflexion sur le sens d'un lieu, l'acoustique, en lien avec la position, c'est quelque chose qui a été traité, qui existe, donc je pense qu'il y a des stratégies qui sont là mais c'est vrai qu'elles ne sont explicitées nulle part.

V : Pour revenir à Hawtin, qu'est-ce qui a fait que ça n'a pas marché ?

A : C'est 1000 trucs, on a fait une première date qui n'était vraiment pas bien. J'étais vraiment furieux. On en a fait une deuxième qui était un peu mieux. Si tu veux lui il a fait son truc. Il a fait son truc mais en termes spatial il n'y avait pas grand-chose d'intéressant. Je veux dire ça sonnait pas mal, il y avait du son, il y avait du son un peu partout, voilà. C'était ce que c'était. En termes d'écriture de l'espace il n'y avait pas grand-chose. Mais ça ne s'invente pas. Nous à l'IRCAM on a une expertise de ça mais c'est très difficile avec des artistes superstar comme ça, qui ne te connaissent pas, de leur dire « eh Ritchie tu ne connais rien à ça je vais t'expliquer comment on fait ». Ce n'est pas possible en fait. Donc ensuite il veut bosser uniquement avec ses outils qui sont *Bitwig*, sur lesquels nous... Enfin bref il y a plein de trucs qui ont fait qu'on ne pouvait pas aller aussi loin, que ce n'est pas très satisfaisant. Mais peut-être qu'il reviendra et qu'on fera un autre truc. Je pense qu'effectivement il y a peut-être une question d'outils et de workflow qui font que ce n'est pas super simple mais là il était très accueilli et il pouvait faire beaucoup d'autres choses.

Je pense qu'il y a aussi une question de culture et d'écoute de ça. Par exemple, quelque chose que je n'ai pas fait c'est de lui faire écouter de trucs. Et en réalité les gens n'écoutent jamais. Nous on est des méga geeks. Je passe ma vie à faire ça mais en fait personne fait ça. Là si je veux aller écouter un truc ambisonique ordre 4 je peux y aller tout de suite et si je veux écouter un truc ordre 7 je demande à Jonathan quand est-ce que je peux squatter l'ESPRO. Ce n'est pas très compliqué mais pour la majorité des gens ce n'est pas possible, y compris des gens comme lui. Donc je pense qu'il y a une question de manque de connaissances. Donc c'est bien que tu effectues un travail comme ça, ça donne de la connaissance et un support un peu théorique sur lequel s'appuyer pour faire ça.

Après nécessairement pour moi il y a un aspect naïf d'utilisation de la spatialisation. Ça je le vois partout, parce que des trucs immersifs il y en a dans tous les sens. De toute façon toutes les boîtes poussent à mort, et il y a 100 trucs super naïfs. Ce n'est pas grave c'est cool mais ce n'est spatialement pas plus intéressant que s'ils l'avaient fait en stéréo. C'est juste qu'il y a un peu du son partout. Mais on a des outils pour le faire et c'est pas mal. Je pense que quand même on manque d'outils, ou en tout cas de workflow d'écriture, ou on a des choses qui font sens qui vont arriver mais je pense que c'est aussi une question de formation des ingénieurs son. Pas au sens apprendre des nouvelles techniques mais au sens de la grammaire du mix. Les références du mix stéréo on les a, ça a une histoire, une esthétique aussi. Là, pour l'instant l'histoire esthétique de la spatialisation je pense qu'elle existe dans le monde de la musique contemporaine européenne mais à part ça c'est ultra niche.

Peut-être que ça va se développer aussi. On va faire beaucoup de naïveté là et puis il va y avoir des trucs plus intéressants. Ça vient après des premières expériences. Si pour toi ou pour moi spécialistes de la spatialisation on trouvait super décevant le truc qu'a fait Richie, même lui il était super content de ça. Parce que c'était la première fois qu'il le faisait, donc du coup lui il trouvait ça trop cool, « j'ai des sons partout, je peux avoir des mouvements ». Même s'ils sont très simples c'était déjà super cool pour lui.

V : J'ai eu la même expérience avec une amie qui a trouvé le concert génial et je n'osais pas trop dire que je ne l'avais pas trouvé extraordinaire. J'avais l'impression d'être rabat-joie, peut-être que j'ai pas vécu la même chose.

A : Mais tu n'as pas vécu la même chose.

V : Il y a la question de l'expérience d'écoute.

A : Ou des attentes. Le truc qu'on peut quand même garder, c'est qu'en termes d'action expressive, Richie, ce n'était pas très intéressant. Il n'y avait pas d'actions expressives là-dessus. Un peu plus le deuxième soir parce que je lui ai un peu rentré dans le chou. En revanche, en termes d'expertise de l'immersion sonore, c'étaient quand même des flux HOA ordre 7 avec une Reverb à l'ordre 7, c'était quand même bien fait.

Ça je pense que Frédéric, on en a parlé ensemble, en est super conscient. On a les outils technologiques pour aller loin, et pour faire en sorte que ça marche. En revanche les outils d'écriture et la grammaire de l'écriture, pour l'instant elle n'existe pas. Enfin elle n'existe pas... Elle n'existe pas dans le monde qui n'est pas la niche Stockhausen, Boulez, Harvey etc.

V : Qu'est-ce que tu entends par grammaire ? Quelle différence avec le vocabulaire ? Qu'est-ce qui existe en stéréo qui n'existe pas en spatialisé à ce niveau-là ?

A : Je pense que les ingénieurs ont plein de manières de traiter la stéréophonie, la gestion des reverbs, la gestion du pan, où est-ce que tu mets les instruments, quand est-ce que tu triches sur la largeur de certains instruments ou pas, il y a plein de techniques sur comment utiliser la stéréo. Qu'est-ce que tu vas laisser vraiment très mono et pourquoi, et on le sait. Et on le fait naturellement. On sait que quand on met une voix à gauche toute seule au début avec du silence et qu'on fait rentrer le reste ça ressemble à un disque des Beatles. On le sait. On sait tel truc, un piano pris hyper près est ultra large, ou alors on sait que quand tu vas à Radio France ils font des couples de 4 km et ça fait des images comme ça, tu le sais, ça existe, il y a des styles. Prise de son classique « à la », prise de son Rock « à la ». Ça on l'a. Comme il n'y a pas ça, toute cette culture là pour le son immersif, du coup les gens tâtonnent un peu plus. Et quand tu tâtonnes tout le monde commence par faire un peu les mêmes trucs. Mais je pense que ça va venir, il y a de plus en plus de Dolby Atmos et compagnie.

A12.6 Entretien avec Antoine Petroff

A : J'ai commencé par de l'immersif, quand j'étais encore étudiant, pour gagner des sous. Je m'étais fait embaucher par une boîte de sono et à l'époque ils faisaient principalement des installations de gros systèmes pour des raves en Suisse. Sur le moment j'installais des systèmes où on ne mettait jamais juste un système stéréo. C'était dans des halles gigantesques, super réverbérantes. Pour essayer d'avoir un son correct on multipliait les points de diffusion, toujours au minimum 8 points, voir beaucoup plus. En revanche le contenu était évidemment stéréo. Et après cette période rave, comme par ailleurs avec des potes on faisait de la musique et que ça m'intéressait de commencer à mixer, j'ai commencé à faire des petits lives avec mes potes musiciens. Je trouvais que ça sonnait pas du tout, j'avais l'habitude des gros machins devant moi [dans les raves, ndlr], et là tu te retrouves dans des petits restos.

Et en parallèle j'étais à l'uni, comme ingénieur. Parce que je m'entendais bien avec le prof d'acoustique il m'a mis sur un projet avec des potes à lui, des vieux de la vieille du free jazz. Ils bossaient sur un nouveau projet d'installation sonore immersive. Le but, ça devait être installé dans un jardin avec une zone centrale avec les gens qui passaient dedans, et dans tout le jardin ils voulaient qu'il y ait du son. L'idée c'est qu'au centre on ait l'impression de rentrer dans une grotte et dans un passage, il fallait vraiment avoir l'impression d'être ailleurs que le visuel, et qu'à l'extérieur l'environnement sonore soit immersif et que les gens se rendent pas du tout compte qu'il y ait des haut-parleurs et qu'il y ait des choses sonores qui ne devraient pas être là. Et c'est là d'ailleurs que j'ai fait mes premières enceintes, parce que la meilleure façon de rentrer dans une grotte... Mettre des haut-parleurs ça ne va pas marcher. C'est là que j'ai contacté le mec qui fait les enceintes PSI, et c'est avec lui que j'ai fait mes premières enceintes. Ma première enceinte faisait 8m de long, il y avait 20 haut-parleurs large bande dedans et c'était un couloir dans lequel les gens rentraient et on avait fait des haut-parleurs pour l'extérieur. C'étaient des pots de fleurs avec des full range et on en avait mis un peu partout.

En fait en faisant ça je me disais que la stéréo ça n'avait aucun intérêt, et ce qui était vachement intéressant c'est qu'à la fois au niveau de l'impression purement sonore... Typiquement quand tu mets des haut-parleurs aléatoires, t'es vraiment dans un environnement sonore, dans une masse sonore, tu es incapable de dire d'où viennent les sons c'est vraiment comme dans la réalité. Et à l'inverse dans le couloir, pareil, vraiment tu étais à 1 mètre de l'entrée du couloir tu avais très peu de son parce que c'était un Line Array proche super directif. Il était à hauteur d'oreille, un de chaque côté, caché plus ou moins, et vraiment quand tu rentrais dans ce truc tu étais dans un autre univers. Et je crois que c'était mixé en 8 zones autour de ça. Et ce qui était génial c'est que sur ce projet les musiciens avaient des idées hyper précises mais il n'y avait personne pour m'aider avec ça. Spat Revolution à l'époque ça n'existe pas.

V : C'était en quelle année ?

A : En 1999.

V : Même aujourd'hui c'est rare d'avoir des musiciens qui ont une vision de l'espace.

A : Justement j'ai énormément appris avec le contrebassiste du groupe. Il y a une partie des tubes qu'on entend à la radio c'est lui qui a fait la musique, il est devenu connu dans les années 60 parce qu'il a fait le groupe yéyé Suisse qui a marché dans le monde et après il s'est lancé dans le Free Jazz et il a vraiment des idées musicales et sonores hyper précises. En revanche la techniques... Mettre un jack il va le mettre dans la prise électrique. *Rires* Et j'ai vraiment appris avec lui parce que tout partait de l'intention et j'ai eu des grandes discussions. J'essayais de me convaincre que la stéréo c'est bien parce que tu as une image qui est la réalité mais ça n'a aucun sens parce que même quand tu prends une photo et que tu regardes quelque chose, de toute façon ce ne sera pas la réalité, tu n'as pas les odeurs, tu n'as pas les trucs qui vont avec. Lui son idée c'est que ça ne sert à rien de vouloir faire un instantané. En revanche c'est de recréer des émotions, c'est l'émotion qui est le principal et après tu réfléchis, avec le son, comment recréer ça. Mais de vouloir recréer notre réalité ça n'a pas de sens

V : Comment ils avaient travaillé leur musique Jazz pour que ça prenne place dans cet espace ?

A : Justement c'était vachement intéressant en plus il y avait plusieurs concepts.

Le premier : utiliser aucun traitement artificiel, tous les sons on est allé les enregistrer dans des lieux qui avaient des acoustiques hyper particulières. On était allé dans une grotte par exemple. On avait passé des nuits à enregistrer dans des bâtiments à Lausanne c'était complètement dingue. Mon premier projet et c'était juste génial, et ils avaient réussi à avoir un partenaire à la radio suisse allemande qui avait pas mal de moyens et ils avaient un studio mobile à dispo. À l'époque c'étaient des DAT, c'était encore plus le bordel qu'ici [en parlant de son studio, ndlr] et à des prix pour moi inaccessible. Génial. On avait enregistré en multipiste à cette époque-là et le mix avait été super compliqué. C'est là où j'avais découvert Max MSP, on s'est dit « on va mixer pour ce truc-là. »

V : Vous avez enregistré directement en multi micro ?

A : Non, la plupart on avait quand même pris en son proche avec plusieurs micros stéréo. Un des trucs qui marchait pas mal : plusieurs paires de micros à des distances différentes. Son idée c'était vraiment de marcher dans la grotte, au niveau narratif qu'il se passe des choses, des choses qui apparaissent et des choses qui disparaissent. Et donc avec ça il y avait moyen vraiment de créer de l'éloignement en montant progressivement la paire qui était la plus loin possible, ce genre de truc.

V : Il y avait une vraie volonté de jouer avec des acoustiques réelles, de les transcrire dans une acoustique qui n'est pas la même, de jouer avec une acoustique présente.

A : Comme c'était à l'extérieur, tu n'as pas d'acoustique réelle, en revanche tu as des sons réels. Typiquement à l'extérieur ce qui marchait bien c'est que tu avais des sons artificiels qui se mélangeaient avec des oiseaux etc. et au bout d'un moment tu as l'impression que les sons artificiels répondent aux sons naturels.

V : Ça me fait penser au Carré Magique de Music Unit à Royaumont.¹¹²

¹¹² <https://www.youtube.com/watch?v=7ucuGnTl2nQ>

A : Typiquement, pareil, on avait fait une installation avec Thibaut à la Villa Médicis à Rome où c'était de la musique générative avec des sons d'espèces d'animaux et des trucs comme ça. Et il y avait des paons qui se baladaient à la villa qui réagissaient avec la bande son, et comme elle était aléatoire il se passait des trucs complètement dingues de temps en temps.

V : Si on veut résumer, ta vision du son immersif c'est faire disparaître le haut-parleur.

A : Ce qui résume tout c'est faire disparaître le haut-parleur, parce qu'à partir du moment où la technique disparaît, ce sont les émotions qui prennent justement. Et après les émotions ça peut être par un hyper réalisme, si tu y arrives, ou par l'abstrait. Mais dans tous les cas ton système n'apparaît pas dans ce que l'auditeur doit percevoir.

V : Ça fonctionne dans des environnements naturels, mais dans l'ESPRO, dans des lieux plus musicaux ?

A : Alors faut en tenir compte. Malheureusement j'en fais beaucoup moins maintenant, mais en Suisse je suivais pas mal de groupe pour du Live, et c'était super varié. Autant de la musique contemporaine c'est écrit pour. Je me souviens on avait créé un collectif qui oscillait entre métal noise et punk et pourtant je me trimbalais tout le temps avec des enceintes pour faire de l'immersif. Parce que simplement, ton mur de guitare, au lieu d'avoir juste une façade c'est distribué dans plein de haut-parleurs, relativement frontal là pour le coup, c'est juste magique. Pareil pour les guitares, quand tu es en studio, souvent j'enregistre avec plein d'amplis différents pour avoir des couleurs. Quand tu es dans le studio tu as une masse sonore démentielle, t'écoutes ça sur tes enceintes stéréo, pffff...

Il y en a qui arrivent à retranscrire ça le mieux possible mais bon. Pareil en live, pourquoi ne pas faire la même chose et y compris utiliser les amplis sur scène comme sonorisation ? Avec ces groupes là il y avait leurs amplis, les enceintes façade, il y avait des endroits où on allait, je rajouter toujours quelques haut-parleurs autour pour rajouter ça. Ça marche mais il faut tenir compte de la salle. Typiquement un 5.1 pour faire ça n'a pas d'intérêt parce que ce n'est pas conçu pour ça.

V : Dans quel sens ?

A : Avoir un central n'a pas d'intérêt, beaucoup plus intéressant d'avoir des amplis n'importe où sur la scène qui te donnent une espèce de centre plutôt qu'une colonne qui te fait un central.

V : Est-ce qu'il ne faut pas que ce soit une relativement petite salle ?

A : Non pas forcément on a fait des festivals avec 1000 personnes, ça fonctionne aussi. J'avais été super intéressé par Queen of the Stone Age. Le niveau sonore si tu as que les amplis sur scène avec le public, tu es déjà quasiment à 100dB donc le niveau sonore des amplis fait vraiment partie du son. Les mecs ils tiennent compte de ça et ce n'est pas pour rien que tu te prends une baffe sonore dans la gueule. Il y a des musiques qu'y s'y prête quoi.

V : Là on reste sur des choses assez frontales.

A : Alors typiquement sur ce truc là j'avais aussi des arrières, souvent, que je ne dirigeais pas sur le public mais que je dirigeais n'importe où pour tous les effets de reverb, delay, ce genre de choses. Ce sont des trucs sur lequel j'ai beaucoup bossé chez Devialet, c'est que les reverbs c'est un des trucs les plus difficile à rendre quand t'as un système ponctuel. Parce qu'en réalité une reverb courte c'est justement le fait que ça vient de toutes les directions que ça fonctionne. Le côté métallique qui vient des reverb et même des très bonnes reverbs c'est aussi en partie du fait que tout est limité à 2 canaux. Là pour ton cerveau, même si t'as 2 sons qui sont proches, tu n'as pas un filtrage en peigne qui se fait parce que ton cerveau il interprète ça comme deux sons différents, là où dès que t'imprimes, que tu bounce entre guillemets, ta notion de spatial disparait, donc t'as une sommation en fréquences, donc tu vas avoir des filtres en peigne, tu vas avoir des phénomènes qui ne sont pas réalistes.

À l'époque, là pour le coup pour des plus petits scènes, je m'étais fabriqué des guirlandes de haut-parleurs full range qui pouvaient faire un peu de niveau avec une dizaine par guirlande. J'alternais un A – B – A – B etc... Dans la salle, dans le public et autour en fonction des salles, je foutais mes guirlandes et mes reverbs je les envoyais là-dedans. Et de fait, vu que c'est que 2 canaux et en multipliant les sources complètement aléatoires tu arrivais du coup à faire disparaître les HP. Et pour certains trucs très frontaux, pour certains éléments précis, souvent je mettais 2 haut-parleurs en plus dirigés contre les murs qui te créent une source sonore beaucoup plus grande.

Ça c'est aussi un des aspects intéressant si tu prends un piano par exemple, il est pas du tout ponctuel, même à une certaine distance il a une largeur qu'un haut-parleur unique ne te donne pas. Alors que si tu prends un haut-parleur et que tu le diriges contre le mur, tu as un spread de fait qui te donne l'impression que la masse est plus grande. C'est pour ça que j'ai rarement utilisé des systèmes standards parce que c'est en fonction des besoins. C'est toujours des espèces de trucs hybrides en partie fait sur mesure.

V : Tu as la volonté de reproduire un champ sonore qui pourrait se rapprocher d'un champ réel.

A : Justement ce qui m'a intéressé dans le fait d'être ingénieur pour pouvoir faire de l'ingénierie du son, c'est pour avoir des systèmes qui se comportent acoustiquement comme dans la réalité. Et après ce que t'envoies dedans là c'est l'émotion qui prime. Si je veux une source qui soit large, j'ai envie d'un système qui va me permettre d'avoir un truc large, si je veux un truc diffus, ça m'intéresse d'avoir ça.

V : Et si tu ne peux pas mettre des haut-parleurs partout ?

A : Tu joues sur des effets de délais, des trucs comme ça. Et il y a certains trucs que tu ne pourras pas faire et puis voilà. Tu en tiens compte, musicalement, tu sais que c'est une impression que tu ne vas pas donner et ce n'est pas grave.

V : Tu favorises l'idée de coupler véritablement la pièce avec la diffusion.

A : C'est super intéressant d'utiliser la pièce. Le nombre de fois où j'arrive dans des salles, je mets une enceinte que je faisais taper dans le mur. On allait manger, je revenais et ils avaient retourné l'enceinte en mode « qu'est-ce qu'il fait il y a un con qui a fait ça. » Non c'est voulu.

Franchement il y a plein d'endroit où les gens me prenaient pour un débile et après quand ils entendaient le résultat « tiens c'est intéressant. »

V : C'est finalement travailler sur l'ASW et le LEV, la reproduction d'une sensation d'auditeur dans un espace acoustique réel.

A : C'est ça. C'est vrai que la démarche, je formalise en réel, c'est en fait d'avoir un système qui peut recréer des conditions physiques d'une réalité. En revanche ce que j'injecte dedans va être purement guidé par l'émotionnel. Mais c'est pour ça que mixer du Atmos à la chaîne je ne peux pas parce que ça m'emmerde et que je préfère complètement mixer pour des trucs bizarres.

V : Pourquoi ça t'emmerde ?

A : *Réfléchis* Parce que c'est trop limité. Même du mix standard je n'en ai pas tant fait que ça parce que ça m'emmerde vite en fait. Justement je préfère nettement le live.

V : La sensation d'espace et d'enveloppement ça se joue sur plusieurs sens et pas que l'audition.

A : Complètement

V : Est-ce que ce n'est pas le live qui s'y prête le mieux justement ?

A : Oui, et le problème c'est que oui l'écoute binaurale au casque, ça marche plus ou moins. Même un système ici qui pour moi est limité, là c'est Atmos¹¹³, qui a ça chez lui ? Personne. Si on diverge, typiquement ce que j'avais essayé de faire chez Devialet, c'est comment rendre cette impression d'espace sonore avec une seule enceinte, avec un seul device. Plutôt qu'essayer de recréer des conditions où tu as un son qui vient de là et d'autres qui viennent de là etc. où en fait ça ne marche pas, c'est plutôt de se dire « certains sons je m'en fous, que la guitare soit là que le piano soit là etc... tant qu'il y a un truc qui est cohérent », de se dire « tous les sons qui sont précis tant pis de toute façon finalement quand tu as un même device ils vont venir du même endroit donc eux on les garde. Ce qui est des sons plus large, genre un mur de gratte ou un truc comme ça, comment on rend la sensation qu'ils sont plus large, qu'ils ne viennent pas d'un point mais qu'ils viennent d'une masse, et tout ce qui est reverb que tu donnes l'impression qu'ils sont partout. » L'idée c'est de décomposer les mixs en fonction de ces 3 critères et après le device diffuse dans l'espace les sons, avec du traitement en fonction de ça.

Typiquement : une voix ou un truc précis, tu l'envoie dans un haut-parleur frontal qui vient vers toi. Pour une masse plus large, ce sera plutôt un haut-parleur qui permet d'utiliser les murs d'une façon limitée pour te donner l'impression truc plus large. Les reverbs sont envoyées complètement aléatoirement dans toutes les directions de la pièce et c'est la pièce qui recrée cet enveloppement.

Chez Devialet on tenait compte de la musique, tu fais une séparation du mix en fonction de ces critères et tu les reproduis différemment en fonction de la nature du son Et de mon point

¹¹³ Mentionne son studio à Paris équipé d'un système Dolby Atmos.

de vue le résultat est beaucoup plus intéressant si tu respectes ça mais à la limite tu ne respectes pas le fait qu'un instrument soit à gauche, droite etc... et t'as un truc beaucoup plus réaliste.

V : Sans modification du travail au mixage.

A : Non, l'idée c'est de garder ça et d'en faire le rendre mieux. Dans un mix par exemple si tu rajoutes de la reverb c'est pour séparer les sources au niveau de leur nature quoi. Alors tu as le cas basique où tu as une voix, tu mets de la reverb sur la voix, pour donner l'impression que c'est une voix dans une salle et pas toute seule, mais par exemple tu as des cas où tu as une reverb à ressort qui a une sonorité particulière c'est pour donner une nature acoustique qui est différente.

De nouveau, toujours, j'étais peu intéressé par le fait qu'au final tu aies une émotion globale qui soit plus intéressante. Enfin intéressante ce n'est pas le bon mot, plus impactante que d'avoir un truc ultra réaliste où tu peux avoir une source qui n'est pas exactement à un endroit donné. Donc y compris quand je concevais des systèmes de live pour des projets, le système peut être complètement disparate et pas du tout homogène et pas ressembler à du Dolby Atmos et voilà, en fonction de ce que j'ai envie de rendre.

V : Et l'émotion dans le mixage, par quoi ça se traduit ? Est-ce que c'est le placement des sources ?

A : Ça je pense, à force par l'expérience tu peux associer certains types de placements à une émotion donnée, par exemple c'est vrai qu'un cas particulier, tu mets une voix, juste une voix, à un endroit, tu vas donner une impression de fragilité. Et à l'inverse la même voix si tu as une reverb longue partout autour tu peux donner une impression de puissance, tu changes presque la narration, juste en rajoutant une reverb. A l'inverse tu mets une reverb courte, tu vas avoir l'impression d'avoir une personne un peu fragile perdu dans un espace clos. En résumant ça va plus être ça

Les guitares électriques c'est vachement intéressant, alors ce n'est pas universel, mais entre guillemets tu as une largeur vraiment... L'impression d'un mur qui t'arrive dessus et si tu la dépasse trop tu perds ce sentiment-là, tu auras plus la puissance, ça devient un enveloppement, tu auras plus un son secondaire. Typiquement un son qui doit avoir un impact, il sera très souvent frontal mais surtout il aura une largeur qui sera limité.

V : Frontal dans un contexte de live ?

A : Ou même au mix. Tu vois même un mix comme ça **montre une session ProTools**, des sons qui doivent avoir un impact vraiment primordial, c'est rare que je les envoie partout, justement ils vont être dans une zone, mais pas non plus dans un seul haut-parleur, à part les voix peut-être...

V : Tu parles de quelque chose de percussif ? De rythmique ?

A : Quel que soit son rôle en fait.

V : Quand il est précis.

A : Non pas forcément précis, mais qui soit puissant, en tout cas qui soit du premier plan mais pas forcément plan acoustique mais premier plan narratif, ou musical etc.

V : Tu vois je pense au son du vin par exemple¹¹⁴, tu n'avais pas fait bouger le kick un peu partout ?

A : Le kick ça m'étonnerait, mais peut-être, après toutes les règles sont... *Rires* Ça c'est l'autre truc que le contrebassiste m'a dit : « toutes les règles sont là pour être... » Et c'était drôle les premières fois où je bossais avec lui il m'a posé des trucs hyper... Et voyant qu'on s'entendait bien il me disait « les règles que je donne là, si tu me surprends en faisant l'inverse vas-y quoi elles sont là pour ça aussi. » *Rires*

V : Et typiquement les morceaux de Romain Barthélémy [qui a également travaillé sur le son du vin, ndlr] qui ont la même esthétique, comment tu travailles quand tu spatialise ses productions ?

A : Justement, dedans, ce que j'ai tendance à faire c'est de différencier ce qui va être des sons ponctuels, donc qui racontent entre guillemets quelque chose, et des sons qui vont être là plutôt pour la texture. Des sons de texture très souvent je vais plutôt avoir tendance à les mettre partout, ou dans des zones beaucoup plus larges, et jamais je vais les mettre sur un haut-parleur parce que je trouve que ça ne marche pas. Ou si tu prends de la musique électronique, les nappes, c'est rare que je les mette comme un mur de guitare électriques, c'est rare que musicalement pour moi les nappes soient là pour mettre de la puissance, elles sont là justement pour créer presque une acoustique musicale. Les nappes de synthé je vais très souvent les répartir sur tous les haut-parleurs en mettant des délais pour perdre la notion du truc, ça devient un espèce de truc abstrait. Dans les sons de Romain il y a certains sons vraiment traités comme ça, d'autres des sons plus en tant que tel mais qui restent dans des zones hyper définies. Et tu as d'autres éléments hyper ponctuels et ceux-là j'aime bien les faire interagir avec des pièces, enfin des rooms, un environnement acoustique ajouté mais qui peut être une reverb ou qui peut être des délais ou qui peut être du flanger avec des délais enfin ça dépend.

V : Par rapport au mix stéréo, est-ce qu'il y a un vrai gain expérientiel et de facilité ?

A : Alors complètement. Déjà je pense que dans les mixs immersifs j'utilise quasiment plus de compresseurs sauf pour un son particulier, j'utilise 10 fois moins d'EQ et c'est 10 fois plus ludique. Et ce qui est intéressant surtout ce que j'ai pu remarquer avec des musiciens qui n'ont pas du tout l'habitude de mixer, quand ils essaient de mettre du mix stéréo très vite ils abandonnent. Alors qu'à l'inverse un mix immersif, tu laisses la main, tu reviens une heure après, tu dis « bon je te laisse faire. » *Rires* Ils ont une meilleure idée que toi, musicale, qu'ils veulent et ils s'en sortent très vite, donc ouai c'est incomparable, vraiment incomparable.

Et en live c'est pareil. Quand tu utilises un L-Isa même en frontal c'est juste génial parce que juste pour séparer les sources tu t'embêtes fais carrément moins pour mixer, ta voix, elle est là et t'as pas besoin de la traiter

¹¹⁴ Je fais référence ici à un projet Ircam Amplify sur lequel nous avions travaillé avec Antoine.

V : Martin Antiphon disait en relativisant qu'il fallait quand même réfléchir à comment donner de la puissance par exemple.

A : Alors plusieurs trucs qui se passent sont intéressant dans la stéréo. Le mastering il est intéressant quand il ajoute une glue mais en fait très souvent il pourri le mix qui a été fait, mais souvent le mec sait ce qu'il fait... Donc c'est fait d'une façon intéressante, donc il se passe un truc intéressant qui est dur de retrouver en immersif parce qu'il n'a pas les outils et dans tous les cas ça n's'y prête pas vraiment.

Après l'autre aspect auquel il faut faire gaffe à pas vouloir mettre des machins partout c'est qu'en effet pour rendre de la puissance on en revient quand même à la façon dont nos oreilles sont faites. On est quand même fait pour être plus attentif à ce qu'il se passe devant. Donc clairement si tu veux rendre de la puissance... Pour m'amuser j'ai essayé d'être à un concert de rock et à un moment de regarder derrière, ça ne marche pas. *Rires* Même en fermant les yeux ça ne fonctionne pas. Est-ce que c'est par habitude ? Je ne sais pas. Effectivement si tu veux de la puissance je pense qu'il faut que ce soit frontal. Moi ça m'arrive souvent, que ce soit live ou que ce soit comme ça, certains éléments, de les mixer en stéréo parce que justement une basse batterie passé dans des compresseurs de mastering stéréo ça a une couleur qui est intéressante. Donc euh ça c'est un autre aspect en tout cas. Je ne mets jamais de règles, de nouveau ça m'est égal.

V : Tu n'utilises pas tes notions de mixage stéréo en fait. Tu t'en défais complètement.

A : Je m'en défais mais je les connais donc des fois ça peut être marrant de les intégrer dedans.

V : Parce qu'en plus tu reçois des stems stéréo.

A : Oui mais avec les stems tu fais ce que tu veux, c'est quasiment du multipiste s'ils sont bien séparés. Après quand je te dis « si t'as une basse batterie je mets toujours en frontal », j'ai aussi essayé autrement. Le premier truc que j'ai fait c'est de les mettre derrière, de les mettre partout. Par expérience c'est vrai qu'envoyer un kick partout ça ne marche pas. Après est-ce que c'est aussi parce que l'interaction avec la pièce réelle dans laquelle t'écoutes fonctionne pas ou est-ce que c'est parce que ça ne marche pas d'envoyer du kick de partout d'un point de vue purement perceptif je n'en sais rien. Je n'ai jamais eu l'occasion de le faire en chambre anéchoïque et je ne le ferais certainement jamais. Mais en tout cas dans des écoutes standard je trouve que ça ne marche pas.

Et un autre truc intéressant c'est que ça ne marche pas pour une batterie acoustique et ça ne marche pas non plus avec batterie électronique, donc ce n'est pas qu'une relation avec ce que t'as l'habitude d'entendre parce que la musique électronique il n'y a pas d'aprioris réalistes dans ce que tu fais donc c'est là où tu te rends compte qu'il y a les mêmes tendances en électronique qu'en acoustique.

V : Est-ce que quand tu travailles les mixages immersifs, tu es avec l'artiste ? Tu échange avec lui ?

A : Ça dépend. J'aurais tendance à dire que la plupart du temps non. **Hésitations** J'aime bien m'mauser justement. **Rires** Après parfois ça ne correspond pas à leurs intentions, ça arrive, en revanche bien sûr on discute avant.

V : Ils ont une vision de ce qui se passe en audio spatialisé ?

A : Ça dépend. Il y en a qui s'en foutent complètement, y compris en stéréo. Ils ont enregistré leur truc, puis il y en a d'autres qui ont une vision très précise de ce qu'ils veulent et dans ce qu'ils ont de très précis, si tu ne leur donnes pas ce qu'ils attendaient, ça ne leur correspond pas, et d'autres tu leur donne complètement l'inverse et que ça sonne ça leur va aussi. **Rires** C'est aussi pour ça que ça m'emmerde de faire des trucs à la chaîne, si je veux le faire, parce que je veux m'amuser avec le contenu quoi. Donc juste faire que ça sonne pour que ça sonne bien à la radio je ne suis pas bon pour ça. **Rires**

V : On parlait du mouvement, est-ce que tu joues beaucoup avec les mouvements ?

A : Je joue pas mal avec les mouvements qui ne sont pas forcément perceptibles. Je m'étais créé pas mal d'automatismes pour faire bouger les sons, par exemple comme ça ou aléatoirement, qu'ils se déplacent de manière complètement aléatoire. Un truc que j'aime bien faire avec les délais, par exemple, j'en ai 5 ou 6 différents et je les paramètre de manière aléatoire. Ça te crée une espèce de masse sonore autour. Donc oui j'utilise les mouvements mais moins pour faire tourner un truc que tu vas percevoir. Après ça dépend avec Romain on avait fait des trucs comme ça. Si tu veux un truc démonstratif effectivement ça marche toujours d'avoir un truc qui tourne, après ça devient vite ennuyant. C'est vrai, c'est pour ça Jean Michel Jarre c'est cool les deux premières minutes mais au bout d'un moment, avoir le même truc qui tourne ...

V : Moi je l'ai plus ressenti avec Richie Hawtin, avec Jean Michel Jarre j'avais moins ce souvenir.¹¹⁵

A : L'autre truc, ça c'est plus pour le binaural, c'est que d'avoir une écoute fixe, dans la réalité, t'as jamais ça ; donc avoir en permanence des trucs qui bougent autour de toi ça participe au réalisme en particulier quand après tu fais un downmix binaural.

V : Ça n'arrive jamais dans la réalité parce que l'auditeur bouge la tête ?

A : Si tu mets un head tracker sur ta tête et que tu regardes tes mouvements, même si tu regardes ton ordi etc. Tu es tout le temps en train de bouger et tu seras toujours en train de bouger, entre autres pour la perception de ce que t'as devant et derrière et pour la notion d'espace, ou simplement pour des sons que tu n'as jamais entendus. Faut savoir que normalement les HRTF ça marche dans ton cerveau que si tu as un son que t'as déjà entendu. Imagine que tu es dans une chambre anéchoïque, on te fait écouter un son que tu n'as jamais entendu de ta vie à un angle particulier, tu auras beaucoup de peine à savoir où il est, simplement parce que ton cerveau, ce n'est pas relatif, parce qu'il a aucune idée si c'est une source qui a un timbre particulier qui est filtré par tes oreilles ou si c'est une source qui est

¹¹⁵ Je fais référence ici aux deux concerts spatialisés à l'ESPRO de la saison 2023/2024 de l'Ircam, celui de Jean Michel Jarre en janvier 2023 et celui de Richie Hawtin en février 2024.

comme ça. En revanche dans la même chambre anéchoïque, si tu bouges un peu la tête, ce n'est plus en absolu que ton cerveau fonctionne, c'est en relatif. Et il sait que le timbre bouge d'une certaine façon et tout de suite il sait que ça va être là. Naturellement on bouge tout le temps la tête parce que c'est ce qui nous permet entre autres d'avoir une notion spatiale, et de le refaire de la même façon en mixage ça marche super bien.

V : Et le binaural tracké qu'on a dans des airpods ?

A : Ça va dans le bon sens après c'est mal fait *Rires* C'est mal fait pour des mauvaises raisons parce que leur head tracking lui-même est quand même un peu foireux, mais bon ils font avec les limitations de l'objet, et après on est limité en nombre de canaux, pour que ce soit réellement efficace. Le head tracking, je crois qu'il est fait sur de l'ambisonique d'ordre 1, ça n'a jamais été très clair leur truc. Tu as un nombre de canaux très réduit. Donc ce n'est pas que tu fais bouger vraiment toute ta scène, là c'est que tes x haut-parleurs virtuels que tu bouges. Si tu fais du Head Tracking avec un pseudo 5.1 ou si tu fais du head tracking avec 20 sources dans l'espace tu n'as pas du tout le même rendu. Mais après, les trucs binaural, on est qu'au début. Parce qu'aujourd'hui il y a plein de limitations qui font que ce n'est pas encore super net.

V : Tu penses que ça devrait tendre vers quoi ?

A : Déjà il faudrait un meilleur head-tracking. La notion des HRTF personnalisée, qui changent quand même un peu la donne. Ça t'amène un réalisme, un naturel supplémentaire, d'avoir plus de canaux, et idéalement il faudrait un format avec certains canaux censés être pour des sources ponctuelles, et certains canaux pour tout ce qui doit être environnement diffus.

V : C'est un peu le bed et les objets non ?

A : Oui et non. Tu vois ça m'irait très bien d'avoir mes fichiers... T'as deux canaux, tu mets dedans tout ce qui doit être diffusé partout dans la pièce. Parce que si tu devais le faire bien il faudrait idéalement, je sais pas 32 canaux juste pour ça, mais finalement ça sert à rien d'avoir 32 canaux parce que le but d'avoir 32 canaux c'est d'avoir un truc complètement diffus, à l'inverse tu gardes juste 2 canaux et c'est ton renderer là qui va te recréer ton côté diffus ou en fonction du nombre de haut-parleur et qui va le faire en fonction du système de diffusion, son but c'est juste qu'il y ait un espèce de diffus qui t'entoure, et ça, ça marche on avait des tests.

Je pense qu'il y a une façon aussi de repenser la notion de canaux qui te permettraient de limiter le débit tout en offrant un truc plus réaliste. Et d'une certaine façon tu peux imaginer : tu as un ambisonique d'ordre... En théorie d'ordre 7, ou tu es plus capable de faire la différence entre la réalité, donc ça fait quand même un bon nombre de canaux. Mais tu peux imaginer un ordre plus faible pour tout ce qui est ponctuel autour de toi et quelques couches de trucs qui devraient être bordeliques, ou plusieurs trucs bordeliques comme ça t'as un renderer qui peut les traiter différemment pour augmenter le réalisme.

Surtout en décorrélant le fait qu'un canal est égal à soit une source soit un haut-parleur, pour moi ça c'est une des erreurs fondamentales dans la façon dont on pense le son. Il faut plutôt penser en termes de couches en fonction de l'intention acoustique que tu veux rendre.

Typiquement pour moi ça n'a aucun sens d'encoder de la même façon une reverb, un piano, ou un instrument. Imaginons une voix centrale : tu la mets en deux canaux ; un piano : tu la mets en 2 canaux et une reverb : tu la mets en 2 canaux, c'est absurde parce que tu utilises la même quantité de données et la notion spatiale n'intervient pas là-dedans.

V : C'est ton prochain projet de développement ?

A : J'ai abandonné parce que se battre contre Dolby... Malheureusement ces gens-là sont pas intéressés par l'émotion que tu vas rendre aux gens. Mais je pense qu'il y a quand même des trucs comme ça qui vont évoluer. Et en particulier je pense que les jeux-vidéo, et ce genre de choses, sont beaucoup plus avancés que la musique par rapport à ça, justement parce qu'eux ont un vrai but de réalisme, d'impression. Du coup ils vont optimiser ce machin pour ce but-là. Alors qu'aujourd'hui la plupart des choix sont fait par des ingénieurs qui ne sont ni musiciens ni ingénier sont et qui font ça pour des questions purement techniques.

V : Après le jeu vidéo ils sont plus avancés mais grâce à cette idée de son à l'image pur.

A : Oui mais de fait, aujourd'hui tu prends l'Atmos, qu'est-ce qui a guidé le choix que le bed ? il est en 7.1.4 ? Bon. C'est juste des habitudes passées de la stéréo puis du 5.1. Mais là, entre guillemets, Atmos c'est encodé en ambisonique. Enfin d'une certaine façon, avoir un truc où c'est vraiment la notion d'espace qui prime ça pourrait être plus intéressant. L'intérêt d'avoir du bed et des objets etc. ce sont juste des problématiques purement techniques.

V : Après le rendu ambisonique est quand même particulier.

A : Alors le rendu ça c'est encore autre chose mais en théorie en ambisonique t'es capable de dire « une source elle vient de là et elle a cette largeur ». Donc normalement en ambisonique tu peux donner l'impression, encoder un haut-parleur qui serait directif vers toi comme tu peux encoder 15 sources différentes qui viennent plus ou moins de cette direction, si tu as l'ordre suffisant.

C'est un des trucs que j'ai pu observer. Plein d'ingénieur aujourd'hui utilisent de l'Atmos, et tu leur parle de la notion de bed et d'objet ils sont déjà un peu pommés, ils ne savent pas trop ce qu'il faut mettre dans l'un ou l'autre, je n'ai pas encore eu trois personnes qui me répondaient de la même façon et aucun qui est vraiment convaincu par sa réponse. Oui l'objet il est fait pour que vraiment il soit optimisé en fonction du système que t'auras au rendering mais alors pourquoi ce n'est pas le cas pour le bed aussi ?

V : Tu favorise quel type d'algorithmes ou de rendus dans tes projets ?

A : En fait la plupart du temps en rien du tout c'est format en fonction du projet. Typiquement tu reprends le premier projet d'installation sonore sur lequel j'avais bossé, le mix c'était un format... Je crois qu'il y avait 16 canaux pour le truc central, 8 canaux pour l'extérieur, sachant que les canaux extérieurs ils étaient dispatchés sur 64 HP je crois. Après là en ce moment, le truc **fait référence à sa session Pro Tools** c'est un enregistrement dans les catacombes. Le but faut que ce soit en Atmos donc je le mixe en Atmos mais dans les derniers trucs pareils j'avais fait quelque chose pour un défilé où il y avait 24 HP dans un escalier. Tu vois c'est un format qui... Je bosse plus souvent avec ce genre de formats qu'avec un truc fixe.

V : Et en élévation, si t'as le choix tu vas travailler avec ?

A : Oui parce que ça te permet d'élargir le truc. C'est vrai que l'élévation te permet ça. Après honnêtement de mettre des objets particuliers vraiment en hauteur je ne l'ai jamais utilisé comme ça. Le seul cas, pour Devialet, j'étais allé enregistrer la fusée Ariane qui décolle en Guyane et l'idée c'est de reproduire la sensation que t'avais. Pour le décollage j'avais vraiment utilisé l'élévation pour un son ponctuel parce que fallait vraiment donner l'impression que le son venait d'en haut. Mais je crois que c'est le seul cas où j'ai vraiment fait un truc réaliste spécialement. Et typiquement il y a certains sons qui se prêtent bien à se mettre en hauteur, je vais plutôt mettre des sons aigus en hauteur parce que c'est là que tu vas percevoir mieux l'élévation. Si je le mets en hauteur c'est que j'ai envie qu'on le perçoive qui vienne de là-haut. Même si c'est inconscient parce que c'est le bordel complet, ou des sons qui sont par exemple des délais ça marche très bien en hauteur. Si tu veux rajouter un délai qui ne perturbe pas le reste là tu peux utiliser l'élévation ça marche vachement bien. Y compris quand tu utilises un délai mais qui n'est pas rythmique, décorrélé niveau tempo mais qui ne perturbe pas trop c'est une bonne façon de séparer.

D'ailleurs, ça aussi, mes fameuses guirlandes ça m'arrivait aussi de les mettre en haut et dans les pieds des gens. Quand t'as du son qui vient d'en bas ça marche vachement bien. À l'inverse tu peux donner les basses et les bas mediums. Est-ce que c'est des a priori ? Je ne sais pas mais en tout cas de les mettre au sol ça marche vachement bien.

V : Ce n'est pas fréquent d'avoir des enceintes en bas.

A : J'avais fait ça pour des instal' mais autrement non. C'est dommage parce que ça t'apporte d'autres trucs. En tout cas ce qui m'a toujours intéressé et c'est pour ça que je suis parti pour être ingénieur, c'est de maîtriser suffisamment la technique pour que quand c'est dans un processus créatif je l'oublie complètement. Dans un processus créatif la technique m'ennuie et je n'ai pas envie de m'en soucier.

V : Et le public est friand de tout ça ?

A : Oui. Pourquoi je m'étais dit que l'immersif c'est vraiment l'avenir ? le truc le plus frappant, je me souviens, c'est quand ma fille qui a débarqué quand elle avait environ 10 ans, elle m'a dit « écoute ça on écoute tous ça à l'école ». C'était le fameux truc « 8D » machin, les gens qui s'amusent à prendre un truc stéréo et à le faire tourner avec un truc binaural basique. Mais n'empêche elle a entendu tout de suite la différence. Et à la différence du cinéma 3D où je pense le problème... Pour moi c'est une aberration de vouloir te dire que tu crées la vraie 3D visuelle mais tu la gardes dans un cadre comme ça, ça n'a aucun intérêt, alors que le principe du son immersif c'est que tu recrées les conditions réelles d'écoute.

V : Et est-ce qu'il y a un manque de démocratisation et d'apprentissage du public ?

A : Alors d'apprentissage du public je ne pense vraiment pas. Je n'ai jamais eu d'expérience du public qui disait « mince ce n'est pas comme d'habitude c'était moins bien ». Jamais. Alors vraiment au contraire, toujours c'est « waouh ». En revanche il y a un vrai souci au niveau technique et financier parce qu'effectivement un pur Line Array Stéréo coute moins cher qu'un

système immersif quel qu'il soit. Après ça me fait toujours rire, tu regardes les prod' de musique, de live. Le pourcentage que représente le coût du système, globalement ce n'est pas prévu dedans, comme les salles qui te disent « ça coute trop cher on ne va pas installer un système immersif ». Quand tu regardes le coût d'exploitation de la salle versus en absolu ce que ça pouvait leur rapporter, si tu réfléchis et que ça te permet de rapporter d'autres public. Il est complètement négligeable. Le problème c'est que la plupart des gens qui prennent les décisions importantes, où que ce soit qui que ce soit dans la société, ce seront les financiers qui n'y comprennent rien. Donc ils ont un regard qui est évidemment nécessaire mais le problème c'est que les décisions ne sont jamais prises de manière intelligente par rapport à ce que le public va ressentir.

Alors maintenant on se retrouve dans des guerres de formats... Franchement le format Atmos c'est une merde sans nom. *Rires* Il a le mérite d'exister mais bon... Et quand Frédéric Amadu¹¹⁶ avait discuté avec celui qui a créé l'Atmos Cinema tu te rends compte que l'Atmos musique maintenant c'est exactement ça, c'est un machin qui a été créé dans un but très précis. Et il y a des financiers qui sont arrivés et qui ont dit « bon les droits du mp3 existent plus donc sortez nous le nouveau machin qui va nous rapporter des sous » et c'est n'importe quoi, c'est mis en œuvre n'importe comment et c'est récupéré par Apple qui ont fait leur merde à leur sauce et aujourd'hui effectivement tu as des ingénieurs qui disent qu'ils n'y comprennent rien, que ça ne marche pas etc. Donc tu as un vrai souci aujourd'hui de format, les outils simples n'existent pas et c'est vrai aujourd'hui, hormis le binaural, tu as aucun moyen pour un musicien dans sa chambre à couper pour mixer en immersif.

Donc ouai il y a encore un sacré travail pour démocratiser tout ça. Mais en revanche au niveau du public je ne pense pas. Public binaural à nouveau oui, et à raison parce que c'est vrai que l'immersion binaurale, ce que j'ai pu tester, c'est que sur la durée, par exemple t'écoutes 2h de son binauralisé spatial, c'est beaucoup plus agréable que d'écouter du stéréo. Mais tu n'as pas un effet « waouh » très perceptible au début, et tu as plein de merdes aussi. Sony qui avait ressorti Space Odity en spatialisé, je ne sais pas qui a fait ce truc là mais ce n'est pas possible.

Donc j'espère que ça va finir par se développer mais c'est vrai que ce à quoi le public a accès n'est pas toujours très qualitatif, et pas pour les bonnes raisons. Les créateurs ils n'ont pas les outils pour et il manque un canal vraiment simple pour diffuser ça.

Quand la stéréo s'est développée, ça n'appartenait à personne et c'était suffisamment simple pour ne pas poser de brevet sur le format en termes de canaux. Là aujourd'hui, là où je pense qu'ils sont vraiment très bêtes tous ces gens c'est qu'ils ne se rendent pas compte qu'à force de faire leur petite guerre, tous ils pourraient gagner beaucoup plus de pognon et ils sont en train de tout foutre en l'air, de limiter l'impact que ça peut avoir, c'est juste bête.

V : Après l'Atmos est déposé mais l'ADM est libre.

A : Oui l'ADM est libre mais dit à un ingénieur qui a fait l'effort acheter un système etc. « - Bon maintenant faut créer un ADM. - Euh ok je fais comment ? » et puis imaginons qu'il

¹¹⁶ CTO d'Ircam Amplify

y arrive et qu'il a son ADM et que le musicien lui dit « mon label veut m'envoyer chez Apple Music je fais quoi ? »

Alors déjà qu'en stéréo ce n'est pas simple... Là où je trouve, justement, là où c'est un vrai problème avec l'immersif, c'est autant avec le mixage, le travail de production s'est démocratisé. Tu prends le début des années 60 les mecs étaient en blouse blanche, il y avait des règles à respecter et tu allais en studio. On a réussi à simplifier tout ça et maintenant avec un ordi portable n'importe quel musicien est capable de créer de la musique. Tu prends Billie Eilish c'est un super bon exemple. Et avec l'immersif on est reparti en arrière ou tu es obligé d'aller dans des studios, certifiés Dolby, et alors ça, c'est refusé parce qu'il faut mettre à -3dB et cocher la case machin truc. Alors que justement ça donne au musicien des libertés créatives de mixer en immersif. Pour l'instant les canaux derrières on est reparti 40 ans en arrière.

Après ce n'est pas simple, créer un format ce n'est pas évident. Après c'est sûr qu'il y a beaucoup de gens qui prennent des décisions là-dedans qui n'y connaissent rien et que leur but ce n'est pas d'améliorer l'expérience du public. Alors que justement le meilleur moyen de gagner des sous c'est ça.

V : Et les systèmes de diffusion immersif qui existent type L-Isa etc. ?

A : Ça c'est bien parce que ça se développe. Pour le live c'est mieux, pas besoin d'avoir un format universel entre guillemets. Après si tu prends le L-Isa, ça m'est arrivé justement de l'utiliser en frontal, je n'ai jamais mixé un truc qui est immersif dessus. Je pense que c'est pas du tout un truc que j'aimerais utiliser parce que c'est vraiment fait pour faire un cercle et l'auditeur est au centre. Et dès que tu veux t'éloigner de ça, ce n'est pas fait pour. Après justement ils ont fait un truc vraiment optimisé pour leurs utilisateurs et eux ils ont bien raison. Après tous ces systèmes marchent, si t'es en config rond ils marchent tous, peut-être plus ou moins bien.

Et par exemple un truc qui, aujourd'hui en live n'existe pas chez des L-Isa, c'est d'intégrer la spatialisation de la même façon que t'as un PAN. C'est d'intégrer dans la table de mixage la spatialisation. Ça n'existe pas. Il y a des consoles qui permettent de contrôler les L-Isa et autres mais ça reste de façon très technique. Et de façon générale, les nouvelles consoles de mix live je ne les aime pas du tout parce que ce sont des machines de labo et ce ne sont pas du tout des machines de live. C'est pour ça que j'aime bien la Allen & Heath. C'est que tu as peu de choses, les accès sont simples, même si tu as un gros groupe tu n'as pas besoin de ces usines à gaz. C'est ce que je trouve dingue, tu prends une grosse table de mix live, ça coute 120 000 euros, par contre ils veulent pas mettre 50 000 euros de son immersif alors que ça ferait beaucoup plus de différence.

V : Martin Antiphon avait fait un VAE pour donner accès aux artistes l'immersion dans le synthétiseur directement.

A : Alors il commence à y avoir des solutions de pan mais ce n'est pas intégré... Pour moi ça devrait être comme le pan, je ne vois pas de raisons à ce que ce soit plus compliqué que ça d'une certaine façon. Un pan et un curseur largeur, ou je ne sais pas en tout cas des macro-paramètres. C'est vrai que par ailleurs dans le son immersif, globalement, il y a plein de gens

qui utilisent des termes qui ne sont pas les bons, ça parle et ça n'aide pas non plus à démocratiser le truc.

Tu vois ça c'est super intéressant, les Acousmoniums c'est génial alors que ça va à l'inverse d'une écoute stéréo audiophile. C'est marrant la première fois que j'avais entendu parler de ça, la stéréo elle est faite pour être écouté sur deux haut-parleurs, et après on était allé voir Pierre Henry qui faisait la 10^e symphonie en Acousmonium et c'est une de mes plus grosses baffes au niveau son. Justement il se passe un truc, parce que c'est intéressant, j'ai eu des discussions sans fin avec des ingénieurs sur des enregistrements classiques : « ouai il faut que ce soit précis », alors que quand tu vas à un concert, la plupart du temps l'écoute elle n'est pas ultra précise. Et pour autant l'émotion elle est juste géniale et tu es vraiment englué dans le truc parce que le plan de la pièce il est fait pour.

Les Acousmonium c'est pas du tout réaliste, mais au niveau des émotions c'est beaucoup plus proche d'un concert acoustique parce que justement tu recrées une façon d'entendre le son qui est beaucoup plus naturel que sur un système stéréo. T'utilise plein d'enceintes différentes, parce que dans la réalité ce n'est pas comme ça. Je m'étais justement amusé à faire des tests, super bon enregistrement de piano. On l'écoutait en stéréo, et sur des enceintes qui séparaient les fréquences partout dans la pièce, 80% des gens ils préféraient l'enceinte éclatée dans la pièce.

Immersif n'est pas forcément lié à la 3D. En fait la 3D participe à l'immersion mais il y a plein de facteurs qui participent à l'immersion. Un exemple, qui m'avait frappé, j'avais fait des enregistrements de musique liturgique dans une église. Le chef d'orchestre c'est un mec super pointu. Je l'ai mixé pour lui et je lui ai fait écouter en studio. Mais il ne se sentait pas là-bas, « je n'y suis pas, il y a un truc qui ne joue pas. » J'ai réfléchi au premier truc que j'avais fait en rentrant. La voix la plus basse était à 100Hz donc j'avais mis un gros coupe bas à 80 pour virer tous les bruits de la rue extérieure qui passaient à travers, sans le dire je lui ai viré le filtre, et là il a dit « ok j'y suis, parfait, je suis dedans je suis là-bas. » En fait pour lui qui connaissait l'église, l'espèce de brouhaha infra basse faisait partie de son immersion à lui.

Et là après ça m'a fait réfléchir, très souvent les subs, les infras etc. participent à te donner cette immersion parce que justement c'est lié au fait que t'as pas cette notion spatiale mais t'as vraiment l'impression d'être dans un truc. Et l'immersif c'est donner l'impression d'être immergé dans un environnement. Donc ça c'est un truc important de bien définir ce que c'est l'immersion et après comment tu le rends, avec des moyens qui sont le son 3D, tu peux faire une impression immersive avec la stéréo, elle sera peut-être plus limitée etc. mais c'est important de bien définir les termes.

En neuroscience on est en train de tester l'impact d'une scène 3D fixe et avec head tracking, quelle est la différence de perception au niveau des patients et aussi au niveau du cerveau. Ce qui est sûr aujourd'hui c'est que typiquement un vrai son spatial active des zones de cerveau différentes que si tu écoutes un mono. Et à priori de la stéréo, mais la stéréo c'est une notion qui peut être très vague, en tout cas mono et 3D ça n'active pas les mêmes zones du cerveau.

V : Ce sont les mêmes zones du cerveau qui traitent les infos spatiales de la vue et du son.

A : La plupart des gens qui font un AVC, au niveau vision, ils perdent la notion gauche de l'espace. Si tu leur donne un livre, ce n'est pas qu'ils voient plus à gauche, ils voient la page de

droite mais pas la gauche, ou ils se rasant que le côté droit du visage et pas le gauche. Et certains mots, ils vont voir partie droite du mot mais pas la gauche. A l'inverse ils entendent bien le son. A priori, ils espèrent qu'avec des sons qui vont de la droite vers la gauche et autres mouvements, ils peuvent réactiver des zones du cerveau.

A12.7 Entretien avec Éric Raynaud

E : J'ai une position très particulière là-dessus [sur le son immersif, ndlr] en fait déjà à la base je ne suis pas ingénieur du son du tout. *Rires* Donc en fait j'aurais certainement une vue un peu plus singulière qu'Hervé Déjardin par exemple.

Moi j'ai une position un peu particulière parce que je suis à la base artiste et en relation avec l'audio spatial. Pour moi c'est d'abord un terrain, un endroit pour faire du créatif, c'est un terrain de jeu pour faire de la création. Et comme tout terrain de jeu on a besoin d'outils et pour moi l'outil ce n'est pas une question secondaire. C'est une question importante parce que c'est très imbriqué. L'outil peut influencer la pratique artistique mais on ne peut pas déconnecter selon moi les outils, le contexte de travail, et la pratique pour laquelle on va utiliser ces choses-là.

Ce qu'on voit beaucoup maintenant dans l'industrie c'est de vendre un software qui est, en gros, une extension d'un domaine stéréo et puis point barre. C'est l'ingé son qui va mixer tout ça et ça va être super. Pour moi ce n'est pas si simple que ça. En fait l'audio spatial, comment dire... C'est comme la VR, ce n'est pas juste une augmentation de l'image frontale. C'est un environnement qui appelle une pratique particulière, c'est un environnement qui appelle une recherche artistique, sur ce qu'on en fait. Pour introduire, c'est un peu situer mon point de vue là-dessus.

Déjà c'est quoi l'audio spatial. Parce qu'il n'y a pas vraiment de définitions très claires. C'est être capable de créer un environnement sonore qui peut se percevoir du point de vue du public. L'audio spatial c'est dehors dans la rue. Et on va dire que la VR c'est pareil, c'est tenter de reproduire des phénomènes naturels de perception sonore en environnement. Et c'est toujours un peu le réflexe de l'humain, c'est recréer artificiellement ce que l'on vit naturellement et on peut même appliquer ça à l'art visuel, et on peut y ajouter la partie artistique, c'est ça qui m'intéresse. C'est à dire qu'on tort la réalité pour la faire vivre autrement. On s'approprie du naturel et on le transforme pour la faire vivre autrement. Ça peut être du 3D, ça peut être quelqu'un avec 2 haut-parleurs qui joue avec les réflexions de l'espace, il n'y a pas une définition claire. C'est sûr que maintenant dans l'industrie pour les enjeux qu'on prend, on vend des soft, on vend de l'activité, donc on va essayer de définir l'audio spatial comme une nouvelle niche de l'industrie mais encore une fois je me tiens un peu à distance de ça.

Ce que j'ai découvert à travers ma pratique c'est comment on construit un espace sonore, qui on va dire, est immersif. Pour moi c'est une pratique artistique d'abord. Si demain on arrivait à une forme d'homogénéité de comment on designe un espace sonore, ça va devenir très chiant. En fait c'est bien que ça reste un peu artisanal quoi. On veut bien avoir des outils génériques mais ça reste artisanal.

Donc j'ai compris très vite qu'il fallait construire ses propres outils ou utiliser une gamme d'outils. Alors dans le son par exemple si tu fais de la prod', si tu utilises un compresseur tu ne vas pas utiliser les mêmes compresseurs pour ta batterie et pour la voix. Et puis après sur ton mix final tu vas utiliser d'autres types de compresseur. L'audio spatial c'est pareil. Encore une fois je ne vais pas taper sur l'industrie mais chaque boîte, que ce soit L-Acoustic, Amadeus, Spat Revolution et Flux ::, nous dit « voilà c'est l'outil ». Mais en fait quand tu veux travailler, faire du travail *meaningful*, qui a de la valeur, on ne peut pas se contenter de dire « c'est l'outil

parfait ». Tu me diras que l'ingé son ou même l'artiste doit avoir une flexibilité. Il peut y avoir des outils qui vont être utilisés différemment selon les contextes. Donc c'est sûr que l'apprentissage des outils il doit être global et il faut avoir une connaissance presque holistique, totale, des possibilités des outils au même titre qu'un bon ingénieur du son va connaître tous les types de compresseurs et comment ils fonctionnent et ce qu'ils peuvent faire.

Alors la troisième question intéressante pour moi c'est « est-ce que c'est une pratique artistique ? » Alors, je ne sais pas quel est ton point de vue. Est-ce que c'est une surcouche finale de mix ? Ça c'est intéressant, parce que finalement ce n'est pas l'ingé son qui va décider d'où on va. Il n'a pas cette légitimité. Si c'est juste positionner des objets sonores dans l'espace, on se prend n'importe qui. Et puis on joue sur la perception, sur l'EQ, les choses comme ça. Mais designer un environnement sonore, j'ai envie de dire qu'il n'y a pas de règles. Par exemple, ce n'est pas l'ingénieur du son qui va dire « tiens moi je vais créer un champ sonore granulaire, je vais faire de la synthèse granulaire 360. » Moi ce que j'aime beaucoup par exemple c'est utiliser une source sonore, la passer dans des effets et nourrir d'autres objets sonores partout. Transformer, créer de la matière sonore. Donc on se rend bien compte que quand on fait ça, ça doit être une pratique artistique. C'est du geste artistique. Ça ne veut pas dire que l'ingénieur du son, en étape finale, ne viendra pas normaliser tout ça pour des questions de diffusion, discuter sur les formats etc. Mais en tout cas,似ilairement que pour un album, la production d'un album ou d'un film c'est au compositeur de décider comment il veut remplir son espace sonore.

V : Avant tu parlais des outils. C'est pour ça que tu as développé XP¹¹⁷ ? Pour tes besoins spécifiques ?

E : Je suis un utilisateur de Max depuis longtemps, et de l'IRCAM Spat. XP, c'est une approche Covid. C'est né parce qu'il n'y avait plus de spectacles et que j'étais en résidence à l'IRCAM. XP est né de l'idée que, en gros, dans un espace tri-dimensionnel, si on aborde le son comme ça, pourquoi pas simplement simuler directement dans l'espace 3D notre scène sonore. Pour l'instant XP c'est un object based avec des points assez classiques, des choses que tu vas retrouver partout, mais à la différence d'autres soft c'est que j'ai mon environnement. C'est un environnement visuel 3D que je peux utiliser, que je peux agrandir avec d'autres fonctionnalités qui n'existeront pas ailleurs.

Par exemple, ça n'existe pas encore dans XP mais on pourrait très bien intégrer un modèle physique de simulation de vecteur de force, de gravité, implémenter des modèles physiques. Les propriétés d'un environnement Open GL, par exemple, c'est le cas. Je voulais combiner ce qui se fait dans le visuel et aller jusqu'au bout de l'idée puisqu'on travaille sur du visuel, on va utiliser les propriétés de ces outils là et voir ce qu'on peut en faire.

Pour l'instant c'est un travail de longue haleine, c'est un peu le module de base mais c'est né de ce truc-là. Et c'est né aussi d'un second truc, c'est qu'effectivement il y a des outils qui existent pour Max, pour d'autres... Encore que... *Rires* L'interface IRCAM dans le Max elle est quand même... Voilà. Et puis il y a les outils élites qui sont à destination des ingénieurs du son : Spat Revolution, L-Isa... Mais dans la communauté artistique il n'y avait personne. Et personne n'utilise ces outils d'élites. On ne va pas se mentir, personne n'utilise ces outils-là

¹¹⁷ <https://www.xp4l.com/>

dans la communauté artistique. D'ailleurs ces boites-là elles ont marqué et elles ont fléché leurs clients : ce sont les ingénieurs. Et je trouve ça super qu'il y ait des outils pour les ingénieurs du son. Mais je trouvais qu'il n'y avait pas cette boite à outils un peu DIY, flexible pour permettre à l'artiste de, lui, créer sa matière sonore. Donc c'est né un peu de tout ça, d'avoir envie d'explorer un autre framework, une autre façon d'aborder le sujet.

Et la communauté XP maintenant il y a des ingés son qui l'ont acheté mais c'est moins la communauté des ingés son que les artistes nouveaux média. D'ailleurs c'était mon objectif, plutôt la communauté trans-média. C'est à dire des gens qui font de la créa, qui sont dans la lumière... voilà c'était plutôt ça la niche de départ. Des gens qui ont une pratique transversale dans laquelle le son spatial intervient, parfois sur des installations, dans des dômes aussi par exemple. Après c'est un tout petit projet, j'ai réussi à créer une bonne communauté mais c'est un petit projet par rapport à la puissance que peut avoir Flux:: ou L-Isa. De toute façon ce n'est pas aussi flexible. Je ne recommanderais pas d'aller faire un projet dans un *stage* ou dans une salle de 3000 personnes, comme on m'a demandé, avec XP. Ce n'est pas fait pour ça, là quand on est dans ce genre d'environnement c'est un outil de Sound Engineering qu'il faut c'est pas un outil de création dont les personnes ont besoin, c'est vraiment fait plutôt pour des formats natifs.

V : Ça a l'avantage d'être sur Live, donc c'est facile de composer et de trouver des idées.

E : Oui parce que si tu prends l'exemple de Flux:: et du Spat Revolution c'est vraiment un outil pensé pour le mix. Donc tu as un software, souvent il faut un deuxième ordi, il faut créer des connexions... Je veux dire que c'est pas réaliste par rapport à la façon de travailler d'un gars qui peut être dans son train, qui va se dire « tiens là j'ai 30 min je vais me générer une petite séquence », explorer aussi ce qu'on peut faire au niveau sonore, créer des trucs, avoir un outil flexible qui permet d'explorer ce qu'on peut faire. Parce que ce n'est pas figé, je l'ai dit en entrée, l'audio spatial c'est ce qu'on en fait. C'est un peu une phrase générique mais il faut y aller, il faut tenter des trucs.

Moi je suis super déçu moi de voir certaines choses... On a des mouvements d'objets *nianiania*, ça suit le train machin mais je suis désolé ce n'est pas ça. L'audio spatial ce n'est pas ça. L'audio spatial c'est un endroit où il faut laisser exprimer son imagination, filtrer des sons, créer des objets. Et encore une fois c'est normal que nous, dans notre communauté artistique, on ait ce genre de questions et c'est normal que l'ingé son ne les ait pas. Donc il y a cet entre-deux qu'il va falloir stabiliser à un moment donné, mais comme il n'y a pas vraiment de normes de productions, il n'y a pas de standard par exemple, comme on voit dans un film ou dans une prod d'album où on sait que le gars va produire et l'ingé son va aider à produire mais ce n'est pas lui qui va dire « tiens on va mettre tel effet. » Donc voilà on n'a pas vraiment de standard dans l'audio spatial parce qu'il n'y a pas vraiment de marché. Et encore on parle de la France. On a une sorte de zoom en France pour des raisons historiques, c'est que c'est un truc de chez nous. Chez nous et à Paris, toutes les boites qui font ça sont ici, c'est juste complètement cinglé. Il y a Holoplot en Allemagne sinon y a peut-être un truc aux *states* mais c'est tout. C'est un truc super français ! Tout le monde est en train de se bouffer le morceau de viande qui reste mais voilà c'est un peu ça quoi.

V : Dans tes créations, tu es dans l'idée de quelque chose de réaliste ou c'est plutôt simuler le réaliste mais aller plus loin ?

E : C'est ce que je disais et j'ai corrigé après, c'est utiliser les modes de perception de l'humain dans sa scène sonore, c'est grossièrement modéliser cette perception là. Mais nous en tant qu'artistes on ne veut pas reproduire le réel non plus, on veut le transformer. Donc on utilise la modélisation. Mais ensuite on veut le *stretcher*, on veut le déformer. Alors ce n'est pas pour tout. J'ai aucun problème, si l'objectif d'une pièce c'est de reproduire ça à l'identique un bord de plage avec le bruit des dunes derrière, c'est super. Pour faire ça, ça peut être les mêmes outils, en tout cas le geste artistique ce ne sera pas le même. Donc c'est comme tout, c'est comme en musique. Moi je viens de la musique électronique expérimentale. Mon langage c'est d'explorer ce qu'on peut faire avec les matières sonores. C'est la transformer, la détruire, la reconstruire. Mais il y a aussi des gens qui vont vouloir faire du Field Recording et recréer un vrai champ sonore très homogène, très compréhensible. Peut-être que là avec tes connaissances d'ingénieur sonore, elles sont plus utiles pour comprendre comment ça fonctionne la perception, mais dans mon cas on veut transformer faire le geste artistique qui va être complètement prépondérant. C'est comme pour la musique, on retrouve les mêmes réflexes.

Je dis ça, j'appuie là-dessus. C'est pour ça qu'XP existe. C'est parce qu'on laisse le champ libre dans le narratif qui est développé pour dire que ce sont des outils d'ingénieur du son. Oui d'accord mais pas que. Il faut qu'avec ces outils on puisse explorer, créer des choses. Donc la modélisation est importante, c'est pour ça que je bosse avec l'IRCAM, parce qu'ils ont un historique de 30 ans là-dessus. Et puis c'est toujours pareil avec les outils, il y a l'algo en bas et l'UI. En gros c'est comment tu vas jouer avec ton algo. C'est un vrai choix. Comment tu vas aller stimuler l'expérience de l'utilisateur pour l'inviter à faire des choses ? Moi je ne modélise pas mais ça m'intéresse de voir quels outils on peut faire pour donner le plus de flexibilité aux artistes pour créer de l'audio spatial. Donc XP est plus fait dans cet esprit-là, vraiment. Et encore je pense que les outils Flux:: Spat Rev ce sont des outils incroyables. Mais je pense que pour plein de raisons ce sont des choses dans lesquelles les artistes se retrouvent pas.

Pour quel type d'expériences.... On a déjà abordé ce sujet parce qu'en fait c'est pareil, entre mixer un chanteur avec un piano et on va dire « tiens on va le mettre en 360 », moi je crois que ça a absolument aucun intérêt. Je vais te dire vraiment mon point de vue : pour la même raison que je trouve qu'il y a des choses qui doivent rester en mono, des albums des années 50 et que c'est super comme ça, il y a aujourd'hui une forme de musique que j'ai toujours envie d'apprécier en stéréo. Il n'y a pas l'extension stéréo, l'immersif pour moi, dans ce contexte, je ne le comprends pas. Je comprends les raisons de business qu'il y a autour de ça évidemment mais au niveau esthétique et pertinence du médium je ne le comprends pas. En revanche, que l'audio spatial soit un nouveau terrain de jeu pour inventer des nouvelles choses, des nouvelles expériences ça c'est sûr, c'est certain.

V : Il y a une différence entre passer en immersif du contenu du Back Catalog et repenser des enregistrements type piano solo en nouveau format, pensé à la base du projet.

E : Il ne faut pas non plus écarter les intérêts de business. Quand je dis ça je vais plutôt contre mes intérêts... Ça ne me gêne pas de dire à des gens que c'est pour faire ça, XP. Tu veux faire quoi ? Non ce n'est pas le bon outil. Qu'un mec qui a investi dans du Dolby et dise « C'est super c'est génial », d'accord mais moi j'ai aussi des expériences de mix dans des studios à Berlin et il n'y a pas d'exploration du medium en fait. Tu arrives, c'est comme de la stéréo

mais il y en a partout, c'est juste de la stéréo partout. Donc c'est quoi l'intérêt du spatial ? Le Spatial c'est créer des contrastes, explorer la profondeur, du coup ça affecte le format d'écriture, de composition, le format d'écriture il va être intégré, parce que quand le compositeur dit « ah ouai j'ai une nouvelle dimension parce qu'avant j'avais l'harmonie le rythme... »

[...]

A12.8 Entretien avec William Russell et Alessandra Denegri

V: What is immersion to you and what does it mean?

W: It's funny, it's a topic of a panel that's happening at Sonaar, and Alessandra is going to be speaking in the panel about this. Maybe this must be coming from you Alessandra.

A: To begin we must make 2 distinctions, and one is about immersive. Then we can enter that conversation about immersive arts.

What is immersive? Nature is immersive. A smell can be immersive. There are things that happen without intention that are immersive. I think it's also kind of a new term, and we don't have a lot of wording for it. We're still trying to define what it is exactly, to be able to break it down and then catalogize things better. In Monom, we try to not refer to things about immersive, but we talk about spatial, how the artwork affects the space and then the space affects you. But the word immersive, from the nature of the word, is what comes to you and you cannot help it but to be engaged with whatever is going on.

Then there's this question about what immersive art is, when an art piece is immersive. There are books that can be immersive, that are created to make you lose this sense of space and time, and you are completely lost in the piece. That is immersive art. And it's so personal. It has to do with your background, who you are, what have been your happiness and your culture. Then maybe something that is immersive for you is not going to be immersive for me. Because I'm different.

W: Yesterday, we were talking about the question: does immersive art have to have some extra sort of technology. Because if you are going to, kind of put it, kind of categorize this in, and talk about it, immersive art is different than looking at a painting. Even if you can be immersed in that, is that called immersive art or is it have to be something where you don't even have a choice, like you're senses are so engaged that it's much more what you used in a technology that can really engaged your senses fully, no matter what. It doesn't matter if you're intellectually engaged in the artwork, it's more about what you feel at the moment in the artwork, so the technology plays a pretty key role.

A: But the technology not in the sense of computer necessarily. Machu Pichu is an immersive experience from the architecture of that temple. You go to Egypt, and you enter this pyramid, it's absolutely immersive, you completely loose the sense of space and time, you can be like anywhere and everywhere at the same time and I think there's a difference when you contemplate the piece and you're not one with the work.

W: Yes, there are contemplative arts, it takes more view to ...

A: Be immersed.

W: And obviously that contemplation can be taken to another level when you add this sort of technology such as spatial sound, which for us is the kind of ultimate tool for immersing people because it's something you can't switch off, you can't shut off your ears. You're going to be feeling this in your whole body. Sound is a thing; we've evolved side by side with vibrations.

It's like sort of our oldest companion. Sound. And Spatial Sound is essentially just the way that we perceive sound, the way that we talk about it, the way we've classified it. Based on how we've evolved to this even. We are trying to create experiences which come and tap into that super highly evolved sense of space through sound to really tap into some places which you don't get from any other kind of medium. If it's purely visual or the sound isn't playing a key role, I think there's something missing. As humans, we need the immersion to be complete. Especially the spatial aspect of it. Which is to say that, when you have mono sound like we are hearing now, it's everything at once. It has an immersion cause we're within it but it's not taking advantages about our ability to perceive sound in space. It's more of a passive experience where spatial sound becomes much more active, the way you tune into different sounds. Even when you listen to mono I'll tune into the drums because I used to play drums, and a bassist will go to the bass, and all the people will go to different things. But with Spatial Sound it takes it to another level. You tune into the space and where that sound is and how that sound has formed, and how that sound interacts with the rest of the space. I think there's a lot of opportunity to transport people physically in a more embodied way rather than other forms. You're totally transported but it's not an embodied feeling in the same way the sound experience things.

V: You defend the way of including the space at the beginning of the compositional act.

W: Yes. As a studio that's where we're exploring, the compositional and the expressive possibilities of spatial sound like an artistic medium, that's where our approach is. It's coming from there, it's coming from "think about spatial sound as its own medium rather than something that accompanies something else", like a film or a live show, whatever situation where spatial sound is typically used in. With Dolby and these other sorts of surround sound, it's usually in a complement of something, it's never really existing by itself. Obviously, you have records in Dolby so you can listen to those, but they are at the end of the process. The artist goes and they record something, generally they're not mixing it for, in Dolby, thinking about space as something as important as timbre and pitch. And these are those things that we do here, thinking about it as part of the compositional process. It's something where you are truly as an equal. That element for us is super important. To educate the artists that come and work with this system, it's not like other sort of systems. Maybe I can talk a little bit about how I'm using the system, and what it's capable of doing. Maybe that will help you understand.

V: You developed your own tools to manipulate the sound in Monom?

W: Actually, 4DSound is the name of the collective, the company. They created the system that we use. The system that's in Monom is the original 4DSound, originally created for an Opera, based on Nicholas Tesla Life. That's why the columns look the way they do. There are newer versions, they just hang the speakers and they're quite small. The idea of this system was to create a sort of sonic staging. Hyper realistic sort of staging and sound. The idea was to create a virtual sonic reality rather than saying it's just for moving sound around and placing sounds in different places in a mix. You really think about it as a new reality, where you can just create a city scape with cars coming through, you can create a forest, whatever. You can just bring up all this staging and then your mind fills in the gaps. It's like a book, but it's more than that because you're feeling it with your body. It becomes a very physical experience as well as an imaginative experience where you are connecting to the sound. Whether it's

hyperreal or abstract sound, you are still connecting in a very immersive way because your brain is thinking that you're in that space. Like a VR experience. The system was developed, and after they realized they created this incredible instrument to basically build sonic world.

[...]

V: Did you work on the Opera?

W: I came in quite a lot later than this. I was just obsessed with the system because I was looking for years for something that could not just be a sound system but be an instrument, when you're able to control sounds in space. And my dream, and the thing in my head, it was essentially like 4DSound had created it and even further than I could have imagined. So, I was very lucky I found them like 8 or 9 years ago. I became obsessed and I'm a musician so I'm not coming from the sound design background, but I had such a deep emotional connection to the system when I first heard it. And it was abstract sound it wasn't even music; it was like white noise. And I remember walking into this room and the lights went off and there was nothing else, there was no other sensory information and then I just heard this white noise coming from a corridor and something coming closer and closer and into the room and then through me and then out the other side and then out in the distance. And it blew everything else away. I've checked everything out there and noting even came close. And because it's not just the system, it's the fact that it's a tool, like an instrument, the works that are created there by the artists are way deeper because the tool itself is so intuitive. In the past 10/15 years it's been developed in collaboration with the artists who have come in and every time an artist comes in the interface becomes more... it integrates new workflows, and new ideas. And new compositional ideas... and it has become this beautiful modular interface that can connect to everything. It's a multimedia interface; you can connect anything to it. It's really a powerful instrument.

V: For your first piece here, what have you tried?

W: The first piece that I did was in collaboration. When we opened the studio, I had already been through a process working with the founder of 4DSound. It was almost a year long work at the Spatial Sound institute, learning the system, learning how it works. Going into the process of how artists are working with it. I was taught by the founder. And, not just on the compositional side but also on the philosophical side, where it's coming from, why they're doing in this way, I learned a lot about the act of listening, that's the core of everything, really the act of pure and really deep spatial listening.

[...]

And the first piece that I really engaged was a work that I did with the artist Varg 2TM. That was the first work that I did from beginning to end fully on my own, sort of my full project that I put myself into. I spent almost a month with them. Back then it took much longer, the software was not where it is now. It was amazing, it was way beyond everything else, but it was still not into where it is now.

That piece ended up by being a sonic film. They had this solo journey that was going on. There was this abstract narration, mainly it was creating the environment and the story over time. They had music infused in it. It was about something. There was a narrative behind it,

you wouldn't know it eventually, but it set my mind. Obviously, it is DNA of the opera. The Story is obviously there. That's what I want to focus on. It is creating these incredibly narrative pieces even if it's not explicit, what I want is to create some sonic films. Because I felt like when you don't have the lights on you really go into this world where you go when you're reading but it's even different, you're really feeling with your body, and you're completely transported in a different way. And now I'm trying to recreate these sorts of complexities, dynamics and natural phenomenon the way that the space works and the room works and then I take that and I go in these zones between hyperreal and surreal, infusing these spatial worlds with different sounds. If you create some wind objects, and put it different sounds through that, like a HiHat through that, it will make this sort of windy HiHat because of the spatiality.

V: So, you use natural recordings, and at the same time working with the classic sounds from music.

W: Yes, exactly. I like to mix it all together. My approach is really to try to create the most organic structures to feed sounds into and it's crazy what happens than, if you put different sounds in different structures.

V: Is there something you put big importance on, for example movements? Energy? Timbre? Is there something most important than others? Specifically, in here?

W: My sound is completely based on the software and it's like a spatial synthesizer. All the sounds that are coming are basic. The question is what happens to that sound and where they go and what they turn into from the spatiality. For example, if I put a mono water sound in, I can split this into hundred particles that can be made into a wave, and then that completely changes what that original sound is and I'm always feeling out what is just one sound, and how I can make that one sound completely immersive. And I start there with the minimum of input possible and then try to pull apart the sound and turn it into something completely different that evokes a real space maybe, or other things. It's not like evoking a real place like a city, it's more evoking the reality of space. Even if it's white noise, it's on the floor, and it's moving like a wave, so it feels like a wave, even if it's white noise. White noise feels like a wave but even if that was a HiHat, it will feel like a wave. You can create these incredibly organic structures from this system because it has been built with this in mind, like to kind of tap into that way that we perceive sound, the way that sound moves, when they go outside. The space is based on how they should, how it would happen, the way the particles are split up, the way they move, it feels realistic. It's a really crazy tool, if you put one sound into it you could spend days, months and years on that, just exploring all the possibilities.

[...]

V: A place like this is not made to be oriented in a way, but now you have a performer in the room, your eyes will focus.

W: Very early on we realize that unless that there was a performative element to what was happening, like a dancer or a violinist, or a real instrument, it didn't make sense to have the artist there. Even though lot of the archive is built from live shows which was done with artists who works with laptops or maybe some small little controllers or things like that which is still totally valid because that live moment is also something very special, but in terms of the actual

spatiality it doesn't add to that it just adds to caring that moment. Maybe lots of artists like to premiere their work like that it's nice but then the piece live as a pure sound installation from then on.

V: Do you encourage people to walk around?

W: In an ideal work yeah, people should be free to work around.

[...]

V: You have residencies all year long. What do you say to the artists that come down there?

W: The system itself leads the way in many ways. We first start with listening so, we listen to different pieces, different styles, start to open the ears to these different possibilities that you would have never had before unless you're in this system. So, there's no other way to start instead of listening and hear what other people have done. That's the best. And then, after that point, the artist is so inspired. The first time for me, your mind just goes crazy like "holly shit the possibilities of my sound in this space". It's the best way to possibility hear sound so they get very excited. So, it stimulates them to start to think about the concept a bit more or they just start playing their sound in the system and start to use the interface. We start going through the different sections of the interface which is all easy laid out you don't have to be a sound engineer or Dolby; you don't have to get some sort of certification to already be able to just start diving. It's an instrument really. It's been built for artists by artists. It has an instinct sort of feeling. I like to think of it like a piano. You can go out to a piano and if you don't know how to play the piano you can still sit down and still play the piano. How then you interact, it depends a little bit on your ear, on you, on your sensibility, but you will be able to play something. And if you learn the piano then you can start expressing on a higher level through it. Like if you learn key and tone, and whatever, rhythm at least, you can start really expressing a lot more stuff. So, some artists are really working with us. Se we are really producing, they just act. They say, "oh the sound is here, the environment feels like this and do this". We are really the one who drives the spatialization. We're producing the space side of things with them, under their directions, like a film sometimes, they're the director on their film and their idea, but we kind of like help them create the world. And then sometimes artists just jump in straight away there on the computer and they're doing it. Most of the time you must have a little bit of a knowledge on Ableton to really like, do that.

V: On Ableton?

W: Yes, because it's integrated, we have this Max4Live device, which is the best, easiest, accessible way to jump in. You don't have to use Ableton, you can use it independently and understand the audio directly. But Ableton, it's well integrated, you can see the colors of the objects. The 3D interface that we're using is something just beyond anything else. It's so easy to build the environment. You can even build it in VR. It's really detailed. Very easy because this way of working is more natural. So, I think of sound like this, I think of sound like I'm trying to condense a whole environment into two speakers, I'm just building sounds like I have experienced sound, which is everywhere equally, omnidirectionally, in all directions.

[...]

V: Artists must be very disappointed to go back to stereo after that.

W: Yes, but generally they don't think about stereo in the same way. There are amazing tools to do spatialization and it's just starting to make you think about it more. Even though you might not be using this, you can use the software on your computer on different setup, that's new, we're now training, and part of the residency is to learn how to use the software so that you can integrate it in your artistic practice. We started to do that more but it's sort of a slow process because we want to educate people, we don't want to be like Dolby. Everyone could have it and you don't really understand that there are more possibilities. And because we're not a commercial company, we are more focus on the artistic side. To create that sort of model and business of selling software to everyone, it's a whole thing. But it is something we want. Our goal is that everyone can have access to this, but it's been in this process and development where it's still such a beast of a thing that we believe that it's important that, when people do come in, they really understand what's behind it because otherwise you get people think that they're using it in a way and then it's pulling it forth not in the right way and people don't understand what they intend to. What they're trying to do here is a different idea about what you can do with just sound and about listening experiences. So, we have a much different goal here than others which is really to help people understand there's a different kind of way you can experience art, music, sound art especially, and that just requires a little bit more care. So, we don't want that anyone could use it and suddenly people are doing what this 4DSound thing "oh I heard a 4DSound" and maybe they didn't have such the experience. But that would change very quickly when people get more savvy or so with this stuff and understand it more and have more of these experiences.

V: How do you think it will evolve in the next years? More places like this?

W: We just open a permanent exhibition space outside, this is mainly use as a studio, but we do these events occasionally. And we do a big event once a year, in April, festival for listening and spatial sound. And that's where we kind of showcase the latest stuff we're working on.

[...]

We have this huge sort of archive of works, which is ready to play, and with this new touring system that we just got, we can come out and we can bring it to different places showing to people what's possible because the only way to do it is to experience it. It's not like other things when you see it on your phone and you say, "oh that's look cool" and I'll check it out, and no. No, because there's so much bullshit in the industry about "it's so much immersive, in 360, oh multichannel and blablabla" it's average. I think a lot of that is just because the tools to create are just not being so, I wish this was more widely available and when that happens these sorts of things become easier more accessible than the content and the experience will become more interesting.

V: What is the 4th dimension here?

W: Good question. Originally, the 4Dsound for what I understand, there were thinking that the 4th dimension was you. As the listener. There's 3D Sound happening but the way that you experienced the work is very dependent of your perspective and your attention. So, the aspect of your listening is part of their DNA let's say. So that was the idea from their side. All this other stuff you see like 8D etc. This is nonsense. There was just the name of the collective and the company and that what they were thinking about.

People usually think about time, the time is the 4th dimension.

V: Are there only woofers on the ground?

W: There are subs under the floor. And they are just basic 15 subs.

V: You didn't put full range speakers on the floor.

W: 16 speakers that are facing up.

[...]

And still have no sweet spot.

V: Here, there are no sweet spot?

W: You think of it more like a real-world environment. If I'm in a forest, and there's a waterfall, either your close to the waterfall or further away it's not better or worse, it's just different. That's sort of how we are considering the compositional environment. You have an infinite continuum to compose in, up down, left right, every direction, every space is equal. So, this is a completely different way of thinking about sound. You're not thinking about it being around you, you're creating a really like a full environment which feels more realistic.

A12.9 Entretien avec Gerriet K. Sharma

G: I think it is very important, when you write something about immersiveness... It's interesting to see where the whole idea of immersiveness came from, and how actually technology has a different language than arts, and literature. Because immersiveness used to be something that happened when you read a book. So derived from that, listening to a record from a mono recording could be immersive as well. And then, again, there's so much bus words now around, and I started to work in what is now called multichannel audio, 20 years ago, and I was not expecting it to become very popular. And it's not popular because people now want it, it's popular because the most successful media company in the world is just pushing it through. I think that's very important to say, that it's not the new hype.

Immersiveness is about 40 million years ago back when we were living in the caves and chanting. At the same time, we were using these words like "natural earing" and so. There's no "natural earing", where should this come from? Because it's all culture, we are socialized in listening to things, like we were socialized in listening in stereo. And if you look at the advertisement, how they advertised stereo, it was the same marketing and wording like we, now, advertise spatial audio. Like "realistic", "authentic", and it's all design, it's not what's happening, nobody can explain the natural earing or listening. Technically and scientifically, you can but it's not happening. So, it's very important to get these things not right, but to a point where... Here, for example we have internship, residents, guessed professors at several places for a long time now, and it's just that, it's more or less defined by what people are talking about in these certain circles, because otherwise is "blabla" we don't have to.

Another question is that my colleague, Johannes from Spæs Lab, asked me the first time we met. That was a long time ago at a conference about space and sound in Dublin, he said: "do you know a sound that is not spatial?" Of course not. So, again is it "blabla" or what does it mean? Mostly it's something that is derived from a technology. From an artistic point of view, this is all not necessary, absolutely not. We accepted as artists the Sweet Spot, which was one of the biggest mistakes we ever made. And now we're looking at the world where actually stereo is happening. It's not happening. It is recorded, because it's working in the studio, but no one is listening in the studio at home, and most people as users have no idea where the difference is between stereo on headphones and loudspeakers. So that's all a big myth, it's makeup. It's a makeup that I appreciate because I really want people to have jobs, and sell stuff, and all my colleagues from the university, after that, all the students that are educated, they need jobs. I think we should differentiate between this and people really working artistically as educated art. And you will become the judge of it tonight when you will listen to a very specific system that's only existing twice in the world¹¹⁸, there's actually no compatibility to anything else, how it is esthetically used, that's the question, not what it is capable of doing. Because we're not asking ourselves the question, when we see a Steinway... It's only what people do with it, that's becoming interesting. So, this is hardly addressed. And I'm concerned.

So that's what's Spæs Lab is all about. We are not called Lab for Spatial Technology because you will see we have a lot. So, we say "there's no excuses left". When I started back

¹¹⁸ Gerriet K. Sharma fait ici référence au studio *Monom*, équipé de la technologie *4D Sound*, dont William Russell est le co-fondateur, et que nous étions sur le point de visiter après noter entretien.

then at the university, I went to Graz in Austria where they are spatialized and became part of this community and I was there for around thirteen years. I have helped from an artistic point of view to develop systems, but back then nobody was interested in what we were doing, we were just a tiny little city that nobody knew about and where something has happened. And at the same time, most of these things were hidden in every town like IRCAM, like MIT, so hardly any artist had ever worked in... Only handpicked. So, the question we were talking about is: what is spatial music? What is music when it's not spatial? And people get very uncomfortable when I start asking these questions. And I don't want people to feel uncomfortable but in Spæs Lab, all artists and software engineer that are both, there's hardly anyone working for us with not a very high level PHD. No one would ever want to work for academia anymore. Not because we hate it, but we made our experiences, so we were really at the point where Spæs Lab really wants... We want substance. We really want to see people doing different things with the system, so that's why we have these open programs where even people from literature, architecture – we have a lot of people from architecture – can address what people call Spatial Audio.

Because if you listen to our archive, we have this archive of multichannel pieces from the past fifteen years... The only reason why we did this is because we want people to be able to learn. And then, if you go through this you will see that lots of things which are praised to be new now are absolutely not capable of becoming 10% of the quality of what they did with analog tape and loudspeakers and mixing desks back then. And that's not because people are not good artists, that's because all these machines that we have at hand now are addressed like machines not like instruments. Back then people were more thinking in a instrumental way, and it's a mindset that we are interested in and not technology. This is how you challenge technology and not just "how look I have nice graphics on my screens, and it gives me the impression that I can make a square here, or I can have a flying object."

Spatial Audio is where you have this Shoe Box here and you put sound in there and you can drive the sound around no matter what sound, it will always make the shape or the direction or something out this shoebox. Nobody with ears ever experienced that, because every sound has a certain trace. You cannot make a spiral with all the same sounds, but this is what people expect because they see it on the screen and they think it's a playback device. And a lot of institutions, also in Paris, actually think that...

Ambisonics, for example, is fit for the reproduction of recordings, precise, realistic, from an engineer's point of view. I don't know any artists who could be really interested in... There's a handful of them; they are very successful of course because they actually play with what is offered by the institution. But people who are really deep in art and want to really develop music and would never accept the position of a loudspeaker. It never happens. See what happened to the guitar. The horrible things they did to the guitar that became a whole culture. Or take the channel. It's not only making the viola da gamba louder, it's just that people were accepting that this is what was given. And it's that problem we have with spatial audio people just accept it. "Ok! So why would I question a decoder?"

So, these are questions that I'm interested to ask. And that's why if you move up, I can show you why we have all our design of loudspeakers, why the walls are the walls, the colors

are the colors, the wheels are the wheels, everything is moveable for example. In our studio everything has to be moveable, so that you don't get the feeling there's a front, there's a back, there's no direction because a lot of mistakes... No, the boring spatial audio stuff that is happening now is still assuming that stage is in front of you. There's no changing in music, artistically, esthetically, nothing is going to happen if you still think the stage is in front of you.

V: So, when you organize events, you can't have people sit, because if you sit, you're looking to somewhere. So, everyone is able to move in the audience?

G: Yes.

V: So, for you the future of spatial audio must not be oriented?

G: "Must not" is not there. I'm really not interested because that's what's happening for so long, and the venues that are now... I'm happy to see that there are venues now that are doing this but there's still this paradigm and they still have the same problem that we had 20 years or 30 years ago that, for example, that the Loudspeakers are at a certain level, and that if you say "ok you can walk around", people at the same time want to sit and then they sit on the ground and the loudspeaker are too high, and there's all the Spatial Audio flying up there with people on the ground. All these problems are so old they're not at rest, and esthetically it's not working.

V: It's only working when there's no musicians at the venue. Because if you have a musician, you will look at him.

G: Even worse. Even worse. If you look at something in a 3-dimensions environment sound, your ears will focus. And you can learn especially with Chion. I think he's still the greatest literature because he is audio visually oriented and you can learn wherever you set visual stimulus against or parallel Spatial Sound, eyes will focus, ears will focus. That's the highest Challenge. 3D Audio and Performers... So, you're losing 80% of spatial audio.

V: And what about Nono who put speakers everywhere in San Lorenzo with musicians all around.

G: But that's not about Spatial Audio, Spatial Audio was part of this. And the whole thing is actually Spatial Audio, so the architecture that was built... Prometeo, all these pieces, that's actually Spatial Audio, and loudspeakers are part of it. So, if you're talking about this kind of Spatial Audio, then we are in a very different history, Mozart was putting loudspeakers sections at some point, or back then, Venice, San Marco, the 2 choirs... On the other side, of the hall, blend in live performance that's different. But if it's recording live performance while they play and threw around the sound it's difficult because we are always stucked to the gestures, that's human. We stick to the gesture of the performer. It's about the empathy thing. The biggest challenge of 3D audio meant to be really strong is when you don't have any performers because then you don't have this empathy with someone playing this. Your ears just develop this performativity.

If once you put someone in there, I've been trying this for years, you have a big problem. For example, I wrote a piece for a certain performer and... From the beginning I felt that when she is playing her recorder, she's doing something with her face, like naturally, and these bruises, that's taking the attention away from what's happening spatially. So, we trained her, we had a theatre performer coming, training her how to use her face not to draw attention, that it became part of the composition to make the spatiality of the piece stronger, or maintain her. But that's not something that people normally would address.

V: But she was still a performer on a stage.

G: Yes, but she had to have certain postures. Like standing with the back to the audience. Playing towards the 3D Audio setup.

V: What about the speakers? Do they draw attention? Do you have an interest in making the technique disappear?

G: That's such a funny question. I already developed a piece in a residency, and when it came to the concert, the light guy came. Super nice guy and he wanted to show me what he could do and the first thing he showed me, and he was so proud, he has these lamps on every loudspeaker, and we had this sky of stars and I was completely shocked. So, I said "No no switch it off switch it off it's horrible". And he was "Why? it's so nice, I have different colors, I can make movements and so", I was like "No!"

For me, I appreciate if they are not visible because it's just always taking away concentration even if you're a really trained listener. The presence is the problem. The presence, and I would go so far, and lot of people doubt this, but I would say surrounding people by loudspeakers is a problem because you feel it, you feel surrounded, you feel in a cage. And it doesn't mean that you cannot enjoy the piece. But the moment you work without a solution where people are not surrounded with loudspeakers, people feel different, feel much more free, freer to move, to concentrate on their body and their relationship to sounds you know. So, there's a lot to explore. It makes a huge difference with the loudspeakers when they are at ear level or above your head, not only sonic, but also from the presence of the machinery. And then of course there's so many questions, you might have a dome of loudspeakers, with levels, and it's so high that you actually cannot really distinctively ear what happens in the sky. Most of these domes have a couple of layers, too many, from most pieces, so there's all of this moves around...

If you listen for example to Chowning, Turenas, I always do this trick when I teach, that I take the piece and I play it in the dome. And then I ask people, mostly when they're beginners, how many loudspeakers are involved. And they are "I could definitely ear this layer and this layer" and I say "- No - Only 8 channels?" I say "No". It's not seldom, that people actually ask to switch off the loudspeakers, and I switch to the up to down and even where it's on the playing, they wouldn't believe that it's only 4 channels. And this is why, because it is very welled composed, and the guy knows about spatialization. Of course, it took him 6 years. But there's a point about this: we produce far too fast, too quickly. Nowadays we're in this production lines, you want to have automation etc. and there's AI for sure, although I'm quite sure AI will come up with sometimes much more interesting results when it comes to spatialization.

But I think, bottom-line is we have a problem with knowledge transfer, so now that lot of people are using this tools, which I appreciate, we see, we feel we hear the repetition of problems that were happily addressed in the 80s and in the 90s. There's a super paper published in the 90s by Jonty Harrison, and it's about diffusion, and the subtitles is something like "is it just throwing around sound?" It's a quote and it's still the question. People are so super amazed by themselves if they can make flying objects but it's... Esthetically there's nothing happening except that somebody is throwing around sounds. Actually this has been done by artists with loudspeakers on big ropes, much cooler a performance I would say.

I don't want to offend anyone, but I want to have a critical discussion, because I see a lot of artists, especially now, getting through... With no critical discussions just because they can use a couple of tools. At the end of the day... It's funny that Eric Raynaud¹¹⁹ made the contact because at the end of the day, people want to use Ableton and then do some cooler moves. But that's artistically super not interesting. And the funny thing is that a lot of people still don't understand XP4L. It's too complicated for most people. So, you know, there you can see there's... Artistically there's a lot of people who don't do real efforts here. What could it mean that you listen to a sound, and there's a sound behind that sound and there's a sound behind that sounds, so like layers, but how's the relationship between them verticality. And playing the natural spatialization we do, because higher frequencies are high and lower frequencies are down, below for us, because that's how we listen to stuff, so how it is related to sound coming from a loudspeaker there. Most people don't know that if they put subwoofer to the ceiling, they will still hear it from the ground. But that's basics. It should be not only for engineer but for artists.

There are these papers from... Amazing papers... Gerriet Kendall. And he's one of the few who was really addressing the problems of spatialization. When he was artist and composer, he was so super clear about that the loudspeakers down, there are not switched on, and then it was really hard to convince them that they are switched on, but if you put sounds down there that you cannot hear from down there because of your natural hearing capability. It's not a problem especially for engineer, most of the systems that we're using are actually proved by moving around one stimulus but compositional wise, it's absolutely nonsense, what's happening if you have two sounds and one is moving like this, and one is moving like that, and you don't have two shapes, it is mostly producing a third shape. But that's scientifically super hard to prove or to investigate so we don't talk about it but that's where music starts, always. It's always more than... There's this funny little odd saying that music is more about the sum of its parts. That's where we should start talking about. Or art is always more than using the tools. If you go in Instagram there's always people pranking around with the gear they're using, "look mama, I'm now in this studio" even in conferences there's professors saying, "I was in this studio in America, I used 150 loudspeakers blablabla.." You think: "Are you kidding me? I thought we were talking about motivation, why are we doing this, what's the difference in using this kind of loudspeakers, why don't you only use 2." That would be so cool.

Now we have this kind of people, at the moment, in our studios and it's kind of scary. You give them 24 loudspeakers and after 2 weeks they only use two. That's the result. And it's not understanding, that's exactly understanding. How we actually, for centuries, developed

¹¹⁹ J'ai eu la chance d'échanger avec Gerriet K. Sharma grâce à Éric Ranaud qui nous a mis en contact.

music. Of course we have the same tendencies, hundred years ago when we developed this huge orchestras. Composers of course said, "Now I have to show them I can do it." You know. But that's so macho. Have you noticed in your studies that if you give, especially to men, lots of loudspeaker, first thing they do is to play very loud. It's completely insane, because especially what Spatial Audio can do, is to play at a level than you cannot do with any other instruments in the world because of the spread, you know, and you can buy the narrative that space itself is telling... Can we make people listening to tiny objects that you could never do in stereo, and also not do with an orchestra. Who's doing this? If you go to IRCAM and listen to their archive, I did, it was ten years ago, I tried to challenge myself again, I want to play as detailed as possible and being strong but not loud. So, I went to the archive and there's hardly any pieces that are not super loud. "I have to show you!"

And then another thing people start talking about immersive audio, is about immersiveness. Mostly what's happening is people are using a lot of sound. Which in itself is immersive. In a lot of systems, if it's not measured in this very basic way like Dolby Atmos, like "Wow, we did some pannings here and there", and a format for that... But huge problem of, for example, ambisonics and the archive of the pieces, there's still not, you can still not archive the subwoofer. So, half the formats like AmbiX, you can send it anywhere in this world, but the grounding of the piece is always in the Sub. Unless there are no lower frequency information, cool. It's so dangerous, especially for ambisonics with this light blur that most of time you have, also in my opinion in higher orders, higher higher orders. Something that is not grounded by Subs is really flying away; you cannot connect to it. And for the Spatial Image, for the composition of Space, you need the ground, because human is being built in space first of all by detecting where the ground is with your feet. That's down, and the rest is up. That's how we are oriented. Maybe not in the future if we are coming up maybe we are not walking anymore but that's still a little bit of time. There are so many myths around. Now we are in this realm of 3D and naturalness. We are back to our natural earing, but we shouldn't forget that immersiveness, first of all, starts with subs. How investigate this?

In SPAES Lab, since 1 and a half years, we're archiving all kind of tools that are out there and we have list like "is it free? is it a program language? is it a plugin?" So we are creating a table out there. "Is it ambisonics? Is it WFS? Is it VBAP?" So that people can compare. If you have an artistic idea, and then you think "which tool?" Or you're a program developer and you go deep inside, and then maybe in this composition, with this programming language, you develop pieces, that's the media art you know, you question a tool, and the result is a piece. But this formatation that behave now like that everyone is telling us how great Dolby Atmos is, you know. Yeah. I'm not against it, it's helping me so much, we both would not talk without Dolby Atmos, because that's how Apple.... And we both know that Apple is not interested in Dolby Atmos.

So, everyone is using everyone, and I like that because that's how Human Being are, but at the end of the day, where... It's interesting. If someone is coming up with an interesting scientific new... Scientific or engineering arts idea, and there's not a lot out there, there's nothing in the horizon. WFS is 200 years old. It's only that there were lacking the CPU but I mean... The math around it, all the theories around waves and propagation is super old. And WFS is for example great but there's not going to be implemented in a lot of pieces. You know we have this super cool loudspeaker system from Berlin, I'm so super happy for this former startup, they made it, but only on the marketing scale, they couldn't say that they sound

better than... And the technology they're using is rather old. And of course, they don't like to hear this but if you look at the fact you can download white sheet for their webpage, it's all there, it's nothing that they invented. Math is out there, science is something that we share, and same with ambisonics, super old. And beautifully, and now we have this history and now we can do things.

[...]

People are not aware of what for example Natasha Barrett did in the beginning of the 2000 which is the nightmare because she is the queen of ambisonics, especially in the beginning of the 2000, that was adventurous. Oh my gosh. Have you been on her website? Downloadable about 30 pieces with explanation, pages and pages. In your lifetime you cannot read all this. Only listen to Chowning, Xenakis, Stockhausen.

V: But these pieces were created for an acoustic, for a specific place. Is it relevant to hear it in another context?

G: Absolutely impossible. But we can learn. It's better to have them in an environment with multichannel, that you can learn from, how you deal with frequencies, with textures. Texture, propagation. But, honestly, most of these pieces are still site specific. And even though we have an ambisonics, we have some rules now, an exchangeable format, they stay kind of site specific. What I composed at IRCAM not sounds the same at ZKM. It's only that we have a format for the trajectories and so, but frequency wise... Yes, you're right. Prometeo is Prometeo. The composition by Olga Neuwirth... We're talking about orchestra and loudspeaker pieces, in collaboration with Markus Noisternig, these pieces will always sound different where we go and that's good, but we should embrace this and not trying to give the impression that these formats authentically capture spaces. It's not happening. And that's what I don't like about it. It's too technical. I think it's almost getting maniac how people are adoring Eignemike recordings. Yes, I know you can do amazing things you know. But there are mono recordings that are more lively. Definitely. For me. And all Jazz recordings, are so live for me, so clear, so vivid, whereas if you take a... Dummy Head, it sounds like... It's precise but it sounds like you're listening threw a metal stick or straw.

So, this is all to be explored and as I said before my concern, and that's why I'm sounding that critical. I couldn't care less that people do whatever they do, just they should, but the reason why I'm adjusting these points is that because we have a huge knowledge transfer problem. I don't like that someone's coming up and telling everyone "This technology is the coolest thing, and you have to do this or that." A good instrument can make a good musician better, but it will not make music. And a good Soundsystem or a good format will not change anything. And I think if we're interested in music, we should be more interested in how music is developed culturally, especially in these days where everyone is looking back and being afraid of moving somewhere for understandable reasons, but we should definitely look at how we could evolve in music instead of telling people that they should believe in marketing. "Oh you need this loudspeaker. Oh you need these headphones" Really? I've seen people, not too long ago, with Dictaphones, making amazing music, like just by recording and overlaying and stuff. It's just how much idea you put in there.

What I like about your project is that people could get ideas in a positive way of how they could use these environments. Because there's so many different... Some of these systems, if you go to Dolby Studio, it must be, it will always be the same, cause otherwise the Dolby consultant won't give you your sticker. But the other systems out there are so different and that's ... that's a challenge.

V: Another question: when you welcome your residents, do you take one day to discuss all this?

G: I have to tell you something, I was absolutely against internships, and I was absolutely against residents. Why? Because I was in a lot of residencies, and I was in a lot of internships. And when Mickael, founding partner of Johannes, came up with this idea, I said no because it was not a possibility. But then we talked about it and what I want to say is we take a lot of care. So, residents come, you will see Matthew, they get... First day is a complete "where's what", switching off switching on, do it again so you are feeling really secure, you cannot break anything, even where's the light, how do you open windows, and then we talk about projects in advance, what do you wanna do, and then we think about what do you need, what kind of adapters, cable, sometimes it's already there, sometimes we try to buy it or to organize it. And I think that every second day we have a conversation. Not by rule, not every second day at 9 o'clock, but every second day. We knock at the door, and we come over, and then we discuss, how is it developing, what are the problems, why don't you want to do it anymore, what are the next plans. We have an ongoing discussion that's important thing. Our resident it's not so much about results, I'm not interested in people finishing pieces because there's so many pieces out there, and there's so many products out there, we don't want products, we want better ideas. And through these conversations, I get ideas of how for example I can make the residency better, or for my own art you know. Someone saying I always hated this software but now I found a very interesting aspects and sometimes I think... Hum maybe I should look into it again. This is how knowledge transfer works. Someone is saying I hate this, and you think "I love it, why?" This is what we're interested in.

Sometimes people use MAX MSP and I say, "Yes but see here's a simple patch that someone did a year ago at Spæs Lab and I can explain you, drop a file here and play around". 2 hours later they come up with an idea, just because the limitation... Or the other way. Then they wouldn't touch Ableton again for half a year. Or sometimes there are hardcore supercollider guys, and I tell them, "Yeah it's cool", I love Supercollider and I support the community.

So, we take this very seriously. And we try to... Most of the residents actually leave something. Like a patch, or a piece. Or a written text. So, what we really want is, we want a kind of community and that's now for years into the process and it seems to work. People come back. I had an email last night from a resident, she was here last summer, and she wants to come back. She's asking what she could contribute, what we need, she's asking if she could work in the library, "Could I setup something?" We have people applying and say, "Ok I could help you with your work." I say, "Ok I have this production coming up with different loudspeaker systems, would you help setting up the software." So that's in the process. But nobody's working at Spæs Lab by without doing something for him or herself, people are mostly working on their own projects but it's not a close shop, so, it's not just you get a key and in the end, someone say, "thank you very much" and "bye bye." I'm not interested in this kind of residency. The only problem that we have is that we have no funding. And that was because I

was against these residencies. Because I never accepted, as an artist, pay for a residency, and the problem we have is that people have to apply for some money so that they can pay a small money for studio time. But that's nowhere near what people want.

V: Is it a specific problem for Spatial Audio that you don't have funding?

G: It's our way of doing Spatial Audio. I think Dolby Atmos Studios won't have the same problem. They're in the realm of pop music and there's enough funding. But for experimental music there's one foot into academic tradition. We're still, I mean, I was addressed 2 or 3 times in the past weeks by several institutions and we're talking about huge university, that they really wanted to understand our institutions, they said it's unheard of that someone has a half sphere, and they all think we are rich kids, and I say "No we're just doing our job", and of course... I think academia was not really... Now we need Dolby Atmos to say "Hey we're doing this for the past 20 years", it should be the other way around.

These residencies and internship are definitely not political, but the idea is to really bring artists in the position that they can connect you know, without having to obey to certain institution. Institution can be something like Dolby Atmos, or a certain school at a certain university. "Oh we only do Super Collider, Oh we only do Max". That's not happening. If you have an artistic idea which is strong enough, then you can connect here. I don't care. I have people that have ever touch any audio software, but they come from literature, and spatial theory, and media theory, and they come up with cool ideas, and then we try to enable them.

A12.10 Entretien avec Lukas Wilke

V: How did you get into immersive audio?

L: I started at the university doing projects. Somehow, they build a studio with a 5.1.4 speaker setup and then I just tried stuff out. We always did immersive microphones setup during the orchestra session. So I did a lot research on that and then I was really lucky that I could use this knowledge in my job right now so, I told everybody that I can do this kind of stuff and then I was again lucky that my boss, he would suggest me doing some immersive productions at the Berlin Philharmonic Hall, for example the pilot project to develop workflow for producing this kind of content and also for distributing it. And right now, we have this distribution network over an app. We distribute Dolby encodes of the mixes or the master whatever you want to call it, which you can play on the Deutschlandfunk app, and then it will be decoded on the operating system layer, so iOS or Android. That's the decoding job. Maybe some more on that later because there's some differences between distribution ways, some do in-app decoding, other OS layer decoding and you get different renderers, there's the spatial audio engine, there's the Dolby renderer and there's the Fraunhofer renderer which is used by Sony 360.

V: You broadcast the ADM file?

L: No, the encodes, so we do the production like. For now, the thing is we do somehow the production and then I print an ADM and then we encode the ADM using the Dolby Media encoder which is a standalone app. You can find webpages about it but it's not available for individual use, it's only available for institutional use. It's just an app so there's nothing very special about it but more elegant way If you have a broadcast network is to use encoding farms like server farms which do the encoding for you, so you don't have to do it on your own. The encodes are done in the Dolby Media encoder and then you have AC3 codec and etc...

These files will be then uploaded, and this is the data you stream on the mobile device. That's veery interesting point, the codec site of the immersive audio because, like the file sizes and the data rates are very important with immersive audio because you are limited in some way if you want to people listen to our stuff you need to distribute it and you can't distribute 60Go for 1 hour of content, I mean you can but who in the world will listen to it. So, you need codecs to get the data rates and file size to be capable of streaming and stuff. And obviously there are some drawbacks and... what's my point. Now. There are different codecs and I did a lot of testing with them. Even the Dolby Codec vs the Fraunhofer codec. Are you familiar with the Fraunhofer immersive audio Mpeg-H? So, these are, I would say, the 2 competitors right now. Nevertheless, Dolby is the... hum... Never mind. Dolby is, I think, more... has more power in the industry and Fraunhofer is also good in marketing their stuff but.... Yeah, never mind. My point is these 2 systems; these 2 codecs work differently and sound differently. They tried to do the same thing, reducing data for immersive audio content and you need to know that Dolby codec, the AC3 one does not transmit binaural settings for example and stuff. You know, what I now recognize, is that I'm drifting away always.

I have so much in mind, maybe I need questions.

V: Maybe we could go back to your immersive audio activity. You do a lot of audio dramas in Dolby Atmos, and classical music?

L: Yes, I also do some pop, jazz pop stuff, but always acoustical music so my work always has to do with acoustical sound source.

V: All of this in the context of the German Broadcast and Service.

L: Yes, it's called Deutschland radio, and the program is actually called Deutschlandfunk Kultur, It's the cultural program with all the "arty arty" stuff.

V: I'm curious about the software you use to work with immersive sound.

L: We use ProTools for post-production. That's our production DAW, and then we use the Dolby Renderer. For the classical music, in terms of workflow, I developed a workflow for live mixing channel based immersive mix, so, I do a 7.1.4 channel-based mix without any renderer of live concert for example, live to tape. So for example a Berlin Philharmonic hall, we have a control room with a nice mixing desk and 7.1.4 speaker setup and then what I do is mixing stereo and 7.1.4 at the same time, so obviously I need to switch between both mixes from time to time, but both mixes are printed right away so, of course I also record the woody track, it's a lot of data going in there but I have this 7.1.4 mix which sounds nice for static sources which you have in classical music most of the time. I would say that a channel-based mix is as good or just as good as any object-based mix right now for speaker playback of course but also for binaural playback I would say. Maybe in the future there might be some advantages in object-based workflows but for now it's a very fast workflow to do this kind of recording and mixing process because you skip the postproduction. Of course, you can do some editing. I did a lot of tests where I do records concerts and then mix them in post, I did some live mixes as well and they ended up at least at 95% of the quality of the postproduction mix. And you're saving a whole day at the studio.

Then I have these 12 channels of audio 7.1.4 and then I need to put it in some sort of distribution format. And the interesting part about this is because there's a lot of ways. For speaker playback it's easy you just have to rout every channel in the right speaker but for distribution purposes you just get the 7.1.2 bit so what do you with the 2 remaining top channels and I tried some stuffs also for the binaural playback and right now I'm happy and it sounds a little bit, counterintuitive but I take this channel based mix, and just create 12 objects at the speaker positions, and then I dial in the binaural settings for each object and this needs some trying and tweaking but I got good results with the front, even more counterintuitive, with the binaural settings turned off and the tops and the surrounds on the mid or far. Usually, I use near to mid for the surround and mid to far on the tops.

V: So, your front channels go directly in the ears.

L: Yes. That's a huge debate if you can do that or you can't. Usually, the immersive experience for binaural playback, I would say, benefits the most for the upper sources and of course the surrounds. and front usually doesn't have this immersive quality for me.

V: Sure, in binaural it's difficult to have a frontal image and to feel it.

L: Yeah, and the drawbacks of the binauralization or the HRTFs are sound coloring, and I would say that the sound coloring is a bigger drawback than not having it a little more spatial. Spatialization usually in classical music isn't that big of a deal. Because sometime, for radio play for example, it's really disturbing if you have a dry voice and then you get this kind of small room spatial sound from the renderer. And that's not nice. But for classical music it's fine I would say. There's some sort of room on top of your roomy recording most of the time that's fine but the sound coloring it's ... I think it's a big drawback. And I tried lots of renderers. Like the spatial audio engine, the Dolby Renderer, the Fraunhofer... Binaural renderers and I would say that all of them have the most drawbacks in sound coloring. Spatialization works nice of course there are some differences, but the sound coloring is the big drawbacks for me so turning off the binaural settings for the front channels saves the sound color. And you also get this immersive quality from the tops and surrounds. But again, that's my empirical tips. It turned out that I like those settings the most, but you can definitely prove me wrong. If there's something better out there I would have to try it out.

So, that's the work I do, and, also, of course I do speaker mixes, and I rely a lot on the control room, on the speaker setups but most listeners of the content I produce will listen to the content on headphones so the binauralization is a big topic. Still not all questions are answered. It's always changing so you need to adapt in the daily bases.

V: Before the rendering, what's the main difference for you between a stereo and an immersive mix?

L: It's a good question because there's a lot of people asking me what's the benefit and why do we need this. So. It's a big debate so I would say that the benefits from stereo to immersive are... From a producer kind of a perspective, you get more creative freedom, you get more resolution, you have a broader frequency range that you can use, the effect of overshadowing of one sound overshadowing another isn't a problem anymore most of the time. You get a huge improvement on resolution in terms of spatial quality. So, the room information is so much more detailed, you also get a lot of improvement in terms of dynamic resolutions. So, you don't need to reduce the dynamic of your content as much because immersive playback systems can reproduce a broader dynamic range. And from the listening and distribution perspective, I would say that adoption of different playback systems works very well. The thing is, with stereo you always need good playback systems, sound reproduction systems to experience the real stereophonic sound experience, and same goes for immersive content but immersive content can be reproduced by wider range of playback systems so... As I mentioned, binaural playback, I think, is, for lot of genres, a big improvement.

I listened to a lot of recordings from all genres and it's really interesting for me to listen to these on Apple Music, or Amazon Music or Tidal and switch back and forth between stereo and immersive mix. When you do this kind of experiences, it's not always the case that immersive mix sounds better, but I would say that an immersive mix which is properly done is always an improvement to the stereo mix on binaural playback. I think those examples where I would prefer the stereo mix are always examples where the immersive mix is not properly done, most of the time it might be just an upmix. And because there's so much immersive material out there which you might not like and then you ask how this happened, who did this and why was it done that way. I don't know but I assumed that a lot of answers to those

questions are “It’s just an upmix”. There was a need for immersive mix, they took the shortcut. Because of course it’s more expensive to do an immersive mix. You always need a stereo mix, immersive mix is always additional to that and actually, remixing stuff from stereo to immersive, I think it’s very easy, I also did this with the sessions which I mixed in stereo, I pulled up and reconfigured it to an immersive mix and it took me, I don’t know, not long, so the stereo mix in terms of sound coloring, like EQ, dynamic processing, worked out of the box, and all I did was like panning and spatialization actually.

V: Not long to go from stereo to spatial because you took your stereo image and you just spread it up? And you used different reverberation of your stereo mix?

L: It very depends on the content. So, the example I mentioned was about radio play. The content itself is a spatial content so you have different spaces, different places different characters, different movements and then you have all these kinds of sound sources, even noises and atmospheric recordings and stuff like that, and then you just need to adopt the movements or the story you wanted to tell in stereo from immersive so you might not take dialog or some movements of little flies or birds from left to right but maybe from front left to back right, stuff like that.

And from music it very depends on the genre. I would say in classical music most things in immersive is done with additional microphone sources. So, of course I use immersive reverb because if I would use a stereo reverb, it would be disconnected from the rest. So, I use immersive reverb instead of stereo one and then I use additional microphone sources so, yeah you need to have those additional microphones signals for sure but I never did a remix of a stereo recording which was recorded with just stereo as the main microphone array of two mikes and maybe some spot mikes, I never did this. But when there’s even a small chance that there will be an immersive mix, I always put up some mikes. So, usually in the Berlin Philharmonic Hall there’s a fix setup of 3D audio microphone array, so you just need to patch them and record them, they’re always there. And when I do production, like 3D production, co-production, sometimes we do co-production with record labels, sessions when we records thousands of takes and you edit it together, usually I have some spare time in the setup process when I can put 4 to 6 more mikes up in the hall. You need to do it of course and of course sometimes time is short but...

V: You think spatial audio during the recording, before going to postproduction, but do you always work in stereo before going into spatial or does it happen to work directly in a 7.1.4 system?

L: During the recording process it’s not always the case that I work in a control room with a 7.1.4 setup. When I do, then, of course every time when I record something there’s always a live mix. The live mix is 90% of the finished product most of the time because everybody relies on it, every musician, every actor, everybody who is in the control room relies on that source. So, it needs to be good right away. So, what I do when I just have stereo speakers and I’m planning to do immersive mix, I pair the mikes in stereo pairs just for a quick listen back and when I have for example 2 surround mikes and maybe 4 top layer mikes, I just listen to them in stereo very quick to check the signal. And I kind of know what to expect from microphones signal so if it’s sounds good for me, I’m pretty confident that it will work in an immersive mix even if I didn’t hear the immersive mix yet.

And for the radio play kind of stuff, that really is interesting because there's so much more you can try out than in music production. We tried some microphones setup, I tried ORTF, that's a French one, with some double MS stuff and now we have a microphone, it's the BP3600 from Audio Technica. It's eight cardioid, or super cardioid capsules on a kind of omni spherical housing. It's kind of 4 ORTF couples. And that's very nice because the workflow is so easy. It's just one mike stand, one cable, 8 channels and you can even use it outside with a little Rycote wind screen on top of each microphone. And this way, you can record 8 channels of immersive microphone signals from a scene of radio play, or whatever. You can also record ambiances or whatever you like.

V: For example, in radio drama, how do you use the extra channels you have in immersive audio compared to stereo. Will you really take advantage of it, and will you use them to localize some direct sources in the rear for example? Do you really use all the dimensions, or do you keep a frontal image?

L: No definitely not. In radio dramas I definitely use all the 360° area. I think you need to be more precise with your sound design. Maybe precise is not the right word, but you need to be more careful, you need to put more work in the sound design because there's so much more space where stuff could possibly happen, even with ambiances. Usually sound archives have stereo ambiances, and stereo ambiances could work in immersive, and you can also use Halo upmix from Nugen or whatever to make it more spatial but it's kind of boring sometimes. There are so much more possibilities. You can design really nice soundscapes. You can use ambiances, you can use... A duck maybe far left and back and there's some car passing by, there might be some...

Even common misconception I heard a lot in radio plays is the rain. Because everybody thinks rain comes from the tops, so I put the rain signal in the top speakers. But when you think about it where does the sound of rain come from. The sound of rain is water hitting a surface and the surface is never in top of you, maybe if you're in some sort of hall but if you're outside the surface is on the ground. So, the very interesting perspective in immersive audio content is to not only put sound sources on top but even under the listener perspective, and that's a thing that Mpeg-H renderer is able to do. Actually, I do not know how good it sounds. I didn't try it out yet because we use the Dolby stuff to produce our stuff.

But usually of course I use height, I use the surround, and I think movement is far more interesting. Of course, you need to be careful not to overdo it, but I had a Radio play in a swimming hall where children were jumping in the water, and it works quite nicely to jump from the top and then you have the splash and water sound. That was a nice effect. And, like fiction stuff where you had some stuff where there was some sort of dragon I would say. He's tall and big, a little bit like Smaug and it worked very well to lift the voice up top; it was on Front but not in the mid layer but between the mid layer and the upper layer and you get a sound which is somebody taller than you. And also, one, possibility is to, for feature production, so, do you know radio features like real stories, like documentaries? You have translations. And original voices from different sites, which doesn't work in stereo that good. In stereo you always need to back the original voice and then you have the over voice, the translation. But in immersive content you can have them side by side at same level and it's really interesting what happens. If you have some sound sources maybe from the front left original voice and from the

back right or rear right the translation at same level, it works out quite nicely and then you can use your brain on which to focus.

V: You won't be necessarily afraid to use an important narrative sound and to put in in the surround or in the rear?

L: No, not at all.

V: What about in music?

L: It's different kind of things.

V: What would you put in the rear, what would you put in the tops?

L: I tried out some stuff. Usually for acoustic music, you have diffuse sound information coming from the rear and the top, but you can use some stuff in classical recordings like far music, far orchestral stuff, when you have musician off stage. For example, it's use in the Heidelberg symphony from Richard Strauss, when you have some musician off stage, and they play, and usually in the concert hall they play somewhere in the foyer, so you get a sense of somehow distant music. And in immersive you can put them like from the top rear which works quite nicely, so you have the orchestral in front and then the distant orchestra top rear and then you have this kind of dialogue between them, and it works very well.

For the jazz pop folk thing, I really enjoyed expanding the circle to the rears maybe almost a little bit to the surround, from the top a little bit threw the upper speakers, so you present a stage, but the stage is wider than the stereo field. It's pretty more like you're almost right up front to a semi-circle of musician. This kind of imaging could work I think, and also like effects, effects returns, the old guitar trick, or other instruments as well but I usually know it from guitar when you put the dry signal on the left and the wet signal to the right in stereo, you can also do it in immersive so you have a dry signal somewhere where it is naturalistically located in this acoustic recording and then you have maybe an artificial reverb, maybe even a plate or a spring reverb or a delay, and put it in the rears or in the tops or in another direction and you kind of get the sense of a realistic spatial experience but all this effect sounds which also contribute to the sound image are played back from other sources, from other speakers. And I would say it enhances the clarity of the mix but at the same time you get the sound quality and the spatial experience you like about the effect of a delay or a nice reverb.

V: What difficulties do you encounter when you work with immersive audio?

L: The most difficult one, it's a two-sided sort of immersive thing because I said that one benefit is the sound reproduction of the same mix in different playback formats, that's a benefit but that's also a difficulty because even if we love to hear our immersive mixes binaurally, of course there's not the same. You also must do drawbacks, and you always need to find the middle between loudspeaker reproduction and headphone reproduction to do a mix which sounds nice everywhere. It's just not the case. I talked to some developers and asked them about studies about the translation from loudspeaker to headphone and everybody I talked to said "Ok for everybody it's the goal but it's just not possible." It's just so different, so you need to adapt your mix to fit the most reproduction environments. And of course, I would say that's

your job as a mixer. I don't mix for myself; I mix for everybody else. If I would mix for myself, I just would sit in my room, and I would listen to the stuff I do. It's not the case; my job is to make as many people happy as I could and so that's definitely difficult. And also, what I said like the sound coloring which is really difficult. when you used to work in stereo you have like correlation of the two channels mixed into your sound sources is most of the time not such a big problem, of course if you have moody mike setup and stuff like that you need to know what you do to get a nice sound of course and in stereo as well. But the experience you have in stereo in terms of sound coloring and even physical integrity, like the transient response and...

Did you hear about the term physical integrity? It describes the physical, the subjective physical feeling of the energy of the content. Like if you have a drum set, you need to experience this physicality during the playback. If there's a dispropense between the content and the physical subjective feeling when listening to the content your physical integrity is not right. So immersive mixes, because of the huge amount of different sound sources in terms of speakers, it's very difficult to get this correlation and the physical integrity because the sweet spot is so small and you have phase cancelation, and comb filtering effect, and in binaural you have all the effects and drawbacks of the HRTF processing. So, to put it short, when I listen to a dance track and I'm automated to move my body, that's always good sign, and I always get that in stereo, but it's harder to get in immersive I would say. You can do really nice mixes and as I said I'm a big fan of immersive mixes, I prefer most immersive mixes to the stereo mixes but this quality of physical integrity sound coloring and... Yeah. I might describe it as punch as well, it's hard to get right. And for stereo I would say on good speakers you put up a nice record, and you could dance to it and if you listen to it on headphones, you also get the same kind of feeling, but immersive mix kind really fall apart in this kind of quality. It's a spatial experience not a physical body experience anymore. I don't know if that makes sense.

V: And did you find tricks to actually feel it again in immersive.

L: Yeah! Use a lot of LFE! *Laughs*

I would say bass frequencies are always welcome, and in immersive you have the resolution in terms of frequency range, in terms of dynamic range, in terms of overall sound pressure level of the sound reproduction system. For a loudspeaker playback but also for binaural. I would say that LFE channel is your friend. I like to use, that's something maybe it's also for pop and electronic mixes, I think it's also common in that part but in acoustic music it's not that common I would say and I learned that in film mixing, subharmonic synthesis is a great tool to get a really defined LFE signal or bass signal for the LFE channel, and I use this, I rather route like bass instruments, like maybe the kick drum or double bass or a low brass, low percussion or whatever, I rather put a send effect to a subharmonic synthesizer and mix the synthesized sub bass signal into the mix than mixing a low passed signal in the sub channel.

I think it's the DBX 120 was the hardware unit for the 80s or 90s, I never used that, but there are some plugins which are modeled after that, and there's this pro subharmonic plugin from Avid which comes with the ProTools license and there's a plugin called Lowender which is also really nice which I use, and there might be some other ones. That's usually my trick to achieve a sufficient bass response.

And there's another topic which I would love to read in some studies or stuff like that is that even in the best control rooms where I listen to stuff, the Sweet Spot is so small. When you have 12 loudspeakers pointed at you and they are time aligned, phase aligned, it's a one-person experience. And for collaborative work it's really difficult because usually I work in teams of 3 to 5/6 people and when you're sitting in the control room and everybody need to get a sense of the mix, and the content, you need to have some reference and, I would say, immersive playback systems depends on the room but, as I said, the best rooms, acoustically nearly perfect rooms, there are a lot of differences between the different places in the room.

V: There are some studies which says that the sweet spot might be bigger for immersive audio playback systems in a cinema or a concert hall venue context.

L: That's definitely the case. I would say like a discrete loudspeaker setup with 12 speakers, the sound of every speaker needs to arrive at the same time at one point and when you move away from the point all the geometry changes so much. So, you're closer to that speaker and far away from... I don't know... Again, I'm happy to be proved wrong there but that's my experience so far.

Actually, just recently, I listen to some 22.2 contents, NHK, it was crazy in a lot of terms because I was in this listening room which was like a reference listening room for research environment. It was so crazy accurate. I never listened to a speaker setup so accurate. The 22.2 playback was so realistic, and at the same time I told to myself I never want to work in that format, it's tedious, it's so much work, like 24 microphones just for a main microphone array or whatever, it's so much work you have to setup so much stuff, and then it's nice if you listen to it in this kind of listening room but who will ever experience this kind of experience.

V: What was the contents?

L: One was a church with people walking around. And an organ.

[...]

That was also really overwhelming but at the same time I thought about maybe one day there will be some listening rooms, public available listening rooms like cinema, which are public available media rooms somehow, which would be also equipped with nice sound reproduction systems, but for audio only content.