



**PRODUCTION MUSICALE ORIENTÉE OBJET :
ENJEUX D'ÉCRITURE ET MIXAGE POUR L'ÉCOUTE AU CASQUE**

Pierre-Yves COLLARD

Mémoire de Master 2

Spécialité Son – Promotion 2023

Directeur de mémoire interne : Alan BLUM

Directeur de mémoire externe : Martin ANTIPHON

Examinatrice : Mireille FAURE

Référent académique : Laurent MILLOT

Coordinateur des mémoires : Corsin VOGEL

Juin 2023

Résumé

Après plusieurs tentatives échouées de développement du multicanal en musiques actuelles, l'audio 3D, par le biais du mixage orienté objet, semble être au cœur de l'innovation audio actuelle et des enjeux commerciaux de l'industrie musicale. L'arrivée des technologies associées à l'approche objet, telles que le Dolby Atmos Music ou Sony 360 Reality Audio, a permis de révolutionner la transportabilité des mixages et ainsi démocratiser l'accès aux formats immersifs pour le grand public. L'écoute au casque associée à la réduction binaurale des mixages objets représente le moyen le plus facile d'accéder à ces formats sur les plateformes de streaming.

Ce mémoire se propose donc de faire un état des lieux de ces nouveaux formats de production musical avec un questionnement préalable simple : le mixage orienté objet peut-il offrir une nouvelle expérience d'écoute musicale au casque ? Pour ce faire, cette recherche aborde dans un premier temps les aspects théoriques et technologiques des outils de mixage immersifs avant de le confronter aux réalités des pratiques des ingénieur·e·s du son à travers différents entretiens menés avec des professionnel·le·s.

Par la suite, un test d'écoute a été mis en place afin d'évaluer les différences entre des mixages stéréophoniques et immersifs de musiques actuelles pour l'écoute au casque. L'hypothèse avait été faite que les morceaux mixés en Dolby Atmos seraient préférés et considérés comme plus immersifs par les sujets tests. Finalement, l'étude perceptive a montré des résultats significatifs dans la préférence et l'association du critère d'immersion aux mixages stéréophoniques en comparaison des mixages immersifs. Ces résultats, associés à notre recherche théorique et entretiens, permettent d'envisager la nécessité d'amélioration des outils de rendu binauraux actuels pour optimiser le passage enceinte – casque et ainsi envisager des nouvelles expériences d'écoutes.

Mots clés : audio orienté objet, binaural, Dolby Atmos Music, mixage musiques actuelles, écoute au casque.

Abstract

After several failed attempts to develop multichannel in contemporary music, 3D audio, through object-oriented mixing, seems to be at the heart of current audio innovation and the commercial stakes for the music industry. The arrival of object-oriented technologies such as Dolby Atmos Music and Sony 360 Reality Audio has revolutionized the transportability of mixes, and democratized access to immersive formats for the audience. Headphone listening combined with binaural reduction of object-based mixes represents the easiest way to access these formats on streaming platforms.

The aim of this research is to take stock of these new music production formats, with a simple preliminary question: can object-oriented mixing offer a new headphone listening experience? To this end, this research first looks at the theoretical and technological aspects of immersive mixing tools, before confronting the realities of sound engineers' practices through various interviews with professionals.

A listening test was then set up to evaluate the perceptual differences between stereophonic and immersive mixes of pop music for headphone listening. The hypothesis was that tracks mixed in Dolby Atmos would be preferred and considered more immersive by the test subjects. However, the perceptual study showed significant results in the preference and association of the immersion criterion with stereo mixes, compared with immersive mixes. These results, combined with our theoretical research and interviews, point to the need to improve current binaural rendering tools to optimize the speaker-headphone crossover and thus consider new listening experiences.

Keywords : object based audio, binaural, Dolby Atmos Music, pop music mixing, headphone listening.

Remerciements

Merci à Martin Antiphon et Alan Blum pour leur expertise, leurs précieux conseils, leur bienveillance et leur disponibilité dans la direction de ce mémoire.

Merci à Corsin Vogel pour son accompagnement et son suivi méthodologique.

Merci à Oscar Ferran, Baptiste Murguiondo, Tristan Mazire, Bruce Keen et Louise Bardet pour m'avoir témoigné de leur expérience et pratique, et pour l'intérêt porté à mon sujet de recherche.

Merci à Florent Fajole pour sa disponibilité et son aide dans la recherche de références bibliographiques.

Merci à l'ensemble des personnes ayant participé à mes tests perceptifs.

Merci à Émile Juin pour ses précieux contacts.

Merci à l'ensemble de la promotion 2023 pour ces trois riches années passées ensemble.

Merci à Sam, Elsa, Philippine, Yann et Vianney pour leur présence rassurante pendant toute cette période de recherche.

Merci à ma famille pour leur soutien pendant ces longues années d'étude.

Table des matières

Résumé.....	1
Abstract.....	2
Remerciements	3
Table des matières	4
Introduction	8
Première partie : État de l'art de la production musicale spatialisée	11
A] Historique de la spatialisation musicale	11
A.1. Des débuts de la diffusion sonore.....	11
A.2. À l'arrivée de la stéréophonie.....	12
A.3. Vers des systèmes non-conventionnels : prémices du son 3D.....	14
A.3.A. 1950 : L'apport de la musique concrète et électroacoustique	14
A.3.B. 1970 - 2000 : Tentatives de nouveaux formats en musiques actuelles.....	15
A.3.C. De 1973 aux années 2000 : le modèle du champ d'onde	15
B] Le développement de la production musicale orientée objet et l'arrivée de l'audio de prochaine génération (Next Generation Audio)	17
B.1. Audio orienté objet : principes et définitions	17
B.2. Vers un format standard pour une compréhension commune : l'ADM.....	19
B.3. Cas du Dolby Atmos Music	21
B.3.A. État des lieux	21
B.3.B. Dispositif et principe de fonctionnement.....	21
B.3.C. Les outils de postproduction.....	23
B.4. L'importance des codecs pour la transportabilité des mixages	27
C] Vers une écoute majoritaire au casque : l'enjeu de l'écoute binaurale.....	30
C.1. Principes et perspectives de l'écoute au casque.....	30
C.1.A. La perception du son spatial	30
C.1.B. L'écoute stéréophonique au casque : problématique de la latéralisation et internalisation	32
C.1.C. L'écoute binaurale : le potentiel de l'externalisation	33
C.1.D. La technologie binaurale.....	35
C.2. Problématiques des HRTF individualisés et non individualisés pour l'écoute musicale en synthèse binaurale	36
C.2.A. Cas des HRTF génériques.....	36
C.2.B. Cas des HRTF individualisés et personnalisés.....	36
C.2.C. La perspective du tracking	37

C.3. L'importance du choix du moteur de rendu binaural	38
C.3.A. Dolby Renderer : AC4-IMS.....	39
C.3.B. Apple Spatial Audio : DD+JOC	40
C.3.C. Autres solutions	42
Deuxième partie : Pratiques et évolution des méthodes de création et postproduction.....	44
A] Des possibilités de construire une nouvelle scène sonore : l'approche théorique....	44
A.1. L'héritage du mixage stéréophonique	44
A.2. Une nouvelle esthétique musicale pour le son spatialisé.....	46
B] Évolutions des pratiques liées à l'arrivée du format orienté objet en musiques actuelles	49
B.1 Réalités actuelles des possibilités esthétiques d'une production musicale orienté objet	49
B.2. Évolutions de la chaîne de production musicale	51
B.3. Évolutions des pratiques de mixages.....	53
B.3.A. La mise en espace des sources sonores.....	53
B.3.B. Vers la fin de la loudness war	55
B.4. Dans la gestion du monitoring	57
C] Conclusion partielle.....	61
Troisième partie : Évaluation perceptive des différences entre mixage stéréophonique et immersif pour l'écoute au casque en musiques actuelles	62
A] Contexte.....	62
B] Méthodologie du test perceptif.....	63
B.1. Définition des paramètres d'étude.....	63
B.1.A. Hypothèses d'étude.....	63
B.1.B. Choix des sujets.....	63
B.1.C. Variables.....	64
B.2 Choix des titres.....	64
C] Réalisation du test.....	67
C.1. Présentation Générale.....	67
C.2. Protocole	68
C.3. Limites du test	71
D] Analyse et interprétation	72
D.1. Analyse du critère préférence.....	72
D.1.A. Préférence sur l'ensemble des titres	72
D.1.B. Préférence individuelle par titre.....	73

D.2. Analyse du critère immersion.....	75
D.2.A. Immersion sur l'ensemble des titres.....	75
D.2.B. Immersion individuelle par titre.....	76
D.3. Profil des sujets	79
D.4. Interprétations.....	79
Conclusion Générale	84
Références	87
Table des figures.....	93
Annexes.....	94
Annexe 1 : Normes de calibration pour un studio Atmos	94
Annexe 2 : Guide d'entretien et formulaire de consentement	95
Annexe 3 : Entretien avec Baptiste Murguiondo.....	97
Annexe 4 : Entretien avec Oscar Ferran.....	104
Annexe 5 : Entretien avec Martin Antiphon.....	111
Annexe 6 : Entretien avec Bruce Keen	116
Annexe 7 : Entretien avec Tristan Mazire	121
Annexe 8 : Mesures des niveaux des extraits utilisés dans le test perceptif	130
Annexe 9 : Questionnaire perceptif.....	131
Annexe 10 : Résultats de l'évaluation du critère préférence.....	133
Annexe 11 : Résultats de l'évaluation du critère immersion	138
Annexe 12 : Résultats d'ANOVA pour le critère « préférence ».....	143
Annexe 13 : Résultats d'ANOVA pour le critère « immersion ».....	146

Introduction

La question de la répartition du son dans l'espace est commune aux différents médiums artistiques sonores et plus que jamais d'actualité depuis ces dernières années. Là où le cinéma a été un champ d'expérimentation pour la spatialisation sonore depuis plusieurs décennies, la production musicale populaire est ancrée depuis longtemps dans le modèle de la stéréophonie. Pour autant, depuis quelques années, ce champ d'expérimentation s'ouvre à la musique avec la notion « d'audio spatial » ou « audio immersif » et à l'arrivée du mixage orienté objet.

Après plusieurs tentatives échouées de développement de nouveaux formats : quadriphonique dans les années 70 (Boren, 2018), puis *surround* dans les années 2000 (Stroud, 2006), l'audio 3D semble être au cœur de l'innovation audio actuelle et des enjeux commerciaux de l'industrie musicale. Le mixage orienté objet offre la possibilité de concevoir un nouvel espace sonore, renouvelant les notions de spatialisation actuelle où chaque canal audio vient alimenter un ou plusieurs haut-parleurs du système de diffusion. Il est désormais possible de venir positionner des sources sonores dans un espace en trois dimensions et d'adapter le rendu aux différents systèmes de diffusions, répondant ainsi aux problématiques de transportabilité rencontrées par les professionnels du mixage. L'évolution du modèle de production musicale et des pratiques de mixage donne lieu à l'émergence d'une nouvelle étape d'*upmix* réalisée en studio et correspondant à l'adaptation du mixage stéréophonique en un mixage orienté objet, en vue d'une diffusion immersive.

Deux principales technologies tentent aujourd'hui de s'intégrer au domaine de la postproduction musicale et sur les plateformes de streaming : Dolby Atmos Music et Sony 360 Reality Audio. Suivant cette dynamique, de plus en plus de studios s'équipent de nouvelles cabines de mixages adaptées et calibrées pour de l'audio immersif afin de répondre aux demandes de production nombreuses. A titre d'exemple, Apple Music propose plusieurs milliers de titres en Dolby Atmos depuis le déploiement du format sur sa plateforme en juin 2021 et ne cesse d'agréments ses collections avec des sorties d'albums proposés à la fois en version stéréophonique et Atmos (Singleton, 2022). Les plateformes de streaming représentent près de 51% du temps consacré à l'écoute de musique dans le monde (IFPI,

2022). Elles sont donc aujourd'hui le principal moyen de diffusion pour ces nouveaux formats de diffusion.

Pour autant, l'accès à un système immersif multicanal composé de plusieurs haut-parleurs, dont plusieurs en élévation, semble aujourd'hui quasi impossible pour une utilisation domestique d'un-e l'auditeur-ice lambda. Devant l'impossibilité de commercialiser de telles configurations d'écoutes, les mixages immersifs disponibles sur les plateformes de streaming ont donc pour finalité, dans la très grande majorité des cas, d'être écoutés au casque dans une version binaurale¹ issue de la réduction du mixage sur enceintes. Grâce à l'adaptabilité et à la possibilité de rendus multiples issus des technologies d'audio orienté objet, la solution d'écoute au casque représente en soit une quasi révolution dans la possibilité d'expérimenter aisément un format immersif pour le grand public.

Ce mémoire se propose donc de faire un état des lieux de ces nouveaux formats de production musicale, en étudiant leur aspect technologique d'une part, en questionnant leurs pratiques et usages dans la chaîne de production d'un titre, et en s'interrogeant sur leur potentiel esthétique dans la restitution du discours musical. Notre recherche sera centrée sur la problématique de la restitution binaurale pour l'écoute au casque que l'on considère primordiale car accessible par le plus grand nombre d'auditeur-ice.

En ce sens, nos questionnements initiaux préalable à l'élaboration de cette recherche se révèlent simples : le mixage orienté objet peut-il offrir une nouvelle expérience d'écoute musicale au casque ? De quelle manière optimiser la réduction binaurale pour un mixage musical orienté objet ?

Dans le cadre de notre recherche et compte tenu du temps imparti pour sa réalisation, il est apparu crucial de se concentrer sur un système en particulier plutôt que de prétendre à une étude exhaustive de tous les formats orientés objet. Nous avons choisi de nous orienter vers le Dolby Atmos Music qui est actuellement la technologie la plus répandue et possédant le monopole de la postproduction immersive dans le domaine des musiques actuelles. Notre

¹ Reconstitution d'une image sonore tridimensionnelle au casque.

recherche se veut à la fois théorique et pratique afin d'envisager un panorama global de ces technologies et de pouvoir se forger un esprit critique vis-à-vis de notre problématique. En complément à notre étude bibliographique, nos propres écoutes et essais techniques, il a été choisi de confronter notre recherche à l'expérience de professionnels par le biais d'entretiens, afin de pouvoir témoigner des réalités de la profession, de leur pratique et expérience vis à vis d'une technologie très récente.

Dans un premier temps, nous aborderons l'approche historique de la spatialisation du son dans le domaine musical afin de mieux envisager l'apparition des nouveaux formats orientés objet. Cette partie nous permettra également d'aborder d'un point de vue technique le fonctionnement des technologies associés au mixage orienté objet et d'envisager les enjeux associés à l'écoute binaurale au casque.

Par la suite, nous examinerons les évolutions spécifiques à l'étape de création et de postproduction engendrées par l'introduction du mixage orienté objet. Nous questionnerons le potentiel esthétique apporté par ces technologies d'un point de vue théorique puis pratique par le biais d'entretiens menés auprès de professionnels. Ces entretiens nous permettront également d'aborder les évolutions des pratiques chez les ingénieur·es du son, et d'envisager de quelle manière la problématique de l'écoute binaurale est considérée dans le *workflow* d'un mixage orienté objet.

Enfin, nous confronterons nos questions de recherche et l'ensemble de nos observations à une expérience d'écoute concrète, au travers de tests perceptifs. Par l'intermédiaire de trois hypothèses issues du travail réalisé dans les deux premières parties, nous tenterons de comparer différents critères d'évaluation entre des mixages stéréophoniques et binauraux issus de mixages immersifs. L'analyse de ces tests nous permettra éventuellement d'identifier des tendances de réponses par rapport à nos questions de recherche et d'ouvrir des perspectives pour des études futures.

Première partie : État de l'art de la production musicale spatialisée

A] Historique de la spatialisation musicale

A.1. Des débuts de la diffusion sonore

L'histoire de la musique a montré que le résultat d'une création artistique dépend de manière significative du contexte sonore dans lequel elle est conçue, ainsi que de l'appropriation des créateurs aux nouvelles technologies qui y sont intégrées (Byrne, 2013). Tout au long du XXème siècle, l'invention de nouveaux formats audio et systèmes de diffusion a permis de développer la recherche tant technique qu'artistique sur les enjeux de spatialisation de la musique. Situé à l'intersection de problématiques partagées entre ingénieurs, producteurs, et artistes, ce développement technologique a toujours eu comme fil rouge de donner à améliorer les qualités d'écoute, et d'offrir un sentiment de réalisme et d'immersion sonore.

Pour autant, l'intérêt des compositeur·ices pour l'espace et le mouvement a débuté bien avant les débuts de la diffusion sonore. La réflexion sur la spatialisation des sources sonores était alors associée au lieu de diffusion de l'œuvre. La musique qui s'est développée entre 1540 et 1610 dans la basilique Saint-Marc de Venise est un exemple de réflexion sur la spatialisation des musiciens au sein d'un espace architectural. Des compositeurs de l'époque comme Adriaan Willaert ont développé un style d'écriture polychorale, où les différents chœurs étaient séparés géographiquement au sein de l'église (*cori spezzati*), de manière à tirer parti des qualités architecturales de l'église et créer des effets de retard, échos, imitation, qu'on pourrait déjà assimiler à un effet stéréophonique (Moretti, 2004). Cette écriture est ensuite poursuivie à Rome au XVIIe siècle pendant l'époque du Baroque Colossal avec des compositeurs tels que Orazio Benevolo (Dixon, 1979). Bien que l'époque classique aseye l'aspect frontal de la scène sonore en positionnant l'orchestre face à l'auditeur·ice, les compositeur·ices de cette période continuent de développer l'intérêt pour la spatialisation des sources. La partition de l'opéra *Don Giovanni* (1787) de Mozart comprend par exemple des parties pour trois orchestres séparés : l'un dans la fosse, un autre sur scène et le dernier

dans les coulisses. Chacun de ces groupes joue des morceaux très différents, avec des mesures distinctes, nécessitant une coordination temporelle extrêmement précise (Begault, 1990). Enfin, l'époque romantique poursuit cette idée de séparation et mise en espace des différents instruments en expérimentant sur des programmes symphoniques pour orchestre et chœur, comme dans les *Requiem* de Berlioz (1837) et Verdi (1874) (Broderick, 2012).

A.2. À l'arrivée de la stéréophonie

Les prémices de la stéréophonie remontent à 1881, où Clément Ader utilisait pour la première fois un système à deux canaux dans son Théâtrophone. Le principe était le suivant : une série d'émetteurs téléphoniques était reliés de la scène de l'Opéra de Paris aux salles de l'Exposition Électrique de Paris. Les auditeurs de ces salles pouvaient alors entendre les spectacles à l'aide d'un récepteur téléphonique pour chaque oreille. Ils percevaient la position des sources sonores grâce aux indices interauraux transmis sur les lignes téléphoniques. La naissance de la stéréophonie en tant que telle pour les supports de diffusion est attribuée à Alan Blumlein, ingénieur chez EMI, en 1931 (EMI Archive Trust, s. d.). Il théorise alors le concept de « binaural audio », appelé par la suite stéréophonie. Il y associe une technique de prise de son et une nouvelle méthode de coupe du disque de gramophone, avec deux sillons pouvant être lus simultanément pour restituer le son stéréophonique, tout en restant compatible avec les disques monophoniques existants. Le brevet de Blumlein a eu un impact significatif en ouvrant un large éventail de possibilités pour les modifications que la technologie stéréophonique pourrait apporter au domaine de l'audio.

Son système fut repris et commercialisé dans les années 1950 par Bell Labs et Westrex Corporation. Les équipements monophoniques, en grande partie des radios, sont alors majoritaires à cette époque. Le développement de la stéréophonie s'est réalisé en partie grâce à une promotion marketing intense et agressive de la part des différents constructeurs et distributeurs du secteur musical. La stéréophonie était alors présentée comme la dernière avancée technologique audio qui met la reproduction sonore sophistiquée à la portée de tous·tes.

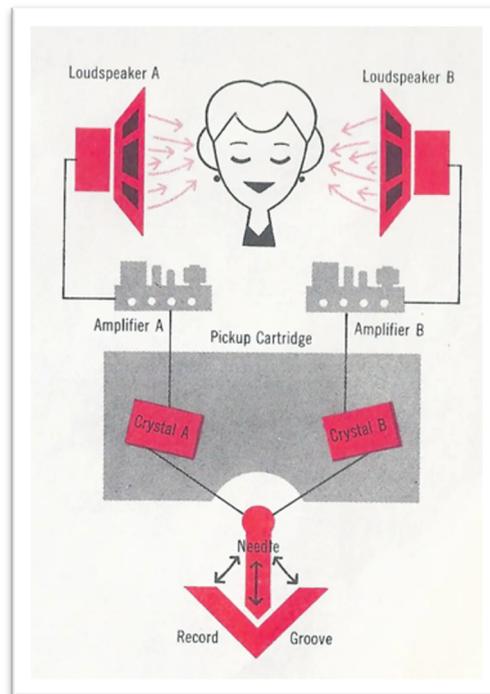


Figure 1 : Principe de fonctionnement de la stéréophonie telle que présentée dans une pochette de RCA

(Collection of Janet Borgerson and Jonathan Schroeder)

Des disques de démonstration ont même été développés par les maisons de disques pour faire percevoir le potentiel de spatialisation : « Sounds in Space » chez RCA ou encore « The Stereo Disc » chez Capitol Records. L'auditeur-ice pouvait alors entendre la voix d'un narrateur passant d'une enceinte à l'autre, un train en mouvement, des paysages sonores, ou encore des enregistrements de musique classique. De plus, pour répondre aux différents systèmes d'écoute dans la période de transition entre monophonie et stéréophonie (de 1950 à 1960), les albums étaient alors commercialisés en deux versions : mono et stéréo. Peu à peu, de la fin des années soixante jusqu'à aujourd'hui, la stéréophonie est devenue la norme, le logo n'était alors plus indiqué sur les pochettes.



Figure 2 : Album de démonstration de la technologie stéréo (RCA Records, 1958)

A.3. Vers des systèmes non-conventionnels : prémices du son 3D

A.3.A. 1950 : L'apport de la musique concrète et électroacoustique

L'organisation du son dans l'espace a été au centre de la réflexion sur la restitution des musiques en concert. Dans le domaine de la musique contemporaine, la seconde moitié du XXème siècle en France a été marquée par une recherche poussée sur l'écriture musicale spatiale, « la notion de relief, de mouvements et d'ubiquités sonores » (Périaux, 2015). Les compositeur-ices ont alors inventé de nouveaux systèmes de diffusion multicanaux. À l'époque, la notion d'espace est intimement liée à la disposition des enceintes dans le lieu de restitution. En 1951, Pierre Schaeffer et Pierre Henry composent *Symphonie pour un homme seul*, une pièce de musique concrète comprenant quatre canaux diffusés respectivement dans cinq haut-parleurs à droite, à gauche, derrière et au-dessus des membres de l'auditoire. Le nombre de haut-parleurs n'a ensuite cessé d'augmenter dans les compositions de cette période. Stockhausen compose la première pièce électroacoustique pour un système quadriphonique en 1960 avec *Kontakt*, à propos de laquelle il dira :

« Changer dans le sens des aiguilles d'une montre ou dans le sens inverse, être à gauche à l'arrière ou à l'avant, ou alterner dans un dialogue entre l'arrière gauche et l'avant droit, sont autant de configurations dans l'espace qui ont autant de sens que les intervalles dans la mélodie et l'harmonie. » (Stockhausen, 1972)

Cette quête de la diffusion multipoint atteint son sommet lors de l'Exposition Universelle de Bruxelles en 1958, avec la diffusion du *Poème Électronique* d'Edgar Varèse au Pavillon Philips avec près de 325 haut-parleurs et 20 amplificateurs. Ces différentes évolutions en matière de composition et diffusion ont été en grande partie rendues possible grâce aux apports scientifiques des groupes de recherches tels que le Groupe de Recherche Musical (GRM) puis plus tard par l'Institut de recherche et coordination acoustique/musique (IRCAM).

A.3.B. 1970 - 2000 : Tentatives de nouveaux formats en musiques actuelles

La première tentative de format autres que stéréophoniques pour les musiques actuelles a vu le jour dans les années soixante-dix. En 1968, le premier système quadriphonique fut développé par Peter Scheiber (Davis, 2003), exploité par les Pink Floyd en 1973 avec la version quadriphonique de *Dark Side of the Moon*. Malgré l'effort marketing mis en place, le format quadriphonique n'a pas connu de succès commercial, entraînant avec lui une perte de confiance généralisée dans l'avenir de la technologie du son 3D (Boren, 2018). Dans les années 2000, l'arrivée de nouveaux formats tels que DVD audio et le SACD permettent le développement du format 5.1. En musique actuelle, des labels et des artistes tels que Bjork ou Peter Gabriel produisent alors des albums dans ce nouveau format. La production en format 5.1 pour la musique n'a pas réussi à se populariser auprès du grand public, principalement en raison des coûts élevés des équipements requis pour une écoute optimale, ainsi que de la nécessité d'un environnement d'écoute approprié pour pouvoir profiter pleinement de l'expérience sonore (Stroud, 2006). Ce format sera vite stoppé par la crise du disque, l'arrivée du format MP3 et la popularité croissante des plateformes de partage de fichiers en ligne telles que Napster. Le 5.1 deviendra par la suite un format niche, avec une majorité de production phonographique pour la musique classique. De 2005 à 2018, les Grammy Awards récompensent les meilleurs albums de musique mixés en 5.1 ou 7.1 dans la catégorie *Best Surround Sound Album*. Depuis 2019, cette catégorie est remplacée par le *Best Immersive Audio Album*, regroupant les mixages surround et orientés objets (Dolby Atmos Music, Sony 360 RA).

A.3.C. De 1973 aux années 2000 : le modèle du champ d'onde

En 1973, suite à l'échec de la quadriphonie, Michael Gerzon, ingénieur britannique, et son équipe de recherche de l'université de York s'emparent du modèle du champ d'onde pour

développer un nouveau format de diffusion : l'ambisonie. Dérivé des travaux de Huygens, ce format permet alors pour la première fois d'être indépendant de la configuration d'enceintes présentes dans le système de diffusion (Gerzon, 1973).

« Des genres de musique tels que l'ambient ou le dub offrent de nombreuses nouvelles possibilités quant à la manière dont les échos et les réverbérations peuvent être gérés sur la scène sonore, et des innovations inattendues dans les styles de production en studio sont à prévoir en réponse à l'utilisation de l'ambisonie. »
(Holman, 2007)

Les sources sonores sont traitées pour être diffusées sur un système de haut-parleurs disposés autour de l'auditeur·ice, généralement en forme de sphère. L'ambisonie permet de créer une expérience d'écoute spatialisée réaliste, et permettant de reproduire les caractéristiques de l'environnement acoustique dans lequel les sources ont été enregistrées. Cette technique est très utilisée aujourd'hui, dans des domaines tels que la production de musique, les jeux vidéo et la réalité virtuelle.

Plus tard dans les années 2000, aux Pays-Bas, naît la WFS, un format où chaque source sonore est reproduite en temps réel à partir d'un ensemble de haut-parleurs disposés sur un plan horizontal ou dans un espace tridimensionnel. La WFS nécessite en théorie une grande quantité de haut-parleurs pour produire une image sonore réaliste et stable. Dans les faits, une bonne stabilité des sources peut être obtenue avec elle un nombre raisonnable d'enceintes comme c'est le cas à Avignon ou à La Scala Paris. La WFS est souvent utilisée pour la sonorisation de grands espaces, tels que les salles de concert ou les théâtres, où elle permet de créer une expérience d'écoute spatialisée immersive pour les spectateur·ice·s. La WFS permet de capturer ou de créer une scène sonore tout en conservant les informations spatiales relatives à la distance et à la direction des différentes sources qui la composent (Berkhout et al., 1993). L'auditeur·ice peut alors être en mouvement sans voir la scène sonore modifiée.

L'ambisonie comme la WFS sont deux techniques de plus en plus utilisées dans les domaines du spectacle vivant et de la création contemporaine. L'espace de projection de l'IRCAM qui a récemment rouvert est un des lieux où ces deux systèmes sont présents.

B] Le développement de la production musicale orientée objet et l'arrivée de l'audio de prochaine génération (*Next Generation Audio*)

B.1. Audio orienté objet : principes et définitions

L'EBU (*European Broadcast Union*) propose deux définitions au principe d'audio orienté objet :

La première, plus générique, caractérise le fait d'associer à de l'audio des métadonnées décrivant son contenu. Cette première définition a émergé de fait de la complexité grandissante du nombre de canaux à gérer dans les nouveaux systèmes audio immersifs. Le format 5.1 présentait déjà différents protocoles pour l'organisation des canaux audio. L'augmentation du nombre de canaux et haut-parleurs associés (formats 6.1, 7.1, 7.1.2 ...) a donc nécessité d'introduire des métadonnées pour labelliser chaque canal audio, et ainsi assurer sa transmission jusqu'au haut-parleur associé.

La seconde définition caractérise un modèle audio précis parmi cinq modèles définis par l'EBU² :

1. **Modèle orienté canal** (*channel based*) : ce modèle caractérise les formats audio les plus classiques, où chaque canal audio est associé directement à un haut-parleur (formats mono, stéréo, 5.1 ...). La configuration du système de restitution est déterminée par le nombre de canaux utilisés. Ce modèle est le plus répandu en raison de sa simplicité d'utilisation et de son fonctionnement transparent.
2. **Modèle orienté scène** (*scene based*) : développé par Michael Gerzon, ce modèle se base sur une représentation du champ acoustique sous la forme d'harmoniques sphériques allant de 1 jusqu'à 7 pour les ordres les plus importants, voir plus élevé encore pour la partie de traitement du signal. Ce modèle est utilisé pour les formats *Ambisonic* et *HOA (High Order Ambisonics)*. Les sources sonores sont placées dans un repère à l'aide de coordonnées *xyz* ou *aed* (azimut, élévation, distance). La résolution spatiale dépend du nombre de canaux mis en jeu : plus l'ordre est élevé, meilleure

² https://adm.ebu.io/background/audio_types.html

sera la résolution (Gerzon, 1976). Dans les faits, les applications disponibles aujourd'hui fonctionnent à l'ordre 2 pour YouTube, 7 pour Protools³ et jusqu'à 7 pour le Spat ou l'archivage des œuvres à l'IRCAM.

3. **Modèle orienté matrix** (*matrix based*) : ce modèle décrit la combinaison de canaux audio pour en générer des nouveaux. Le format *Mid Side* (MS) est un bon exemple de ce type de combinaison : le canal central (*Mid*) correspond à la somme des canaux gauche et droit, alors que les canaux latéraux correspondent à la différence des canaux gauche et droit.
4. **Modèle orienté binaural** (*binaural based*) : ce modèle correspond au format binaural pour une écoute au casque (les principes de l'écoute binaurale sont abordés plus loin dans cette étude).
5. **Modèle orienté objet** (*object based*) : c'est le modèle qui nous intéresse dans le cadre de notre étude, il caractérise l'association de propriétés spatiales à une source sonore. Le principe de l'audio orienté objet a d'abord été développé par l'Université de Delft (Sporer et al., 2001) puis, d'un point de vue commercial au cinéma, avec Dolby Atmos comme format le plus connu dans les années 2010. Dans le cas du format objet, une source sonore peut être placée dans un espace en trois dimensions, indépendamment du système audio utilisé pour sa restitution. À chaque source sont associées des métadonnées caractérisant ses propriétés spatiales (position, taille, dispersion). Chaque source devient ainsi un élément unique de la scène sonore, pouvant être traitée de manière indépendante les unes des autres. L'un des avantages de ce modèle réside dans son adaptabilité aux différents systèmes d'écoutes. L'algorithme de rendu, appelé *renderer*, a pour rôle de transcrire les caractéristiques spatiales de chaque source au système audio de restitution. Un seul fichier est utilisé pour s'adapter aux différents systèmes d'écoute (haut-parleurs, casque...), permettant ainsi une indépendance et une flexibilité dans la restitution.

³ Pro Tools 2023.6

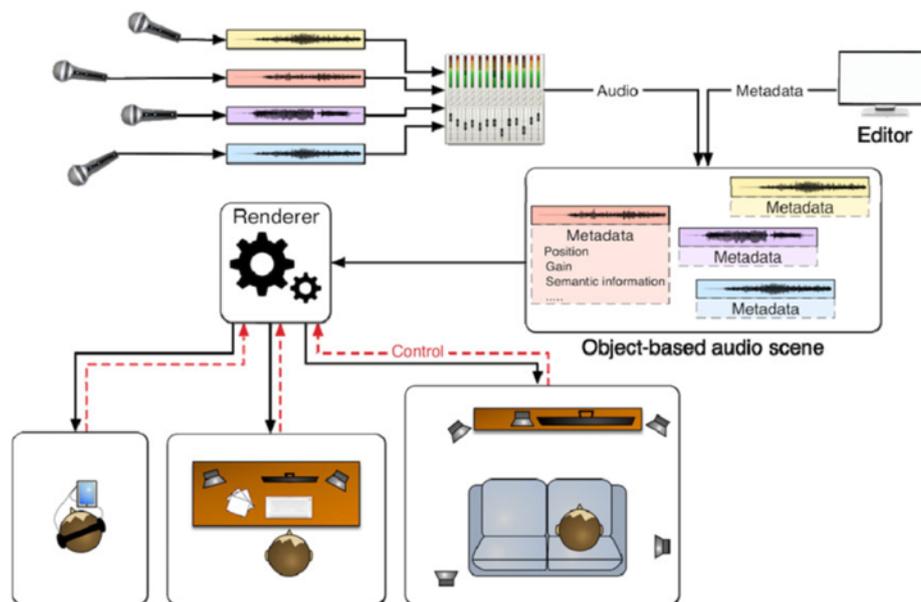


Figure 3 : Représentation de l'adaptabilité du modèle orienté objet (Colin, 2019)

B.2. Vers un format standard pour une compréhension commune : l'ADM

Nous avons pu voir les cinq modèles correspondant aux nouveaux types d'audio définis par l'EBU et la multiplicité de technologies qu'ils impliquent. L'une des questions importantes qui émerge au vu de ce nombre important de modèles est : quel format utiliser pour assurer une transportabilité des différents contenus audio et la bonne compréhension des métadonnées par les différents algorithmes de rendu ?

Pour tenter de répondre à cette problématique, l'EBU a développé avec le soutien de la BBC, France TV, Radio France et Fraunhofer, un format open source nommé ADM : *Audio Definition Model*⁴. Ce standard européen, *open-source*, vise uniquement à formaliser la description des contenus audio (les métadonnées). Il ne définit pas un format permettant de transporter l'audio. L'ADM se présente sous la forme d'un fichier XML, contenant deux sections :

- Le contenu (*content*), décrivant ce que contient l'audio (langue des dialogues, volume sonore, etc.),

⁴ Recommendation ITU-RBS.2076-1,2017

- Le formatage de métadonnées (*format*), décrivant la nature technique de l'audio afin qu'il puisse être décodé ou rendu correctement.

Le schéma ci-dessous illustre la division en deux parties du modèle ADM (contenu et format) ainsi que les liens entre les différents éléments faisant partie de ces sections.

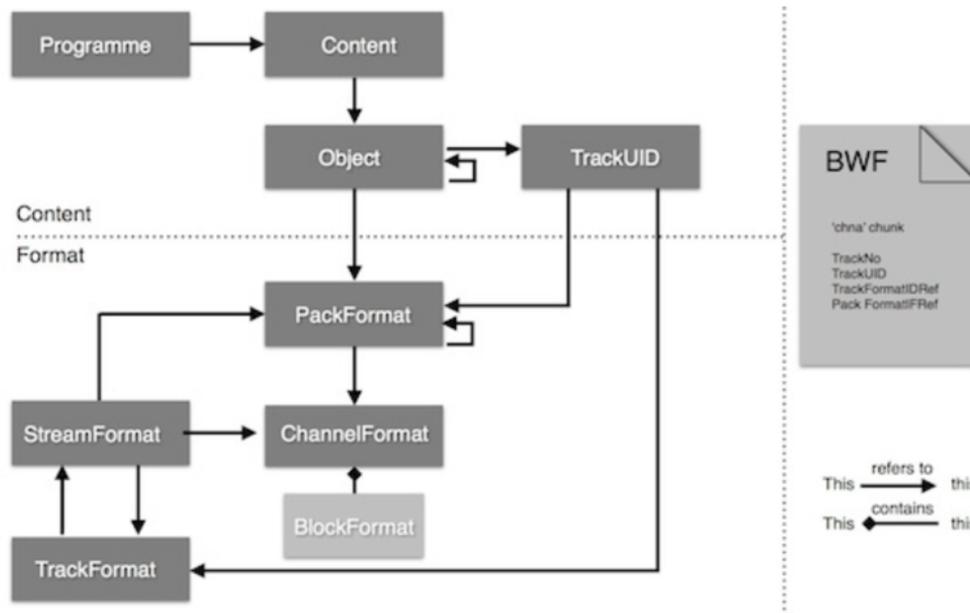


Figure 4 : Représentation du modèle ADM (Marston, 2015)

Une subtilité nécessaire à comprendre est que l'ADM ne donne pas d'instruction sur la manière dont doit être rendu l'audio. En effet, cette mission fait partie du rôle du *renderer* : déterminer, à partir des informations contenues dans le fichier ADM, comment transcrire le mieux possible l'audio au système d'écoute comme pour le passage d'un mix en 7.1.2 à une réduction binaurale par exemple. L'ADM est donc un ensemble de métadonnées rassemblées dans un fichier XML (nommé '*axml*' chunk) associé aux formats audio avec lesquels il est compatible : le BWF (*Broadcast Wave Format*) et BW64. Il est important de noter que les cinq modèles définis dans la partie précédente sont combinables entre eux dans l'ADM, mais les formats propriétaires n'admettent qu'une base multicanale et un certain nombre d'objets dépendant des systèmes (Messonnier et al. 2022). À titre d'exemple, les données de rendu

binaurales renseignées dans le format ADM de Dolby ne pourront pas être lues à travers le Spat de l'IRCAM.

B.3. Cas du Dolby Atmos Music

Cette partie se consacre à la technologie Dolby Atmos, qui représente l'une des applications de l'audio orienté objet la plus utilisée en postproduction pour les musiques actuelles.

B.3.A. État des lieux

D'abord introduite au cinéma en 2012, la firme Dolby a lancé en 2019 Dolby Atmos Music, ouvrant ainsi la voie à la commercialisation du format audio 3D sur les plateformes de streaming (Apple Music, Tidal, Amazon Music). On compte désormais plus de 400 studios équipés de cette technologie dans le monde⁵, un nombre qui ne cesse d'augmenter au vu du déploiement actuel de cette technologie sur les différents services de streaming. Depuis 2021, Apple Music propose l'un des plus gros catalogues d'audio 3D, avec des chiffres de consommation reflétant la rapide évolution du format Atmos. A titre d'exemple, en 2022, 40% des plus grosses sorties sur Apple Music sont disponibles en audio spatial, et la plateforme a enregistré une augmentation de 50% du nombre d'auditeurs écoutant des titres en audio spatial (Singleton, 2022).

B.3.B. Dispositif et principe de fonctionnement

Cette technologie se base sur l'augmentation du dispositif surround 7.1, avec l'ajout d'enceintes en élévations. Bien qu'une multiplicité de configuration d'enceintes soit possible en fonction de la taille de la salle de mix, Dolby recommande un minimum de douze canaux : 7.1.4 (Dolby, 2021). Ce dispositif correspond à la configuration suivante : sept enceintes dans le plan horizontal, un canal de *sub* et quatre enceintes dans le plan vertical. Concernant le calibrage d'un tel dispositif, Dolby requiert un niveau de 85 dB SPL au point d'écoute pour chaque haut-parleur avec au moins 20 dB de *headroom*⁶. Le calage se fait principalement par alignement temporel et égalisation des différents haut-parleurs. L'ensemble des détails concernant le calibrage et les dimensions minimum pour une cabine de mixage certifiée Atmos sont renseignés en Annexe 1.

⁵ <https://professional.dolby.com/music/dolby-atmos-music-studios/>

⁶ Marge disponible entre le niveau maximal du signal (niveau crête) et le 0 dBFS

Le *panning*⁷ des sources à travers l'espace en trois dimensions se fait grâce à un algorithme basé sur la technologie LBAP (Niklasson, 2021) : *layer based amplitude panning* (panoramique d'amplitude par couche). Le principe du LBAP repose sur une superposition de couches de VBAP 2D (*vector based amplitude panning*) selon l'élévation des haut-parleurs. Pour une même élévation (une même couche), le panning de la source est réalisé en manipulant le gain des signaux acheminés vers les deux haut-parleurs les plus proches d'une source sonore virtuelle. Pour le cas du panning d'une source entre différents niveaux d'élévation, un fondu est appliqué entre les deux couches les plus proches. Le LBAP a probablement été choisi par Dolby car il offre un meilleur rendu que le VBAP pour des systèmes non uniformes⁸. Il est important de noter que le Dolby Atmos, à l'inverse d'autres technologies orientées pour le live telles que L-ISA de L-Acoustic, ne propose pas l'utilisation d'un effet de salle (*room engine*)⁹. Il s'agit donc principalement d'un moteur de rendu utilisant un panoramique d'amplitude, dépourvu d'effet de salle. L'ajout de réverbérations, éventuellement multicanales, devient donc nécessaire pour créer des effets de distance notamment.

La technologie Dolby Atmos fonctionne avec deux types de sources audio : les *bed* et les objets. Les *bed* correspondent au modèle orienté canal : chaque source est associée à un canal alimentant un des haut-parleurs, leur position est donc fixe dans l'espace. Les objets correspondent à une source sonore monophonique ou stéréophonique, pouvant être placée dans tout le champ acoustique, indépendamment de la position des haut-parleurs. Une source objet peut être statique ou dynamique (en mouvement), et jouer sur une ou plusieurs enceintes à la fois. Pour une même position et niveau donnés, il n'y a pas de différence sonore perceptible pour une source traitée comme un objet ou un *bed*. La différence et les avantages de l'un par rapport à l'autre se jouent sur l'étape d'encodage et de décodage. Comme nous l'avons vu dans la partie précédente, des métadonnées sont associées à chaque type de source audio pour permettre à l'algorithme de rendu, le *Dolby Renderer*, d'optimiser le

⁷ Placement d'une source dans l'espace sonore

⁸ https://doc.flux.audio/en_US/spat_revolution_doc/Spatialisation_Technology_Panning_Algorithms.html

⁹ Simulation d'un espace acoustique virtuel en fonction du positionnement de la source dans la scène sonore

contenu audio au dispositif d'écoute. L'ingénieur-e du son a la possibilité d'utiliser un total de 128 pistes (jusqu'à 118 objets et un *bed* 10 canaux).



Figure 5 : Schéma bloc simplifié des modèles audio utilisés dans Dolby Atmos

(Dolby, 2013)

B.3.C. Les outils de postproduction

Deux principaux outils sont utilisés dans la chaîne de production Dolby Atmos : le *music panner* et le *renderer*. Nous détaillerons ici brièvement les fonctionnalités de chacun afin de donner les clés de compréhension à leur utilisation pour la suite de cette étude.

Le Dolby Atmos Music Panner

Ce plug-in, intégré directement dans la station de travail audio numérique, permet de positionner les différentes sources sonores dans un espace en trois dimensions. C'est le premier plug-in Dolby développé exclusivement pour une intégration dans un *workflow* de production musicale. Il prend en compte des sources monophoniques ou stéréophoniques et permet de définir leurs positions selon trois axes x y z, respectivement pour le plan horizontal, vertical et l'élévation. La taille de chaque source (*size*) est également paramétrable, cela influe sur l'encodage des sources comme nous le verrons dans la partie suivante. L'ensemble des paramètres spatiaux peuvent être automatisables et sont ensuite intégrés dans les métadonnées de chaque objet sonore.

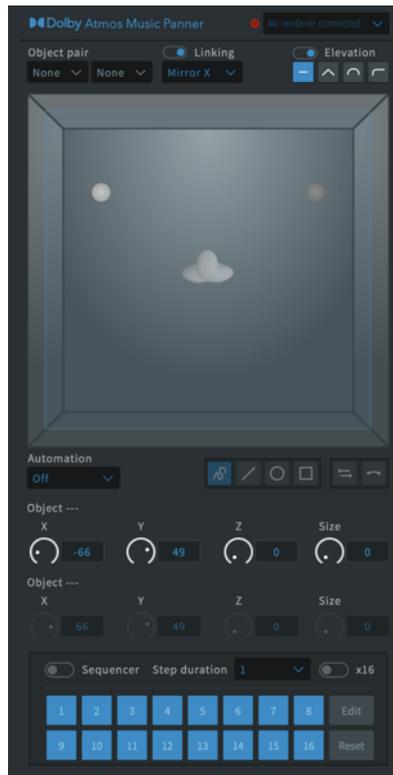


Figure 6 : Illustration du Plugin Dolby Atmos Music Panner

Le *Dolby Atmos Renderer*

Il intègre l'algorithme de rendu utilisé par la technologie Dolby Atmos. Il représente l'élément central de la chaîne de production. Le *Dolby Atmos Renderer* est une application *standalone* utilisée en parallèle de la station de travail audio numérique sur laquelle travaille l'ingénieur-e du son. Il permet d'interpréter les métadonnées associées aux sources audio afin de les restituer sur un système d'écoute donné et d'exporter le mix final.

L'interface du *Dolby Renderer* comprend plusieurs fenêtres de *monitoring*¹⁰ :

- Un contrôle visuel sur l'état des 128 canaux et leur localisation dans la scène sonore (partie gauche et inférieur de l'interface logiciel)
- Un contrôle du niveau par canal de diffusion et de *loudness*¹¹ général pour le mix (partie centrale supérieure de l'interface logiciel)

¹⁰ Contrôle du signal audio

¹¹ Niveau sonore tel que perçu par l'être humain

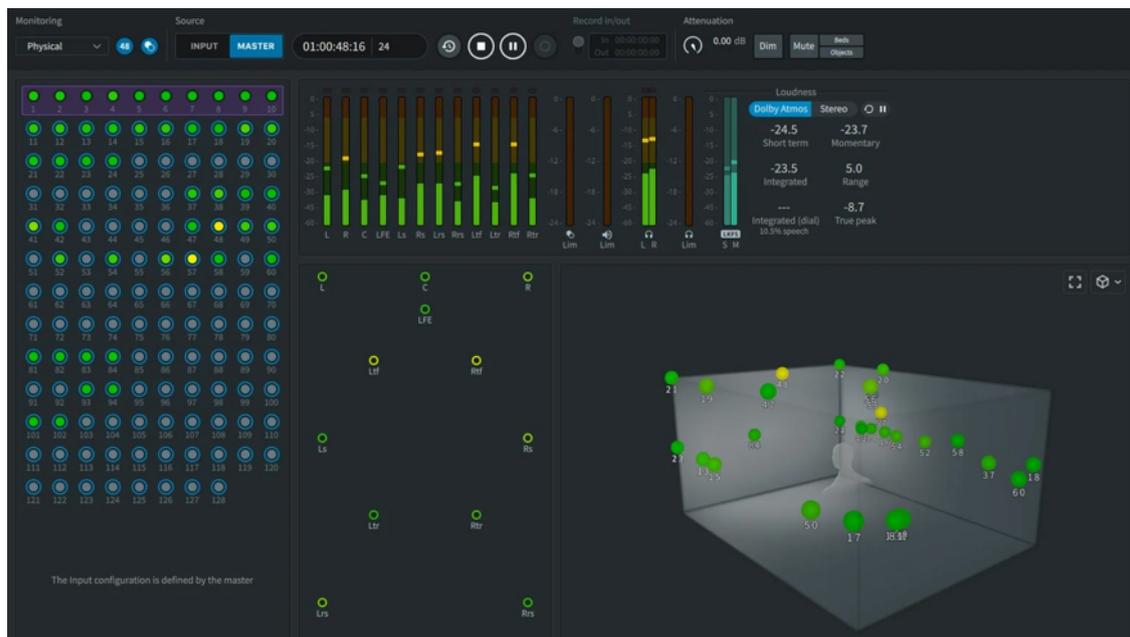


Figure 7 : Interface du Dolby Renderer (Dolby, 2023)

Le *Dolby Renderer* permet également de définir les paramètres de rendu binaural (*Binaural Render Mode*) qui seront associés au master Atmos, puis dans l'encodage du fichier ADM. Ces métadonnées ne sont prises en compte par le *renderer* que dans le cas d'une écoute au casque où le master Atmos est downmixé dans une version binaurale. Il existe quatre types de paramètre : *off* (désactivé), *near* (proche), *mid* (moyen), *far* (lointain). Chacun d'eux correspond à une valeur de distance virtuelle entre la tête de l'auditeur-ice et la position de l'objet. Ces paramètres peuvent être à la fois associés à l'un des 128 objets, et aux canaux physiques, à l'exception du canal LFE. Cette gestion des paramètres binauraux permet à l'ingénieur-e du son d'avoir la possibilité de contrôler la position perçue pour chaque objet et ainsi monitorer le rendu global de son mix pour une écoute au casque.

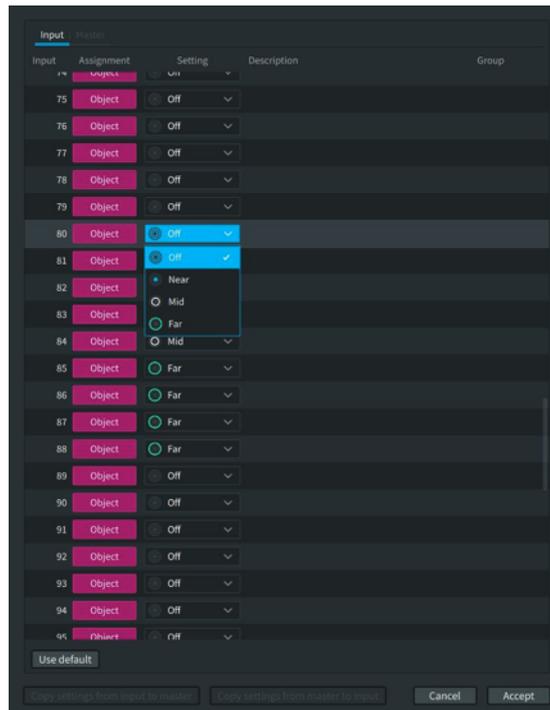


Figure 8 : Fenêtre d'attribution des paramètres de rendu binauraux par objet dans la suite de production de Dolby (Dolby, 2023)

Trois types d'exports pour un master Dolby Atmos sont possibles (Dolby Hybrik, s. d.) :

1. Le Dolby Atmos Master File (DAMF, .atmos), ce fichier propriétaire contient trois types de sous-fichier :
 - Un fichier *.atmos* décrivant entre autre le nombre d'objets et *beds* utilisés.
 - Un fichier *.atmos.metadata* décrivant la position de chaque source audio temporellement.
 - Un fichier *.atmos.audio* contenant l'audio jusqu'à 128 canaux.
2. ADM BWF (.wav), ce fichier au format non propriétaire, mais intégrant des métadonnées propriétaires, représente l'export principal pour la diffusion commerciale. D'après la norme Dolby, cet export correspond au format attendu pour la diffusion d'un mix Atmos sur les plateformes de streaming. Il contient les informations des fichiers *.atmos* et *.atmos.metadata*.
3. Des fichiers de type 9.1.6/7.1.4/5.1.4 PCM et même stéréophonique ou binaural, pour une diffusion orientée canal, aussi appelés re-render.

Pour résumer, le mixage est d'abord tourné en temps réel à la suite duquel le fichier DAMF est créé. Le DAFM est ensuite exporté en un fichier ADM BWF. Des *re-render* sont ensuite possibles aux différents formats décrits ci-dessus. Depuis Protools 2022, il est désormais possible d'exporter des ADM en *offline*.

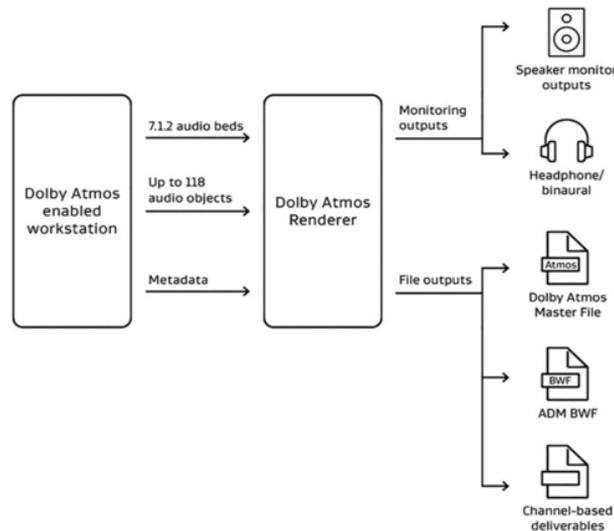


Figure 9 : Exemple d'une chaîne de production Dolby Atmos (Dolby, s. d.)

B.4. L'importance des codecs pour la transportabilité des mixages

Nous avons vu dans la partie précédente les grandes lignes du fonctionnement de l'outil de postproduction Dolby Atmos Music. L'étape d'encodage qui suit l'export d'un master est déterminante pour la transportabilité du mixage. En effet, à titre d'exemple, pour un mixage d'une durée de trois minutes, comprenant 128 pistes, réalisé à une fréquence d'échantillonnage de 48 kHz, à une quantification de 24 bits, la taille et le débit d'un tel fichier seraient les suivants :

$$Taille = (48000 \times 24 \times 128 \times 180) \div 1024^2 = 2,5 \text{ Gb}$$

$$Débit = (48000 \times 24 \times 128) \div 1024^2 = 140 \text{ Mbps}$$

On remarque que la taille et le débit du fichier sont beaucoup trop importants pour des usages de type streaming qui ont un ordre de grandeur situé entre 320 kbps et 990 kbps

(respectivement pour des formats stéréophoniques compressés et *lossless*). Il est donc nécessaire de compresser le fichier pour sa diffusion sur de telles plateformes : c'est le rôle du codec, un algorithme ayant pour but de coder et décoder un fichier. Le développement de la technologie NGA (*Next Generation Audio*), ayant pour but de développer les futurs formats audio pour le broadcast, a permis de définir des codecs capables de gérer une approche orientée objet. Il existe quatre codecs associés à cette technologie : MPEG-H, Dolby AC-4, DTS-UHD Audio et DTS:X. Les principaux codecs présents sur les plateformes de streaming sont les codecs propriétaires de Dolby (AC-4 IMS et DD+JOC) ainsi que son principal concurrent : le MPEG-H 3D Audio (développé par Fraunhofer IIS) et utilisé dans la suite de production Sony 360 RA. Ces codecs sont un format hybride, ils peuvent prendre à la fois en compte des sources sonores objets et des canaux (des *bed* pour le Dolby Atmos).

Dans le cadre du Dolby Atmos Music, la compression d'un master se décompose en deux étapes : l'encodage spatial et le regroupement. Lors de l'étape d'encodage spatial (*spatial coding*), à chaque élément, *bed* et *object*, est attribuée une position et des métadonnées associées. L'encodage spatial est utilisé pour réduire les 128 canaux de *bed* et d'objets en 12 ou 16 éléments appelés clusters.

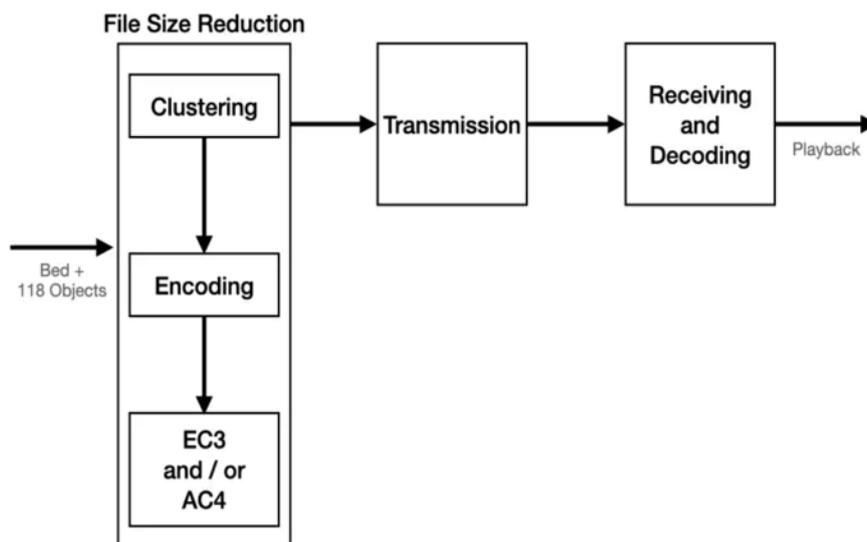


Figure 10 : Schéma bloc : encodage sous un codec Dolby (Nair, 2021)

Le regroupement spatial (*clustering*), réunit des objets possédant une position similaire dans la scène sonore en trois dimensions. Cela permet de réduire la quantité de données associées aux bed et objets et réduit ainsi la quantité de positions disponibles à un instant donné. La taille de chaque source (paramètre *size*) influe sur le nombre de clusters auquel sera rattachée la source sonore lors de cette étape. L'étape de regroupement spatial est un processus qui peut être monitoré à travers le *Dolby Renderer* à la fin d'un mix.

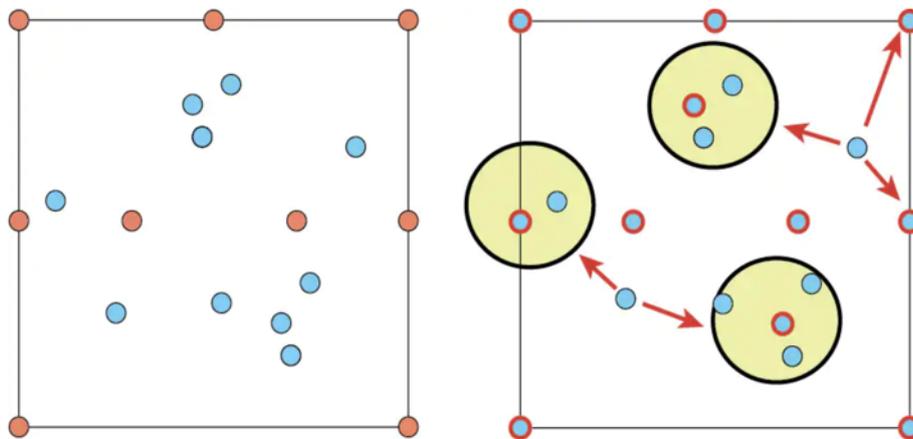


Figure 11 : Principe du clustering : passage de 20 à 12 channels (Nair, 2021)

L'étape d'encodage prend en compte les différents clusters d'objets et réduit ensuite la taille globale du fichier selon le codec choisi : DD+JOC ou AC4-IMS. Aucun des éléments audio n'est supprimé pendant ces étapes. Le débit de données minimum est de 384 kbps (avec 12 éléments) ou 448 kbps (avec 16 éléments). Lorsque ce flux atteint un appareil compatible Atmos, le *renderer* utilise les métadonnées attribuées aux objets pour les restituer de manière appropriée selon l'environnement d'écoute du consommateur. Nous détaillerons dans la partie C.3 les différences entre chaque codec et leur importance dans la restitution fichier en fonction du moteur audio associé.

C] Vers une écoute majoritaire au casque : l'enjeu de l'écoute binaurale

C.1. Principes et perspectives de l'écoute au casque

Un certain nombre de recherches et de mémoires consacrés à la perception binaurale ont déjà exposé de manière complète les principes de la perception spatiale (Blauert, 1996). Il s'agira donc ici de présenter un bref rappel sur cette thématique, avant d'approfondir notre recherche.

C.1.A. La perception du son spatial

La perception des sons dans l'espace pour l'être humain est reliée à trois dimensions spatiales : le plan horizontal (l'azimut), le plan vertical (l'élévation) et la distance. Pour localiser une source sonore dans l'espace, nous faisons appel à la différence de perception entre nos deux oreilles. Pour déterminer la position de la source dans le plan horizontal, deux principaux indices acoustiques sont utilisés :

- *L'ITD : Interaural Time Difference*

Il correspond à la différence temporelle de l'arrivée du signal sonore entre les deux oreilles. Cet indice est utile pour les sources sonores de fréquences inférieures à 1500 Hz (Blauert, 1996).

- *L'ILD : Interaural Level Difference*

Il correspond à la différence d'intensité sonore perçue entre les deux oreilles. L'ILD est pratiquement inexistante lorsque les fréquences sont inférieures à 1500 Hz. En revanche, pour des fréquences supérieures à 1500 Hz, l'ILD devient un indicateur efficace (Blauert, 1996).

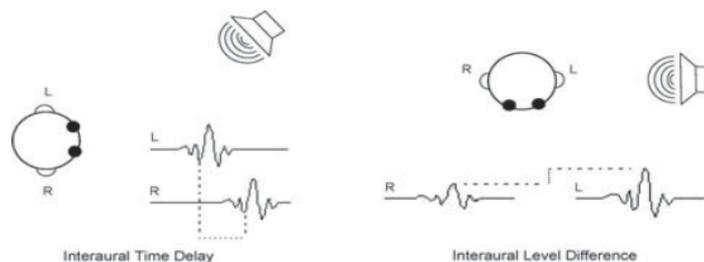


Figure 12 : Exemple d'ITD et ILD (Kapralos et al., 2008, p.2)

La position de la source dans le plan vertical est principalement déterminée par la forme du pavillon et la réflexion des ondes sonores sur le torse et les épaules. Les interférences induites par ces réflexions entraînent des altérations spectrales telles que le renforcement (pics spectraux) ou la dégradation (trous spectraux) de certaines fréquences. Tout au long de sa vie, l'être humain mémorise une variété de fonctions de transfert qui correspondent à différentes directions pour les sources sonores. La mémorisation de ces fonctions joue un rôle essentiel en dissipant l'ambiguïté de localisation entre l'avant et l'arrière, ainsi que le haut et le bas. L'absence de pavillon limite la précision de la localisation des sources sonores et réduit l'impression d'externalisation des sources (Gardner, 1973).

Pour déterminer la distance de la source, trois principaux indices sont utilisés :

1. **La variation d'intensité** : le niveau de la source augmente lorsqu'elle s'approche, et inversement lorsqu'elle s'éloigne
2. **Le rapport entre champ direct et champ réverbéré** : plus la source est proche, plus le champ direct est présent ; plus elle est loin, plus les réflexions (et donc le champ réverbéré) sont présentes
3. **La différence de composition spectrale du signal** : plus la source est loin plus il y aura d'absorption en hautes fréquences.

L'ensemble de ces indices sont dépendants de l'apprentissage de chaque individu. Notre perception est pour autant sujette à quelques imperfections dans la localisation de sources sonores dans l'espace :

- Le pouvoir séparateur de l'oreille dans le plan vertical est plus performant à l'avant qu'au-dessus et à l'arrière de la tête. Il est également beaucoup moins performant sur le plan vertical que sur le plan horizontal. (Damaske et Wagner 1969)
- Il existe un cône de confusion entre le plan avant/arrière et haut/bas pouvant provoquer une inversion de position de la source. Cette zone correspond à des ILD et ITD quasi égaux. Des micro mouvements de la tête peuvent alors aider à diminuer cet effet en changeant le rapport ILD / ITD.

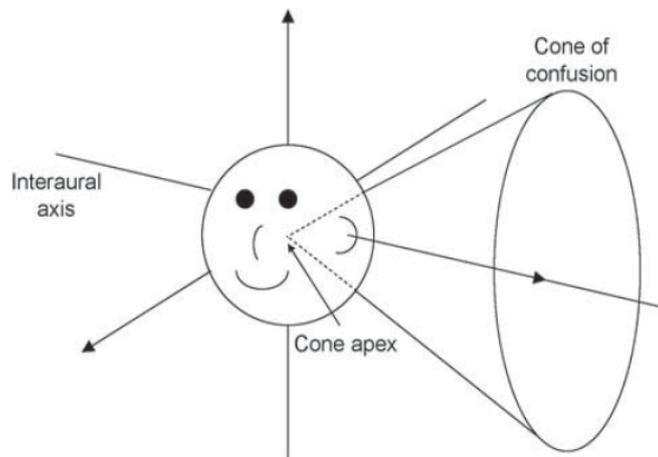


Figure 13 : Cône de confusion (Kapralos et al., 2008, p.3)

C.1.B. L'écoute stéréophonique au casque : problématique de la latéralisation et internalisation

Dans le cadre d'une écoute stéréophonique sur haut-parleurs, un certain nombre de phénomènes acoustiques interagissent avec la source diffusée dans la pièce : réflexion, absorption, diffraction. Comme nous l'avons vu dans la partie précédente, ces différentes interactions créent les indices psychoacoustiques nécessaires à la bonne localisation des sources sonores sur le système de diffusion. Chaque haut-parleur diffuse à un angle de trente degrés par rapport à l'auditeur-ice, le signal diffusé par l'enceinte gauche interagit avec le signal diffusé par l'enceinte droite. Cette interaction permet l'apparition d'un centre fantôme situé entre les deux haut-parleurs, un phénomène psychoacoustique créé lorsqu'un signal arrive au même moment et avec la même intensité aux deux oreilles de l'auditeur-ice.

Lors d'une écoute stéréophonique au casque, la proximité des hauts parleurs aux oreilles de l'auditeur-ice crée un phénomène d'individualisation de chaque canal audio : c'est la latéralisation. Le signal perçu par chaque oreille est identique au signal délivré par chaque haut-parleur, il n'y a pas d'interaction entre les canaux. Ce type d'écoute a une conséquence importante sur la restitution de la stéréophonie : l'image sonore est restreinte en comparaison d'une écoute sur haut-parleurs. La scène sonore résultante est constituée de trois lobes sonores isolés dans la tête de l'auditeur-ice, plutôt que comme une scène sonore

complète et réaliste présentée en face de l'auditeur·ice : c'est ce qu'on appelle l'internalisation (Blauert, 1996).

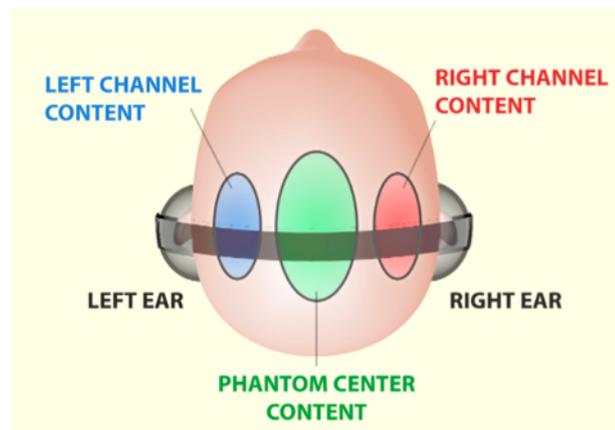


Figure 14 : Problématique de la latéralisation lors de la reproduction d'un contenu stéréophonique au casque (Usmani, 2015, p.1)

Cette latéralisation peut donc s'avérer problématique dans la restitution d'un contenu stéréophonique. L'auditeur·ice perçoit alors une image stéréophonique latéralisée de façon non naturelle (Mushendwa, 2009). Pour autant, l'apprentissage perceptif continu d'un·e individu·e au long de sa vie et l'adaptabilité de la perception humaine font qu'une écoute musicale au casque peut tout de même être appréciée (Gresinger, 1990).

C.1.C. L'écoute binaurale : le potentiel de l'externalisation

À l'inverse de l'écoute stéréophonique qui est limitée par le déplacement de sources sur le plan horizontal, l'écoute binaurale utilise les trois dimensions de la perception spatiale humaine pour reconstituer une image sonore "fictive" tridimensionnelle et externaliser les sources sonores en-dehors de la tête. L'écoute binaurale est nécessairement réalisée à l'aide d'un casque audio afin d'affranchir l'auditeur de son environnement sonore (lieu d'écoute) et d'optimiser la séparation de chaque canal audio. La technologie binaurale repose donc sur la reproduction virtuelle du rayonnement d'une source sonore en champ libre et de son interaction avec les éléments anatomiques de l'auditeur·ice : diffraction et réflexion sur la tête, les épaules et le pavillon de l'oreille. Cette interaction est modélisée par les filtres appelés HRTF (*Head-Related Transfer Functions*, représentation fréquentielle) ou HRIR (*Head-Related Impulse Responses*, représentation temporelle). Ces filtres comprennent les indices

de localisation (ITD, ILD et spectraux) propres à chaque individu·e. Les filtres qui décrivent les fonctions de transferts sont composés de paires de filtres (gauche et droite) qui décrivent le processus de filtrage d'une source sonore à une position spécifique par rapport au corps. Ces filtres HRTF dépendent principalement de la morphologie de l'oreille (taille, dimensions des cavités internes), ainsi que d'autres paramètres physiques propres à chaque individu (Schönstein & Katz, 2010).

Plusieurs études ont prouvé que l'externalisation lors d'une écoute au casque permettait d'atteindre une meilleure qualité sonore et une expérience d'écoute plus agréable (Begault, 1990). Pour autant, plusieurs inconvénients existent, le processus d'externalisation interfère souvent avec les effets utilisés (échos, réverbération) lors de la phase de mixage par l'ingénieur·e du son (Liitola, 2006). Dans la littérature scientifique, la comparaison entre l'écoute stéréophonique et binaurale au casque a déjà été abordée pour plusieurs technologies multicanales et différents types de résultats ont été présentés. D'après les résultats de Fontana, Farina et Grenier (2007), il n'y a aucune différence notable dans la préférence des auditeurs pour un mixage musical binaural en comparaison de mixages stéréophoniques. Cela peut être expliqué par le fait que les auditeur·ice sont habitués aux contenus stéréophoniques. Rumsey, Zielinski, Kassier & Bech (2005) ont suggéré que la nouveauté associée à la musique en 3D pourrait produire des effets positifs sur la qualité de l'écoute, mais que l'effet « wow » produit par un tel contenu pouvait influencer la préférence. Enfin, Abehsera Morell et Lee (2021) ont montré qu'une binauralisation de toutes les sources sonores d'un mix ne conduit pas nécessairement à une expérience plus immersive par rapport à un mixage stéréophonique traditionnel. Un mélange entre des sources binauralisées et non binauralisées pourrait contribuer à améliorer l'immersion (par exemple une batterie en stéréo et des guitares en binaural, ou toutes les sources en stéréo avec une réverbération binauralisée). Ces recherches serviront de base à la réalisation de notre test perceptif pour ce mémoire.

C.1.D. La technologie binaurale

La technologie binaurale repose sur deux grandes pratiques : le binaural natif et la synthèse binaurale.

Le binaural natif correspond à la captation d'une scène sonore à l'aide de microphones placés au niveau du conduit auditif. La pratique la plus courante est d'utiliser une tête artificielle (par exemple le modèle Neumann KU100) avec un microphone omnidirectionnel par conduit auditif. La scène sonore est ensuite reproduite au casque en faisant coïncider les signaux captés (droite et gauche) à chaque haut-parleur du casque. Dans le domaine de la production musicale, cette technique est le plus souvent utilisée pour des captations de musiques filmées. Radio France propose par exemple un certain nombre de contenus binauraux natifs sur sa plateforme Hyper Radio. Cette technologie fixe la localisation des différentes sources de l'image sonore par rapport au dispositif de captation, elle ne sera pas approfondie dans le cadre de cette étude.

La synthèse binaurale correspond à une reproduction virtuelle de l'image sonore. Les signaux (mono, stéréo, multicanaux ou ambisoniques) sont filtrés par les HRTF permettant de restituer à l'auditeur l'ensemble des informations de localisation. Cette technique permet une plus grande flexibilité dans la spatialisation des sources sonores. Elle représente la principale technologie utilisée en réalité virtuelle (VR), ainsi que pour la réduction binaurale de mixages objets.

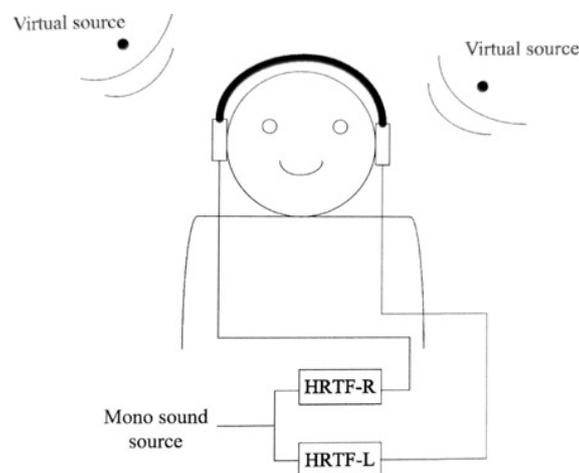


Figure 15 : Principe du rendu binaural (Sunder, 2022)

C.2. Problématiques des HRTF individualisés et non individualisés pour l'écoute musicale en synthèse binaurale

L'une des problématiques intrinsèques à la synthèse binaurale réside dans le choix des HRTF. La composante spectrale des filtres choisis pour chaque auditeur·ice influe grandement sur la sensation d'élévation et sur la capacité de localisation de sources sonores situées sur le cône de confusion. Deux solutions sont possibles dans le choix d'une paire de filtre : travailler avec des HRTF individualisées (propre à chaque auditeur·ice) ou alors des HRTF génériques. Cette partie se propose de discuter ces deux solutions afin de choisir celle qui sera utilisée pour la suite de ce mémoire.

C.2.A. Cas des HRTF génériques

Comme nous l'avons vu précédemment, la synthèse binaurale se base sur l'utilisation de filtres HRTF correspondant aux indices de localisation nécessaires pour restituer une source sonore virtuelle. L'utilisation de HRTF génériques en synthèse binaurale amène des erreurs plus importantes dans la localisation des sources sonores : des ambiguïtés avant-arrière, des erreurs de localisation en élévation et un manque d'externalisation des sources sonores virtuelles (Bahu, 2016). La qualité de la scène sonore reproduite s'en voit donc grandement dégradée. L'utilisation de HRTF génériques peut s'avérer suffisante lorsqu'elles sont associées à un stimulus visuel associé au sonore comme dans le cas de la VR (Berger et al., 2018). Dans le cadre de notre étude, cette stimulation visuelle est inexistante. L'autre solution possible serait de travailler sur la plasticité du système auditif humain, en développant l'apprentissage perceptif pour des filtres non-individualisés (Wenzel et al., 1993). Cet entraînement demande un nombre de tests et un temps important. Il n'est donc pas envisageable dans le cas d'une utilisation grand public comme des plateformes de streaming.

C.2.B. Cas des HRTF individualisés et personnalisés

L'utilisation de HRTF individualisées offre un niveau de rendu quasi similaire à une perception en champ libre. Certains défauts persistent tels que la confusion avant-arrière ou un manque de précision dans la localisation en élévation. Pour autant, la mesure d'une HRTF individualisée nécessite un dispositif très lourd. En effet, les fonctions de transfert sont mesurées en chambre anéchoïque et nécessitent un nombre très important de stimuli pour obtenir une décomposition de l'espace suffisamment précise. Plusieurs bases de données ont

été développées par différents organismes de recherches à partir de mesures réalisées en chambre anéchoïque : Listen (Ircam) ou *CIPIC* (UC Davis). Le format *SOFA* (Spatially Oriented Format for Acoustics, standardisé par l'AES) propose de regrouper ces différentes bases de données dans un seul même format donnant accès aux mesures et métadonnées de manière publique.

Une solution envisageable pour pallier aux contraintes de réalisation d'une HRTF individualisé est l'utilisation de HRTF personnalisés. Cette personnalisation peut être réalisée de différentes manières :

1. En réalisant un filtre personnalisé à partir d'autres HRTF modifiées pour correspondre à un individu spécifique. Cette méthode demande à l'utilisateur·ice un effort long et conséquent pour trouver les filtres appropriés et est trop sujette aux variations d'algorithmes utilisés.
2. À partir de photos de l'individu·e permettant la mesure des différentes parties du corps. Les indices morphologiques sont ensuite traités par un algorithme associant une HRTF correspondante. Cette solution prometteuse fait partie d'un des axes de recherche du projet Bili (Ircam & Orange Lab) pour le développement de HRTF personnalisées. Cette méthode commence déjà à être commercialisée tant pour les plateformes de streaming : Sony (Sony 360 Reality Audio Headphones), Apple (Apple Personalized Audio), que pour les logiciels de production : Dolby (Dolby pHRTF), Steinberg (Immerse Ambi Decoder). Il suffit dans la plupart des cas d'effectuer une série de photo ou vidéo proposant tous les deux de prendre en photo le pavillon auditif pour réaliser une HRTF personnalisée.

C.2.C. La perspective du *tracking*

Le *head-tracking* cumulé à la synthèse binaurale représente une bonne perspective pour lever certaines ambiguïtés de localisations (Stitt, 2016). Apple propose déjà cette fonctionnalité sur son moteur binaural Spatial Audio pour ses derniers casques et écouteurs. L'idée est d'avoir une image sonore stable, externalisée, qui reste fixe lorsque l'on tourne la

tête. Il faut pour ça des latences très faibles afin que notre système auditif ne sente pas cette mise à jour de position. Pour autant, au-delà des avantages du point de vue de la perception binaurale, on peut se questionner sur l'intérêt esthétique de fixer une scène musicale. Si l'on prend l'exemple de la musique pop, les voix sont en générale placées à l'avant et au centre dans le mix. Dans la perspective d'une utilisation avec *head-tracking*, cela signifierait que lorsque l'auditeur-ice tourne la tête, la scène sonore serait déplacée autour de lui : les voix au centre se retrouveraient à l'extrémité droite ou gauche selon son mouvement, et le point de focus « en face » de l'auditeur-ice serait centré sur les éléments sonores initialement placés aux extrémités (reverb, synthés...). De plus, dans le cas du *tracking* associé aux équipements d'Apple, lorsque l'auditeur-ice garde la tête tournée de manière prolongée, la scène sonore vient se repositionner dans l'état initial, créant ainsi un mouvement perceptible de toute l'image sonore. Le risque avec cette possibilité de déconstruction spatiale du mix est alors de créer une impression sonore inhabituelle pour l'auditeur-ice, voire un effet gadget. Les possibilités perceptives offertes par le *head-tracking* se rapprochent plus d'une expérience sonore vécue en live. Son utilisation associée à un enregistrement de concert, voire à un support visuel, serait alors pertinente et bénéfique (Andre, 2017). L'intérêt du *head-tracking* dans le cadre de la diffusion de contenus musicaux est donc discutable et nécessite d'être approfondi dans le cas d'une recherche future.

C.3. L'importance du choix du moteur de rendu binaural

Le choix du moteur binaural est central pour la gestion du rendu d'un mixage à l'auditeur-ice. C'est lui qui a pour but de transcrire, dans le cas d'un *downmix* binaural, les informations de localisation des différentes sources sonores positionnées sur haut-parleurs lors du mixage par l'ingénieur-e du son. Chaque moteur audio a son propre fonctionnement et ses secrets industriels associés. À l'heure actuelle en 2023, il n'y a pas de solution unique : chaque plateforme de streaming définit son propre moteur de rendu et les codecs associés. Chacun-e peut donc développer un avis critique et un ressenti différent en fonction de la plateforme de streaming utilisée. Des questions émergent : pourquoi, à l'inverse d'un contenu stéréophonique, un mixage Atmos serait perçu différemment selon la plateforme que l'on utilise pour un même système d'écoute donné ? L'importance de connaître le fonctionnement de ces outils est donc double : pour l'ingénieur-e du son d'une part, contrôler la compatibilité de son mixage en fonction de l'outil de rendu, et pour l'auditeur-ice d'autre

part, obtenir la meilleure reproduction binaurale possible par rapport au mix pensé par l'ingénieur-e du son. Cette partie détaille les caractéristiques des deux principaux moteurs de rendus disponibles sur le marché des plateformes de streaming : Dolby Renderer, Apple Spatial Audio.

C.3.A. Dolby *Renderer* : AC4-IMS

Nous avons pu voir le principe du logiciel Dolby *Renderer* dans le cadre d'un mixage avec la suite de production Dolby (partie B.3 de cette étude). C'est le même outil de rendu qui est présent dans un appareil compatible Dolby Atmos (portable, tablette). Le contenu audio streamé est interprété par le *renderer*, puis optimisé pour le système d'écoute en question. Le moteur de Dolby est utilisé comme outil de rendu pour les titres mixés en Atmos et disponibles sur les plateformes Amazon Music et Tidal (Kempe, 2023).

Comme nous l'avons vu, deux codecs propriétaires sont associés au moteur Dolby : le DD+JOC (aussi appelé E-AC3) et l'AC4-IMS. Le premier, développé en 2014 lors de la commercialisation du format Atmos pour le cinéma, est utilisé pour la diffusion de contenus Atmos sur enceintes (pour un home cinéma par exemple) et ne prend pas en compte le mode de rendu binaural du mixage (Binaural Render Mode - BRM définis en partie B.3). Le second, plus récent, est optimisé pour la restitution sur des appareils mobiles, casques audio et donc pour le rendu binaural. L'AC4-IMS prend en compte les paramètres BRM définies par l'ingénieur-e du son pendant la phase de mixage. En ce sens, les contenus audio Atmos disponibles sur les plateformes Amazon Music et Tidal restent fidèles aux choix de mixage de l'ingénieur-e du son pour l'écoute au casque. Le décodage diffère selon le type d'appareil et la plateforme utilisée (Gray, 2023) :

- Pour Tidal, le décodage de l'AC4-IMS est effectué par l'appareil sur les produits Android, et en amont de l'appareil pour les produits Apple.
- Pour Amazon Music, le décodage de l'AC4-IMS est effectué en amont de l'appareil

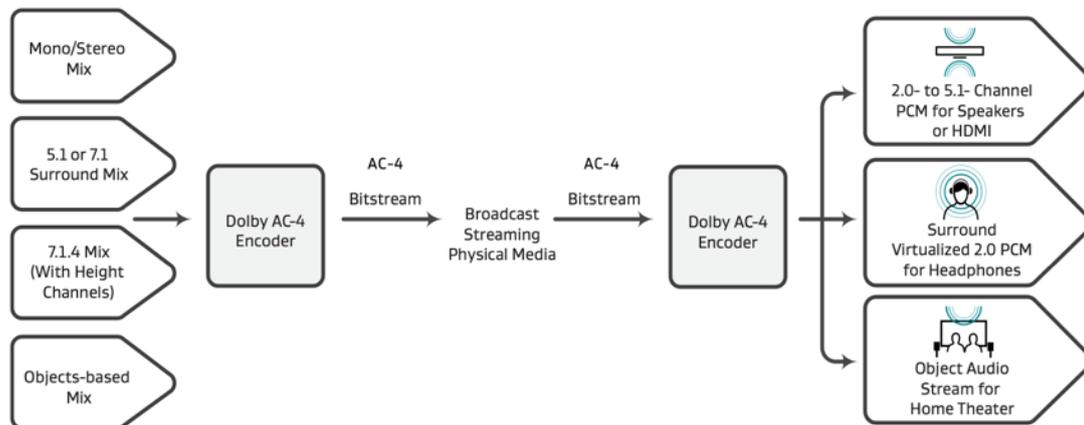


Figure 16 : Principe de diffusion d'un contenu audio Atmos en streaming au format AC4 (Dolby, 2015)

C.3.B. Apple Spatial Audio : DD+JOC

Apple Spatial Audio est le moteur de rendu binaural utilisé par la plateforme Apple Music depuis le lancement de son partenariat avec Dolby en 2021. À l'inverse de Amazon Music et Tidal, Apple a choisi d'utiliser son propre algorithme de rendu propriétaire et donc d'interpréter le fichier ADM différemment de l'algorithme de Dolby. Le point le plus important réside dans le fait que la plateforme n'utilise pas le codec AC4-IMS pour les contenus audio de sa plateforme, mais le DD+JOC généré à partir du Dolby Media Encoder. Cela implique un élément fondamental pour la restitution des contenus audio au casque : les paramètres de rendu binaural (*mid*, *near*, *far* et *off*) définis par l'ingénieur-e du son ne sont pas pris en compte dans la restitution finale à l'auditeur-ice. Apple se retrouve donc avec un fichier multicanal mais pas de données binaurales. Pour corriger cela, la firme part du principe que les utilisateur-ice-s équipé-e-s avec les moyens d'écoutes de la marque (*Airpods*, *Airpods Pro*, *Airpods Max*) peuvent avoir leur HRTF personnalisé à partir d'un appareil IOS, ce qui est suffisant pour percevoir correctement les signaux et la binauralisation dans leur modèle. Par contre, pour les utilisateur-ice-s d'Apple Music n'étant pas équipé-e-s d'un appareil affilié Apple, il n'y a pas de HRTF personnalisé. Apple a donc récemment intégré ses propres filtres HRTF différentes de celles de Dolby et utilise donc ses propres paramètres pour transcrire la position de chaque objet à leur moteur de rendu binaural (Gray, 2023). La réduction binaurale écoutée sur la plateforme diffère donc du master binaural original perçu par l'ingénieur-e du

son. Apple vend une expérience spatiale unique mais est en réalité très flou sur la manière dont il crée sa propre virtualisation pour l'écoute au casque.

L'ingénieur du son Baptiste Murguiondo m'a confié que l'algorithme de rendu binaural d'Apple évoluait régulièrement, créant de légers changements dans la perception des mix sur la plateforme. Concernant l'évolution du *renderer* d'Apple, Baptiste m'a indiqué qu'au tout début de l'Atmos sur Apple Music, la sensation d'une pièce (effet *room*) était vraiment perceptible, et qu'il n'y avait pas de perception de hauteur. Aujourd'hui, l'amélioration du *renderer* a permis d'obtenir un rendu beaucoup plus agréable et beaucoup plus correct en termes de sensation de hauteur et de représentation du centre.

L'une des problématiques centrales pour l'ingénieur-e du son devient donc de pouvoir contrôler les différents types de rendu disponibles sur les plateformes de streaming afin de s'assurer de la compatibilité de son mix. Peu de solutions existent dans le cas de l'Apple Spatial Audio :

1. Exporter au format MP4 du fichier Dolby Atmos Master File, puis l'écouter sur un appareil Apple (Dolby, 2021). Cette méthode non optimale peut s'avérer chronophage car elle ne permet pas un *monitoring* temps réel et nécessite un export à chaque fois que l'on souhaite écouter le résultat.
2. Le logiciel Logic permet de monitorer à la fois les rendus de Dolby et d'Apple directement dans le logiciel de mixage (Apple, 2023).

Dans la réalité, les ingénieurs du son avec lesquels j'ai eu l'occasion d'échanger ont souligné que, malgré ces solutions de contrôle de rendu, il leur était impossible d'ajuster leur mixage en se basant sur cette restitution telle qu'elle existe actuellement, étant donné qu'elle pourrait différer à l'avenir.

Plateforme	Formats disponibles	Moteur de rendu binaural	Codec associé pour l'écoute au casque	Prise en compte des BRM
Apple Music	Dolby Atmos	Apple Spatial Audio	DD+JOC	Non
Amazon Music Tidal	Dolby Atmos	Dolby Renderer	AC4-IMS	Oui
Spotify Deezer	Stéréo	/	/	/

Figure 17 : Résumé des codecs associés aux moteurs de rendu binauraux sur les plateformes de streaming

C.3.C. Autres solutions

Tout comme Dolby, Sony propose sa propre suite de production et son moteur de rendu binaural dédié, connu sous le nom de Sony 360 Reality Audio. Contrairement à Dolby Atmos Music, qui mixe l'utilisation de *channel-based* et *object-based*, la technologie de Sony repose entièrement sur l'utilisation d'objets. La spatialisation des sources sonores est réalisée dans une sphère, et le moteur de rendu binaural de Sony s'appuie sur le codec MPEG-H. Cette technologie est disponible sur les plateformes de streaming telles que Tidal et Amazon Music. Néanmoins, suite aux retours reçus de la part d'ingénieurs du son concernant la qualité de la réduction binaurale de Sony, ainsi que sur la base de notre propre expérience, nous avons décidé de ne pas approfondir cette technologie dans le cadre de notre étude. Nous avons choisi de nous concentrer exclusivement sur les moteurs de rendu binauraux associés à l'utilisation de Dolby Atmos Music, jugés plus performants et jouissant du monopole dans le domaine de la production musicale.

D'autres outils de *downmix* et moteurs de rendu binauraux sont actuellement en développement et pourraient apporter une alternative aux solutions de Dolby, Apple et Sony. C'est le cas de l'outil MyBino, issu de la collaboration entre le service audiovisuel du

Conservatoire de Paris et différents partenaires. Un travail important a été réalisé pour que la sonorité des HRTF utilisées dans l'algorithme de réduction se rapproche le plus possible de l'équilibre spectral du signal stéréophonique tout en optimisant la perception spatiale de la scène sonore (Messonier et al., 2022).

Deuxième partie : Pratiques et évolution des méthodes de création et postproduction

A] Des possibilités de construire une nouvelle scène sonore : l'approche théorique

A.1. L'héritage du mixage stéréophonique

Avant de s'intéresser aux perspectives de mixages dans le contexte d'une production musicale orientée objet, il est tout d'abord nécessaire de s'arrêter un moment sur les esthétiques de production apportées par le mixage stéréophonique. Comme nous l'avons vu dans la première partie de cette étude, la stéréophonie fait office de référence pour le mixage musical depuis près d'une soixantaine d'années. L'utilisation de la stéréophonie sur une telle durée et, de fait, sur un nombre incalculable de titres, implique nécessairement un impact dans la perception collective de ce format de production.

Plusieurs études ont été menées pour tenter de donner des clés d'analyse et d'écoute critique des enregistrements en musique populaire. L'une d'entre elles, menée par Allan Moore, introduit le concept de « *sound-box* » (Moore, 2001) pour caractériser la scène sonore d'un enregistrement musical stéréophonique : « Un « espace textural virtuel », envisagé comme un cube vide de dimensions finies, changeant en temps réel ». Ce concept permet de définir le cadre spatial dans lequel la performance musicale est perçue par l'auditeur-ice lors de l'écoute d'un titre au casque ou sur haut-parleurs. La structuration des différentes sources audio dans l'espace est régie par trois dimensions associées : profondeur, largeur et hauteur et peut être sujette à évolution au cours du temps. Moore et Dockwray ont travaillé sur la période de transition entre les mixages monophoniques et stéréophoniques au milieu des années soixante (Dockwray et Moore, 2010). De cette étude, ils ont pu faire émerger deux grandes tendances de mixages qui ont découlé de l'arrivée de la stéréophonie : le mixage « triangulaire » et mixage « diagonal ».

Le mixage « triangulaire » est associé à un éclatement des différentes sources sonores dans l'espace, avec une latéralisation à l'extrême de chaque côté. Par exemple, dans *Purple*

Haze de Jimi Hendrix, la voix est présente uniquement dans le canal droit, la guitare dans le canal gauche et la batterie et basse au centre. Ce type de spatialisation triangulaire a surtout été utilisé dans les mixages des années 1966 et 1967, au tout début du passage de la monophonie vers la stéréophonie. Les raisons possibles à ce choix de spatialisation et à cette esthétique de production si particulière sont d'abord techniques pour les ingénieurs du son de l'époque : latéraliser à l'extrême permettait d'éviter une surreprésentation des sources au centre lors de la lecture d'un titre stéréophonique sur les systèmes d'écoute monophonique, utilisés alors par la majorité des auditeurs à l'époque. De plus, les potentiomètres de panoramiques des consoles de l'époque ne possédaient souvent que trois crans : gauche - centre - droit, rendant les choix de spatialisation plus radicaux.

Progressivement jusqu'en 1969, l'esthétique du mixage triangulaire se voit peu à peu remplacée par le mixage diagonal : les sources sonores principales (voix, batterie, basse) sont placées au centre de l'image sonore. Cette évolution dans la spatialisation des sources découle d'une réflexion des producteurs sur le lien entre la performance musicale, l'arrangement et le mixage. Comme Moore le décrit :

« Le déplacement de la voix vers le centre appelle une plus grande attention de la part de l'auditeur, l'identifie comme le destinataire personnel de l'expression du chanteur, et facilite même l'identification avec le chanteur que certaines chansons recherchent. » »

(Dockwray et Moore, 2010, p.187)

Là où les mixages triangulaires résultaient d'une crainte de la compatibilité monophonique et des réticences des ingénieurs du son face au nouveau format stéréophonique, les mixages diagonaux se sont mis au service de l'écriture musicale pour exploiter au fur et à mesure tout le potentiel de la stéréophonie. Cette esthétique de mixage a prédominé jusqu'à aujourd'hui la grande majorité des mixages de musique populaire.

Il est important de s'arrêter sur une autre caractéristique fondamentale et commune à la description des scènes sonores des productions musicales stéréophoniques. Moylan (2015) décrit la manière selon laquelle nous avons été conditionnés pendant toutes ces

années de production stéréophonique à une représentation frontale de la scène sonore. L'auditeur·ice se retrouve nécessairement placé·e devant une scène virtuelle, à la manière d'un·e spectateur·ice lors d'un concert. Deux dimensions spatiales sont alors associées au concept de *sound stage* que Moylan décrit : la largeur, permettant la localisation des sources dans l'image stéréophonique grâce à la latéralisation, et la profondeur, permettant une sensation de distance apportée par les réverbérations et délais entre autres. Comme nous l'avons vu dans la partie 1.C.1 de cette étude, dans le cas d'une écoute stéréophonique au casque, la frontalité de l'image sonore devient limitée et les sources sonores placées au centre sont perçues au centre de la tête. La perspective d'un accès à l'écoute binaurale généralisée offre donc des perspectives, tant pour les ingénieur·e·s du son en phase de mixage que pour les auditeur·ice·s d'une meilleure représentation des scènes sonores lors de l'écoute au casque.

A.2. Une nouvelle esthétique musicale pour le son spatialisé

Nous avons pu voir dans la partie précédente quelles ont été les évolutions à la fois esthétiques et techniques induites par le passage de la monophonie à la stéréophonie. Ces évolutions ont été le fruit d'une réflexion et d'un ensemble de contraintes sur lesquelles se sont penché·e·s les producteur·ice·s - ingénieur·e·s du son et artistes pour prendre en main tout le potentiel de créativité de la stéréophonie. Dans un sens, nous pourrions comparer l'enjeu de transition technologique qui a façonné la production musicale entre les années 1960 et 1970 à celle que nous vivons actuellement avec l'arrivée des technologies de mixage orienté objet. Une question se pose alors : quelles perspectives esthétiques vont émerger en vue de l'utilisation grandissante de ces technologies ? Nous avons analysé dans la partie précédente de quelle manière la scène sonore des contenus stéréophoniques en musique actuelle a été représentée de manière frontale et centrée. La démocratisation de l'accès aux outils de production pour l'audio 3D, auparavant réservé à une élite académique et aux spécialistes de ces technologies, ouvre la voie à de nouvelles possibilités de création. N'importe quelle personne, artiste ou ingénieur·e du son, peut aujourd'hui réaliser un mixage orienté objet avec comme seul moyen un casque et une station audionumérique. C'est dans ce contexte d'ouverture, d'accessibilité et de renouveau technique que de nouvelles formes de compositions peuvent apparaître (Burne, 2012). Le paradigme d'écoute instauré par le marché actuel de la musique fait que nous consommons de la musique à la manière d'un

spectateur assistant à une performance sur scène. L'intérêt du grand public pour ces nouveaux formats de diffusion ne pourra être développé que si les compositeur·ice·s s'approprient ces technologies, en proposant une nouvelle écriture musicale périphonique, telle que définie par Gerzon dans les années 70. Le concept de périphonie introduit l'idée d'une écoute non plus dans un plan triangulaire comme dans la stéréophonie, mais dans deux plans verticaux et horizontaux formant un cône autour de l'auditeur·ice.

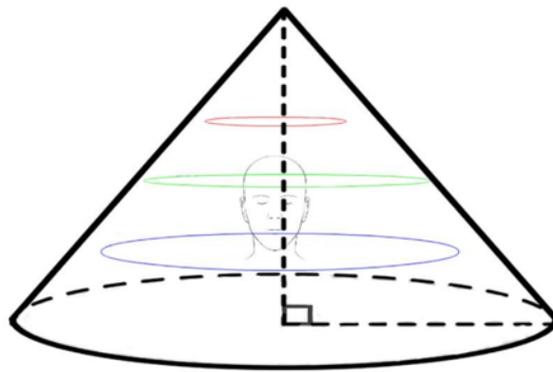


Figure 18 : Principe du cône d'écoute périphonique (Lord, 2022)

Dans ce sens, J. Lord (2022) propose dans sa thèse plusieurs nouvelles possibilités de techniques de production et d'écriture musicale pour s'affranchir du modèle de scènes sonores frontales utilisées dans les productions stéréophoniques. Elle démontre qu'une approche non frontale peut être utilisée pour souligner le sens de la narration et le contenu sonore. Pour autant, ces nouvelles approches nécessitent certaines précautions pour ne pas créer un mix déséquilibré, incohérent ou inhabituel. La voix humaine se prête particulièrement bien à différentes expérimentations de spatialisation car elle est plus facile à localiser de fait de notre perception, et sa composante en hautes fréquences permet d'utiliser le plan d'élévation. De plus, son rôle fondamental et son omniprésence dans les musiques populaires donnent un argument de poids à son expérimentation. Dans le cadre de la spatialisation des voix, où nous avons pour habitude qu'elles soient centrées, Lord introduit deux concepts comme exemple de perspectives de production : l'omnimophonie et la polypériphonie. Le premier décrit le cas où la voix ne serait pas localisée en un point central,

mais dans une multitude de directions, selon une répartition spatiale par ambitus par exemple, plaçant ainsi l'auditeur au cœur de l'expérience sonore. Le concept d'écoute polypériphonique propose d'expérimenter le changement de focus en divisant les phrases du texte dans différentes directions spatiales.

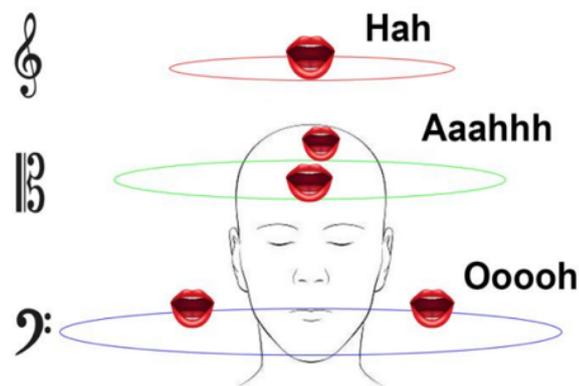


Figure 19 : Exemple d'omnimonophonie (Lord, 2022)

Ces différentes perspectives de spatialisation associent donc l'écriture musicale aux évolutions techniques. L'utilisation du casque en écoute binaurale offre à l'ingénieur-e du son plusieurs plans de spatialisation, i-el peut ainsi doser pour chaque source sonore le niveau d'externalisation ou d'internalisation souhaité. Ces techniques offrent donc une amélioration conséquente sur les possibilités de composition qui ne peuvent pas être réalisées dans le cadre d'une production stéréophonique.

B] Évolutions des pratiques liées à l'arrivée du format orienté objet en musiques actuelles

Afin d'affiner notre recherche par rapport à la réalité des pratiques dans le monde professionnel, nous avons souhaité confronter l'approche théorique évoquée dans les parties précédentes à l'expérience concrète d'ingénieur·e·s du son sur le terrain. Pour cela, nous avons mené puis transcrit cinq entretiens avec des mixeurs ayant une solide expérience du travail de mixage en Atmos : Baptiste Murguiondo, Oscar Ferran, Martin Antiphon, Bruce Keen et Tristan Mazire (Annexes 3 à 7). Les rencontres se sont à la fois réalisées en présentiel et distanciel pour des raisons pratiques, et un guide d'entretien avait préalablement été réalisé afin d'adresser les mêmes thématiques aux différents mixeurs rencontrés (cf. Annexe 2). Ce travail de synthèse nous permettra de faire ressortir les grandes tendances, pratiques et réflexions issues du terrain en vue de la partie pratique de ce mémoire.

B.1 Réalités actuelles des possibilités esthétiques d'une production musicale orientée objet

Nous avons tout d'abord souhaité confronter les mixeurs à l'approche esthétique théorisée en 2.A. Deux questions leur étaient posées dans ce sens : quelles sont les possibilités esthétiques apportées par le mixage objet, et quelles seraient les caractéristiques d'un projet écrit pour l'Atmos ?

Concernant les possibilités esthétiques apportées par le mixage objet, l'ensemble des mixeurs s'accorde sur le fait qu'elles sont très limitées par le fonctionnement de la chaîne de production d'un titre. Actuellement, le mixage objet représente une nouvelle étape de mixage et un coût supplémentaire après le mixage stéréophonique, pour lequel le projet n'a pas nécessairement été pensé pour, et qui potentiellement n'intéresse plus l'artiste qui a suffisamment donné dans les étapes précédentes. Les ingénieurs du son rencontrés m'ont d'ailleurs confié qu'il était extrêmement rare d'avoir l'artiste présent lors de la phase de mixage Atmos, à l'inverse de la stéréophonie.

« Pour moi la façon actuelle de faire de l'Atmos, que je trouve artistiquement complètement stupide, c'est de faire un mix stéréo, puis de faire un mix Atmos. Artistiquement, je la trouve complètement stupide parce que ça veut dire qu'on réadapte un mix qui a été pensé en stéréo en immersif, et en général sans artiste qui vient donner son avis et en général sans personne, ni producteur, ni artiste, pour te valider quoi que ce soit. Sur les 15 disques que j'ai faits, j'en ai peut-être deux ou trois que les artistes ont écouté et encore. »

Extrait d'un entretien réalisé avec Oscar Ferran (Annexe 4)

L'étape de mixage immersif reste donc limitée par la nécessité de délivrer un mixage cohérent avec le mixage stéréophonique. Pour autant, Baptiste Murguiondo rappelle que cette technologie est nouvelle et que nous ne sommes qu'aux débuts de ce qui est possible de faire esthétiquement. Les limitations esthétiques sont en partie dues à la volonté de ne pas vouloir choquer ou changer radicalement les manières de mixer trop rapidement. De nombreuses visions de mix et possibilités techniques émergent de l'arrivée de ces nouvelles technologies, chaque mixeur·euse expérimente différemment.

« Ce qu'on est en train de vivre comme évolution, c'est carrément exponentiel en termes de créativité et de possibilités. Pour moi, c'est presque aussi important que l'invention de l'enregistrement. [...] Quand on a inventé l'enregistrement, on a inventé un nouveau rapport à la musique. [...] L'audio spatial, en termes de créativité de ce que ça disrupte de nos pratiques musicales, est presque aussi important que ce qui s'est passé pour moi à ce moment-là dans les enjeux et dans la façon dont il faut repenser nos façons de faire. »

Extrait d'un entretien réalisé avec Martin Antiphon (Annexe 5)

Pour l'ensemble des ingénieurs du son rencontrés, le véritable essor de la technologie et des possibilités esthétiques ne pourra se concrétiser que lorsque les compositeurs s'investiront davantage, en prenant en main les technologies disponibles et en concevant leurs projets dès le départ en vue d'une restitution dans un format immersif.

« Un projet écrit pour l'Atmos, ça veut dire que lors de la phase d'enregistrement et pré-production, tu ajoutes le format Atmos dans la pensée du projet et forcément là pour le coup tu légitimes la possibilité de mettre des sources en haut, derrière, sur les côtés, où tu veux, car c'est pensé pour. Un bon producteur, il pense en permanence quand il ajoute un élément à qu'est-ce que ça va rajouter fréquemment et dans la dynamique. Là, si tout d'un coup tu lui dis « tu n'as pas juste une stéréo mais tu peux jouer sur l'ensemble du spectre et partout » et bien évidemment qu'un prod, ça lui donne plein d'autres idées et là tu as matière à mixer. »

Extrait d'un entretien réalisé avec Oscar Ferran (Annexe 4)

Les solutions pour intégrer l'Atmos dans des logiciels de pré-production et de composition ne cessent de se développer à l'instar de Logic¹² par exemple. L'utilisation du potentiel esthétique du mixage objet réside donc dans le fait que le projet soit pensé dès le début du projet de création.

B.2. Évolutions de la chaîne de production musicale

L'arrivée du Dolby Atmos dans le domaine de la production musicale a modifié le schéma de production des titres et albums et défini le métier de mixeur immersif. Comme nous l'avons vu, le mixage immersif représente une nouvelle étape de mixage qui survient après le mixage stéréophonique. À travers nos entretiens, nous avons pu voir que le mixage objet est réalisé dans la plupart des cas à partir des *stems*¹³ pré-masterisés issus du mixage stéréophonique. Soit l'ingénieur·e du son récupère les *stems* exportés par l'ingénieur·e du son stéréo, soit il a la possibilité de repartir de la session stéréo d'origine, lui permettant ainsi de travailler avec les outils de traitement de l'ingénieur·e stéréo.

À terme, l'arrivée du format ADM pourrait également remettre en question l'organisation de la chaîne de production mixage-mastering. Jusqu'à l'arrivée des technologies immersives, il n'y avait qu'une seule version disponible sur les plateformes, le master stéréophonique, éventuellement disponible dans différents débits (mp3, *lossless*...).

¹² Logic Pro propose un moteur de rendu Atmos intégré au sein même du logiciel (<https://www.apple.com/fr/logic-pro/>)

¹³ Le terme « stems » est utilisé pour décrire la combinaison d'une ou plusieurs pistes en un seul fichier audio.

Avec le mixage objet et la concurrence des différentes technologies, une multiplicité de versions d'un même titre devient accessible sur les plateformes de streaming (stéréo, Atmos, Apple Spatial Audio, Sony 360 RA). Le fichier ADM permet à partir d'un seul fichier de s'adapter à de multiples rendus, dont un *downmix* stéréophonique. Il existe donc un intérêt économique certain pour les labels à vouloir économiser l'étape de mixage stéréophonique en obtenant le master stéréo à partir de la réduction issue du *renderer* :

« Ce dont j'ai peur [...], c'est qu'à un moment les labels se disent « on ne peut plus se permettre de payer les deux mix, mais on a besoin de faire de l'Atmos parce que tout le monde s'y est mis, donc faites un mix Atmos et après faites-nous des réductions stéréo, et ce sera ça le mix stéréo ». Aujourd'hui, le fait de faire le mix stéréo d'une certaine manière et d'adapter le mix Atmos de l'autre, je trouve ça artistiquement stupide, mais je trouve que ça marche bien en termes de son. Si en tant que mixeur tu dois faire un mix stéréo et après l'adapter en Atmos tu perds trop de temps. Ce n'est plus rentable si on te paye juste pour un mix et que tu fais les deux. Mais je sais que c'est ça que les labels veulent, Universal nous le dit clairement, leur objectif c'est qu'on vende des mix Atmos et qu'ensuite il y ait des réductions stéréo. Ça veut dire qu'ils n'ont aucune conscience de ce qu'est un mix et une manière de travailler en stéréo, c'est complètement absurde. »

Extrait d'un entretien avec Oscar Ferran (Annexe 4)

D'après Oscar Ferran, ce changement de modèle de postproduction dans le domaine du mixage musical pourrait avoir lieu si la plateforme Spotify se met à diffuser des contenus immersifs. La plateforme représente le plus gros service de streaming avec près de 515 millions d'utilisateur·ice·s (Spotify, 2023). Le format Atmos pourrait ainsi devenir la norme.

L'introduction des technologies immersives remet également en question l'utilisation du matériel analogique au sein de la chaîne de production musicale, tant pour la phase de mixage que mastering. Son utilisation était déjà extrêmement restreinte dans les pratiques de mixage 5.1 et semble impossible car trop coûteux dans le cas du Dolby Atmos Music qui présente encore plus de canaux de diffusions. Pour Baptiste Murguiondo, tant que le matériel analogique sera utilisé, l'étape de mixage et de mastering en stéréophonie existeront en tant

que tel, indépendamment des possibilités de réduction stéréophoniques offertes par les *renderer* :

« Il y a un truc en musique qu'il ne faut pas sous-estimer, c'est l'utilisation des machines analogiques. Alors j'ai bossé sur des machines analogiques, c'est très cool, c'est très agréable, mais le mixeur qui dit qu'avoir une machine analogique ça sauve son son, moi ça me pose problème et en fait ça coûte trop cher de faire de l'analogique multicanale. Donc, tant que tu auras des gens qui réfléchissent comme ça, le mix stéréo existera toujours. Il ne faut pas oublier un truc, c'est qu'aujourd'hui se passer du stéréo ça veut dire tuer le mastering. C'est aussi pour ça que beaucoup d'ingénieurs sont complètement contre. »

Extrait d'un entretien avec Baptiste Murguiondo (Annexe 3)

B.3. Évolutions des pratiques de mixages

B.3.A. La mise en espace des sources sonores

Comme nous l'avons vu, l'arrivée des technologies orientées objets pour la musique offre un tout nouveau champ de spatialisation et d'outils associés pour les ingénieur·e·s du son. De multiples questions émergent, les sources sonores peuvent dorénavant dépasser le cadre spatial de la *sound-box* (cf. 2.A.1) et être positionnées de manière précise dans un espace complexe en trois dimensions. La possibilité de séparation des différents éléments qui composent cette nouvelle scène sonore permet un grand avantage : faciliter le démasquage fréquentiel entre les sources et diminuer l'utilisation des outils d'égalisation. Comme nous l'avons vu dans l'exemple de l'omnimonophonie (cf. 2.A.2), l'élévation des sources devient corrélée à leurs composantes spectrales : des sources avec une majorité de hautes fréquences seront plus facilement perçues en élévation. Pour autant, cette spatialisation se doit d'être réalisée sans compromettre le discours musical. La tentation de la division des sources à travers l'espace est forte lors de l'utilisation de ces outils : panning avant et arrière, en élévation, trajectoires dynamiques d'objets, mais peut nuire à la performance mise en jeu, voire distraire l'auditeur·ice. Cette problématique était déjà présente dans les mixages surround des années 2000. Prenons l'exemple de la spatialisation d'une section rythmique qui serait éclatée dans l'espace, la perte de fusion entre les sources sonores contribuant à un

même pattern (poly)rythmique peut entraîner une perte d'efficacité du groove. Les possibilités proposées par ces nouveaux outils panoramiques ne doivent donc pas délaissier l'importance du centre et de la fusion des sources qui étaient jusqu'ici au cœur des mixages stéréophoniques. Dans ce sens, deux visions de spatialisation semblent émerger, l'une cherchant à être la plus fidèle au mixage stéréophonique, et l'autre profitant du potentiel offert par la technologie. Ces tendances de mixages nous ont également été décrites en entretien par l'ingénieur du son Oscar Ferran :

« Pour moi il y a un peu deux écoles pour faire de l'Atmos : il y a ceux qui essaient d'en faire trop et ceux qui essaient d'en faire le moins possible. Moi depuis le début [...], il fallait qu'il y ait un respect absolu du travail du mixeur stéréo, du travail de l'artiste et de la pensée de l'artiste. Il y a des mix Atmos que j'ai écouté de gens qui mettent une guitare acoustique en haut à gauche, les cordes juste derrière et des trucs où on reconnaît à peine le mix stéréo quand tu passes de l'un à l'autre. En général, les maisons de disque aiment bien parce qu'elles ont l'impression qu'elles ont donné de l'argent pour quelque chose qui valait le coup. »

Extrait d'un entretien réalisé avec Oscar Ferran (Annexe 4)

La restitution du fichier ADM pour l'écoute au casque implique une restitution binaurale et donc de nouvelles prises en compte dans la spatialisation des sources sonores. La musique classique permet de capter les indices de localisation binauraux, puisque la prise de son est globale avec par exemple l'arbre Linderg¹⁴. À l'inverse, en musiques actuelles, la prise de son est pour la grande majorité réalisée à proximité, voir ultra-proximité de la source sonore. La reproduction binaurale pour de telles sources implique alors l'utilisation de réverbérations multicanales et la définition de paramètres de rendu binauraux dans le cas du Dolby Atmos. Les possibilités, et les *templates*¹⁵ de sessions de mix, deviennent alors multiples en termes d'utilisation de combinaisons de réverbérations stéréophoniques et multicanales : utilisation pour un groupe d'objets audio, ou à l'ensemble du mix, spatialisation

¹⁴ L'arbre « 2L-cube » permet une configuration de prise de son avec des microphones placés dans un format cubique en 7.1.4. <https://www.merging.com/news/use-cases/morten-linderg-2l-norway>

¹⁵ Modèle de structure de la session comprenant tous les outils nécessaires à la réalisation du mix et permettant ainsi de gagner du temps.

des reverb en mode objet dans l'espace en trois dimensions... L'écoute binaurale pour la restitution au casque permet également une approche beaucoup plus réaliste et proche d'une perception réelle en termes de *panning*. En effet, la technique d'écoute stéréophonique classique au casque peut donner une sensation de « *popping* », c'est-à-dire un passage très rapide du son du centre aux canaux droit et gauche, avec une sur-latéralisation. À l'inverse, l'écoute binaurale permet d'externaliser les sources et offre une perception plus fluide, avec une sensation de focus et d'enveloppement sur les différentes sources sonores.

B.3.B. Vers la fin de la *loudness war*

L'audio orienté objet nécessite donc un changement dans les techniques de mixage et un réapprentissage par rapport aux pratiques couramment utilisées en stéréophonie. Cette nouvelle approche est à la fois nécessaire pour la spatialisation des sources sonores, mais pour la gestion de la dynamique également. La norme établie par Dolby pour la diffusion des fichiers ADM sur les plateformes de streaming est de -18LUFS¹⁶ (Dolby, 2023). La *loudness* des masters Atmos, c'est-à-dire l'intensité sonore telle que perçue par l'oreille humaine, est donc largement inférieure aux masters stéréophoniques prompts à la *loudness war*¹⁷ (on peut trouver dans le commerce des masters aux alentours de -9 LUFS). Les masters Atmos présentent donc beaucoup moins la sensation d'écrasement, de surcompression, *limiting*¹⁸ et distorsion que des masters stéréophoniques.

« Si deux choses s'entrechoquent, je les éloigne et ça marche. Je n'ai pas besoin d'égaliser et de compresser les choses. Je peux vraiment conserver l'intégrité de la musique. Je peux conserver la dynamique. Je dispose d'un tout nouvel ensemble d'outils et d'un tout nouvel espace pour transmettre l'émotion de la chanson. »

Extrait d'une interview de Steve Genewick, ingénieur du son à Capitol Studio

La norme établie par Dolby a nécessairement entraîné une modification des techniques de gestion de dynamique chez les ingénieur·e·s du son. Par exemple, en musiques

¹⁶ Unité de mesure du niveau sonore tel que perçue par l'être humain

¹⁷ Pratique courante dans l'industrie du disque où la dynamique sonore est réduite au profit de l'augmentation de la sensation subjective de volume

¹⁸ Effet dû au limiteur correspondant à l'écrtage des niveaux les plus forts.

actuelles, l'utilisation d'effets sur un *mixbus*¹⁹ (égalisation, compression pour un effet « *glue* », *limiting*) est très utilisée dans la majorité des mixages stéréophoniques en musiques actuelles. Ce type de chaîne de traitement permet d'augmenter la sensation de *loudness* perçue, sans jamais atteindre le 0 dBFS : c'est le principe de la *loudness war*. Le niveau de *headroom*, c'est-à-dire la marge disponible entre le niveau de crête maximal du titre et le niveau de saturation à 0 dBFS, est donc très réduit. En mixage orienté objet, cette méthode n'est pas applicable puisqu'un objet ne peut pas aller à travers un *mixbus*. Seuls des traitements peuvent être appliqués au *bed*, et donc à un ensemble de sources. Le niveau de *headroom* devient donc bien plus grand. Il devient même essentiel si l'on reprend en compte le principe du format ADM : un seul fichier pour de multiples systèmes de diffusions. Le rendu à chaque haut-parleur est déterminé en temps réel pendant le *playback*. Il faut donc une marge de manœuvre suffisante pour être compatible sur les différents systèmes et ainsi s'assurer que les calculs de rendu ne produisent jamais un niveau de signal indésirable. La technologie Atmos permet donc d'envisager la fin de la *loudness war* et la possibilité d'un retour à plus de dynamique et de transitoire.

« En général, les gens aiment bien écouter en Atmos parce que le niveau est fixé à -18 dBLUFS, donc que tu le limites ou pas, tu l'entendras au même volume et j'aime bien le son non limité en fait. Quand j'ai fait le premier album de AIR, *10 000 Hz*, c'était bien écrasé quand même tu vois et quand on l'a remixé en Atmos tout le monde m'a dit « c'est ouvert ton mix, il n'y a pas de limiting en fait ». C'était un peu l'époque dans les années 2000 où c'était la course à l'écrasement des masters. C'est vrai que c'est plaisant d'entendre avec un peu plus de dynamique, c'est un peu plus ouvert. C'est l'un des points forts de l'Atmos, c'est que tu as une plus large dynamique que le -14 dBLUFS qu'on t'impose sur Spotify ou Tidal. »

Entretien mené avec Bruce Keen (Annexe 6)

On peut observer sur les deux figures ci-dessous les différences de dynamique décrites par Bruce Keen entre les deux versions de mixage Atmos et stéréophonique du titre *As It Was* de Harry Style. Il est intéressant de noter que le mixage Atmos présente de

¹⁹ Piste dans laquelle plusieurs sources sont sommées afin d'être affecté simultanément par un ou plusieurs traitements ou effets.

nombreux transitoires et une plage dynamique importante : variation de niveau faible et fort. À l'inverse, le mixage stéréophonique présente une forme d'onde typiquement écrêtée par un limiteur. La plage dynamique semble beaucoup réduite que pour le mixage Atmos, il n'y a quasi pas de variations de niveau une fois passée l'introduction (début de la forme d'onde).



Figure 20 : Forme d'onde de la réduction binaurale du mix Atmos (Morceau : « As It Was » - Harry Style)

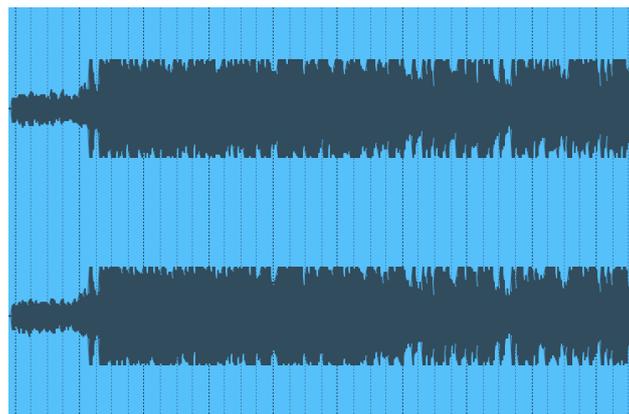


Figure 21 : Forme d'onde du mixage stéréo (Morceau « As It Was » - Harry Style)

B.4. Dans la gestion du *monitoring*

La question du *monitoring* devient essentielle lors du mixage d'un titre en audio orienté objet afin de déterminer quel système de diffusion fait office de référence pour l'ingénieur-e du son : les enceintes ou le casque ? Le premier point important à rappeler est que le fichier ADM sert d'unique master contenant toutes les informations de spatialisation pour être décodé en temps réel sur les différents systèmes d'écoutes (enceintes, casques, barre...). Comme nous l'avons vu en 1.C.3, il faut ensuite ajouter à cette multiplicité de formats de restitution les différents types de décodage selon le codec et la plateforme de

streaming utilisés. La problématique centrale pour les ingénieur·e·s du son devient donc de pouvoir s'adapter à cette multiplicité de possibilités de restitution en session de mixage.

Concernant le système d'écoute faisant référence pour le mixage, les mixeurs que nous avons rencontrés nous ont témoigné de l'importance de mixer dans un studio calibré et sur enceintes. Il est certes devenu essentiel de pouvoir monitorer la sortie encodée en binaural dans le cadre d'un processus de mixage immersif. Pour autant, le mixage sur enceinte fait office de référence et ne peut pas être compromis au profit du binaural :

« Je vérifie que ça se passe bien en binaural mais ce que je dis tout de suite à mes clients c'est : « Attention par contre le timbre, faut que tu écoutes celui des enceintes car c'est lui qui a raison. » Le casque c'est un rendu binaural lambda qui va changer. »

Extrait d'un entretien réalisé avec Martin Antiphon

Le *monitoring* devient donc un exercice complexe de passage constant entre le mixage stéréophonique original faisant référence, le mixage objet sur enceintes et la réduction binaurale. Comme nous l'a indiqué Baptiste Murguiondo, au vu des problématiques inhérentes à la réduction binaurale lors de l'écoute au casque, les mixeur·euse·s tentent en réalité de rapprocher le binaural du multicanal. De plus, pour la plupart d'entre elle-eux, le travail au casque est associé à une fatigue auditive importante et ne peut être une solution d'écoute viable pour plusieurs heures de mixage par jour. Les ingénieurs du son rencontrés m'ont également témoigné de la nécessité d'un processus d'apprentissage pour comprendre comment l'écoute sur haut-parleurs est transposée à celle au casque en binaural. Cet entraînement passe notamment par l'appréhension de la définition des tailles et positions de sources sonores dans l'espace et est sujet à la variabilité du fonctionnement des HRTF génériques de Dolby sur chaque individu·e. Baptiste Murguiondo, par exemple, a éprouvé des difficultés significatives avec les courbes de Dolby en ce qui concerne la localisation des sources au casque. Pour résoudre ce problème, il a découvert que Virtuoso, un logiciel de binauralisation, était la solution la plus convaincante pour lui lors de son travail. Dans un sens, la problématique du *monitoring* et de la compatibilité reste similaire à la stéréophonie : l'ingénieur·e du son ne sait jamais sur quel support va être écouté le titre. Dans le cas de la

stéréophonie, la compatibilité entre appareils résulte principalement d'un travail dynamique et fréquentiel, mais très peu sur les panoramiques. Dans le cas d'un mixage objet c'est l'inverse : la compatibilité entre équipements résulte de la capacité à transcrire la spatialisation des sources entre les différents systèmes d'écoute.

De plus, à l'heure actuelle, les différences entre le mixage objet diffusé sur enceintes et la réduction binaurale au casque sont encore importantes (Puremix, 2021). Il est nécessaire d'accepter la destruction de certains éléments du mixage lors de l'écoute au casque, par exemple le passage à l'avant d'une source spatialisée à l'arrière sur enceintes. Pour Andrew Sheps (Puremix, 2021), le paramètre le plus compliqué à gérer dans la compatibilité enceintes-casque est la gestion de ce qu'il nomme le « *front wall* », correspondant aux trois canaux centraux LCR. Le problème réside dans la gestion du centre et des sources fantômes. Prenons le cas d'un *stem* stéréo spatialisé dans les canaux LR d'un système Atmos, la source produit un niveau égal sur les deux enceintes créant ainsi un centre fantôme (une grosse caisse au centre par exemple). Dans le cas de la reproduction de cette même source au casque en binaural, la création du centre fantôme est beaucoup plus complexe à réaliser d'un point de vue algorithmique. Le *renderer* doit calculer comment deux points sources, situés dans deux points espaces différents, doivent sonner de la même manière pour recréer un centre fantôme. La gestion du rendu est beaucoup plus adaptée pour un seul point source mono situé dans le canal du centre. D'après Scheps, cette problématique a émergé lorsqu'une grosse partie du catalogue d'Apple Music, au lancement de leur partenariat avec Dolby, a été réalisée en remixant rapidement des *stems* stéréo : grosse caisse, caisse claire, voix. Sur enceinte, cela ne pose aucun souci, mais lors d'une écoute en binaural, des problèmes de phases et de spatialisation sont perceptibles. Plusieurs ingénieur·e·s du son préconisent ainsi de ne pas affecter de paramètre de rendus binauraux sur les trois canaux centraux (LCR) pour éviter ces problèmes. Pour autant, les modes de rendu binauraux et les algorithmes de décodage sont continuellement en train de changer et de s'améliorer avec le temps. Rappelons que la réduction binaurale d'un titre est réalisée en temps réel lors de la lecture et du décodage du fichier par l'appareil de lecture. Une nouvelle version binaurale n'est possible que si l'entièreté du fichier ADM est réexportée et donc supprimée des plateformes de streaming. Ne pas affecter les paramètres binauraux de peur du rendu actuel peut donc aussi être dangereux. Si l'on se projette dans un avenir proche, il y a un risque d'avoir un effet daté

« 2020 » des mixages où l'on remarquerait qu'il n'y avait pas d'info binaurale sur les fichiers en question. Cette tendance nous a été confirmée lors de nos entretiens par plusieurs mixeurs :

« Beaucoup de mixeurs mettent le LCR en off effectivement sinon ils ne retrouvent pas leur mix au casque. Donc là, se pose très vite une question qui est : est-ce que tu préfères mettre tout en *mid*, avoir un mix très englobant et compenser (enlever des aigus remettre du grave) et potentiellement dans cinq ans, ou même dans trois ans, l'algorithme change et d'un coup, tout ce que tu as fait de compensation est éclaté. Ou, est-ce que tu te dis : sur mes enceintes c'est bien, au casque en *off* c'est à peu près bien, mais si les algorithmes changent et s'améliorent, mon mix Atmos sonnera beaucoup moins immersif que ceux qui ont tout mis en *mid*. Très vite, j'ai tranché et j'ai mis le LCR en *off* parce que je ne pouvais pas me résigner à écouter des trucs qui ne vont pas du tout. »

Extrait d'un entretien réalisé avec Oscar Ferran (Annexe 4)

D'autres problématiques dans le passage des enceintes au casque nous ont également été évoquées au cours de nos entretiens : les problèmes de balance tonale, déphasage et filtrage en peigne liés à des ITD très courts dans les mesures binaurales. La perception des sources situées à l'arrière et en élévation reste également très floue voire imperceptible.

« Il y a des choses qui se passent fréquemment (à l'écoute au casque). Des expériences que j'ai eu sur notre système d'écoute, ça sonnait très bien. Dans le passage en réduction binaurale, il y avait des artefacts, des choses qui se passaient dans le haut médium et les aigus qu'il fallait corriger pour que ce soit cohérent en binaural. Donc un problème de timbre clairement. »

Extrait d'un entretien réalisé avec Tristan Mazire (Annexe 7)

C] Conclusion partielle

La confrontation de l'approche théorique à la réalité de l'expérience de terrain nous a permis de constater à quel point le potentiel artistique et technique des technologies telles que le Dolby Atmos se retrouvait limité dans les faits. Les perspectives esthétiques semblent à la fois très vastes en termes de potentiel de créativité mais limitées par le manque de projet pensé pour ces technologies. Nous avons également pu constater que la transportabilité des mixages associée au fichier ADM représente un intérêt économique certain et semble dessiner des changements possibles dans la chaîne de production musicale. Les nouveaux outils de mixage proposent des évolutions concrètes dans la spatialisation des sources sonores et la gestion dynamique. Pour autant, pour les ingénieur-s du son que nous avons rencontrés, le passage de l'écoute sur enceintes à l'écoute en réduction binaurale ne semble pas encore suffisamment satisfaisant et abouti. Si la plupart des mixeurs rencontrés contrôlent leur mixage au casque en réduction binaurale, l'écoute sur enceinte fait office de référence pour les critères de timbre et spatialisation. La diffusion sur enceintes semble donc apporter une expérience nouvelle en matière d'écoute, mais la concurrence du format stéréophonique semble encore rude pour l'écoute au casque. Dans cette optique, nous avons développé un test perceptif visant à évaluer la perception des différences entre mixage stéréophonique et immersif pour l'écoute au casque, à partir de titres issus du répertoire de musiques actuelles.

Troisième partie : Évaluation perceptive des différences entre mixage stéréophonique et immersif pour l'écoute au casque en musiques actuelles

A] Contexte

Notre recherche a jusqu'ici tenté de dresser un panorama sur l'évolution des techniques et esthétiques des productions musicales induites par l'arrivée du mixage orienté objet. Nous avons pu aborder la complexité d'appréhension de ces technologies ainsi que les nouveaux champs de possibilités qu'elles offrent pour les mixeur·euse·s. Nous avons également décidé de focaliser notre étude sur les enjeux d'écoute au casque, puisqu'elles représentent aujourd'hui le moyen le plus simple et répandu d'avoir accès à ces nouveaux formats immersifs (Léon et al., 2014). L'évaluation perceptive que nous menons dans le cadre de ce mémoire a pour but de déterminer si le mixage orienté objet en musique actuelle peut offrir une nouvelle expérience d'écoute musicale au casque. Pour autant, il est important de souligner que ces tests sont réalisés au tout début de la commercialisation de ces technologies. Des progrès certains, notamment pour l'écoute binaurale sont à prévoir dans les années à venir.

Notre évaluation perceptive prend appui sur le travail réalisé par Simon et Luck (2018). Les deux auteur·ice·s ont travaillé sur les différences de perception et les préférences entre des mixages binauraux et stéréophoniques pour les musiques actuelles. Les mixages binauraux ont été effectués grâce à une station de travail audio VR (3DSoundLabs) puis convolués en binaural avec des HRTF génériques (KEMAR) et individuelles. Les résultats de cette étude montrent une préférence significative pour les mix 3D utilisant la technologie binaurale par rapport à la stéréophonie. Cette préférence est encore plus importante lors de l'utilisation de HRTF individuelles. Dans les perspectives de leur étude, les auteur·ice·s soulignent l'importance d'approfondir leur recherche à d'autres technologies. Notre mémoire s'intéresse au cas concret de la production musicale orientée objet en musiques actuelles. Pour cette évaluation perceptive, nous nous focaliserons donc sur les technologies utilisées pour la diffusion sur les plateformes de streaming (Dolby Atmos, Sony 360 RA), et leurs

algorithmes de rendus associés. Nous tenterons de déterminer s'il existe une préférence notable pour le format binaural transcrit du mixage objet en comparaison du mixage stéréophoniques. De plus, comme nous l'avons vu en 1.B.3, au vu des différences de décodage des fichiers ADM pour l'écoute binaurale selon les moteurs de rendu, nous analyserons s'il existe une différence perceptible, voir une préférence, pour un renderer en particulier ou non.

B] Méthodologie du test perceptif

B.1. Définition des paramètres d'étude

B.1.A. Hypothèses d'étude

Hypothèse 1 : D'après les résultats de Simon et Luck (2018), l'écoute au casque des morceaux mixés en audio orienté objet avec réduction binaurale est préférée aux morceaux mixés en stéréophonie.

Hypothèse 2 : D'après les résultats de Simon et Luck (2018), l'écoute au casque des morceaux mixés en audio orienté objet avec réduction binaurale procure une sensation d'immersion plus importante que les morceaux mixés en stéréophonie.

Hypothèse 3 : D'après notre état de l'art sur le choix du moteur de rendu binaural, pour un même morceau, il existe une différence perceptible entre les algorithmes de rendu associé à chaque plateforme de streaming.

B.1.B. Choix des sujets

L'enjeu de notre étude est de tester nos hypothèses à la fois sur un public initié au domaine du son et un public non averti. Dans le premier cas, cela nous permettra de confronter un panel de sujet aguerri à l'écoute critique et analytique de pièces sonores. Dans le second cas, cela nous donnera la possibilité d'établir des données statistiques sur le public le plus majoritairement visé par la diffusion des nouvelles technologies immersives sur les plateformes de streaming. Pour des raisons de faisabilité, le test a été soumis aux étudiants et enseignants de l'École Nationale Supérieure Louis-Lumière.

B.1.C. Variables

Variable inter-sujets

Tous les sujets du test ont été soumis aux mêmes extraits mais dans un ordre différent : l'ordre des titres et des versions de mix a été mélangé de manière aléatoire. L'écoute des différents extraits a été effectuée dans le même lieu et dans les mêmes conditions d'écoute.

Variable intra-sujets

Deux variables dépendantes ont été retenues pour l'évaluation des différentes versions de mixage : la préférence et la sensation d'immersion. Le critère de préférence est lié à l'appréciation d'une version de mixage par rapport aux autres versions et a pour but de montrer s'il existe une tendance significative pour un mixage en particulier. Bien qu'il existe un manque de consensus sur la définition du critère d'immersion dans la littérature scientifique, l'utilisation de cette variable dans le cadre de notre étude est rattachée à son sens commun : évaluer la sensation d'enveloppement et l'idée d'être entouré par le contenu sonore chez le sujet.

B.2 Choix des titres

Initialement, neuf titres issus du commerce ont été choisis dans autant de genres différents afin de ne pas biaiser l'expérience en privilégiant un genre donné. Cette sélection permettait également d'être représentatif de la variété de mixages immersifs présents sur les plateformes d'écoute. L'attribution d'un genre à un titre a été définie d'après la catégorisation de l'album sur les plateformes de streaming. Il a été choisi de ne sélectionner que des titres issus d'albums sortis après l'arrivée de la technologie Dolby Atmos Music sur les plateformes de streaming (2021-2023). Bien que les conditions de productions et mixages ne soient pas connus dans les détails pour la majorité des titres choisis, ce choix est justifié par le fait que, dans de telles conditions, il est plus probable que les artistes aient été incités à écouter et valider le mixage objet plutôt que des mix de bas de catalogue où, dans la plupart des cas, l'artiste n'a pas été consulté. De plus, il a été décidé de sélectionner différents ingénieurs du son pour chaque extrait, dans le but d'obtenir une plus grande diversité de pratiques de mixage et de ne pas favoriser un-e mixeur-euse en particulier.

Parallèlement, un mixage personnel a été réalisé en stéréophonie puis en Atmos afin de se familiariser avec le flux de travail de la technologie de Dolby et dans le but d'être intégré à la sélection de stimuli pour le test perceptif. Le mixage stéréophonique a été effectué en premier lieu à partir de l'enregistrement d'un trio de jazz que j'ai réalisé moi-même en 2022 à l'École Louis Lumière. Étant donné que ce projet n'avait pas été initialement conçu pour l'Atmos, le mixage stéréophonique a été réalisé en premier. Les pistes issues du mixage stéréophonique ont ensuite été exportées pour la réalisation du mixage Atmos sur Protools. L'une des problématiques majeures auxquelles j'ai été confronté est l'absence d'accès à une salle équipée en Atmos, ce qui m'a contraint à réaliser intégralement le mixage objet en utilisant un casque. Les principales difficultés ont été rencontrées dans la perception des plans d'élévation et de l'arrière, ainsi que dans les modifications des timbres des sources induites par la réduction binaurale.

Finalement, parmi les dix titres initialement prévus pour la phase de tests, il a été décidé de n'en conserver que huit afin de réduire l'effet de fatigue chez les participants. En effet, après avoir effectué plusieurs essais du test, il est apparu que la durée dépassait généralement les trente minutes, ce qui était jugé trop long par les participants. Il a donc été décidé de retirer deux titres de l'étude, dont notre mix personnel qui avait été réalisé exclusivement au casque. De plus, en comparaison du reste de la sélection et de notre expérience très récente dans la technologie Atmos, il paraissait plus cohérent d'utiliser uniquement des mixages professionnels issus du commerce.

Stimuli	Genre	Artiste	Morceau	Album	Année	Mixage Stéréo	Mixage Atmos	Extrait
1	R&b / Soul	Aloïse Sauvage	Croc Top	Sauvage	2022	Mathieu Gibert & Elio Agostini	Oscar Ferran	0'00 - 1'10
2	Electro	Bjork	Victimhood	Fossora	2022	Bergur Þórisson, Jake Miller	Heba Kadry & Bergur Þórisson	0'00- 1'15
3	Folk	Boygenius	Not Strong Enough	the record	2023	Mike Mogis	Nick Rives	0'00- 1'09

4	Rap	Damso	MEVTR	QALF infinity	2021	Nikola Fève	Peter Philips	0'44-1'38
5	Pop	Harry Style	As It Was	Harry's House	2022	Spike Stent	Mike Miller	0'00-1'12
6	Rock	Paramore	This Is Why	This is Why	2023	Manny Marroquin	Kevin Madigan	1'10 - 2'13
7	Variété	Stromae	L'enfer	Multitude	2022	Lionel Capouillez	Peter Philips	0'00-1'17
8	Alternatif	Tears for Fears	River of Mercy	The Tipping point	2022	Charlton Pettus	Steven Wilson	3'41-4'30

Figure 22 : Résumé des titres sélectionnés pour le test perceptif

Quatre versions de chaque titre seront présentées aux participants de l'étude :

1. Le master du mixage stéréophonique.
2. La réduction binaurale du fichier ADM Atmos provenant du catalogue d'Apple Music et décodée en DD+JOC au format Apple Spatial Audio.
3. La réduction binaurale du fichier ADM Atmos provenant du catalogue d'Amazon Music, décodée en AC4-IMS et étant fidèle à la restitution perçue par l'ingénieur-e du son en sortie du Dolby Renderer.
4. Un *upmix* Atmos réalisé par nos soins à partir du master stéréophonique et à l'aide du *plugin* Halo Upmix de Nugen Audio. Pour ce faire, le fichier stéréo a été *upmixé* au format Atmos en utilisant le *preset* Dolby Atmos 7.1.2 du *plugin* Halo Upmix. Le *bed* a ensuite été réduit au format binaural grâce au Dolby Renderer. L'interface du *plugin* est présentée ci-dessous. Cette version nous permettra de quantifier la différence entre le rendu binaural du mix objet et l'*upmix* stéréo.



Figure 23 : Interface du plugin Halo Upmix de Nugen Audio

Étant donné que le catalogue Sony 360 RA offre peu de contenu commun avec les titres disponibles en Dolby Atmos, et afin d'éviter de surcharger le test perceptif en versions de mix, ce format sera exclu de cette étude.

C] Réalisation du test

C.1. Présentation Générale

Le test a eu lieu dans les locaux de l'ENS Louis Lumière du 9 au 12 mai 2023 dans le laboratoire de techniques audio. L'évaluation perceptuelle étant uniquement réalisée au casque, jusqu'à trois sujets pouvaient effectuer en simultanément et indépendamment les tests d'écoute étant proposés sur trois ordinateurs différents. Le modèle de casque DT990 Pro 250 Ohm a été choisi pour tous les sujets afin de ne pas créer de différences du point de vue du système d'écoute. Ce casque est un modèle assez couramment rencontré en mixage musical

voir chez les auditeur-ice-s également. Chaque casque était relié à une interface AVID Mbox connecté à un Macbook Pro. Le niveau était calibré pour une écoute confortable, choisi de manière arbitraire mais identique pour chacun des trois postes.

C.2. Protocole

Le test se composait de huit titres proposant chacun quatre versions de mixage d'une durée moyenne d'une minute pour un total de trente-deux stimuli. L'ensemble des stimuli a été placé dans une session Reaper où le numéro de chaque titre était indiqué par un marqueur de 1 à 8 et où chaque version de mix était assignée à un stimulus numéroté de 1 à 4. Le participant n'avait aucun moyen de discerner la version du mixage qu'il écoutait. Les quatre versions de mix pour chaque titre étaient présentées en même temps sur l'interface de test et synchronisées afin de faciliter la comparaison. Le sujet avait la possibilité de passer d'une version à l'autre sans décrochage à l'aide d'un contrôleur midi assigné aux solos de chaque piste. Les versions de mixages présentés sur les pistes « stimulus » ont été randomisées. De plus, différentes sessions ont été mises en place afin de randomiser l'ordre des titres et ainsi limiter l'effet d'ordre.

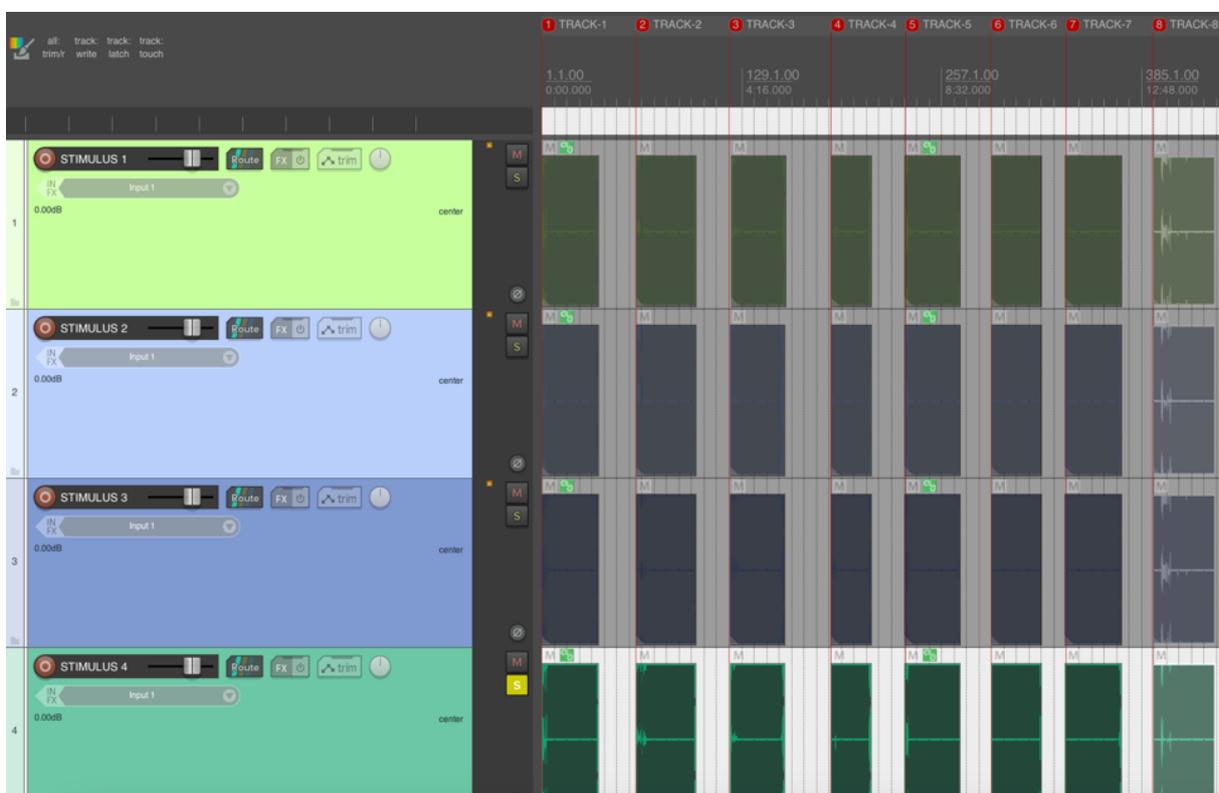


Figure 24 : Session de test sous Reaper

L'ensemble des stimuli ont été normalisés à un niveau de -18dB LUFS intégrés afin de minimiser les différences de niveau. Pour autant il est important de noter que des différences de *loudness range*, niveau LUFS *short-term* et *momentary* étaient tout de même existantes, dues aux différences intrinsèques aux versions de mix et aux pratiques de gestion de dynamique entre les masters stéréophonique et Atmos.²⁰

Avant de démarrer le test, un certain nombre de consignes ont été données au sujet afin de lui présenter le déroulé de l'expérience. L'ensemble de ces consignes étaient présentées sur un document annexe accessible pendant tout le déroulé du test (cf. Annexe 10). Le sujet devait avoir écouté l'ensemble des quatre versions de mixage (stimuli 1, 2, 3, et 4) avant de passer à la phase de notation. Le sujet pouvait écouter l'extrait autant de fois qu'il souhaitait. Une fois la notation terminée, il pouvait ensuite passer à l'extrait suivant. La grille d'évaluation soumise aux sujets comprenait deux échelles indépendantes pour évaluer les variables « préférence » et « immersion » :

- L'échelle de notation utilisée pour évaluer le critère de préférence était continue, de 0 à 5, orientée horizontalement, et comprenait trois points descriptifs (0 = je n'aime pas du tout, 2.5 = pas de préférence, 5 = aime beaucoup). Le concept de préférence a été aisément compris par tous·tes les participant·e·s, et aucun·e d'entre elle·eux n'a eu besoin d'explications supplémentaires.
- L'échelle de notation utilisée pour évaluer le critère d'immersion était continue, de 0 à 5, orientée horizontalement, et comprenait deux points descriptifs (0 = pas immersif, 5 = très immersif). Le concept d'immersion a été redéfini à chaque sujet avant le début du test d'après la définition présentée en 3.B.2. Elle a été aisément comprise par tous·tes les participant·e·s, et aucun·e d'entre elle·eux n'a eu besoin d'explications supplémentaires.

²⁰ Le récapitulatif des mesures de niveau est renseigné en Annexe 8

Il était indiqué au sujet que plusieurs stimuli pouvaient avoir la même note. Les évaluations devaient être effectuées simultanément et les échelles étaient présentées l'une en dessous de l'autre dans le but de réduire la durée du test et d'éviter une fatigue excessive chez les participants. Cependant, le fait d'avoir deux échelles présentées en même temps a pu entraîner des interactions entre les évaluations, même si les attributs mesurés n'étaient pas directement liés. Il a donc été indiqué aux participant-e-s que les deux critères étaient notés de manière indépendante : par exemple, une sensation importante d'immersion n'impliquait pas nécessairement une note haute sur l'échelle de préférence et réciproquement. Un espace libre de commentaire était présenté à la suite de chaque échelle de notation afin de laisser l'opportunité aux participant-e-s de préciser leur choix de classement des différentes versions de mixages ou de renseigner un élément qu'ils-elles auraient jugé-e-s pertinent d'évoquer.

Enfin, afin de placer le sujet dans une position de confiance vis-à-vis du test et de la difficulté de comparaison de certains stimuli, il était précisé au participant qu'il n'y avait pas de mauvaises réponses : si aucune différence n'était perceptible entre les différentes versions, il lui était possible de le signaler. Une fois l'évaluation de l'ensemble des titres réalisée, un court questionnaire était soumis au sujet afin de d'obtenir des informations sur son profil : âge, profession, habitudes d'écoutes musicales, familiarité avec l'écoute de musique au format immersif.

Une fois le test fini, un temps était pris avec chaque sujet pour réaliser un compte rendu de l'expérience. Cela a permis d'expliquer aux participants le sujet de l'étude, les hypothèses et les résultats attendus dont ils n'étaient pas informés préalablement. Cela a également été une opportunité de leur fournir des informations supplémentaires sur la technologie Dolby Atmos, l'utilisation du binaural en mixage musical, et les perspectives de recherche à poursuivre après cette étude.

C.3. Limites du test

Le choix des variables préférence et immersion a été préféré à des critères plus précis d'écoute critique (timbre, dynamique, spatialisation des sources) afin de pouvoir laisser l'opportunité à des sujets non-initiés à l'audio de participer au test perceptif. Pour autant, il est important de noter que ces deux variables regroupent des critères assez larges et sujets à discussion au sein de la communauté scientifique comme pour le cas de l'immersion (Simon & Luck, 2018). De nombreux facteurs peuvent influencer les variables préférence et immersion, on peut par exemple citer le genre musical, les préférences personnelles pour un morceau plutôt qu'un autre, l'orchestration... L'objectif de ce mémoire est de comparer l'écoute au casque d'un mixage stéréophonique à celui réalisé en audio orienté objet avec réduction binaurale. Par conséquent, les effets liés aux différences entre les titres n'étaient pas directement privilégiés dans le cadre de cette recherche. Pour autant, ces paramètres pourront être discutés dans l'analyse des données obtenues afin de les confronter aux résultats du test perceptif.

Par ailleurs, il a été choisi de privilégier des extraits mixés par des professionnels aguerris plutôt que par nos soins afin de minimiser les différences avec la réalité de la production musicale orientée objet accessible sur les plateformes de streaming au grand public. On a également préféré choisir une pluralité stylistique des mixages en sélectionnant des ingénieur·e·s différent·e·s pour chaque titre. Néanmoins, les extraits testés n'ont pas suivi un processus de production similaire et nous n'avons pas pu avoir accès aux partis pris de mixage par les ingénieur·e·s du son. La diversité et la qualité des extraits présentés représente donc aussi une des limites de notre protocole. Dans ce sens, il convient de souligner que notre analyse des résultats ne vise pas à les généraliser pour un genre musical spécifique, mais plutôt à fournir des tendances.

D] Analyse et interprétation

Vingt-neuf personnes, âgées de 22 à 50 ans, ont passé le test perceptif : vingt-six sujets considérés comme initiés au domaine du son, et trois sujets considérés comme non-initiés. Avant la phase de test, il avait été envisagé de séparer l'analyse des résultats en deux groupes selon le niveau d'expertise des sujets. Au vu du nombre trop peu significatif de sujets non-initiés, cette séparation n'a pas abouti. De plus, au vu de la cohérence de notation des variables préférence et immersion chez les sujets non-initiés par rapport aux sujets initiés, il a été choisi de tout de même intégrer leurs résultats à l'analyse globale et de ne pas poursuivre la différenciation d'expertise. Une analyse de la variance (ANOVA) a été réalisée afin de vérifier la significativité des variables. L'analyse statistique détermine le seuil de significativité (p) pour les variables intra-sujets (mixage et titres écoutés), permettant de déterminer si leur impact sur les variables dépendantes (préférence et immersion) est statistiquement significatif. Lorsque p est inférieur à 0,05, une valeur est considérée comme significative, indiquant ainsi que la différence perçue par le sujet n'est pas due au hasard. Il est important de préciser que la méthode d'analyse ANOVA ne permet pas de conclure de manière définitive sur une préférence spécifique, mais plutôt de mettre en évidence des tendances. Au préalable, il a été vérifié que l'ensemble des données suivait une loi normale. Afin de ne pas alourdir la lecture, les termes *stéréophonie* et *stéréophonique* seront simplifiés par *stéréo*.

D.1. Analyse du critère préférence

Les résultats présentés dans cette partie correspondent à l'hypothèse de recherche suivante : l'écoute au casque des morceaux mixés en audio orienté objet avec réduction binaurale est préférée aux morceaux mixés en stéréo.

D.1.A. Préférence sur l'ensemble des titres

L'analyse de variance effectuée pour analyser les données a révélé des effets significatifs du facteur mixage sur l'ensemble des titres ($p=0$) ce qui signifie qu'il y a une influence globale de l'évaluation du critère de préférence en fonction de la version de mixage écoutée. Une légère tendance semble se dégager pour la préférence du mixage stéréo et de l'*upmix* réalisé à partir du master stéréo. Pour autant, aucune différence significative n'a été décelée entre ces deux versions ($p=0,7$). Une différence significative est notable entre le mix stéréo et les mixages Atmos issus d'Apple Music ($p=0,002$) et Amazon Music ($p=0$). Enfin,

aucune différence significative n'a été perçue entre les deux décodages Atmos issus d'Amazon Music et Apple Music ($p=0,01$). Une différence significative a été relevée entre l'upmix et les mixages Atmos d'Apple Music ($p=0,006$) et d'Amazon Music ($p=0$).

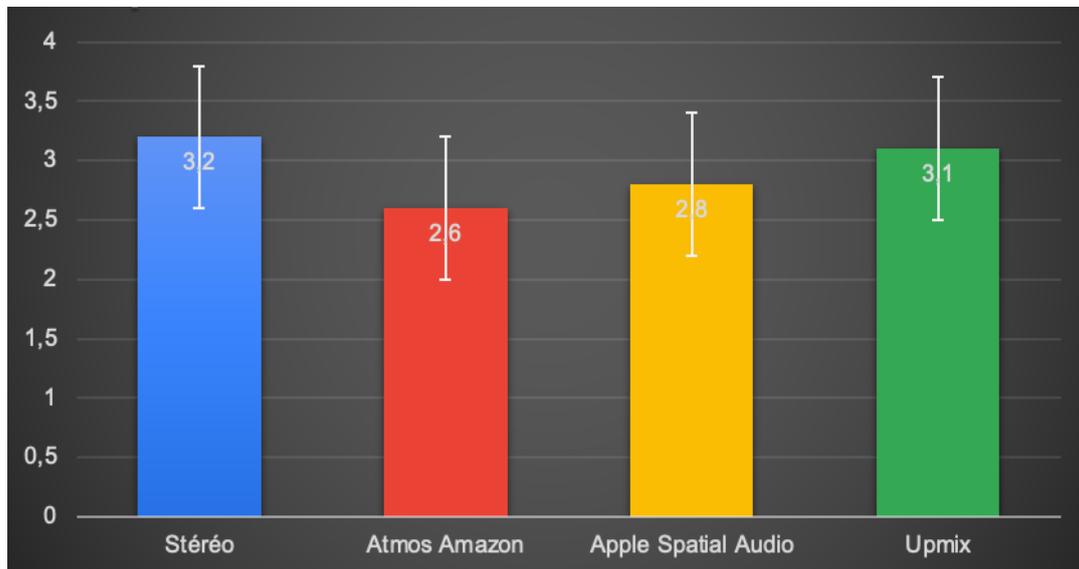


Figure 25 : Moyenne et écart-type de la variable préférence sur l'ensemble des 8 titres

D.1.B. Préférence individuelle par titre

L'analyse de variance effectuée pour analyser les données a révélé des effets significatifs du facteur mixage sur cinq des huit titres ce qui signifie qu'il y a une influence globale de l'évaluation du critère de préférence pour les morceaux de Boygenius, Damso, Harry Style, Paramore et Tears For Fears. En revanche aucune différence significative n'a été décelée pour les morceaux de Aloïse Sauvage, Bjork et Stromae.

Aloïse Sauvage	Bjork	Boygenius	Damso	Harry Style	Paramore	Stromae	Tears For Fears
0,07	0,3	0	0,03	0	0,005	0,2	0,01

Figure 26 : Résultats des différences significatives pour le critère préférence en fonction du genre

Concernant le morceau folk de Boygenius, une tendance significative semble se dégager pour la préférence du mixage stéréo et de l'*upmix*. Aucune différence significative n'a été décelée entre ces deux versions ($p=0,35$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple Music ($p=0,002$) et Amazon Music ($p=0$). Aucune différence significative n'a été perçue entre les deux mixages Atmos issus d'Amazon Music et Apple Music ($p=0,1$). Enfin, une différence significative a été perçue entre l'*upmix* et les mixages Atmos ($p=0,001$).

Concernant le morceau rap de Damso, une tendance significative semble se dégager pour la préférence de l'*upmix* par rapport aux mixages stéréo ($p=0,037$), Atmos issus d'Apple ($p=0,017$) et d'Amazon ($p=0,007$). Aucune différence significative n'a été décelée entre les mixages stéréo, Atmos issus d'Apple Music et Amazon Music ($p=0,97$).

Concernant le morceau pop de Harry Styles, une tendance significative semble se dégager pour la préférence du mixage stéréo et de l'*upmix*. Aucune différence significative n'a été décelée entre ces deux versions ($p=0,5$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple ($p=0,004$) et Amazon ($p=0,0002$). Aucune différence significative n'a été perçue entre les deux mixages Atmos issus d'Amazon Music et Apple Music ($p=0,8$). Enfin, une différence significative a été perçue entre l'*upmix* et les mixages Atmos ($p=0,0001$).

Concernant le morceau rock de Paramore, une tendance significative semble se dégager pour la préférence du mixage stéréo. Une différence quasi significative a été décelée entre la mixage stéréo et l'*upmix* ($p=0,05$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple Music ($p=0,04$) et Amazon Music ($p=0,0005$). Aucune différence significative n'a été perçue entre les deux mixages Atmos issus d'Amazon Music et Apple Music ($p=0,1$). Enfin, aucune différence significative n'a été perçue entre l'*upmix* et les mixages Atmos ($p=0,16$).

Concernant le morceau alternatif de Tears for Fears, une différence significative a été décelée entre le mixage stéréo et le mixage Atmos issu d'Amazon Music ($p=0,0007$). Aucune

différence significative n'a été décelée entre le mixage stéréo et le mixage Atmos issu d'Apple Music ($p=0,1$). Aucune différence significative n'a été décelée entre le mixage stéréo et l'*upmix* ($p=0,07$). Aucune différence significative n'a été perçue entre l'*upmix* et le mixage Atmos issu d'Apple Music ($p=0,1$). Aucune différence significative n'a été perçue entre les deux mixages Atmos ($p=0,1$). Aucune différence significative n'a été perçue entre l'*upmix* et le mixage Atmos issu d'Amazon Music ($p=0,08$).

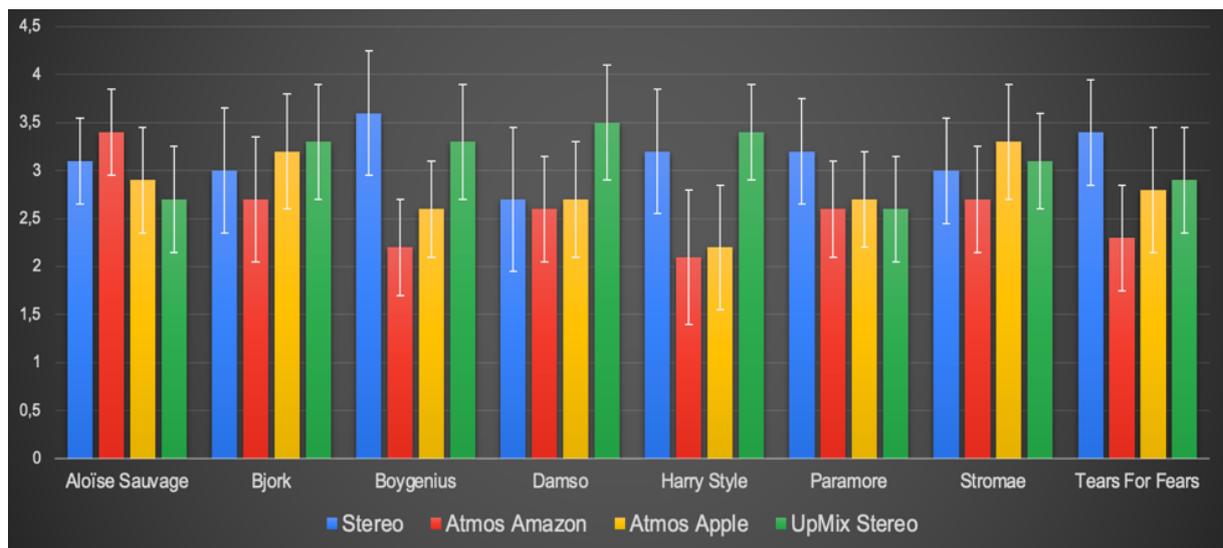


Figure 27 : Moyenne et écart-type des résultats obtenus pour la variable préférence

D.2. Analyse du critère immersion

Les résultats présentés dans cette partie correspondent à l'hypothèse de recherche suivante : l'écoute au casque des morceaux mixés en audio orienté objet avec réduction binaurale procure une plus grande sensation d'immersion.

D.2.A. Immersion sur l'ensemble des titres

L'analyse de variance effectuée pour analyser les données a révélé des effets significatifs du facteur mixage sur l'ensemble des titres ($p=0$) ce qui signifie qu'il y a une influence globale de l'évaluation du critère d'immersion en fonction de la version de mixage écoutée. Une tendance semble se dégager pour une sensation d'immersion plus importante sur le mixage stéréo. Sur l'ensemble des titres, une différence significative a été décelée entre le mixage stéréo et l'*upmix* ($p=0,0003$), les mixages Atmos issus d'Apple Music ($p=0$) et

Amazon Music ($p=0$). Une différence significative a été perçue entre l'*a* et le mixage Atmos issu d'Amazon Music ($p=0$) et d'Apple Music ($p=0$). Aucune différence significative n'a été perçue entre les deux mixages Atmos ($p=0,8$).

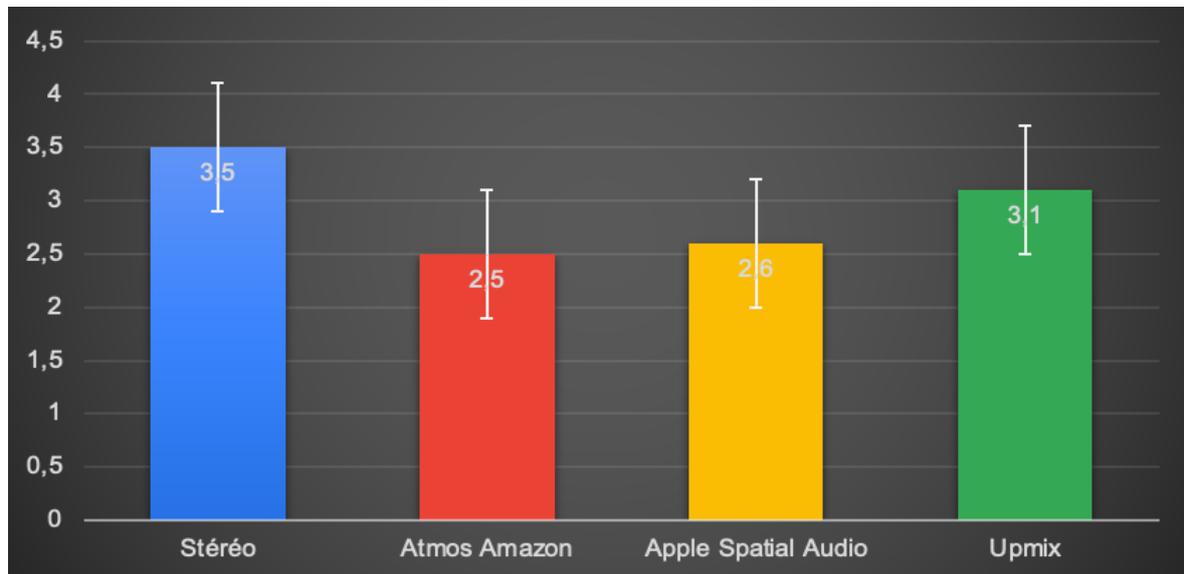


Figure 28 : Moyenne et écart-type de la variable immersion pour l'ensemble des huit titres

D.2.B. Immersion individuelle par titre

L'analyse de variance effectuée pour analyser les données a révélé des effets significatifs du facteur mixage sur six des huit titres ce qui signifie qu'il y a une influence globale de l'évaluation du critère d'immersion pour les morceaux folk, rap, pop, rock, variété et alternatif. En revanche aucune différence significative n'a été décelée pour les morceaux classés dans les genres R&B et électro.

Aloïse Sauvage	Bjork	Boygenius	Damso	Harry Style	Paramore	Stromae	Tears For Fears
0,6	0,1	0	0,0007	0	0,0007	0,001	0,0004

Figure 29 : Résultats des différences significatives pour le critère immersion en fonction du genre

Concernant le morceau folk de Boygenius, une tendance significative semble se dégager pour la sensation d'immersion du mixage stéréo. Une différence significative a été décelée entre le mixage stéréo et l'*upmix* ($p=0,0006$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple Music ($p=0$) et Amazon Music ($p=0$). Aucune différence significative n'a été perçue entre les deux mixages Atmos issus d'Amazon Music et Apple Music ($p=0,48$). Enfin, aucune différence significative n'a été perçue entre l'*upmix* et les mixages Atmos ($p=0,36$).

Concernant le morceau rap de Damso, une tendance significative semble se dégager pour la sensation d'immersion pour l'*upmix* et le mixage stéréo. Aucune différence significative n'a été décelée entre ces deux versions ($p=0,17$). Une différence significative a été décelée entre l'*upmix* et les mixages Atmos issus d'Apple ($p=0,0003$) et d'Amazon ($p=0,0004$). Une différence quasi significative a été décelée entre le mixage stéréo et les mixages Atmos issus d'Apple ($p=0,05$) et d'Amazon ($p=0,06$). Aucune différence significative n'a été décelée entre les mixages stéréo, Atmos issus d'Apple Music et Amazon Music ($p=0,95$).

Concernant le morceau pop de Harry Styles, une tendance significative semble se dégager pour la sensation d'immersion des mixages stéréo et de l'*upmix*. Aucune différence significative n'a été décelée entre ces deux versions ($p=0,22$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple ($p=0$) et Amazon ($p=0$). Une différence significative est notable entre l'*upmix* et les mixages Atmos issus d'Apple ($p=0$) et Amazon ($p=0$). Une différence quasi significative a été perçue entre les deux mixages Atmos issus d'Amazon Music et Apple Music ($p=0,05$), avec une légère tendance pour la version d'Amazon Music.

Concernant le morceau rock de Paramore, une tendance significative semble se dégager pour la sensation d'immersion du mixage stéréo. Une différence significative a été décelée entre le mixage stéréo et l'*upmix* ($p=0,0008$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple Music ($p=0,004$) et Amazon Music ($p=0,0004$). Aucune différence significative n'a été perçue entre les deux mixages Atmos issus

d'Amazon Music et Apple Music ($p=0,44$). Enfin, aucune différence significative n'a été perçue entre l'*upmix* et les mixages Atmos ($p=0,7$).

Concernant le morceau de variété de Stromae, une tendance significative semble se dégager pour la sensation d'immersion des mixages stéréo et de l'*upmix*. Aucune différence significative n'a été décelée entre ces deux versions ($p=0,2$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple Music ($p=0,004$) et Amazon Music ($p=0,0007$). Une différence quasi significative est notable entre l'*upmix* et le mixage Atmos issu d'Apple ($p=0,06$) et significative entre l'*upmix* et le mixage Atmos d'Amazon ($p=0,01$). Enfin, aucune différence significative n'a été perçue entre l'*upmix* et les mixages Atmos ($p=0,66$).

Concernant le morceau alternatif de Tears for Fears, une tendance significative semble se dégager pour la sensation d'immersion du mixage stéréo. Une différence significative a été décelée entre le mixage stéréo et l'*upmix* ($p=0,01$). Une différence significative est notable entre le mixage stéréo et les mixages Atmos issus d'Apple Music ($p=0,0001$) et Amazon Music ($p=0,004$). Aucune différence significative n'a été perçue entre les deux mixages Atmos issus d'Amazon Music et Apple Music ($p=0,32$). Enfin, aucune différence significative n'a été perçue entre l'*upmix* et les mixages Atmos ($p=0,19$).

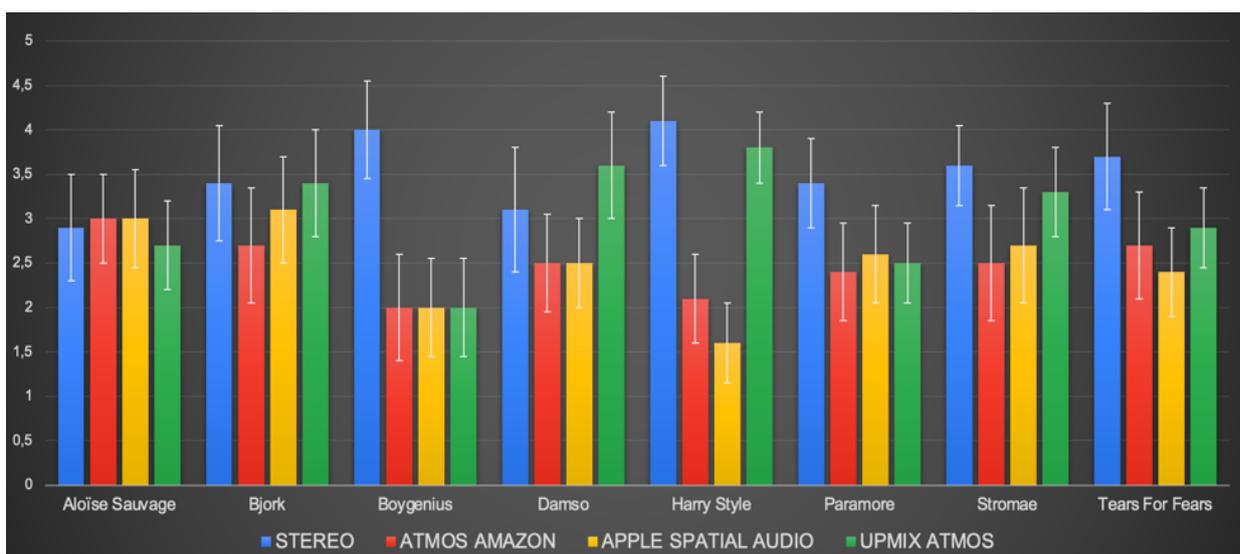


Figure 30 : Moyenne et écart-type des résultats obtenus pour la variable immersion

D.3. Profil des sujets

L'âge des sujets ayant participé au test est compris entre 22 et 50 ans. Concernant les systèmes de diffusion pour l'écoute musicale, 75% des sujets ont pour habitude d'écouter de la musique au casque, et 61% sur enceintes. Le temps d'écoute musical moyen hebdomadaire est de 10,5 heures par semaine. Concernant les supports de diffusion, 93% des sujets consomment de la musique sur les plateformes de streaming et 14% sur format physique (CD ou vinyle). Enfin sur l'ensemble des sujets, 36% des participant·e·s étaient familier·e·s et avaient expérimenté l'écoute musicale au format immersif (Dolby Atmos, Apple Spatial Audio, Sony 360RA).

D.4. Interprétations

À travers un test perceptif comparatif entre différentes versions de mixage d'un même titre, cette étude avait pour but d'étudier trois hypothèses concernant la préférence, la sensation d'immersion et la possibilité de différenciation entre les différents codecs de restitution binaurale. Il s'agira ici de confronter nos résultats aux hypothèses émises en 3.B.1 afin de fournir une analyse pertinente et de dégager des tendances. Afin d'étayer notre propos, l'ensemble de cette analyse sera complété par les commentaires laissés par les participant·e·s sur le questionnaire perceptif.

Concernant le critère de préférence, l'hypothèse avait été formulée à partir des résultats de Simon et Luck (2018) : la réduction binaurale des mixages objets est préférée au mixage stéréo. Au vu de nos résultats, cette hypothèse ne semble pas être confirmée dans le cadre de notre étude. Une différence significative existe dans l'évaluation du critère de préférence pour l'ensemble des titres, mais une tendance significative semble être accordée au mixage stéréo et à l'*upmix*. En revanche la réduction binaurale des deux mixages orientés objets réalisés en Dolby Atmos ne présente pas de tendance significative pour la préférence et ne semble donc pas être privilégiée par les sujets. Les commentaires laissés par les sujets ont pu guider notre analyse sur l'évaluation du critère de préférence dont nous pouvons faire émerger plusieurs facteurs influents.

Le niveau sonore intrinsèque à chaque mixage est l'un des facteurs qui a pu influencer l'évaluation du critère de préférence. L'influence du niveau a été minimisée pour ce test en normalisant l'ensemble des extraits à -18dB LUFS intégrés. Pour autant, une normalisation

parfaite était compliquée à réaliser et des différences existantes entre les différents extraits ont pu influencer la perception de la dynamique : niveau RMS, transitoires, contenu fréquentiel... Deux sujets ont indiqué ressentir une sensation plus « punchy » sur les mixages stéréo et moins puissante pour les mixages Atmos sur les titres de Harry Style et Tears For Fears. Trois participants ont également rapporté une perception de différence de niveau sonore pour le titre de Stromae entre le mixage stéréo et l'*upmix* par rapport aux mixages Atmos. Dans les faits, une fois les titres normalisés, les mesures que nous avons réalisées présentaient de faibles différences dans les mesures de loudness (Cf. Annexe 8) :

TITRE	<i>Loudness Range Unit – Mix Stéréo</i>	<i>Loudness Range – Mix Atmos</i>
Harry Style	1,9 LU	2,3 LU
Tears For Fears	1,9 LU	2,3 LU
Stromae	9,9 LU	11,8 LU

Figure 31 : Mesures de LRA

À l'exception de ces extraits, aucune autre disparité de niveau n'a été commentée par les participant·e·s. Ces indications quant à la préférence d'une version de mixage par rapport à une autre correspondent aux discussions que nous avons eues avec des mixeurs lors de nos entretiens où il a été confirmé que la concurrence de la stéréo était forte et que les mixages Atmos étaient perçus moins fort car moins compressés.

Le timbre semble être un autre facteur influant dans les choix de préférence d'une version de mixage à une autre. De nombreux sujets ont indiqué avoir préféré les mixages stéréo aux mixages Atmos en raison de la différence de timbre entre les deux versions. Aux mixages Atmos sont associées une sensation de basses fréquences importante et un côté « *boomy* », en particulier sur les titres de Harry Styles et Damso.

Le timbre de la voix semble également privilégié sur les mixages stéréo, où sur les mixages Atmos il est décrit comme « flou » et « filtré », sur les titres de Bjork, Damso et Paramore notamment. Cette description correspond également aux retours que nous avons pu collecter lors de nos entretiens et à la problématique du détimbrage lors du passage des enceintes au casque pour les mixeurs.

Concernant le critère d'immersion, l'hypothèse avait été formulée à partir des résultats de Simon et Luck (2018) : la sensation d'immersion est plus importante sur les réductions binaurales des mixages objets que pour les mixages stéréo. Au vu de nos résultats, cette hypothèse ne semble pas être confirmée dans le cadre de notre étude. Une différence significative existe dans l'évaluation du critère d'immersion pour l'ensemble des titres, mais une tendance significative semble être accordée au mixage stéréo. En revanche la réduction binaurale des deux mixages orientés objets réalisés en Dolby Atmos ne présente pas de tendance significative pour la préférence et ne semble donc pas être privilégiée par les sujets. Au même titre que pour le critère de préférence, les commentaires laissés par les sujets ont pu guider notre analyse sur l'interprétation de l'évaluation du critère d'immersion.

Pour la plupart des sujets, le critère d'immersion a été corrélé à des ressentis de largeur, enveloppement, frontalité, proximité et présence associés au mixage stéréo. La sensation rapportée par les participants est en accord avec l'écoute en stéréo au casque, qui offre une séparation parfaite des signaux entre les deux oreilles, une capacité qui n'est pas atteinte avec le binaural. Par ailleurs, lors des évaluations des mixages Atmos, les participants ont signalé une externalisation de certaines sources centrales, telles que la voix, avec un effet de recul et un positionnement plus flou par rapport aux mixages stéréo, comme dans le titre de Tears For Fears. Certains sujets pour lesquels les HRTF ont le mieux marché ont réussi à percevoir des sources sonores dans le plan arrière. Cependant, cette externalisation a provoqué une réaction inverse en termes de sensation d'immersion chez les sujets qui y ont associé un ressenti "perturbant", en ayant l'impression que quelqu'un était situé devant ou derrière eux. De plus, la sensation de profondeur et la précision de spatialisation ont été évoquées pour caractériser les mix Atmos à l'inverse des mixages stéréo. Pour autant, ces critères n'ont pas été associés à une sensation d'immersion, à l'inverse de l'aspect frontal de la stéréo. Ainsi, la latéralisation, associée à la stéréo, semble être privilégiée par rapport à l'externalisation, associée aux mixages Atmos, lors de l'évaluation du critère d'immersion. Cependant, comme nous l'avons évoqué dans les limites du test, la notion d'immersion englobe également d'autres critères qui n'ont pas été évoqués par les sujets dans les commentaires du test perceptif : réalisme de la scène sonore, implication émotionnelle... Il est intéressant de noter que l'évaluation de la sensation d'immersion par les sujets s'est donc basée en grande partie sur des caractéristiques spatiales de l'image sonore.

Plusieurs facteurs peuvent venir expliquer l'évaluation des trois titres (R&B, Électro, Variété) où aucune tendance significative n'a été décelée. Plusieurs sujets ont en effet indiqué que leur préférence pour une version d'un mixage pouvait changer en fonction du passage de l'extrait et de l'arrangement. Cette remarque a particulièrement été soulignée sur l'extrait de Stromae, où l'hésitation se caractérisait entre les chœurs présents dans le passage introductif et le premier couplet, et également sur le titre de Bjork où l'ensemble de l'arrangement électronique est très évolutif. Pour le titre d'Aloïse Sauvage, l'évaluation du critère de préférence et d'immersion a été jugé très difficile au vu de la proximité des mixages remarqué par la plupart des sujets. Il est intéressant de mettre en perspective ces résultats avec l'entretien que nous avons réalisé avec Oscar Ferran, mixeur Atmos de ce titre, où il nous indiquait vouloir coller au maximum le mixage stéréo lors du mixage Atmos. En ce sens, les résultats perceptifs obtenus sont cohérents avec la pratique de mixage de l'ingénieur du son.

Enfin, l'évaluation des deux critères préférence et immersion devait être réalisée de manière indépendante par les sujets. Pour autant, au vu des résultats, les tendances sur l'ensemble des titres et sur les titres analysés individuellement semblent être similaires pour ces deux critères. Il semble y avoir une corrélation entre la préférence et la sensation d'immersion pour le format stéréo. Dans cette optique, plusieurs participants ont indiqué avoir évalué leur préférence en fonction de leur expérience d'immersion. De plus, plusieurs titres ne présentant pas de différence significative entre les versions de mixage sont communs aux deux critères (R&B et variété), permettant ainsi d'identifier les extraits pour lesquels les différences entre mix stéréo et Atmos sont le plus flou.

Pour conclure, bien que nos résultats ne prétendent pas démontrer l'hégémonie du mixage stéréo par rapport au mixage Atmos pour l'écoute au casque, il est intéressant de voir que des tendances significatives existent. L'issue de ce test ne confirme pas les résultats trouvés par Simon et Luck (2018) pour le cas du Dolby Atmos Music. L'analyse des tendances confirme les problématiques intrinsèques à la réduction binaurale que nous avons pu mettre en avant dans les recherches et entretiens préalablement menés à ce test perceptif : détimbrage, localisation du centre flou, mauvaise perception des plans arrière et en élévation. Par ailleurs, il convient de souligner que l'externalisation, qui constitue un avantage inhérent au binaural, ne semble pas être corrélée à une perception d'immersion, contrairement à la

latéralisation associée à la stéréo. Il est également intéressant de noter que les titres qui se démarquent le plus dans les tendances de préférence et d'immersion pour la stéréo sont associés à des arrangements musicaux comprenant en majorité des sources instrumentales acoustiques (rock, folk, pop). Les titres les moins significatifs en termes de différence comprennent eux un arrangement beaucoup plus électronique (R&B, électro, variété). On peut faire l'hypothèse que les arrangements électroniques semblent se prêter le mieux à un mixage Atmos pour l'écoute casque, en raison des filtres HRTF qui sur de l'électro apportent une couleur différente et non un changement de timbre d'une source acoustique. De plus, notre étude n'a pas pu dégager de tendance significative sur les différences entre les algorithmes de rendu associés à chaque plateforme de streaming. Il conviendrait de réaliser un test perceptif centré sur cette problématique afin de faire émerger des tendances éventuelles.

Enfin, afin de pouvoir interpréter plus précisément nos résultats par rapport aux pratiques des mixeurs, il aurait été intéressant de connaître à quel point et de quelle manière chaque ingénieur du son a travaillé la compatibilité de la réduction binaurale pour l'écoute au casque. Alors que l'expérience dans le domaine de la stéréo est considérable et que la réalisation de ces mixages est maîtrisée par de nombreux professionnels, l'élaboration de mixages spatiaux de qualité est plus complexe. La maîtrise du rendu ADM reste l'un des défis auxquels font face les ingénieur-es du son, avec des limitations et des incertitudes quant au résultat final.

Conclusion Générale

Notre recherche a permis de dresser un état des lieux de la production audio orientée objet dans le domaine des musiques actuelles. Ces nouvelles technologies émergent dans un contexte de production caractérisé par une diversité croissante des supports, des modes de consommation et des utilisations des contenus audio, offrant ainsi un terrain propice à l'innovation. Par un coût important, tant dans le développement de nouvelles cabines en studio que dans l'allongement des durées de postproduction, les technologies telles que le Dolby Atmos Music représentent un pari sur l'avenir. L'arrivée du format ADM et la possibilité de fournir l'ensemble du marché avec un seul mix représente un changement économique probable dans la chaîne de postproduction d'un titre en musique actuelle. Par sa possibilité d'adaptabilité au système d'écoute des utilisateurs, l'audio orienté objet se différencie de la quadraphonie et du format 5.1 qui ont échoué à ne pouvoir être commercialisé au grand public. Dans cette perspective, alors que le multicanal était auparavant réservé à une minorité d'utilisateur·ice·s, ce mémoire a choisi d'aborder la problématique de la restitution du mixage objet au casque qui constitue actuellement le moyen le plus accessible pour un large public d'accéder à une expérience de mixage immersif.

Notre étude a permis de montrer que les solutions de rendu binaural développées par Dolby et Apple sont encore insuffisantes pour pouvoir proposer une nouvelle expérience d'écoute musicale au casque. La concurrence de la stéréophonie est encore trop rude et la technologie Atmos n'est pas encore suffisamment au point pour pouvoir rivaliser sur l'expérience d'écoute au casque. Les algorithmes de rendu binauraux ne sont pas assez performants sur la sensation d'immersion et l'expérience vécue pourrait se résumer à une stéréophonie de moins bonne qualité, qui paraît moins forte car moins compressée. Une amélioration des moteurs de rendu binaural semble donc nécessaire pour pouvoir faire concurrence au format stéréophonique.

Néanmoins, nous nous sommes heurtés à certaines limites. Le champ de notre recherche s'est vu volontairement large pour dresser un panorama suffisant des pratiques et problématiques actuelles associées au mixage orienté objet et à l'écoute au casque. Notre partie pratique associée aux tests perceptifs a dressé des tendances significatives dans la différence entre le mixage stéréophonique et Atmos. Pour autant, elle aurait pu gagner en

significativité en restreignant le test à une comparaison plus simple pour les sujets. De même, le nombre d'entretiens réalisés et le manque de temps que nous avons rencontré ne nous a pas permis d'élaborer une analyse précise des pratiques des ingénieur·e·s du son. Nous aurions gagné en termes de représentativité à généraliser notre grille d'entretiens à plus de professionnel·les. Enfin, nous nous sommes confrontés au fait que le Dolby Atmos Music soit une technologie récente. D'une part dans les entretiens, où les mixeurs nous ont parfois avoué ne pas avoir suffisamment de recul sur la technologie. D'autre part, nous aurions souhaité pouvoir obtenir plus de transparence de la part des constructeurs afin de mieux cerner le fonctionnement des différents outils associé à leur technologie.

Finalement, les perspectives à cette recherche sont multiples. Deux axes d'ouverture principaux se dégagent de notre étude : l'évolution des conditions de restitution au casque pour le consommateur et la prise en compte du potentiel esthétique associé au mixage orienté objet dès la phase de création. Concernant l'amélioration des moteurs de rendu pour l'écoute au casque, de nombreuses pistes d'évolutions sont multiples : amélioration des bases de données HRTF et des moteurs de convolution, développement des solutions de *head-tracking*, individualisation et personnalisation de l'écoute binaurale. De plus, les compétences et techniques de mixages vont également évoluer. L'avenir de ces technologies récentes est encore inconnu et dépendra de leur réception au sein du grand public. En ce sens, l'amélioration des réductions binaurales pour l'écoute au casque est un défi majeur que les constructeurs devront relever pour proposer une nouvelle expérience d'écoute. Concernant les perspectives pour la phase de création, nous avons vu à quel point les enjeux esthétiques associés aux nouvelles technologies de mixage objets dépendaient de la phase d'écriture et de pré-production. Plusieurs champs de recherches sont alors possibles dans ce sens : développement d'outils, plugin et instrument intégrant des solutions de spatialisation orienté objet dans les logiciels de création.

Le mixage orienté objet représente donc un nouveau terrain de jeu tant pour les ingénieur·e·s du son que les créateur·ice·s. Les champs de possibilités, de recherche et d'engagement sont vastes. Il faudra pour autant des évolutions importantes, aux différents niveaux de la chaîne de production d'une œuvre, pour que le modèle de consommation se

détache du système orienté canal et que puissent émerger de nouvelles propositions artistiques.

Références

ABEHSERA MORELL, P., & LEE, H. (2021). *Binaural Mixing of Popular Music*. AES: Journal of the Audio Engineering Society.

ANDRE, G. (2017). *Adaptation rapide à une spatialisation binaurale non-individualisée : Cas de la réalité virtuelle appliquée au jeu vidéo*. Mémoire de fin d'études, École Nationale Supérieure Louis Lumière.

APPLE, (2023) *Formats de monitoring Dolby Atmos dans Logic Pro*. Consulté 10 avril 2023, à l'adresse <https://support.apple.com/fr-fr/guide/logicpro/lgcp179f27c1/mac>

BAHU, H. (2016). *Localisation auditive en contexte de synthèse binaurale non-individuelle*. Traitement du signal et de l'image. Université Pierre et Marie Curie.

BEGAULT, D. R. (1990). *The Composition of Auditory Space : Recent Developments in Headphone Music*. Leonardo, vol. 23, pp. 45-52.

BERGER, C. C., GONZALEZ-FRANCO, M., TAJADURA-JIMÉNEZ, A., FLORENCIO, D., & ZHANG, Z. (2018). *Generic HRTFs May be Good Enough in Virtual Reality. Improving Source Localization through Cross-Modal Plasticity*. Frontiers in Neuroscience, vol. 12, pp. 21.

BERKHOUT, A., VRIES, D., & VOGEL, P. (1993). *Acoustic control by wave field synthesis*. The Journal of the Acoustical Society of America, vol. 93, pp. 2764-2778.

BLAUERT, J. (1996). *Spatial Hearing : The Psychophysics of Human Sound Localization*.

BOREN, B. (2017). *History of 3D Sound*. In *Immersive Sound* (1^{re} éd., p. 40-62). Routledge.

BYRNE, D. (2013). *How Music Works* (Main édition). Canongate Books Ltd.

COLIN, E. (2019). *L'audio orienté objet : Une solution face à la multiplicité des formats*. Mémoire de Master « Cinéma et Audiovisuel », Aix-Marseille Université.

DAVIS, M. (2003). *History of spatial coding*. Journal of the Acoustical Society of America, 51, 554-569.

DIXON, G. (1979). *The Origins of the Roman 'Colossal Baroque'*. Proceedings of the Royal Musical Association, vol. 106, pp. 115-128.

DOLBY LABORATORIES INC., (s. d.) *Dolby Atmos for Content Creators*. Consulté 30 mars 2023, à l'adresse

<https://professional.dolby.com/content-creation/Dolby-Atmos-for-content-creators/2>

DOLBY LABORATORIES INC, (2021). *Dolby Atmos Music Room Configuration. Best practices for music room configuration*. V.11.2021.

DOLBY LABORATORIES INC, (s. d.). *Dolby Atmos Music Studios*. Consulté 10 avril 2023, à l'adresse <https://professional.dolby.com/music/dolby-atmos-music-studios/>

DOLBY LABORATORIES INC, (s. d.). *Dolby Atmos Renderer*, Consulté 10 avril 2023, à l'adresse <https://professional.dolby.com/product/dolby-atmos-content-creation/dolby-atmos-renderer/>

DOLBY HYBRIK, (s. d.). *Dolby Atmos—Hybrik Tutorials*. Consulté 10 avril 2023, à l'adresse <https://tutorials.hybrik.com>

DOLBY LABORATORIES INC, (2015). *Dolby AC-4: Audio delivery for next-generation entertainment*.

DOLBY LABORATORIES INC., (2013). *Authoring for Dolby Atmos Cinema Sound Manual*.

DOLBY LABORATORIES INC., (2021) *How do I QC my Dolby Atmos mix?* Consulté 10 avril 2023, à l'adresse https://professionalsupport.dolby.com/s/article/How-do-I-QC-my-Dolby-Atmos-mix?language=en_US

DOLBY LABORATORIES INC., (2023) *Is there a loudness target for music mixed in Atmos?* Consulté 10 avril 2023, à l'adresse https://professionalsupport.dolby.com/s/article/Is-there-a-loudness-target-for-music-mixed-in-Atmos?language=en_US

EBU ADM Guidelines, (s. d.). *Types of Audio*. Consulté 10 avril 2023, à l'adresse https://adm.ebu.io/background/audio_types.html

FONTANA, S., FARINA, A., & GRENIER, Y. (2007). *Binaural for Popular Music : A Case of Study*. Ecole Nationale Supérieure des Télécommunications, TSI Paris, France.

FLUX AUDIO, (s. d.) *Panning Algorithms*. Consulté 10 avril 2023, à l'adresse https://doc.flux.audio/en_US/spat_revolution_doc/Spatialisation_Technology_Panning_Algorithms.html

GARDNER, B., GARDNER, S., (1973). *Problem of localization in the median plane: Effect of pinnae cavity occlusion*. Journal of the Acoustical Society of America, vol. 53, pp. 400–408.

GERZON, M. (1973). *Periphony : With-Height Sound Reproduction*. AES Journal Forum.

GERZON, M. A., & WORAM, J. M. (1976). *Blumlein Stereo Microphone Technique- and Author's Reply*. Journal of the Audio Engineering Society, vol. 24, pp. 36-38.

GRAY, J., (2023). *Mastering Dolby Atmos Music with Justin Gray—Approaches To Immersive Object-Based Mastering*. YouTube.

<https://www.youtube.com/watch?v=FcZWOCMVkWE&list=WL&index=3>

GRESINGER, D. (1990). *Binaural Techniques for Music Reproduction*. AES: Journal of the Audio Engineering Society.

HOLMAN, T. (2008). *Surround Sound : Up and running*. Routledge & CRC Press.

IFPI, (2022). *La consommation de musique dans le monde - Enquête 2022*

ITU, (2019) *BS.2076 : Modèle de définition audio*. Consulté 10 avril 2023, à l'adresse <https://www.itu.int/rec/R-REC-BS.2076/fr>

KAPRALOS, B., JENKIN, M., & MILIOS, E. (2008). *Virtual Audio Systems. Presence Teleoperators & Virtual Environments*, Association for Computer Machinery, vol. 17, pp. 527-549.

KEMPE, L., (2023) *TIDAL's new wave of music*. Dolby Professional. Consulté 10 avril 2023, à l'adresse <https://professional.dolby.com/music/tidal-ac4-atmos/>

LÉON, C., FÉLICIE DAVID, & ENGUERRAND, DU R. (2014). *Écoute de musique amplifiée, analyse des comportements chez les 15-35 ans : résultats du baromètre santé*. Santé Publique France.

LIITOLA, T., (2006). *Headphone Sound Externalization*. Helsinki University Of Technology Department of Electrical and Communications Engineering

LORD, J., (2022). *Redefining the spatial stage : Non-fronte-oriented approaches to periphonic sound staging for binaural reproduction*. In *3D Audio* (Routledge).

LORD, J., (2022). *The Phenomenal Rise of Periphonic Record Production, A practice-based, musicological investigation of periphonic recomposition and mix aesthetics for binaural reproduction*. University of West London.

MARSTON, D., (2015). *Audio Definition Model Software*. BBC R&D. Consulté 10 avril 2023 à l'adresse <https://www.bbc.co.uk/rd/publications/audio-definition-model-software>

MERGING TECHNOLOGIES, (s. d.). *Use Cases : MORTEN LINDBERG 2L - NORWAY*. Consulté 10 avril 2023, à l'adresse <https://www.merging.com/news/use-cases/morten-lindberg-2l-norway>

MESSONNIER, J.-C., LYZWA, J.-M., & LING, A., (2022). « Méthodes de prise de son en multicanal / Objets ». Dans *La perception du son spatialisé : Un son propre pour un sens figuré*. Cahier Louis-Lumière n°15, 2022.

MOORE, A. F., (2001). *Rock : The Primary Text : Developing a Musicology of Rock*. 2nd edn Aldershot, Ashgate

MORETTI, L. (2004). Architectural spaces for music: Jacopo Sansovino and Adrian Willaert at St Mark's. *Early Music History*, vol. 23, pp. 153-184.

MOYLAN, W., (2015). *Understanding and Crafting the Mix : The Art of Recording*. Routledge & CRC Press.

MUSHENDWA, S., (2009). *Enhancing Headphone Music Sound Quality*. Aalborg University – Institute of Media Technology and Engineering Science.

NAIR, S., (2021). *After the Mix : Encoding and Delivering Dolby Atmos Music*. Consulté 10 avril 2023, à l'adresse <https://www.avid.com/fr/resource-center/encoding-and-delivering-dolby-atmos-music>

NIKLISSON, F., (2021). *Spat Revolution and Dolby Atmos workflow*. FLUX:: Immersive. . Consulté 10 avril 2023 <https://www.flux.audio/2021/01/28/spat-revolution-and-dolby-atmos-workflow>

PAUL, S., (2009). *Binaural Recording Technology : A Historical Review and Possible Future Developments*. *Acta Acustica*, vol. 95, pp. 767-788.

DOCKWRAY, R., MOORE, A. F., (2010) *Configuring the sound-box 1965–1972.*, Cambridge University Press 2010

PÉRIAUX, B., OHL, J.-L., & THÉVENOT, P. (2015). *Le son multicanal : De la production à la diffusion du son 5.1, 3D et binaural*. Dunod.

PUREMIX, (2021). *Andrew Scheps Mixing with Dolby Atmos*. Youtube. https://www.youtube.com/watch?v=PCS3_vxtBUE

PUREMIX, (2021). *Dolby Atmos Discussion Panel w/ Andrew Scheps Talks To Awesome People*. Youtube. <https://www.youtube.com/watch?v=cnX4pVrxqq0>

EMI ARCHIVE TRUST, (s. d.). *Alan Blumlein and the invention of Stereo*. Consulté 10 avril 2023, à l'adresse <https://www.emiarchivetrust.org/alan-blumlein-and-the-invention-of-stereo/>

RUMSEY, F., ZIELIŃSKI, S., KASSIER, R., & BECH, S., (2005). *Relationships between experienced listener ratings of multichannel audio quality and naive listener preferences*. The Journal of the Acoustical Society of America, vol. 117, pp. 3832-3840.

SCHÖNSTEIN, D., & KATZ, B. (2010). *Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale*.

SIMON, A., & LUCK, G. (2018). *Perceptual differences and preferences between binaural and stereo mixes of music*. University of Jyväskylä.

SINGLETON, M. (2022). *Apple Music's Spatial Audio Strategy Is Paying Off With More Listeners – Billboard*. Consulté 10 avril 2023, à l'adresse <https://www.billboard.com/pro/apple-music-spatial-audio-strategy-listeners-releases/>

SPORER, T., PLOGSTIES, J., & BRIX, S. (2001, mai 1). *CARROUSO - An European Approach to 3D-Audio*. Audio Engineering Society Convention 110. <https://www.aes.org/e-lib/browse.cfm?elib=10014>

STITT, P., HENDRICKX, E., MESSONNIER, J.-C., & KATZ, B., (2016). *The Role of Head Tracking in Binaural Rendering*. 29th Tonmeistertagung - VDT International Convention.

STOCKHAUSEN, K. (1972). *Lecture 5—[Part 2/3] Karlheinz Stockhausen—Four Criteria of Electronic Music (KONTAKTE), (1972)*. YouTube. <https://www.youtube.com/watch?v=ZzDhg-MoHlo>

STROUD, R. (2006). *Reference vs. Preference : The 'Surround' of Surround Sound*. Stroud Audio Inc.

SUNDER, K. (2022). *Binaural audio engineering*. In *3D Audio* (Routledge).

USMANI, M. H., CEPEDA, R., SULLIVAN, T., & RAJ, B. (2015). Improving headphone spatialization for stereo music. *Journal of the Acoustical Society of America*, 137, 2330-2330.

WENZEL, E., ARRUDA, M., KISTLER, D., & WIGHTMAN, F. (1993). *Localization using non individualized head-related transfer functions*. The Journal of the Acoustical Society of America, vol. 94, pp. 111-123.

Table des figures

FIGURE 1 : PRINCIPE DE FONCTIONNEMENT DE LA STEREO TELLE QUE PRESENTEE DANS UNE POCHETTE DE RCA	13
FIGURE 2 : ALBUM DE DEMONSTRATION DE LA TECHNOLOGIE STEREO (RCA RECORDS, 1958)	14
FIGURE 3 : REPRESENTATION DE L'ADAPTABILITE DU MODELE ORIENTE OBJET (COLIN, 2019)	19
FIGURE 4 : REPRESENTATION DU MODELE ADM (MARSTON, 2015)	20
FIGURE 5 : SCHEMA BLOC SIMPLIFIE DES MODELES AUDIO UTILISES DANS DOLBY ATMOS	23
FIGURE 6 : ILLUSTRATION DU PLUGIN DOLBY ATMOS MUSIC PANNER	24
FIGURE 7 : INTERFACE DU DOLBY RENDERER (DOLBY, 2023)	25
FIGURE 8 : FENETRE D'ATTRIBUTION DES PARAMETRES DE RENDU BINAURAUX PAR OBJET DANS LA SUITE DE PRODUCTION DE DOLBY (DOLBY, 2023)	26
FIGURE 9 : EXEMPLE D'UNE CHAINE DE PRODUCTION DOLBY ATMOS (DOLBY, S. D.)	27
FIGURE 10 : SCHEMA BLOC : ENCODAGE SOUS UN CODEC DOLBY (NAIR, 2021)	28
FIGURE 11 : PRINCIPE DU CLUSTERING : PASSAGE DE 20 A 12 CHANNELS (NAIR, 2021)	29
FIGURE 12 : EXEMPLE D'ITD ET ILD (KAPRALOS ET AL., 2008, P.2)	30
FIGURE 13 : CONE DE CONFUSION (KAPRALOS ET AL., 2008, P.3)	32
FIGURE 14 : PROBLEMATIQUE DE LA LATERALISATION LORS DE LA REPRODUCTION D'UN CONTENU STEREO AU CASQUE (USMANI, 2015, P.1)	33
FIGURE 15 : PRINCIPE DU RENDU BINAURAL (SUNDER, 2022)	35
FIGURE 16 : PRINCIPE DE DIFFUSION D'UN CONTENU AUDIO ATMOS EN STREAMING AU FORMAT AC4 (DOLBY, 2015)	40
FIGURE 17 : RESUME DES CODECS ASSOCIES AUX MOTEURS DE RENDU BINAURAUX SUR LES PLATEFORMES DE STREAMING	42
FIGURE 18 : PRINCIPE DU CONE D'ECOUTE PERIPHONIQUE (LORD, 2022)	47
FIGURE 19 : EXEMPLE D'OMNIMONOPHIE (LORD, 2022)	48
FIGURE 20 : FORME D'ONDE DE LA REDUCTION BINAURALE DU MIX ATMOS (MORCEAU : « AS IT WAS » - HARRY STYLE)	57
FIGURE 21 : FORME D'ONDE DU MIX STEREO (MORCEAU « AS IT WAS » - HARRY STYLE)	57
FIGURE 22 : RESUME DES TITRES SELECTIONNES POUR LE TEST PERCEPTIF	66
FIGURE 23 : INTERFACE DU PLUGIN HALO UPMIX DE NUGEN AUDIO	67
FIGURE 24 : SESSION DE TEST SOUS REAPER	68
FIGURE 25 : MOYENNE ET ECART-TYPE DE LA VARIABLE PREFERENCE SUR L'ENSEMBLE DES 8 TITRES	73
FIGURE 26 : RESULTATS DES DIFFERENCES SIGNIFICATIVES POUR LE CRITERE PREFERENCE EN FONCTION DU GENRE	73
FIGURE 27 : MOYENNE ET ECART-TYPE DES RESULTATS OBTENUS POUR LA VARIABLE PREFERENCE	75
FIGURE 28 : MOYENNE ET ECART-TYPE DE LA VARIABLE IMMERSION POUR L'ENSEMBLE DES HUIT TITRES	76
FIGURE 29 : RESULTATS DES DIFFERENCES SIGNIFICATIVES POUR LE CRITERE IMMERSION EN FONCTION DU GENRE	76
FIGURE 30 : MOYENNE ET ECART-TYPE DES RESULTATS OBTENUS POUR LA VARIABLE IMMERSION	78
FIGURE 31 : MESURES DE LRA	80
FIGURE 32 : ANOVA CRITERE « PREFERENCE » - ENSEMBLE DES 8 TITRES	143
FIGURE 33 : ANOVA CRITERE « PREFERENCE » - ALOÏSE SAUVAGE	143
FIGURE 34 : ANOVA CRITERE « PREFERENCE » - BJORK	143
FIGURE 35 : ANOVA CRITERE PREFERENCE - BOYGENIUS	143
FIGURE 36 : ANOVA CRITERE « PREFERENCE » - DAMSO	144
FIGURE 37 : ANOVA CRITERE « PREFERENCE » - HARRY STYLE	144
FIGURE 38 : ANOVA CRITERE « PREFERENCE » - PARAMORE	144
FIGURE 39 : ANOVA CRITERE « PREFERENCE » - STROMAE	144
FIGURE 40 : ANOVA CRITERE « PREFERENCE » - TEARS FOR FEARS	145
FIGURE 41 : ANOVA CRITERE « IMMERSION » - ENSEMBLE DES 8 TITRES	146
FIGURE 42 : ANOVA CRITERE « IMMERSION » - ALOÏSE SAUVAGE	146
FIGURE 43 : ANOVA CRITERE « IMMERSION » - BJORK	146
FIGURE 44 : ANOVA CRITERE « IMMERSION » - BOYGENIUS	147
FIGURE 45 : ANOVA CRITERE « IMMERSION » - DAMSO	147
FIGURE 46 : ANOVA CRITERE « IMMERSION » - HARRY STYLE	147
FIGURE 47 : ANOVA CRITERE « IMMERSION » - PARAMORE	147
FIGURE 48 : ANOVA CRITERE « IMMERSION » - STROMAE	148
FIGURE 49 : ANOVA CRITÈRE « IMMERSION » - TEARS FOR FEARS	148

Annexes

Annexe 1 : Normes de calibration pour un studio Atmos

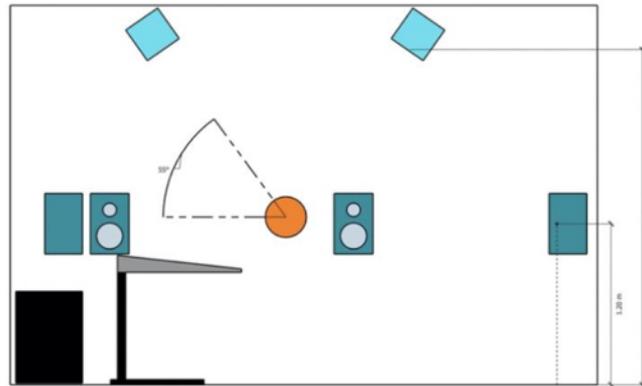


Figure 4 Dolby Atmos Music Room Elevation showing Minimum Dimensions

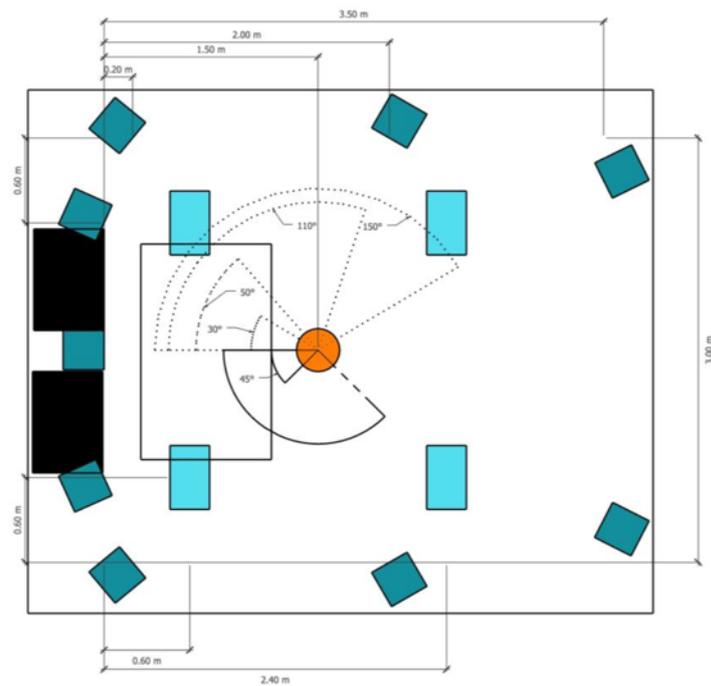


Figure 3: 9.1.4 Dolby Atmos Music Room Layout showing Minimum Dimension

Annexe 2 : Guide d'entretien et formulaire de consentement

1. Est-ce que tu peux te présenter (nom, profession, parcours, années d'expérience) ?
2. Comment définirais-tu ton métier de "mixeur immersif" à quelqu'un qui ne le connaît pas ?
3. Quelles sont tes expériences associées au mixage orienté objet ? (Technologies, types de projets / demandes de labels, contraintes de temps)
4. Apport du mixage objet :
 - Quelles conséquences les évolutions apportées par l'arrivée du mixage objet ont-elles eu sur tes pratiques de mixage ?
 - Qu'est-ce que tu pourrais dire des possibilités esthétiques qui sont apportées par le mixage objet, s'il y en a ?
5. *Monitoring / Casque*
 - Comment contrôle-tu les différents types de décodage et formats de restitution ? (Plateformes, rendu enceintes et casque)
 - Ton utilisation du casque pour mixer a-t-elle évolué avec l'arrivée du mixage objet ?
 - As-tu rencontré des problèmes / difficultés lors du passage du mixage sur enceintes à la réduction binaurale au casque ? Si oui lesquelles ?
6. D'après toi, le mixage orienté objet peut-il offrir une nouvelle expérience d'écoute au casque ? De quelle manière ?
7. Quelles évolutions techniques pourraient permettre d'améliorer ta méthode de travail dans le futur ?
8. Comment vois-tu la pérennité de cette technologie dans les années à venir ?
9. Autres points que tu souhaiterais aborder ?

FORMULAIRE DE CONSENTEMENT PRÉALABLE À L'ENTRETIEN

Présentation du projet :

Cet entretien se tiendra dans le cadre de la rédaction du mémoire de fin d'étude de l'ENS Louis Lumière de l'étudiant Pierre-Yves COLLARD (direction interne : Alan BLUM, direction externe : Martin ANTIPHON) qui traite de la production musicale orientée objet et des enjeux d'écriture et mixage pour l'écoute au casque.

Déroulé de l'entretien :

Votre participation prendra la forme d'un entretien pouvant durer entre 30 et 90 minutes, dans lequel on vous posera une série de questions sur vos expériences et vos pratiques professionnelles. L'entretien sera enregistré sous format audio en vue d'une retranscription écrite, et ne sera jamais diffusé. La participation à l'étude étant volontaire, aucune indemnisation n'est prévue pour y participer.

Confidentialité :

Vos réponses lors de cet entretien ne seront utilisées que dans le cadre de cette étude. Vous ne serez pas anonymisé, votre nom et votre fonction seront révélés dans le cadre de la présentation des résultats de l'analyse. Vous pouvez toutefois choisir d'être anonymisé.e, auquel cas votre nom ne sera pas cité et vous serez désigné.e par un numéro.

Vos droits :

Vous pourrez décider de vous retirer de cette étude à tout moment sans donner de raison. Si vous décidez de vous retirer, il vous sera demandé ce que vous voulez faire des données déjà collectées.

Vous pouvez poser des questions au sujet de la recherche à tout moment (avant, pendant et après votre participation) en me contactant par mail (py.collard09@gmail.com) ou par téléphone (0619984148).

A remplir par le participant :

Date, Nom, Prénom :

- J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré la participation à cet entretien ainsi que sa captation audio à des fins de recherche. J'accepte que mon identité et ma fonction soient révélées dans le cadre de la présentation des résultats de l'analyse.

Signature :

- J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré la participation à cet entretien ainsi que sa captation audio à des fins de recherche. Je n'accepte pas que mon identité soit révélée dans la présentation des résultats de l'analyse et je souhaite être anonymisé(e).

Signature :

A remplir par l'expérimentateur,

Le / / 2023 à

Pierre-Yves COLLARD



Annexe 3 : Entretien avec Baptiste Murguiondo

Est-ce que tu pourrais te présenter ?

Alors je m'appelle Baptiste Murguiondo, j'ai travaillé pendant 4 ans au studio Guillaume Tell à peu près. On faisait principalement de la stéréo et du 5.1 dans le studio A et du 5.1 orienté musique de film dans le B. À la suite de quoi s'est posé la question de l'Atmos réellement parce qu'on a vu un intérêt pour l'économie du studio et puis pour l'économie au global et que ça allait potentiellement démarrer. Donc on a monté la cabine Atmos au sein du studio. C'est moi qui ai été mis en charge entre guillemets de superviser toute la partie installation, routing et setup. Alors évidemment on a fait ça avec des entreprises. On a fait ça avec CTM Solutions qui sont venues avec une idée globale du projet, du routing et de l'intégration du setup dans la cabine actuelle. À la suite de ça on a mis en place avec eux toute la partie routing et toute la partie intégration et on a bossé. J'ai longuement parlé avec Dolby sur toute leur intégration, tout leur système, ce qui fait que je commence à avoir une bonne vision même s'il y a un certain nombre de choses qui ont été éclaircies ces derniers temps, notamment sur le binaural. Aujourd'hui j'ai quitté le studio, je me suis lancé un indépendant. Je continue toujours à faire du mix Atmos, notamment en binaural, et je travaille maintenant quasiment essentiellement pour pas dire exclusivement mais essentiellement avec ça parce que j'ai quand même des systèmes d'écoute multicanal, mais c'est des écoutes beaucoup plus consumer-end et non professional-end.

Comment toi tu définirais ton métier à quelqu'un qui ne connaîtrait pas le rôle du mixeur immersif ?

Aujourd'hui je dirais que le métier de mixeur immersif en musique n'est rien d'autre qu'un grand panner. Aujourd'hui le travail, sur le cas de l'immersif, c'est de faire des pan parce que la contrainte technologique est qu'aujourd'hui on nous demande quand même de fournir des mix stéréos en amont du mix Atmos. Le mix Atmos vient dans un second temps. Donc en fait, on repart du mix stéréo, ou des stems, et on va venir spatialiser. Alors oui il y a du travail de mix au sens où évidemment le fait de spatialiser t'oblige à travailler les niveaux, le fait de séparer les canaux t'oblige aussi à travailler fréquemment, parce que mine de rien on fait énormément de cut et de coupe fréquemment et dynamiquement dans le son pour le faire rentrer sur deux canaux (en stéréo). Là, on a beaucoup plus d'espace en immersif. Donc on a du travail d'ajustement mais aujourd'hui c'est très très rare de prendre un mix stéréo et de partir directement en immersif. C'est d'ailleurs assez compliqué de mon point de vue.

Du coup tu résumerai ça plutôt un travail de spatialisateur ?

Ouais, il y a d'abord tout un travail de mixage en stéréo et ensuite il y a cette deuxième étape où le mix immersif aujourd'hui se résume globalement à spatialiser et résoudre les problèmes de spatialisation.

Et du coup dans ce prolongement, moi je me demandais sur quel type de projet tu avais été amené à travailler en particulier ?

J'ai fait un peu de tout. Quand j'étais au studio, et même un peu maintenant, j'ai bossé avec Universal et Warner. 90% du temps, pour ne pas dire 100% du temps, on récupérait des stems et on faisait un

mix Atmos. On a eu des clients indépendants où on est parti de zéro pour faire le mix immersif et on a extrait la stéréo du mix immersif à travers le down mix. Concernant les contraintes de temps, quand je parlais des stems, j'avais plus ou moins deux heures montre en main pour récupérer le système et faire l'export. Alors globalement je prenais mon temps et je disais « bon on fait les deux heures pour finir le mix et l'export je le ferai de mon côté derrière ». Sachant qu'aujourd'hui je suis resté à peu près sur cette contrainte de temps alors je peux aller à 3h, mais globalement sur un morceau bien fait, bien pensé, bien travaillé de 2-3 minutes, tu n'as pas besoin de beaucoup plus de temps je trouve.

Je me demandais dans cette contrainte de temps j'imagine qu'en fait à aucun moment tu as le temps de dialoguer avec l'ingénieur-e stéréo ?

Non après globalement 90% des projets ce qu'on demande c'est d'être cohérent vis-à-vis du stéréo. Donc en fait tu as déjà une première étape : récupérer tes stems et checker que c'est cohérent par rapport au master stéréo, ce qui est rarement le cas puisqu'ils sont exportés pré-master évidemment. Donc déjà tu passes un tout petit peu de temps en stéréo ou en immersif à retrouver la couleur du Master. Et puis après tu vas venir spatialiser, mais ça va très très vite. Dans ce contexte-là, t'as pas forcément besoin de la vision de l'artiste. Après j'ai aussi fait des mix avec l'artiste mais alors là c'était plus dans le cas de l'électro où l'artiste était là. Mais évidemment là ça a pris plus de temps, on en avait eu pour 10h à peu près.

A part ces cas-là, qui sont plus rares, la majorité du temps tu n'as pas de validation de l'artiste, ça passe plus par le label ?

Ouais, et encore pour le label... Ça commence à changer petit à petit mais les gros cas de figure qu'on avait c'est que globalement personne n'avait moyen d'écouter en multicanal donc ils écoutaient le binaural. Sachant que le gros de quand on a bossé pour te dire, Dolby n'avait même pas intégré leur HRTF personnalisé. Donc comment dire que le binaural en tant qu'écoute critique c'était compliqué... Sachant que bon, de toute façon, les retours qu'on pouvait avoir vis-à-vis du QC (quality control) par rapport au binaural, on ne pouvait pas en faire grand-chose parce que de toute manière on n'avait pas forcément envie de sacrifier le multicanal. Globalement, les labels faisaient tourner des fiches en interne sur comment ils voulaient que leur session soit organisée en Atmos : vérifier qu'il n'y ait pas des objets vides qui se baladent, que ton fichier ne soit pas trop lourd, que tu aies bien des courbes HRTF sur tous tes paramètres binauraux pour être sûr que ça soit vraiment binaural indépendamment de si ça sonne bien ou pas. Et en général ça s'est arrêté là.

Ok. Concernant le monitoring, il y a à la fois cette question de la possibilité de rendu qui est propre au format objet qui fait qu'à partir d'un seul fichier tu peux t'adapter aux multiples formats de restitution et il y a aussi le fait qu'aujourd'hui, selon que ton fichier soit uploadé sur Apple Music ou sur Tidal, le décodage ne va pas être identique. Donc je me demandais : comment tu fais justement pour t'adapter à cette multiplicité de possibilités de restitution du mix ?

Globalement ce qu'on fait c'est que nous on bosse le multicanal. Au studio où j'étais on était dans des conditions pour savoir que ce qu'on faisait était cohérent. On rapprochait le binaural du multicanal globalement. Parce que de toute manière il y a des trucs qui sont très bizarres dans la manière dont Dolby interprète même leur binaural. Un des exemples les plus aberrants à mon sens c'est

typiquement que ton LFE (canal de sub) prend 5dB et demi quand tu passes en binaural. Il y a plein de choses sur lesquelles on n'a pas la main. Après pour parler de la différence de cas entre Apple et Tidal, on n'en fait pas grand-chose.

Concernant les codecs, Apple a pris le DD+JOC, Tidal a pris DD+JOC et l'AC4. Le gros problème c'est que les métadonnées binaurales sont encodées sur l'AC4 et non pas sur le DD+JOC. La problématique qui s'en est suivi de ça c'est qu'Apple se retrouve avec un fichier multicanal mais pas de données binaurales. Donc en fait ce qu'ils ont fait c'est qu'ils sont partis du concept de faire un downmix de ton ADM Atmos pour le virtualiser avec leur propre room model. Le binaural que tu as dans le casque Apple est virtualisé à partir de la room model d'Apple de leur 5.1.4 : c'est la version facile. La version un petit peu plus évoluée c'est que Apple, ils partent du principe que quand tu écoutes sur leur matériel (AirPods...) tu as fait ton HRTF personnalisé à partir d'iPhone ce qui est suffisant pour percevoir correctement les signaux et la binauralisation dans leur modèle. Par contre, ce que j'ai découvert c'est que si tu écoutes de l'Atmos au casque sur Apple music sans un appareil affilié Apple, il n'y a pas d'HRTF personnalisée et donc ils prennent en compte les metadatas que toi tu rentres dans ton Dolby Renderer (near, mid, far) et ils recréent eux-mêmes trois courbes HRTF propre à leur room modèle qu'ils incorporent dans leur room. Donc finalement il y a quand même des metadatas binaurales prisent en compte, mais différentes de celles de Dolby car ce ne sont pas les mêmes courbes HRTF.

Et comment Apple accède aux métadonnées binaurales du fichier ADM ?

Car l'encodage se fait au niveau de leur partie serveur. Toi tu leur livres un ADM, et du coup ils font leur encodage, alors ils n'ont pas accès au codec de l'AC4, mais ils ont quand même accès à l'ADM. Donc en fait encore une fois là pour le coup on rentre dans la supposition sur le pourquoi du comment ou autre. Mais il semblerait que ce soit ça.

Est-ce que tu perçois des différences entre ces deux plateformes ?

Oui, en plus Apple change leur rendu quasiment toutes les semaines, ce qui fait que d'une semaine à l'autre on n'entend pas forcément la même chose ou autre. Le rendu s'est vraiment amélioré parce qu'au tout début de l'Apple Music t'entendait vraiment une pièce (room). C'était ultra bizarre sans parler du fait que tu n'avais pas de hauteur. Maintenant ça s'est calmé. C'est quand même beaucoup plus agréable et beaucoup plus correcte, tu commences à percevoir de la hauteur, du centre. Donc voilà après indépendamment de dire que c'est bien ou ce n'est pas bien tu n'as pas la main dessus. Tu peux le monitorer à travers Logic mais tu ne vas pas orienter ton mix à travers le rendu tel qu'il est aujourd'hui sachant que demain ça ne sera pas le même. A priori le rendu binaural de Dolby, qu'il soit bien ou pas il ne va pas trop bouger normalement.

A quel moment et comment décides-tu d'affecter les paramètres de rendu binauraux ?

Globalement j'ai une vague idée de l'image que je veux rendre donc j'ai un peu une vague idée de comment je vais faire mais globalement je le fais à la fin du mix parce que pour moi le mix binaural est un plus, mais mon objectif reste le mix multicanal. A terme, j'aime à croire que c'est quand même lui qu'on va essayer d'écouter même si ça ne sera pas forcément le cas en fonction des coûts. Mais c'est quand même mon objectif principal.

Est-ce que tu penses qu'à l'avenir, les améliorations d'algorithmes de décodage affecteront les choix de rendu binaural que tu fais aujourd'hui ?

La volonté de Dolby sur l'évolution des codecs c'est de faire des codecs de moins en moins compressés. Parce que mine de rien aujourd'hui on avait fait le test entre une écoute Apple Music multicanal et l'ADM, sur les aigus sens quand même la compression du codec. Dolby va quand même dans cette mouvance essayer de moins compresser le plus possible, le codec, le fichier. Je ne sais pas si tu as fait gaffe mais aujourd'hui on commence même à avoir des Atmos sur des Blu-ray, donc en non compressé. Mais je ne pense pas qu'on va sentir une certaine couleur dans l'évolution du binaural parce qu'aujourd'hui on utilise certains paramètres. Je pense, à moins que les paramètres binauraux changent complètement, que dans 40 ans il y aura encore des mecs qui utiliseront du near, mid, far.

Est-ce que tu as rencontré des problèmes dans le passage de ton mix entre écoute sur enceintes et écoute au casque ?

En fait tu en as pas mal. Déjà tu as le problème du sub, c'est une vraie problématique et tu n'as pas vraiment de solution. Le canal centre reste aussi un vrai problème, on n'arrive pas à le binauraliser. A moins que les maisons de disque nous obligent à mettre un paramètre binaural, on préfère laisser le centre en off, parce que la virtualisation du centre en Binaural est ultra compliquée et ça te fout plus le bazar qu'autre chose. Cela dit, même sur enceintes multicanales, je mixe très très peu dans le centre physique, car c'est hyper dépendant des installations, de comment il est placé, s'il y a un écran ou autre. En général je préfère utiliser le centre fantôme. Après dans les grosses différences que tu as quand même moins de sensibilité au niveau de la perception arrière même si tu sens qu'il y a des choses, ça reste quand même beaucoup plus flou et d'une manière générale la hauteur. Je trouve qu'Apple fait plutôt du bon travail là-dessus. Ce n'est pas forcément ce que toi tu voudrais, ce n'est pas forcément ce que toi tu fais, mais tu as quand même une perception qui existe et en ça j'en suis reconnaissant. Aujourd'hui je pense que c'est ça qui est important : que ça fonctionne, plutôt qu'on soit parfaitement raccord avec le mix qu'on voudrait faire.

Est-ce que tu as noté une modification de ton utilisation du casque dans ta méthode de mixage par rapport à ce que tu faisais au stéréo ?

Dans le studio où j'étais je ne l'utilisais presque jamais car l'acoustique était quasi parfaite. Maintenant que je me suis mis à mon compte, je bosse maintenant en stéréo au casque, ne serait-ce que pour entendre le grave. Là je vais être amené à bosser à distance donc avec un ordi portable j'aurai des retouches de mix à faire et la grande question que j'avais c'était que je ne vais pas avoir mes enceintes, j'aimerais bien emmener mon Audix mais dans un train, c'est un peu compliqué, donc je vais me retrouver à devoir bosser avec des AirPods pro. En fait la seule solution que j'ai trouvée c'est Virtuoso, un soft de binauralisation globalement la plus probante que j'ai trouvé pour travailler. Ça m'aide parce que moi les courbes de base de Dolby je n'entends rien c'est mon crâne doit être à l'opposé de la courbe par défaut donc c'est vrai que pour moi, ça a été le jour et la nuit.

Est-ce que tu as remarqué des évolutions dans tes pratiques de mixage avec l'arrivée du mix objet ?

Ouais ça a changé, pas mal de trucs. Après je t'avoue globalement je m'éclate bien plus en faisant de l'électro en Atmos que sur d'autres genres parce que tu peux faire d'autres trucs. Pour le reste c'est quand même limité par cette vision très traditionnelle d'avoir le groupe devant toi ou autre, ce qui n'est pas forcément plus mal. Parce que pour avoir entendu des essais. J'ai pris des réflexes un peu, mes overhead de batterie je peux me permettre de les mettre en hauteur ça amène une dynamique qui est intéressante. Pour le rock, je me sers un peu du plafond et des arrières, mais je travaille surtout sur une stéréo étendue, en profitant du point 9 avec les wide qui font vraiment la liaison entre tes sides. Ça crée vraiment un globe qui fonctionne vraiment bien pour le rock, pour créer un peu ce mur de guitare.

Qu'est-ce que tu penses des possibilités esthétiques que peut apporter l'Atmos, s'il y en a ?

Dans les faits, je pense qu'on est quand même globalement très au début de ce qui est possible de faire esthétiquement parce que c'est une nouvelle technologie, qu'il ne faut pas choquer, il y a toujours plein de visions de mix, chaque mixeur bosse différemment. Voilà mais je pense qu'il y a beaucoup de choses à faire techniquement parlant et je pense qu'on va vraiment décoller quand les compositeurs vont commencer à s'y intéresser. On commence de plus en plus à avoir des solutions pour intégrer de l'Atmos dans des logiciels de pré-prod et de composition.

Ça serait quoi pour toi un projet écrit pour l'Atmos ?

Je ne suis pas sûr qu'on écrive différemment, mais si les musiciens, compositeurs, arrivent déjà avec une idée de comment ils veulent spatialiser et de ce qu'ils veulent rendre comme effet spécifique ou quoi que ce soit, ça peut être intéressant. Un exemple tout con, quand on avait des groupes qui venaient enregistrer pendant trois quatre mois au studio, ils répétaient en cercle et ils avaient l'habitude d'entendre leur instrument de cette manière-là. Alors, est-ce que ça ne pourrait pas être intéressant de reproduire cette image du cercle tel qu'eux, ils l'entendent pour le spectateur ? Il y a des idées à aller chercher. Après c'est au risque et aux dépens de beaucoup de problèmes, on ne va pas se mentir. Mais d'un autre côté on fait de la musique, on ne fait pas de la technique. La musique ça a toujours évolué parce que des gens ont tenté des trucs que tout le monde pensait complètement fou, incroyable ou stupide, et ça l'était jusqu'au jour où ça marche. Pourquoi ça changerait avec cette technologie-là ? Quand on est arrivé du mono à la stéréo, on a fait plein de daube en stéréo, jusqu'au moment où on est arrivé à trouver un truc, qui cette est une stéréo très monoisée quand même aujourd'hui, mais qui reste quand même une stéréo où t'as le meilleur des deux mondes. Je pense qu'avec l'Atmos on peut trouver un truc qui est le meilleur des deux mondes.

L'ingé avec qui je bossais, Denis Caribaux, avait donné une interview où il avait eu un argument qui avait été très mal compris qui disait que l'intérêt de l'Atmos à travers le binaural, ou même à travers des enceintes pas chères, c'est qu'on peut avoir une émotion musicale qui est de bien meilleure qualité qu'avec des enceintes stéréo bas de gamme. Et en fait, il disait : là où tu auras besoin pour avoir une émotion musicale d'avoir des enceintes stéréo très haut de gamme, tu peux avoir la même chose avec des enceintes moyen de gamme en Atmos. Vu que tu as moins d'éléments pour chaque enceinte, tu n'as pas besoin d'avoir la même qualité d'enceinte, d'être capable de parfaitement définir

chaque élément. Évidemment, si tu as de l'argent pour avoir 14 enceintes à 10000 balles l'unité tant mieux ! Mais même si tu n'as pas cet argent-là tu peux retrouver une émotion qui est peut-être potentiellement plus prépondérante plus stéréo ? Et pour moi l'Atmos et le Binaural c'est ça que c'est censé apporter. Je ne suis pas dans un film à être synchro à une image et à devoir donner la sensation que l'image est cohérente par rapport à ce qui se passe dans la scène. Moi je fais de la musique j'ai envie juste que les gens ferment leurs yeux et entendent ce que je fais. Si j'arrive à faire un mix Atmos où ça oblige les gens à fermer leurs yeux pour être capable de pleinement se concentrer sur ce qui se passe autour d'eux et créer une expérience musicale, je considère que mon pari est réussi.

Est-ce que pour toi le mixage orienté peut offrir soit maintenant ou soit au futur une nouvelle expérience d'écoute au casque ? Et si oui, de quelle manière ?

C'est des discussions qu'on a depuis quasiment le début de l'Atmos de se poser la question de si ça a de l'intérêt ou pas. Ce dont je te parle c'est d'après mes discussions et mon entourage, mais on a un certain nombre de gens qui ne connaissent rien à la musique, qui écoutent de l'Atmos et qui s'en rendent compte pour certains, pas tous, qu'il y a un truc qui est différent, positif ou négatif mais au moins ils s'en rendent compte. Apple disait en juillet dernier que l'Atmos représentait 60% des écoutes, pour 80% d'écoutes au casque environ je crois. Donc même si c'est mis par défaut, les gens peuvent le changer. S'ils le laissent activé, soit c'est qu'ils ne pigent vraiment rien à la techno, mais sinon c'est que ça plaît je pense. Après apporter quelque chose au casque, à terme pourquoi pas ? Aujourd'hui...non, parce que la concurrence avec la stéréo est trop rude et la technologie n'est pas encore au point sur ce concept-là dans le sens où c'est compliqué de limiter le binaural sans limiter ton master indépendamment pour qu'il puisse concurrencer la stéréo. Et pour l'instant dans la mesure où ce n'est pas ultra performant sur la sensation d'immersion, tu as juste une stéréo de moins bonne qualité, un peu moins forte, qui paraît moins dynamique parce que c'est moins compressé donc voilà aujourd'hui ça me paraît un peu compliqué. A terme j'aime à croire que ça sera bien, surtout si après tu as des technologies, comme le head-tracking qui fonctionnent vraiment bien. Sur Apple, outre les problèmes de désynchronisation que tu as avec le tracking et qui fait que ton mix se retrouve coincé sur un côté, globalement tu as vraiment cette sensation d'immersion qui existe et donc c'est intéressant. Quand tu es dans un train ou autre, que tu ne bouges pas trop et que ça suit légèrement tes mouvements de tête quand tu regardes à droite à gauche, ça fonctionne bien.

Donc si je résume un peu c'est potentiellement une meilleure expérience d'écoute au casque si les algorithmes de décodage tendent à s'améliorer ?

Oui, il faut que les algorithmes s'améliorent. Il faut qu'en même temps que les outils des professionnels fonctionnent réellement pour le binaural. Et tout ça doit se passer de manière assez rapide car si les premières expériences binaurales les gens écoutent sont mauvaises, ils n'y retourneront pas. J'aimerais y croire mais il y a tellement de trucs qui peuvent se passer et qui sont interdépendants...

Et du coup tu dois comment la pérennité de cette techno dans les années à venir ?

Ça me paraît difficile qu'elle disparaisse. Ce qu'a fait Dolby avec l'Atmos est vachement intelligent et tombe à un moment particulièrement opportun parce qu'on sort du Covid, les gens ont fait beaucoup

d'investissement musical. Proposer une expérience sans surcoût derrière ce n'est pas forcément stupide. Bon après sur le plan professionnel ça crée des tensions au niveau de la chaîne de production parce que voilà pour le coup c'est un coup en plus mais bon il y a réellement quelque chose. On arrive à l'époque où on est capable de faire des enceintes en multicanal sans tirer des câbles partout donc c'est plus facile à faire accepter dans les foyers. On est capable de faire de la virtualisation d'enceinte à travers les barres de son même si on en est encore qu'aux balbutiements. Un Amazon Echo Studio, je ne vais pas dire que tu as de l'arrière, mais tu as cette sensation d'ampleur qui se crée donc bon pourquoi pas, même si le sweet spot est minuscule. Ce qui m'étonne et ce que beaucoup de professionnels ne prennent pas en compte c'est qu'à travers un mix objet tu as la possibilité d'un down et up mix cohérent, quel que soit le format. Les trois quarts des gens quand ils pensent à l'Atmos, ils pensent que c'est seulement bouger des instruments alors qu'on peut fournir tout le marché avec un seul mix.

Tu penses qu'à terme le fichier stéréo pourrait découler du fichier ADM ?

À terme, dans 20 ans peut-être je n'en sais rien. Après je ne suis pas sûr, dans la mesure où il y a un truc en musique qu'il ne faut pas sous-estimer, c'est l'utilisation des machines analogiques. Alors j'ai bossé sur des machines analogiques, c'est très cool, c'est très agréable, mais le mixeur qui dit qu'avoir une machine analogique ça sauve son son moi ça me pose problème et en fait ça coûte trop cher de faire de l'analogique multicanale. Donc tant que tu auras des gens qui réfléchissent comme ça le mix stéréo existera toujours. Il ne faut pas oublier un truc, c'est qu'aujourd'hui se passer du stéréo ça veut dire tuer le mastering. C'est aussi pour ça que beaucoup d'ingénieurs sont complètement contre.

Est-ce qu'il y a une étape de mastering en Atmos ?

Le gros problème c'est qu'en vrai on a tous la même norme de diffusion à -18 LUFS. Donc le mastering pour gagner du volume ça n'existe plus. Il te reste le mastering pour unifier les couleurs des morceaux ? Mais aujourd'hui vu que l'Atmos, c'est vu comme un truc secondaire, c'est le même mixeur qui fait généralement tous les morceaux Atmos. Puis aujourd'hui, le seul outil que tu as en mastering de réellement sérieux, c'est le Dolby Assembler, qui se résume à un EQ et un limiteur, donc tu ne vas pas faire grand-chose.

Quelles évolutions techniques pourraient te permettre d'améliorer ta méthode de travail dans le futur ?

Ce que j'aimerais bien c'est qu'ils écoutent un peu plus les ingénieurs. Le renderer, ça fait un an et demi qu'on râle pour avoir le contrôle du sub en binaural. C'est plein de petits détails pratiques, du type expliquer concrètement ce qui se passe au moment de l'encodage.

Annexe 4 : Entretien avec Oscar Ferran

Est-ce que tu pourrais te présenter ?

Alors j'ai fait l'ISB à Brest et après j'ai été un an assistant aux Studios Saint-Germain puis un an assistant à l'IRCAM. Je me suis mis en freelance et maintenant je bosse 60 à 70 % du temps en studio, le reste à l'IRCAM où je fais des mix ambisoniques, et aussi un peu avec La Blogothèque pour de la prise de son et du mix à l'image.

Comment toi tu définirais ton métier à quelqu'un qui ne connaîtrait pas le rôle du mixeur immersif ?

Moi le truc principal quand même que je voulais bien dire avant ça c'est que je ne suis pas complètement persuadé que cette technologie soit très intéressante. En studio je trouve que c'est une expérience assez folle et c'est pour ça que j'aime bien en faire. Au casque je ne suis pas encore complètement persuadé donc je n'arrive pas vraiment à la surprendre à mes potes ou à des gens à qui j'ai besoin de l'expliquer. Mais pour quelqu'un qui ne connaîtrait pas, ce que j'essaierai de lui expliquer c'est l'idée d'avoir une image sonore moins frontale, plus englobante et donc plus immersive. Ça correspond à la sensation que la scène sonore est tout autour et pas juste devant.

Quelles sont tes expériences associées au mixage objet ?

Je m'y suis mis il y a un peu plus d'un an et demi en septembre 2021. Au départ c'était principalement pour travailler avec un mixeur pour qui j'étais longtemps assistant qui s'appelle Stan Neff qui n'était pas du tout intéressé par le mix multicanal et qui du coup m'a proposé de faire tous ses disques en Atmos. J'ai un peu appris tout seul au début parce que c'était un peu le début de la grosse vague où tout le monde voulait s'équiper et tout le monde commençait à en faire donc j'ai fait 4-5 albums très vite là j'en suis à une quinzaine de disques en Atmos.

Après ce qui m'intéresse et qui me plaît le plus c'est de travailler en hybride avec de l'analogique, avec des sommateurs et une chaîne master. Et du coup pour moi le gros problème et qui fait que la différence de façon de travailler est énorme entre le mix stéréo et Atmos c'est qu'on ne va pas commencer à avoir 150 machines pour faire une chaîne master et que ce sera trop compliqué de commencer à intégrer de l'analogique dedans. Et donc bah pour moi la façon actuelle de faire de l'Atmos, que je trouve artistiquement complètement stupide, c'est de faire un mix stéréo puis de faire un mix Atmos. Artistiquement, je la trouve complètement stupide parce que ça veut dire qu'on réadapte un mix qui a été pensé en stéréo en immersif et en général sans artiste qui vient donner son avis et en général sans personne, ni producteur, ni artiste, pour te valider quoi que ce soit. Sur les 15 disques que j'ai faits, j'en ai peut-être deux ou trois que les artistes ont écouté et encore. Si tu veux mettre un tout petit peu de grains et un tout petit peu de son analogique quelque part, c'est bien de repartir du stéréo, tu vois ce que je veux dire ? Donc je trouve que l'approche est très différente, rien que par le fait que tu mixes dans la boîte ou pas dans la boîte.

Est-ce que les artistes sont venus écouter avec toi ? Est-ce que tu as eu un dialogue avec le mixeur stéréo sur les projets que tu as fait ?

C'est là que j'ai pas mal de chance, en travaillant avec Stan, même si je me contredis un peu avec ce que j'ai dit avant. Quand c'est lui qui a fait les enregistrements, tout est nickel, tout est passé dans pas mal de machines. Il mixe beaucoup sur sa SSL avec pas mal d'inserts analogues. Quand il sait qu'il va y avoir de l'Atmos, il fait un peu attention. En général je ne repars pas de stems de sa part, je repars des sessions freezeées, il faut que j'aie tous ses plugins de sa chaîne. Ça me permet de remonter un maximum dans les pistes séparées pour ne pas avoir les mains liées. C'est beaucoup plus de temps après pour le mix que si tu as des stems mais c'est là que tu commences à faire des mix Atmos intéressants. Là où je me contredis c'est que du coup, c'est que quand il freeze, il ne prend pas forcément le temps de passer par toutes ses machines analogues. Il met pleins de plugins et plein de trucs qui font que ça reste quand même une session qui a été mixée en stéréo et donc qui permet de travailler sur le grain en stéréo. Et bon c'est quand même un peu différent de partir d'une session de zéro et de l'intégrer en Atmos. Tout ça pour dire que c'est la méthode idéale. Après, Artistic Palace, qui est un autre studio où je travaille, pour le coup c'est plus une prestation où je reçois des stems. L'avantage des stems là pour le coup c'est que tout normalement est passé par leur chaîne master donc vraiment c'est le son du truc et le grain du truc ce qui est quand même cool mais des fois avec 8 stems tu ne peux rien faire.

C'est quelle temporalité, tu passes combien de temps ?

J'essaie de faire un album de 12 titres en trois jours. J'essaie de faire quatre titres par jour en sachant qu'en général j'en fais 6 le premier jour, 6 le deuxième, et le troisième j'exporte les masters. Jusqu'ici normalement si j'ai tout bien reçu, je bosse sur les mix non limités ce qui est plus simple. Quand j'intègre les masters, j'ai une demi-journée où je réécoute tout avec les masters stéréos et j'ajoute ma chaîne master pour coller un peu plus au master stéréo. En général on travaille pas mal pour Universal et Sony qui sont très pointilleux sur la manière de délivrer les mix Atmos en particulier. Ils ont un cahier des charges hyper précis et si tu t'oublies le moindre truc tout ton album est refusé. Ils ne demandent aucun objet vide et en même temps idéalement ils demandent que tous les objets se suivent et qu'il n'y ait pas de trou.

Est-ce que tu as perçu une évolution dans ta manière de mixer ?

Oui et non. Avant ça je veux bien faire une petite parenthèse sur ma manière de mixer de l'Atmos parce que pour moi il y a un peu deux écoles pour en faire. Il y a ceux qui essaient d'en faire trop et ceux qui essaient d'en faire le moins possible. Moi depuis le début que je suis avec Stan, on a réfléchi à notre manière de voir les mix. C'est ses projets, on était d'accord sur le fait que si on faisait des mix Atmos, il fallait qu'il y ait un respect absolu de son travail de mixeur, du travail de l'artiste et de la pensée de l'artiste. L'idée n'était pas de faire des trucs de fou, de pas commencer à mettre des trucs derrière des trucs sur les côtés complètement what the fuck. Moi j'ai tendance à faire quand même des mix Atmos assez frontaux dans le sens où évidemment la snare, la basse, la voix restent plutôt sur le LR et un peu dans le centre. Les pianos, les guitares, les trucs comme ça restent quand même assez devant, un peu plus large et avec un peu de reverb à l'arrière. C'est sur les synthés, sur les effets sur les backing vocals que là tu peux vraiment mettre des effets derrière, un peu des mouvements. Puis, il y a des mix Atmos que j'ai écouté de gens qui mettent une guitare acoustique en haut à gauche, les

cordes juste derrière et des trucs où on reconnaît à peine le mix stéréo quand tu passes de l'un à l'autre. En général, les maisons de disque aiment bien parce qu'elles ont l'impression qu'elles ont donné de l'argent pour quelque chose qui valait le coup.

Sur les évolutions des pratiques de mixage, sur la dynamique, moi je trouve que globalement c'est très très agréable de moins avoir à compresser, à limiter même si je le fais. Ça compresse un peu moins parce que c'est un peu mieux réparti. Je trouve ça vraiment très agréable, en particulier au casque je trouve que ça fait du bien d'avoir plus de dynamique. Sur le démasquage, je le ressens, même si comme je te dis vu que je mix très frontal, il y a encore quand même pas mal de trucs qui sont au même endroit. Le gros problème entre la restitution binaurale et le son sur enceintes, quand tu commences à mettre des instruments un peu larges, tu ne sens plus le côté glue sur enceintes, il y a des trucs partout. Quand tu écoutes au casque, tu n'as pas du tout cette sensation-là, parce que tu n'entends plus du tout les trucs derrière et sur les côtés tu entends juste un truc effectivement qui se masque moins. Donc au casque, je trouve ça assez réussi. Sur enceinte vu que mon but est d'essayer de sonner très proche du mix stéréo mais plus, je dirais qu'évidemment ça se masque moins mais ça ne me fait pas moins travailler.

Au-delà des possibilités techniques, est-ce que toi tu considères qu'il y a de nouvelles possibilités esthétiques ?

L'aberration actuelle du format c'est qu'on s'est déjà habitué à un truc complètement débile qui est qu'on fait un mix stéréo et après si on a l'argent et que on veut être bien vu par Apple, qu'on doit être dans les clous, on fait une version Atmos qui n'intéresse plus personne parce que l'artiste il a déjà assez donné, il n'a pas du tout pensé le projet comme ça. Donc voilà, mon but est d'essayer de faire en sorte que ce soit cohérent. Mais évidemment que si un artiste demain me disait "j'ai envie de faire un disque en priorité Atmos, produit, écrit et enregistré Atmos" ça serait super, mais ça ne m'est pas encore arrivé.

C'est quoi pour toi un projet écrit pour l'Atmos ?

Ça veut dire que lors de la phase d'enregistrement et pré-prod, tu ajoutes le format Atmos dans la pensée du projet et forcément là pour le coup tu légitimes la possibilité de mettre des sources en haut, derrière, sur les côtés, où tu veux car c'est pensé pour. Un bon producteur, il pense en permanence quand il ajoute un élément à qu'est-ce que ça va rajouter fréquemment et dans la dynamique. Là si tout d'un coup tu lui dis "tu n'as pas juste un stéréo mais tu peux jouer sur l'ensemble du spectre et partout" et bien évidemment qu'un prod, ça lui donne plein d'autres idées. Et là tu as matière à mixer.

Concernant le monitoring, il y a à la fois cette question de la possibilité de rendu qui est propre au format objet qui fait qu'à partir d'un seul fichier tu peux t'adapter aux multiples formats de restitution et il y a aussi le fait qu'aujourd'hui, selon que ton fichier soit uploadé sur Apple Music ou sur Tidal, le décodage ne va pas être identique. Donc je me demandais : comment tu fais justement pour t'adapter à cette multiplicité de possibilités de restitution du mix ?

Pour faire un petit historique, j'ai commencé à mixer de l'Atmos quasi exclusivement sur enceinte et je checkais au casque. Mon tout premier mix, tous les réglages binauraux étaient sur mid, trucs de base, et c'était n'importe quoi : beaucoup trop d'aigus et plus de bas. Et là tu flippes, donc là tu essayes, tu passes tes trucs en near et ce n'est toujours pas fou. Donc je prends la décision très vite, j'appelle Dolby et je lui dis bon, qu'est-ce qui se passe ? On ne peut pas faire de binaural en fait ? Et ils me répondent que beaucoup de mixeur mettent le LCR en off effectivement sinon ils ne retrouvent pas leur mix au casque. Donc là se pose très vite une question qui est : est-ce que tu préfères mettre tout en mid, avoir un truc très englobant et compenser (enlever des aigus remettre du grave) et potentiellement dans 5 ans ou même dans trois ans l'algorithme change et d'un coup tout ce que tu as fait de compensation est éclaté. Ou est-ce que tu te dis : bon bah sur mes enceintes c'est bien, au casque en off c'est à peu près bien mais si les algorithmes changent et s'améliorent bah mon mix Atmos sonne beaucoup moins immersif que ceux qui ont tout mis en mid. Très vite, j'ai tranché et j'ai mis le LCR en off parce que je ne pouvais pas me résigner à écouter des trucs qui ne vont pas du tout quoi. A cette époque je n'avais pas d'airpods, donc Apple je parlais du principe que ça allait bien se passer. Le gros souci était que la différence entre mes mix Atmos et le mix stéréo n'était pas énorme au casque. Du coup j'ai commencé à créer un template beaucoup plus fourni en disant ok bon alors il y a des trucs qu'il faut que je laisse dans le bed ou le timbre est hyper important. Par contre les synthés, si ça détimbre ce n'est pas très grave donc eux je les sors du bed. Maintenant je continue à penser que ma manière numéro une de mixer doit être soit enceinte. De manière très personnelle, mixer au casque toute la journée, je trouve ça horrible donc je préfère mixer sur enceinte. Et vu que les algorithmes sont pourris, il y a un moment où pour moi l'enceinte fait foi. Mais j'ai adapté de plus en plus mon template pour qu'au casque il y ait une différence qui commence à être notable.

Comment se passe le quality control, tu as un retour des artistes, des labels ?

J'essaye d'envoyer à chaque fois aux labels les mix ADM et les réductions binaurales pour qu'ils puissent. Je ne sais pas s'ils les écoutent mais en tout cas en tout cas ils retrouvent à peu près leur petit. J'aimerais bien en discuter avec eux, évidemment c'est des questions que tu ne leur poses pas forcément, mais essayer de savoir s'ils sont contents, s'ils entendent la différence parce que quand même ils mettent plusieurs milliers d'euros de plus pour ça tu vois. Est-ce que c'est que du marketing ? Je pense. Ou est-ce qu'ils y trouvent un intérêt et quand ils écoutent les mix, ils se rendent compte que ça sonne mieux ?

C'est quoi les principaux problèmes que tu as remarqué entre le passage sur enceinte et le passage en casque ?

Trop d'aigus, en binaural. Évidemment quand tu es en far, ça rajoute une réverbération horrible donc je ne mets quasiment rien en far. Il y a aussi la balance tonale et le déphasage typiquement sur les cordes. Au casque ce que j'ai besoin de savoir c'est est-ce que ça sonne mieux que le stéréo et est-ce que ça sonne suffisamment différent.

Est-ce que pour toi le mix objet peut apporter une nouvelle expérience d'écoute au casque ?

Du fait que j'essaye de me rapprocher du mix stéréo le plus possible, pour un mixeur, pour un artiste qui s'y connaît en son, s'il fait le AB, il va préférer mes mix Atmos. Mais c'est que technique si tu veux,

c'est un peu plus profond, qui se masque moins ça c'est sûr. Est-ce que c'est une nouvelle expérience de spatialisation sur enceinte ? Oui ! Le casque pas vraiment. La seule fois où j'ai constaté une nouvelle expérience sonore c'est la fois où ça a été pensé dès le départ et enregistré avec un micro ambisonique d'ordre 1. Tu ressens les sources d'une manière beaucoup plus floue que si c'était juste un micro stéréo ou mono. Moi je me suis dit la première fois ça va faire un truc super bizarre, mais vu que c'est le même micro et que c'est dans une même pièce, en binaural ça marche super bien. Avec Apple Music ça marche limite encore mieux parce que tout est en mid. Dans ce cas, c'est une nouvelle expérience de son spatialisé, où n'importe qui qui écoute ça avec des AirPods Max il se dit « ah bah ça je n'avais jamais entendu, j'ai l'impression que le son est partout ». Avec ma manière de faire de l'Atmos, je ne peux pas vraiment revendiquer le fait que les gens ont l'impression que les sons sont partout en binaural.

Est-ce que tu penses que c'est de l'ordre du perceptible au casque la différence stéréo - Atmos pour la majorité des gens ?

Je ne trouve pas. En faisant le AB, les gens commencent à avoir l'impression que c'est un peu plus englobant et encore parce qu'il y a une telle présence à l'avant. Ça pose la question de qu'est-ce que tu privilégies ? Parce qu'évidemment si je ne faisais pas du tout ma méthode, en envoyant rien dans le bed en off, et que je faisais du pan d'objets mono sans stems un peu partout, si l'algorithme était bien fait bah là évidemment oui les gens auraient la sensation que le son est partout. Mais tu perds tellement d'autres choses et sur enceintes, c'est tellement n'importe quoi que je ne puisse pas me résoudre à ça tu vois. Du coup il y a un peu les deux écoles, ceux qui mixent sur enceintes et ceux qui mixent au casque et qui s'en moquent que sur enceintes ce soit n'importe quoi. Mais mon mix n'est plus du tout le même et je suis obligé de compenser des trucs parce que tout est un objet, toute la balance fréquentielle n'est plus bonne, et là si tu fais venir ton ingé son stéréo ou ton artiste en studio, ce qui peut arriver, et que tu lui fais le AB, il fait : « qu'est-ce que je suis en train d'écouter tu es malade ou quoi ? ».

Ça n'arrive jamais le cas où tu fais d'abord écouter au casque la version à l'artiste ?

En fait, c'est là que le truc devient absurde, en fonction de comment la personne va pouvoir écouter : si tu vas pouvoir faire venir l'artiste, s'il va checker au casque, si l'ingé son stéréo vient écouter et va vouloir reconnaître son mix... Tous ces trucs font que finalement, tu peux te retrouver à faire une méthodologie différente à chaque mix, ce qui est possible, mais du coup tu ne t'améliores pas vraiment parce que tu changes complètement à chaque fois de manière de faire. Ma méthodologie c'est que les enceintes font foi parce que les algorithmes binauraux sont trop bizarres. Les algorithmes binauraux ne peuvent que s'améliorer et du coup mes mix sonneront de mieux en mieux. Le jour où les algorithmes binauraux seront vraiment bien, peut-être que je pourrais passer des trucs en near ou en mid et que ça sera bien. Pour l'instant je ne veux pas le faire et peut-être que du coup dans 10 ans mes mix Atmos d'aujourd'hui ne sonneront pas assez "Atmos" par rapport aux nouveaux, mais au moins pendant ces dix années-là ce ne sera pas n'importe quoi tu vois. En fait on est en train d'habituer des gens à écouter de l'Atmos n'importe comment, avec une technologie nouvelle et mal faite. Nous, en tant que mixeurs, notre travail c'est d'essayer de faire en sorte que ça reste cohérent et que ça soit bien fait. Le jour où la technologie sera mieux, on s'adaptera. Tous mes collègues qui font de l'Atmos, le premier truc que je leur demande, c'est s'ils mettent leur bed en off, et tout le monde le fait.

Quelles évolutions techniques pourraient t'aider dans ta méthodologie de travail ?

Franchement moi personnellement je ne leur demande qu'une chose, c'est d'améliorer leur algorithme binaural. Il faut faire en sorte que quand tu écoutes en binaural, si tu mets tout en mid, ça sonne bien, mais je pense que ça ce n'est pas près d'arriver. Après Apple est quand même pour l'instant la plateforme la plus utilisée en Atmos et je suis agréablement surpris. Je préfère mes mix sur Tidal parce qu'évidemment ils sonnent comme le renderer et c'est ça que j'ai pu checker le plus.

Tu perçois quoi comme différence justement entre Apple Music et Tidal / Amazon?

La différence est énorme par le fait que je mets mon bed en off sur l'un (*Tidal*) et sur l'autre il est mis en mid (*Apple*). Donc tout est plus flou sur Apple Music en binaural je trouve. Sur des trucs un peu percussifs, t'entends une reverb un peu bizarre.

Comment vois-tu la pérennité de cette technologie ?

J'aimerais que ça reste un truc un peu alternatif pour les artistes et les labels qui ont l'argent et l'envie d'innover et de faire des trucs fait pour ça et intéressant dans ce format. C'est très naïf et très utopique. Ce que je n'aimerais surtout et ce dont j'ai peur qui arrive, c'est qu'à un moment les labels se disent "on ne peut plus se permettre de payer les deux mix, mais on a besoin de faire de l'Atmos parce que tout le monde s'y est mis, donc faites un mix Atmos et après faites-nous des réductions stéréo, et ce sera ça le mix stéréo". Ça c'est ma plus grande crainte et c'est la plus grande crainte de tous les ingés qui font du mix stéréo, parce que de un ça va obliger tout le monde à s'équiper, et de deux c'est la mort du mix stéréo telle qu'on le connaît. Si tu commences par faire du mix Atmos pour ensuite sortir un mix stéréo ça change complètement la manière de travailler. Encore une fois aujourd'hui le fait de faire le mix stéréo d'une certaine manière et d'adapter le mix Atmos de l'autre je trouve ça artistiquement stupide, mais je trouve que ça marche bien en termes de son. Si en tant que mixeur tu dois faire un mix stéréo et après l'adapter en Atmos tu perds trop de temps. Ce n'est plus rentable si on te paye juste pour un mix et que tu fais les deux. Mais je sais que c'est ça que les labels veulent, Universal ils nous le disent clairement, leur objectif c'est qu'on vende des mix Atmos et qu'ensuite il y ait des réductions stéréo. Ça veut dire qu'ils n'ont aucune conscience de ce que c'est qu'un mix et d'une manière de bosser en stéréo et c'est complètement absurde.

Tu trouves que le downmix stéréo du renderer n'est pas convaincant ?

Je le trouve pas mal mais ça ne fait pas faire les mêmes choix. Ce serait intéressant de faire un test, de faire un mix stéréo d'une manière et de faire un mix Atmos et de le réduire en stéréo. C'est forcément biaisé par lequel tu as commencé à faire mais ce serait intéressant de voir les choix que tu as fait en Atmos et est-ce que du coup en stéréo ça devient les bons choix ou des mauvais choix ? Mais en tout cas, la manière de travailler un mix stéréo : le masquage, faire une glue, organiser ta session parce que tu n'as que deux enceintes et que du coup c'est le bordel, si tu bosses en Atmos ben j'ai peur que tu n'arrives pas à faire une réduction stéréo si bonne que si tu avais dès le départ pensé tout ton mix en stéréo. Typiquement moi ma seule expérience de ça, c'est le mix 5.1, quand tu fais une réduction stéréo tu reprends un peu tout de zéro. Alors après, est-ce que finalement ce temps-là perdu à

réadapter ton mix 5.1 en stéréo n'est pas le même que commencer par stéréo et le faire en Atmos ? En vrai, peut-être. Moi en tout cas si un jour on me demande de commencer par faire un mix Atmos sur un truc où l'artiste n'a pas du tout pensé le titre en Atmos mais en stéréo j'aurais tendance à quand même faire un mix stéréo et après faire un mix Atmos mais. Si tu es payé le prix d'un mix pour faire les deux dans tous les cas c'est chiant parce que quand tu fais un mix 5.1 et que tu fais une réduction stéréo, tu es payé pour faire tes réductions. Après ça va vraiment dépendre d'une chose c'est est-ce que Spotify s'y met ou pas ? Si c'est le cas ça devient, ça deviendra la norme.

Annexe 5 : Entretien avec Martin Antiphon

Est-ce que tu pourrais te présenter ?

Martin Antiphon, ingénieur du son, spécialisé dans l'audio 3D et dans la spatialisation du son depuis de nombreuses années malgré le fait que je continue à faire de la bonne stéréo traditionnelle en analogique. Je suis aussi musicien. Je suis un peu des créations, beaucoup en audio spatial évidemment.

Comment tu définirais ton métier de mixeur immersif à quelqu'un qui ne le connaîtrait pas ?

Je ne sais pas si je lui expliquerai très différemment de ce que c'est que de mixer en stéréo en fait. J'expliquerai ce que c'est le mixage tu vois ?

Et à quelqu'un qui s'y connaît en son ?

C'est comme en stéréo, c'est-à-dire que finalement un ingénieur du son, il peut être amené à intervenir à plein de degrés différents puisque c'est lui qui est en charge du reste de la restitution du rendu final et que finalement il peut être amené à faire juste une bonne balance ou à revoir complètement la copie de l'artiste ou du réalisateur en disant « écoute, artistiquement, ton truc que ça groove pas un caramel » donc en fait on va changer les plans, on va changer les perspectives, on va rajouter des effets, on va rajouter des reverb, on va créer des plans. Si on fait de l'audio immersif, au lieu de faire tout ça sur de la stéréo on applique ça à une full sphère ce qui implique plein d'autres contraintes et plein d'autres limitations artistiques et des limitations technologiques. D'où l'intérêt d'avoir un workflow propre et qu'on continue tous de challenger de faire évoluer.

Est-ce que du coup tu pourrais me parler un peu plus de tes expériences associées du coup au mixage objet ?

J'ai fait pas mal de musique classique avec Alexandre Parrot, Gautier Capuçon et Vanessa Wagner notamment. En musiques actuelles j'ai fait le live de Souchon et après j'ai fait quelques projets électro notamment avec le label 3D Sound Records et mon projet Ninja Cyborg. Et puis là j'ai plein d'autres trucs en cours.

Est-ce que sur les projets Atmos où tu n'avais pas fait le mixage stéréo, tu as eu un dialogue avec l'ingé qui l'avait fait ?

Dans mon cas, systématiquement. Je n'ai pas voulu travailler autrement et je n'ai pas eu l'opportunité non plus de travailler autrement ou alors quand j'aurais pu avoir l'opportunité de le faire j'ai refusé parce que c'était pour faire du bas de catalogue.

Et dans ce cas-là est-ce que toi tu prenais la session de mixage stéréo avec des plugins déjà freezés ?

C'était une continuité finalement ouais, à chaque fois c'était des sessions dans lesquelles c'était voilà c'était négocié avec le réal ou le mixeur de récupérer un certain nombre de choses.

Quelles conséquences pour toi les évolutions apportées par le mix objet ont eu sur tes pratiques de mixage ?

Oui ! Là où je dis qu'il y en a eu, c'est parce que j'ai eu l'opportunité de faire des produits finis. Le problème c'est que tant que tu n'as pas d'obligation de livraison et que tu n'as pas d'obligation de résultats, tu ne te mets pas la pression et donc du coup tu ne vas pas au bout d'une décision finale. Là ça m'a amené à devoir faire des releases avec des contraintes de budget, de temps, d'esthétique, de culture parce qu'on pouvait se permettre des trucs mais pas trop etc. Donc oui parce que ça m'a donné l'opportunité d'en faire beaucoup alors qu'avant j'avais fait des mix spatiaux au studio dans Ableton Live avec le Spat révolution et donc pas la même ergonomie de travail.

Et comment qualifierais-tu les évolutions d'un point de vue plus concret ?

Alors je dirais que pour l'outil c'est régressif par rapport à par rapport à ce que tu peux faire avec le Spat en ambisonique clairement. Par contre en termes de workflow avec Pro Tools c'est quand même génial, tu peux faire pleins de choses pratiques comme du bounce offline etc.

Quel temps passes-tu sur un mix Atmos en comparaison d'un mix stéréo ?

C'est pareil, c'est à dire qu'en fait, ça dépend de ce que tu fais. Quand je fais un disque de musique classique je le fais dans la journée : conformation des tracks, export en temps réel en 96kHz et en 48kHz. Je livre aussi systématiquement les mesures de loudness les rendus binauraux. Donc voilà pour un album, avec une scène fixe c'est une journée. Le live de Souchon qui durait quand même deux heures sur lequel je voulais qu'on se prenne la tête on l'a fait en trois jours, j'aurais pu le faire en deux ou en un. J'aurais pu mais j'aurais pris beaucoup moins de décisions. Par contre sur certains mix de variétés où là on voulait vraiment faire de la spatialisation et chercher des trucs, c'est une journée par titre. Après moi quand je mix en stéréo, je mix un à deux titres par jour. Donc en réalité si je voulais avoir le même degré de perfection dans un mix spatial, il faudrait pareil un à deux par jour.

Est-ce que tu dirais qu'il y a des nouvelles possibilités esthétiques qui ont émergé suite à l'arrivée de cette technologie ?

Pour moi c'est encore plus important que le passage du mono à la stéréo. Et encore, pour moi, le passage du mono à la stéréo il a été noyé dans l'évolution des techniques de production c'est-à-dire qu'on a mis tout le même panier. Il y a les Beatles qui sont arrivés en même temps que la stéréo, donc certes ils se sont éclatés sur la stéréo mais ils ont aussi inventé la prod. Donc, est-ce qu'on aurait eu le même effet « waou » si on avait eu cette révolution dans la prod sans la stéréo, et que la stéréo était arrivée après ? Pas certain. Par contre là ce qu'on est en train de vivre comme évolution, c'est carrément exponentiel en termes de créativité et de possibilités. Pour moi c'est presque aussi important que l'invention de l'enregistrement. Bon, c'est un peu pompeux, mais ce que je veux dire, c'est que quand on a inventé l'enregistrement on a inventé un nouveau rapport à la musique. Je ne décrirais pas lequel mais le rapport à la musique a changé : le format c'est devenu 3 minutes parce que c'était les durées des galettes, etc.

L'audio spatial, en termes de créativité de ce que ça disrute de nos pratiques musicales, est presque aussi important que ce qui s'est passé pour moi à ce moment-là dans les enjeux et dans la façon dont il faut repenser nos façons de faire.

Est-ce que tu pourrais préciser sur quels aspects ?

Sur tous les aspects de composition dans l'espace, c'est-à-dire qu'à la fois on a la possibilité en musique acoustique de créer des prises de son beaucoup plus immersives et réalistes, même si on restera toujours dans quelque chose d'artificiel parce qu'on ne peut pas reproduire le niveau. Pour la musique production on va dire, donc tout ce qui est électro, variété, on a un champ des possibles qui pour moi est à inventer tu vois, il y a énormément de choses à inventer.

Concernant le monitoring, il y a à la fois cette question de la possibilité de rendu qui est propre au format objet qui fait qu'à partir d'un seul fichier tu peux t'adapter aux multiples formats de restitution et il y a aussi le fait qu'aujourd'hui, selon que ton fichier soit uploadé sur Apple Music ou sur Tidal, le décodage ne va pas être identique. Donc je me demandais : comment tu fais justement pour t'adapter à cette multiplicité de possibilités de restitution du mix ?

Et ben je vais te dire c'est très simple ce n'est pas mon problème. Je vais m'expliquer. La seule chose que je vérifie quand je mixe en stéréo c'est que je n'ai pas de problème de phase quand je suis en mono et que quand on va le graver sur un vinyle ça ne va pas être trop large et qu'il ne va pas y avoir un problème de faire sauter le diamant. Mais à part ça, que les gens écoutent sur une petite paire d'enceinte ou sur une sono démentielle, dans une voiture ou autre, moi mon mix il sonne partout, et d'ailleurs ce n'est pas au mastering que ça se fait c'est au mixage. En audio spatial c'est pareil, c'est-à-dire que moi je vais faire un mix ADM qui est bien et après charge à celui qui l'écoute de l'écouter dans des conditions qui correspondent à ses contraintes. Alors je vais quand même pondérer ce que je dis. Si jamais je fais un mix spatial dans lequel je vais décider de mettre la grosse caisse devant et la caisse claire derrière, si tu diffuses ça dans un dans un club qui a une jauge de plus de 200 personnes ou tu vas te retrouver avec 20 ou 30 mètres entre les enceintes, tu vas perdre tout ton groove. Donc je dirais que la pratique du mixage peut évoluer en fonction de ses contraintes. Mais ça c'est une vraie question et c'est vrai que moi j'y fais attention quand je mix. Quand je dis que je n'en ai rien à faire, j'y fais quand même attention, comme quand je vérifie la compatibilité mono. Je vérifie que ça se passe bien en binaural mais ce que je dis tout de suite à mes clients c'est : "attention par contre le timbre, faut que tu écoutes celui des enceintes car c'est lui qui a raison. Le casque c'est un rendu binaural lambda qui va changer". Après mes enceintes elles peuvent changer mais il n'y a pas de filtre, alors que le binaural il y a un filtre.

Est-ce que les paramètres de rendu binaural tu les anticipes dès le début de ton mix ?

Alors, il y a des habitudes que j'ai. Je passe beaucoup de temps à mixer en binaural. Je vais quand même faire attention à plein de choses dans ma vie binauralisation qui sont liées au fait qu'on binauralise soit des objets, soit des speakers, donc en fonction de ce que je vais décider de mixer, de comment je vais décider de mixer : objet ou chanel based, je ne vais pas chercher à contrôler la même chose quand je vais écouter en binaural. Mais je vais mettre de côté l'effet de filtrage des HRTF parce

que ça je sais que ça peut changer. Par contre, le fait de binauraliser une écoute, que ce soit des haut-parleurs ou des objets, ça ce n'est pas la même chose, je n'écoute pas la même chose.

Qu'est-ce que tu observes comme difficultés dans le passage entre ce que tu écoutes sur enceinte et ce que tu écoutes en réduction binaurale ?

Il y a des défauts qui sont incontournables donc soit on les apprivoise, il y en a qu'on peut essayer de corriger donc on les corrige, il y a tous les problèmes de filtrage en peigne qui sont liés par les ITD qui sont très courts dans les mesures binaurales. Il y a toute la calotte frontale qui est floue. Il y a énormément de problématiques mais malgré tout, c'est quand même assez exceptionnel ce qu'on arrive à faire aujourd'hui. Je dis qu'on n'y est pas, mais je préfère ça à rien. L'autre paramètre qui change tout c'est le tracking. Ça change absolument tout, même sans faire de mouvements spectaculaires.

Pour toi le mixage orienté objet tel qu'il existe aujourd'hui en musique actuelle, donc globalement avec Dolby Atmos, est-ce que tu penses qu'il peut offrir une nouvelle expérience d'écoute au casque ou est-ce que tu penses que ce n'est pas encore suffisamment satisfaisant ?

Ce n'est pas encore satisfaisant mais ça permet clairement d'offrir une nouvelle expérience d'écoute, surtout avec le tracking. Après c'est un débat que j'avais encore en interne aujourd'hui avec mes associés et que j'avais eu avec des labels quand j'ai fait des Atmos en leur disant que le moteur de Dolby c'est chiant parce qu'il est fait automatiquement. Et on avait fait beaucoup de demandes pour pouvoir nous-mêmes livrer le binaural. Je pense qu'un binaural bien mixé et bien maîtrisé, pour l'instant il sera quand même mieux qu'un mauvais downmix avec des mauvaises personnalisations de HRTF. Mais ça va évoluer. Mais je me méfie beaucoup de ça, de la personnalisation et tout ça. C'est-à-dire que nos métiers de mixeur on a appris à faire des trucs qui sonnent pour tout le monde donc il n'y a pas de raison qu'on ne soit pas capable de le faire encore aujourd'hui. La personnalisation de la HRTF à outrance fait que, si on ne compense pas avec la pratique de mixage ou la pratique d'encodage, on risque de se retrouver avec des problèmes. C'est à dire que des gens qui vont avoir certaines fréquences prédominantes bah en fait c'est leur filtre, ils ont l'habitude, donc ils ont l'habitude d'écouter un mix avec la voix toujours un peu plus forte. Maintenant si on leur met des HRTF qui correspondent à leur écoute et qu'on n'a pas bien maîtrisé, je suis dubitatif là-dessus. Je pense que des fois un bon binaural natif c'est quand même plus convaincant. Mais le problème du binaural natif c'est que du coup il ne permet pas d'avoir du tracking et que quand même le tracking c'est vachement bien.

Est-ce que tu penses que pour toi c'est de l'ordre de perceptible ?

C'est une bonne question. Pour moi, une bonne stéréo peut complètement enterrer un mauvais mix binaural, surtout qu'un mix stéréo sera nécessairement plus large puisqu'il a la capacité de décorrélérer absolument les signaux des 2 oreilles, ce qu'on ne peut pas faire en binaural. Maintenant, ce que je trouve d'intéressant, moi, c'est la transportabilité des formats. C'est à dire qu'aujourd'hui tu fais un bon ADM, il ne sera peut-être pas bien décodé, il n'est peut-être pas bien mixé donc les gens vont préférer le stéréo, mais en fait ça dépend avec quel moteur et ça dépend comment c'est fait, tu vois

? Dans les écoutes comparatives c'est très compliqué de ne pas se faire avoir par le niveau. Nous, on a fait un benchmark si tu veux de notre solution My Bee Knows, on a fait passer quand même une plus d'une centaine de sujets et sur lequel on a volontairement fait des écoutes, où les niveaux de notre solution, celle d'Apple ou Dolby était randomisé pour justement shuffler la problématique du niveau. Entre la perception de la dynamique, le RMS, les transitoires, le contenu fréquentiel, c'est très difficile de dire lequel est le plus fort. Par contre, ce qui est sûr, c'est que c'est le plus fort qui sonne le mieux à la fin. Des bons mixes stéréo tout le monde sait en faire alors que des bons mix spatiaux, sachant qu'on a aucune maîtrise du rendu ADM et c'est un problème. Là avec Music Unit on est en train de lancer un label d'audio 3D et en fait donc tout le débat, c'est est-ce qu'on fait des ADM quand même ou pas ? Mais ce qui est sûr, c'est qu'on va distribuer des stéréos binauraux mixés avec notre solution. En fait, le gros problème, c'est que le filtre entre le mixage et le rendu chez le spectateur est tellement différent et tellement loin de la proposition du mixeur que nécessairement, ça ne peut être que moins bien. Tu vois un mixeur qui a peaufiné son mix à 0,2dB près, forcément si tu changes des trucs, c'est moins bien. Et le mixage ADM, avec une écoute binaurale ça fait ça.

Tu la vois comment la pérennité de cette techno dans les années à venir ? Quelles évolutions éventuellement permettre d'améliorer ta méthode de travail ou les moyens de consommation pour le public ?

En fait, il y a plusieurs leviers. Si tu veux, il y a d'un côté des outils de création pour éviter aux artistes de devoir spatialiser des stems, car cela pose problème de conservation de timbre, de distribution d'énergie. Donc, il y a des outils de création pour les artistes, des outils de mesure et des outils pour mieux maîtriser le mixage. Lorsque je parle d'outils, je parle aussi de culture, de compréhension. Et d'un autre côté, pour les spectateurs si l'on parle de ton sujet, c'est-à-dire l'écoute du mixage avec un casque, je dirais deux choses : clairement une amélioration des moteurs binauraux, que ce soit les bases de données HRTF ou les moteurs de convolution, ensuite le tracking évidemment, car le suivi d'Apple est vraiment très réussi, mais il a une latence et si tu veux que ton cerveau fonctionne, il ne doit pas y avoir de latence.

Annexe 6 : Entretien avec Bruce Keen

Quelles conséquences les évolutions apportées par l'arrivée du mixage objet ont-elles eu sur tes pratiques de mixage ?

Quand on reçoit les éléments, pour faire un mix Atmos, la première chose dont je m'occupe, c'est la balance entre les éléments du morceau pour que ça matche avec le mix stéréo. Je passe beaucoup de temps dessus, je passe peut-être deux heures, deux heures et demi à matcher précisément non seulement la balance entre le morceau mais aussi le mastering parce que quand tu fais un mix Atmos, le mastering fait partie de ton mix. En fait il n'y a pas de mastering de l'Atmos, souvent quand on mixe un morceau on va le donner au mec qui masterise. Nous, en Atmos, une fois que tu as fini, c'est le produit que tu livres. Il va directement sur le support final.

Une fois que j'ai fait ce boulot-là, c'est-à-dire une fois que j'ai préparé mon mix Atmos, là je me concentre sur la spatialisation de chaque élément. En général j'essaie de voir ce que j'ai comme élément sur la batterie, sur les percussions, de voir si je peux envisager de panner ce qui est devant derrière, par exemple des toms ou des percussions. En général j'essaie de mettre ce qui est aigu, notamment des tambourins ou des flûtes, des strings, solina ou des tenues, j'essaie de les mettre plutôt en haut. Chaque élément j'essaie de trouver sa position dans un mix Atmos. En général, j'essaie de laisser le beat devant. Surtout sur IAM, ce n'était pas vraiment le propos de mettre une casse claire derrière parce qu'on veut quand même garder cet impact du mix qui vient de devant. Quand je bossais sur AIR, ce qui était intéressant c'était justement d'avoir des éléments du beat que je pouvais faire soit glisser devant derrière ou sur le côté par le biais d'objets. En fait tu ne peux pas panner un instrument qui est en 7.1.2 au-dessus. C'est à dire que tu es obligé d'utiliser un objet pour faire tout ce qui est pané. Tu peux le paner dans un 7.1.2 mais c'est beaucoup moins fluide qu'avec un objet.

Donc après ce qu'il faut aussi prendre en considération c'est que les éléments que tu vas mettre en objet ne passent pas par ton bus principal. Quand tu as un bus de mixage, la compression est dépendante des éléments que tu envoies dans ton bus principal. C'est un changement qui est assez fort, comme ta source ne va plus être dans ton bus principal, il ne va pas compresser de la même manière. C'est important quand tu joues avec la compression d'avoir un bus commun à tout le monde. Quand tu sors quelque chose en objet, il faut aussi que tu le recompresse parce qu'il est plus compressé. En général j'ai un EQ et un compresseur et chaque objet je le recopie dessus.

Après toute la problématique c'est le panning en fait, normalement tu peux imaginer que tu es en train de regarder un groupe et de ressentir un peu les éléments qui arrivent sur le côté, donc en général j'utilise des délais notamment le slapper, le spanner qui sont mes éléments principaux quand je mixe en Atmos. En général quand je fais un mix Atmos j'essaie de noter en début de sessions dans la page de commentaire où je dois panner chaque élément.

Est-ce que les artistes étaient avec toi quand tu mixais ?

Les artistes ne sont pas là, donc je ne veux pas qu'ils disent que le morceau est différent. Ils ont passé du temps à mixer leur version stéréo. Ce que je veux qu'ils retrouvent, c'est la même chose en Atmos. C'est important de respecter les choix qu'ils ont fait à l'époque. Après, c'est tout un apprentissage

pour faire tout ce qui a été fait en analogique, en digital. Comme je t'expliquais, j'ai deux autres studios à Telegraph qui sont équipés avec beaucoup de hardware en analogique et moi je suis plutôt quelqu'un qui vient de l'analogique et ça m'a pris un certain temps pour passer au digital. Ça m'a pris à peu près un an pour trouver les outils nécessaires pour pouvoir retrouver ce que moi je ressens en analogique en digital. Ce n'est jamais exactement pareil en digital, c'est toujours un peu différent mais il faut utiliser aussi la force de chaque support. Le digital c'est très clean, c'est propre, ça crunch moins, il y a moins de saturation. Il vaut mieux en avoir un peu moins d'ailleurs, il ne faut pas essayer d'avoir de la compression comme en analogique.

Comment gères-tu la problématique du monitoring au casque ?

Je suis dépendant de la pièce. C'est un truc important pour moi. C'est l'endroit où tu bosses, c'est l'acoustique du lieu et je ne le ressens pas avec aucun casque. Je n'ai jamais retrouvé au casque l'effet que je peux avoir dans une pièce. Il y a des gens qui adorent écouter au casque d'ailleurs quand tu mixes en stéréo, tu dois forcément écouter au casque. Tu vas entendre des petits détails que tu n'avais pas forcément entendus sur tes enceintes. Il faut que l'écoute soit fluide pour que ce que tu rendes, ça marche ailleurs en fait.

Comment contrôle-tu les différents types de décodage et formats de restitution ?

Une fois que tu es relativement content avec ton mix, ici j'ai une fonction qui me permet d'aplatir mon mix Atmos en stéréo. Comme ça je peux comparer avec le mix original. Je peux vérifier que tous les effets que j'ai mis ça n'a pas augmenté trop un élément. Si j'ai mis trop de reverb je peux me rendre compte comment ça matche avec la version stéréo. Une fois que tu es relativement content avec ça, tu peux aussi te dire que ça ne rend pas pareil en Atmos. Souvent, ça m'arrive quand je réécoute un mix Atmos de changer un peu le mix parce qu'on t'entend plus de choses en Atmos. Des fois, un truc qui luttait un peu quand tout est devant, tu peux l'élargir un peu derrière. Souvent on a tendance à éclaircir les cordes pour les entendre, en Atmos je les mets derrière et tu n'as pas besoin de les éclaircir tu les entends très bien parce qu'ils ne sont pas devant.

Les artistes sont toujours venus à la fin écouter ?

Ouais, toujours. Sauf si c'était à l'étranger, auquel cas je leur envoie. En général l'artiste passe à la fin pour donner son avis.

Et toi tu as eu cette expérience de pouvoir à la fois mixer en stéréo et faire la phase de mix Atmos ou est-ce qu'à chaque fois tu travaillais dans format où tu récupérais des stems ?

Non, souvent je fais du bas de catalogue donc je suis obligé de tout remixer. C'est-à-dire que je reprends les bandes originales et je suis obligé de tout remixer. C'est tout un travail pour retrouver en digital un mix qui a été fait en analogique de s'approcher au plus près de ce qui existe en analogique en digital. Ce n'est jamais exactement pareil mais il y a suffisamment d'outils aujourd'hui pour faire un peu ce que font les consoles SSL, pour retrouver un peu les reverb 480 qui étaient beaucoup utilisées à l'époque, les H3000. C'est un assez long travail en fait de comparaison.

En général, quand je travaille, j'ai une première étape où je suis vraiment focalisé sur l'équipement, qui est le truc le plus important en fait, la balance spectrale. Une fois que je suis à peu près content de la balance spectrale, après je me concentre vraiment sur la compression. J'ai du mal à faire les deux simultanément en fait.

En général, les gens aiment bien écouter en Atmos parce que le niveau est fixé à -18 dBLUFS, donc que tu le limites ou pas, tu l'entendras au même volume et j'aime bien le son non limité en fait. Quand j'ai fait le premier album de AIR, *10 000 Hz*, c'était bien écrasé quand même tu vois et quand on l'a remixé en Atmos tout le monde m'a dit « c'est ouvert ton mix, il n'y a pas de limiting en fait ». C'était un peu l'époque dans les années 2000 où c'était la course à l'écrasement des masters. C'est vrai que c'est plaisant d'entendre avec un peu plus de dynamique, c'est un peu plus ouvert. C'est l'un des points forts de l'Atmos, c'est que tu as une plus large dynamique que le -14 dBLUFS qu'on t'impose sur Spotify ou Tidal.

Qu'est-ce que tu dirais quoi des possibilités esthétiques associées à l'Atmos ?

C'est très dépendant de la musique, en fait toute la musique ne s'y prête pas. J'ai l'impression qu'un groupe de rock, je préfère l'écouter en stéréo en fait. Quand j'ai mixé les albums de AIR, c'était ultra flagrant pour moi que c'était évident que cette musique-là elle devait être mixée en Atmos parce que beaucoup d'effets, beaucoup de délais, beaucoup d'effets de voix, d'orchestre qui s'y prêtaient vraiment bien. C'est vrai que quand j'ai fait Lomepal, c'était un peu différent. Quand j'ai fait *L'Ecole du micro d'argent*, j'étais un peu sceptique au début. Puis finalement avec tous les scratches, tous les samples, on a pu vraiment paner autour. Le groupe a apprécié. C'était plaisant. Après des trucs auxquels je ne pensais pas, ça se prête aussi bien. Quand on a fait l'album de Rone, je me suis surpris. Ça rend vraiment bien avec un truc synthétique, acoustique. C'est vrai que des fois tu as envie d'entendre le truc devant en fait, tu n'as pas envie d'entendre le truc derrière. La difficulté justement comme je te disais quand on est en groupe de rock, c'est que tu arrives à le rendre devant avec un petit peu d'arrière. T'es pas obligé que ce soit toujours 360 tu peux avoir 180 c'est très bien ça s'ajuste comme je te dis ça dépend des moments avec la musique. Des fois juste un élément qui pan ça suffit c'est ça peut être correct ce n'est pas la peine de faire de l'Atmos pour de l'Atmos en fait. Si c'est juste pour faire une démo ça ne sert à rien. J'ai assez vite compris que faire des pannes pour faire des pannes, ce n'était pas très intéressant.

Qu'est-ce que tu as remarqué comme principal défaut entre ton écoute sur enceinte et le passage en restitution binaurale ?

C'est un sujet délicat parce que je trouve que, malheureusement, on écoute avec des AirPods qui n'ont pas vraiment un bon son. Je trouve que la façon dont on utilise Apple, elle n'est pas très probante. En fait, l'audio spatial que propose Apple n'est pas la même chose que ce que tu livres. Tous les rendus binauraux que tu as pu faire en Dolby Atmos, ils ne sont pas pris en compte. Donc, ça c'est un truc que moi je ne maîtrise pas. Moi, ce que je maîtrise, c'est ce qui sort des enceintes. La façon dont ils ont décidé de le rendre sur leur plateforme, elle est indépendante de moi. Je ne peux pas changer le mix que moi j'ai fait. Comme je te disais tout à l'heure, ce qui est intéressant, en fait, c'est d'avoir un mix qui est plutôt devant et de spatialiser un petit peu derrière. Ça a tendance à rendre mieux au casque.

Au casque, tu as tendance à plus sentir quand tu as une reverb qui s'élargit de temps en temps, de façon continue. Je t'avoue que je devrais peut-être passer plus de temps au casque, parce qu'il y a des gens qui font que ça. Moi, ça ne m'intéresse pas trop. Ça passe par là, je le fais. Quand je finis un album, je passe toujours une journée à essayer de voir comment je peux l'améliorer un peu.

Les BRM, les paramètres de rendu binaural, tu les mets à la fin de ton mix, ou c'est dès le début ?

J'essaie de les mettre après. J'essaie de faire un mix cohérent avec des enceintes, et après de faire un mix binaural. Je ne fais pas l'inverse. Je ne commence pas par un mix binaural, et après sur des enceintes, ça ne marche pas pour moi. Je ne peux pas me dire, "Ah ouais, c'est super au casque," et puis après essayer de le mettre sur les enceintes. Ça ne fonctionne pas.

Est-ce que tu as perçu des différences entre Apple Music et les autres plateformes comme Tidal et Amazon Music ?

Je ne sais pas, je ne vais jamais écouter Tidal. Moi comment je procède en fait c'est qu'une fois que j'ai mon mix je l'exporte dans Logic. Dans Logic tu peux passer de Dolby Atmos à Apple. C'est ma seule façon pour moi de comparer, c'est d'écouter avec Logic, avec les supports qui sont proposés dans Logic. Après, moi je livre, après ce qu'ils en font, ce n'est pas moi qui maîtrise le bout.

Est-ce que tu penses que ça peut offrir une nouvelle expérience d'écoute au casque ?

Ça dépend des gens. Il y a des gens qui ne l'entendent pas du tout, il y a des gens qui l'entendent beaucoup. Comme je t'ai dit, ça dépend vraiment de comment c'est mixé, si t'as rien qui bouge, tu ne vas pas l'entendre, c'est pour ça que je te parlais beaucoup d'effets de panning, de reverb et de délai autour d'un élément. C'est quand même un truc que tu vas ressentir au casque. D'où l'utilité d'utiliser Spanner pour moi en tout cas. Spanner de toute façon c'est un logiciel qui est un peu incontournable quand tu viens du film et qui en l'occurrence pour l'Atmos c'est vraiment crucial pour nous. Le gros défaut de l'Atmos c'est que la restitution du neural est médiocre. Je pensais qu'ils pourraient faire mieux que ça. Des fois tu as l'impression que tu fais ça juste pour vendre des AirPods.

Et du coup, tu la vois comment la pérennité de la technologie dans les années à venir ?

Il y a tout un article dans Gearspace : 25% des gens pensent que ça va survivre, 75% disent que ça ne sert à rien. Moi, j'espère que ça va survivre, même si c'est qu'un marché de niche. L'autre jour, j'étais avec un ingénieur du son qui m'a dit : "C'est une belle recherche archéologique ce que tu fais, en fait, tu te redécouvres d'une autre manière le son." Et effectivement, quand tu vas à la géode, il y a un kiff, tu ne peux pas le voir ailleurs. C'est un peu ce truc-là que j'ai essayé de viser. Le travail que moi je fais, c'est aussi un travail de restitution numérique d'un truc analogique. Ça m'oblige de pousser un peu plus en profondeur tous les outils digitaux, d'essayer de comparer mon approche analogique de l'approche numérique. Je trouve qu'on arrive à faire des super trucs en digital aussi. Toujours une petite préférence pour l'analogique, mais ça, je pense que c'est inhérent à mon âge.

Est-ce que toi il y a des évolutions en technique qui n'existent pas encore qui pourraient te permettre d'améliorer ta méthode de travail ?

La vraie révolution aujourd'hui, ce serait plus dans le rendu binaural, en fait, c'est-à-dire d'avoir un système binaural qui fonctionne vraiment. Je trouve que là, on n'y est pas. Nous, on avait pas mal bossé sur la VR360 avec les casques Oculus et tout ça. À l'époque, j'avais bossé avec certains produits binauraux qui marchaient bien mieux que ceux qui sont proposés actuellement sur les téléphones. Alors, je trouve que s'il y a vraiment un progrès, il doit être là. Je trouve que les mix sont bien, je trouve qu'ils sont mal rendus et que la technologie oblige. Ça marchera seulement si on arrive à faire un meilleur rendu. Je trouve ça un peu dommage qu'il y ait une dissension entre Apple et Dolby. Je trouve que Dolby fait plutôt un bon job, le système est bien développé. Apple ne joue pas le jeu en fait, t'as plus l'impression qu'ils veulent vendre des AirPods en faisant des pubs Spatial Audio plutôt que d'essayer de vraiment développer l'Atmos

Puis le problème c'est qu'il y a un problème financier. Moi je crois que je n'ai jamais gagné d'argent avec l'Atmos. Le temps que je passe, je peux le faire parce que notre société tourne avec d'autres sujets, elle tourne principalement grâce à la pub et au cinéma. Mais l'Atmos c'était quelque chose en plus qui nous sert un peu de vitrine mais si je devais en vivre, je serais perdant. Ça prend énormément de temps, quand je repars de zéro sur un projet, c'est à dire que je dois remixer entièrement le projet, je dois faire en sorte qu'il soit identique à l'original, je dois le masteriser, je dois faire le mix Atmos, ça prend un temps de fou. Sur IAM j'ai dû passer deux mois et demi dessus. Ce qui me prend le plus de temps, c'est de faire en sorte que ça ressemble au mix.

Annexe 7 : Entretien avec Tristan Mazire

Est-ce que tu pourrais te présenter ?

Ça fait 18 ans que je suis ingénieur du son. J'ai fait l'ISTS en 2004 je crois. Et suite à ça, j'ai fait toujours plein de choses différentes : du studio du live...Je suis parti en tant qu'assistant dans un studio qui s'appelait le Studio de la Reine à l'époque, qui était un des précurseurs de cette mode qui est apparue par la suite du matos vintage extrêmement haut de gamme. J'ai eu accès à des machines incroyables, ça m'a formé très vite sur des machines hyper intéressantes. En parallèle à ça, j'ai fait beaucoup de live. Je trouve que c'est cool de faire de tout : le live ça forme énormément sur des choses que tu peux faire en studio et l'inverse aussi. De plus en plus maintenant, je crois qu'il faut faire du live, du studio, et de l'audiovisuel parce que tous les secteurs sont transverses et se recoupent. Enfin, il y a trop de choses à faire et tu vois avec le mix objet il y a vraiment des trucs à expérimenter là dedans.

Par rapport justement à ta pratique en tant que mixeur objet qui est une pratique quand même hyper récente, toi comment tu définirais ton métier de mixeur immersif à quelqu'un qui ne le connaîtrait pas ?

Je pense que bizarrement le mix objet c'est plus des gens qui ne connaissent pas le son qui peuvent le plus le comprendre. Tu leur dis que tu recrées un espace, tu places les objets, les sons dans l'espace, et ils comprennent. Pour la stéréo, l'idée de comment faire cohabiter toutes les sources, c'est plus compliqué à expliquer. Donc je pense que c'est plus facilement palpable et explicable au grand public pour le mix objet. Après pour les gens qui connaissent le son, il y a des gens qui sont complètement réfractaires à ça et qui disent que ça n'a aucun intérêt, puis les autres qui voient un vrai intérêt pour certaines applications, notamment pour les live. Franchement honnêtement j'ai mixé des concerts en Atmos c'est quand même juste extraordinaire. Le mec de Dolby m'avait passé un morceau de Norah Jones qui était vraiment très jazz et tu avais l'impression d'être avec le groupe c'était fou. Donc même sur un genre musical comme ça où ça ne doit pas se balader dans l'espace et où tu n'as pas vraiment d'effet à faire je trouve qu'il y a un super intérêt.

Donc pour toi ton métier de mixeur immersif c'est plus un travail de spatialisation ? Tu l'associe plus à ça ?

Ouais après de ce que tu as dit tout à l'heure sur la compatibilité binaurale quand je mixe, je fais toujours moitié du temps au casque, moitié du temps sur les enceintes et tu te rends compte que la spatialisation influe aussi ta manière de mixer en égalisation et en compression. Si tu veux que ce soit compatible, il faut faire pas mal de traitement.

Ok on va revenir juste après sur ça. Toi c'est quoi tes expériences concrètes associées au mixage en Dolby Atmos : quel type de projet ? Quel type de demande en général ? Tu as combien de temps ?

On a installé l'Atmos ici parce qu'on m'a demandé de mixer l'album de Grand Corps Malade *Mesdames* en Atmos. Ils s'en sont suivis plein de projets, de live et de mix pour des labels.

Grand Corps Malade était avec toi quand tu as mixé ?

Oui il est venu contrôler et il a été assez surpris. Même pour tous les artistes avec qui j'ai travaillé, ça pousse des perspectives, il y a vraiment des choses à faire. Là sur les types d'albums que j'ai faits il y avait beaucoup de choses électroniques, plein d'éléments plein d'effets donc il y avait un vrai intérêt à le faire en Atmos, c'était vraiment intéressant.

Et est-ce que toi tu as dialogué avec le mixeur stéréo ? Ou est-ce que c'est toi qui avais fait le mix stéréo aussi ?

Oui j'ai fait les deux du coup c'est moi qui ai sorti les stems (*pour le mix Atmos*). Je pense qu'il faut sortir des stems. Le truc c'est que j'ai une manière de bosser en stéréo qui fait que je peux difficilement prendre cette session là et la retranscrire en Atmos parce que je mets énormément de traitement sur mon bus master. J'ai un rearbus où en gros tout interagit, je ne peux pas éclater tout de suite les choses dans la pièce, donc il faut que je fasse des stems.

Si tu étais dans la situation où tu n'avais pas fait le mix stéréo, est-ce que tu penses que ça compliquerait les choses de recevoir uniquement les stems et que ça soit deux personnes différentes qui mix ?

Non j'ai fait des projets où on m'a envoyé les stems, ce n'était pas tellement un souci. Je demande les stems par contre avec et sans effet parce qu'effectivement quand il y a des reverb sur les éléments ça ne marche pas en Atmos. Il faut vraiment des choses dry à mon avis (*sans reverb*) et ensuite moi je rajoute des reverb multicanales. Ça ne marche pas des masses de spatialiser des choses déjà réverbérées ça fait des trucs très étranges.

Est-ce que de la même manière où entre l'étape de mix et mastering il y a un dialogue entre le mixeur et l'ingé mastering, tu as eu aussi un dialogue avec le mixeur stéréo ?

Non pas des masses sur les projets que j'ai faits. On me donnait vraiment les stems et c'était à moi de me débrouiller. Après je comparais toujours son mix évidemment pour ne pas trop m'éloigner de ce qui avait été fait.

Ok. Du coup pour rentrer un peu plus dans tes pratiques de mix, d'après toi, quelles conséquences les évolutions apportées par cette technologie ont eu sur tes pratiques de mixage par rapport à ce que tu faisais en stéréo ?

Du temps ? Je ne sais pas exactement. Je ne dirais pas que ça a fait évoluer mes pratiques de mix mais que ça ouvre des perspectives. Comme je t'ai dit tout à l'heure, moi c'est principalement de la spatialisation que je fais. Et du coup c'est un peu la même manière de mixer que je retranscris en Atmos et que je spatialise. Ça n'a pas fait changer radicalement mes processus de travail je pense.

Même en termes de workflow ?

Non pas tellement. En tout cas, sur les projets que j'ai faits, je travaille exactement pareil. Si j'avais demain des projets vraiment écrits pour de l'Atmos je changerais mes manières d'enregistrer et de mixer mais pour l'instant c'est exactement la même chose pour moi.

Et du coup ce serait quoi d'après toi un projet écrit pour l'Atmos ?

Déjà dès la prise : mettre de la multi microphonie, d'avoir cette vision dès la prise, je pense que c'est ça le plus important, encore plus que juste du mix. C'est vraiment à la prise que les choses se jouent. Il y a moyen de faire des choses intéressantes. Sur le mix Atmos de Grand Corps Malade que j'ai fait au Zénith de Lille, on avait un setup de micro d'ambiance énorme. Je ne l'ai pas encore mixé en Atmos et je vais le faire même juste pour moi. Mais là ça fait vraiment sens parce qu'il y a moyen de se placer vraiment au centre du public. Il y a des micros partout qui sont à peu près placés comme dans une configuration Atmos et donc il y a moyen de jouer avec ça. Je pense que de la prise c'est important.

Ça c'est dans le cadre plutôt d'un disque de concert mais tu vois aussi ça pour le studio ?

Oui complètement.

Qu'est-ce que tu penses des possibilités esthétiques apportées par le mixage objet ?

Il y a évidemment des possibilités artistiques. Là pour l'instant tout ce qu'on m'a demandé, c'était quand même de la retranscription de mix stéréo en Atmos. Donc j'ai pris certaines libertés sur les projets que j'ai mixé moi-même en stéréo et en Atmos, mais il a clairement des choses à faire. Même la spatialisation est ultra créative, les possibilités d'effets à mettre, le placement, l'intention... Franchement le mix Atmos est super créatif. Je trouve clairement encore plus que la stéréo parce que tu te confrontes moins aux problèmes techniques liés à la réduction stéréo de pleins de sources. Là, tu peux vraiment éclater les choses. Tu dois à mon avis moins égaliser fin, tu travailles différemment. C'est à mon avis beaucoup plus artistique, en tout cas oui clairement.

Tu parlais tout à l'heure que l'une des évolutions que tu aimerais avoir c'est du temps ?

Franchement mixer un titre en Atmos ça va vite. C'est juste que ça rajoute du travail quand on est déjà pas mal occupé. C'est des choses encore en plus à faire rentrer dans un planning mais c'est super intéressant. Et là je pense que ouais les prochains projets avec la boîte qu'on veut monter c'est réfléchir les projets dès le début en Atmos pour pouvoir mettre un processus en forme de la prise jusqu'au mix. Et quand je dis mix je parle de mix en Atmos en binaural et en stéréo directement.

Aujourd'hui les labels donnent le même temps pour un titre mixé en stéréo et pour un titre mixé en Atmos ?

Tous les labels réduisent tout. Je sais plus combien je faisais de titres Atmos par jour mais il est possible d'en faire deux, trois, voire quatre selon la complexité des titres. Quand les stems sont bien faits ça va beaucoup plus vite. Généralement moi pour un mix stéréo je mets une journée minimum par titre et je pense que si on partait d'une session d'enregistrement qu'on doit mixer en Atmos, ce serait à peu près la même chose au minimum.

J'aimerais aborder la thématique du monitoring, et ton rapport à l'écoute au casque. Si l'on prend l'exemple du Dolby Atmos, il y a déjà deux plateformes de streaming différentes qui décodent différemment ton mix donc tu as déjà deux versions de ton mix possible. Et ensuite la propriété du

mixage orienté objet c'est que d'un seul fichier tu t'adaptes aux différents systèmes de diffusion. Du coup ma question quand tu mixes, comment tu t'adaptes à cette multiplicité de format ? À la fois en termes de possibilités de moyens de restitution (casques, enceintes, bar atmos, voiture...) et aussi par rapport aux différentes plateformes (Apple Spatiale Audio ou par exemple le Dolby présent sur Amazon et Tidal). Comment toi tu gères tout ça ?

Ce qui est extrêmement positif avec ce format-là, c'est que tu vois, ça se décode à la lecture et du coup ce n'est pas comme du 5.1 ou c'est figé. Là au moins, et c'est pour ça que ça marche, c'est que n'importe qui, selon son setup peut écouter un mix immersif, d'une qualité peut être moindre mais en tout cas c'est beaucoup plus ouvert. Pour l'instant, en tout cas en studio, j'ai vraiment fait des contrôles entre le système d'écoute ici et le casque en binaural. Après je suis allé vérifier sur Apple évidemment. Mais je n'ai pas directement fait de contrôle en temps réel sur les plateformes. Je ne sais pas s'il y a un plugin qui fait ça comme en stéréo parce que ça c'est super intéressant.

Je sais que c'est possible maintenant sur Logic depuis tout récemment.

Ça c'est super intéressant. Je sais qu'on le fait beaucoup surtout quand on fait du mastering ici avec ADPTR : tu peux écouter directement ton mix avec les différents algorithmes et ça en Atmos il faut absolument le faire parce qu'il faut se rendre compte directement de ce que ça fait. Il y a tellement de choses de phases, d'EQ, surtout dans les aigus. Il y a beaucoup de choses qui se passent.

Et est-ce que toi par rapport à ton utilisation du casque en monitoring en session de mix tu sens que ça a évolué entre ce que tu fais en stéréo et ce que tu fais maintenant au casque ?

La réduction de Dolby je trouve ça phénoménal la qualité. J'utilise beaucoup le casque en Atmos, je l'utilise quasiment peu en stéréo, sauf au mastering pour vraiment aller chercher des points précis.

Pourquoi tu l'utilises peu en stéréo ?

Je ne sais pas c'est une habitude que j'ai tout simplement. En fait, j'ai vraiment l'habitude des enceintes, du recul. Je trouve que le casque me fait trop corriger de choses. Je vais trop loin dans les EQ.

Est-ce que tu as rencontré, si je reprends le contexte du mix Atmos, des difficultés sur le passage des enceintes au casque ?

Ce ne sont pas des difficultés mais il y a des choses qui se passent fréquemment. Des expériences que j'ai eu sur notre système d'écoute, ça sonnait très bien. Les réductions binaurales il y avait des artefacts, des choses qui se passent dans le haut médium et les aigus qui fallait corriger pour que ce soit cohérent en binaural.

Donc essentiellement un problème de timbre ?

Ouais un problème de timbre clairement. La spatialisation, les reverb je n'ai pas trop rencontré de problèmes. Sinon, ce qui se passe dans le grave, la réduction, je la trouve incroyable ! Il y a une

profondeur dans le grave qui est vraiment impressionnante, beaucoup plus qu'en stéréo, c'est super bien fait.

Une question un peu de perspective : est-ce que toi d'après toi cette technologie de mix orienté objet peut réinventer l'écoute au casque. Et si oui, de quelle manière d'après toi?

Je ne sais pas si ça va révolutionner. De toute façon tout le monde écoute avec des intra-auriculaires donc le binaural, il y a un vrai intérêt pour ça. Les gens sont plus ou moins sensibles à ça, mais il y a un vrai intérêt pour le volume, pour la spatialisation, pour la profondeur. Je pense que ça ne va pas révolutionner mais il y a une vraie valeur ajoutée.

Est-ce que tu penses que c'est de l'ordre du perceptible pour la majorité des gens ou est-ce que ça reste ?

Honnêtement, je pense que pour la majorité des gens, s'il n'y a pas de gros mouvements, des gros effets de gauche à droite, ils ne s'en rendent pas compte si ce n'est pas expliqué. Ça dépend de certaines personnes. Il y a des gens plus ou moins sensibles, mais qui globalement ne s'en rendront pas compte.

Et ce que toi tu as remarqué des différences entre ce que tu écoutes sur Apple music, et ce que tu peux écouter sur les autres plateformes comme Tidal et Amazon ?

Leur effet de reverb chelou là sur Apple Music ? C'est un peu étrange mais je pense que ça va changer. Ils vont faire évoluer l'algorithme, ça va être de mieux en mieux, mais c'est étrange d'ajouter quelque chose que les mixeurs ou les artistes n'ont pas choisi ça c'est quand même un truc un peu limite quoi tu vois ? C'est comme si sur une photo ou une peinture tu mettais des couleurs qui n'étaient pas là. Il y a une question quand même de propriété intellectuelle je trouve. C'est un peu loin ce que je dis mais tu vois en même temps c'est toucher au travail de quelqu'un.

Ok et est-ce que d'après toi tu as des genres musicaux qui s'y prêtent le plus ? Ou est-ce que tu penses que c'est valable pour tous en fait ?

Franchement, c'est vraiment valable pour tous. À mon avis tu vois, il y a plein de personnes qui disent non, c'est bien que pour l'électro quand il y a des choses synthétiques et tout ça... Oui, effectivement, tu peux faire voyager des choses. Mais il y a un vrai intérêt pour tous les styles. Rien que la spatialisation d'éléments, les reverb, mettre l'auditeur dans une pièce avec les musiciens, c'est super intéressant. On a écouté un truc assez fou dimanche avec Bob Clearmountain. C'était le groupe de Bob Dylan the band. Les musiciens étaient tous dans la même pièce en cercle et donc là l'Atmos on était au milieu de ce cercle. Et même pour ça c'était super intéressant tu vois, alors que c'est un truc qui a été fait dans les années 60-70. Tu penses que l'Atmos ça n'a aucun intérêt là et en fait c'était génial ! T'étais entouré de musique tu avais l'impression de faire partie de la performance, c'était extraordinaire. Je pense que tous les styles ont un intérêt !

En termes de mix, est ce qu'il y a des choses héritées des pratiques des mixeurs de 5.1 ?

Il y a quand même des choses, après en qualité sonore je trouve que c'est vraiment différent. La différence entre le bed et les objets de l'Atmos c'est le jour et la nuit. Les objets ça sonne vraiment bien. Il n'y a aucun problème de phase quand tu places les éléments, même juste sur des enceintes dans la pièce, c'est bluffant la qualité sonore. Donc oui, je pense qu'il y a un héritage du 5.1 pour la réflexion des projets mais la qualité sonore des mix objets c'est bluffant.

L'une des perspectives aussi c'est qu'à terme le format ADM ça soit le format référence comprenant les formats objets et le format stéréo puisque le renderer peut le faire. De ce que j'ai entendu, la réduction stéréo du Dolby Renderer ne serait pas encore convaincante. Quelle est ton expérience ? Quand toi tu monitor la référence stéréo tu le fais dans le Dolby Renderer ?

Je suis d'accord. La réduction stéréo c'est particulier, après il y a des décennies de mix stéréo donc c'est normal. On a tellement travaillé la stéréo qu'il va falloir encore un petit peu de temps je pense avant d'avoir aussi facilement une réduction qui stéréo intéressante. Et c'est ce que je te disais tout à l'heure en plus, les mixeurs qui mixent des albums en stéréo, ils ont une manière de travailler qui n'est pas encore celle de l'Atmos dès le début du projet donc c'est un peu lié à ça je pense.

Par rapport à tes pratiques, quelles évolutions techniques pourraient permettre d'améliorer ta méthode de travail dans le futur ?

De l'intégrer à tous les logiciels vraiment, dans le Protools maintenant c'est le cas. D'avoir effectivement la possibilité d'écouter directement ce qui se fait sur les plateformes, ça c'est quand même très bien. Peut-être, des plugins en Atmos de compression, d'égalisation, encore plus poussées et des réverbérations. Ici on utilise flux c'est bluffant, ça marche vraiment super bien. Je sais qu'il commence là y avoir des compresseurs des limiteurs là. Après dans les process et tout c'est quand même plutôt bien pensé je trouve.

Est-ce que tu as l'impression qu'il y a une perspective de sortir un peu de la loudness war ?

On a toujours fait attention depuis le début à ne pas mixer trop fort surtout depuis les mesures LUFS, les limitations de Spotify de YouTube et tout ça. Donc en fait je trouve que ça va dans la bonne direction de faire des mix à -18LUFS, c'est une très bonne chose. Généralement les recommandations de l'AES pour l'Internet c'était -16LUFS donc on était entre -14LUFS -16LUFS. -18LUFS ce n'est pas si éloigné donc c'est plutôt cohérent. C'est bien qu'ils aient pris ce choix-là. Je crois que c'est pour leur laisser de la marge pour les réductions binaurales et stéréo. C'est une bonne chose d'habituer les gens à ça.

Et toi tu l'as senti dans tes mix la possibilité de retrouver cette dynamique ? Car dans les faits, vu que c'est le mixeur stéréo qui délivre stems, il n'y a pas déjà toute une chaîne de traitement optimisée pour la stéréo et du coup potentiellement très compressée ?

Honnêtement ça fait quand même un certain temps que je n'ai pas reçu des stems ou j'ai écouté des mix vraiment super compressés, super limités. C'est plus du tout le cas maintenant, même les choses

qui sonnent très très fort. Ça dépend quoi, mais en tout cas sur les projets auxquels j'ai collaboré ces derniers temps, il n'y a pas eu ce problème-là. Même moi je commence à enlever mon compresseur sur mon master, c'est terrible mais... Donc ouais.

Est-ce qu'il y a vraiment une étape de mastering sur l'Atmos ?

Moi de tout ce que j'ai rencontré il n'y en avait pas. Il n'y avait pas d'étapes de mastering et je pense, je ne sais pas, les mecs de mastering te le diront mais je ne suis pas sûr que ce soit fondamental. Si on part des stems, enfin voilà, et en même temps je n'ai jamais fait. Donc si ça se trouve il y a une vraie valeur ajoutée c'est super important. C'est vraiment une question en suspens encore. Mais il y a des gens qui font déjà du mastering en Atmos.

Bah justement, je sais qu'il y a des gens qui sont crédités pour ça mais dans les faits...

Il y a clairement un vrai intérêt surtout sur un album, la même chose que le mastering en stéréo, de lisser, de donner un projet global surtout pour des titres qui n'ont pas été enregistrés dans les mêmes endroits, de donner une couleur et tout ça. Ouais, si on le fait pour la stéréo il faut le faire pour l'Atmos aussi clairement. Il y a quelque chose de l'ordre aussi même des normes. En fait les normes en Atmos elles sont quand même super importantes : le volume des subs, le volume des sides tout ça. Qu'il y ait une personne qui contrôle ça à la fin d'un projet, je pense que c'est une bonne chose, donc si c'est l'ingé mastering qui le fait c'est une bonne chose.

Et qu'est-ce que tu penses du fait justement que ça soit Dolby qui est le monopole sur tout ça ?

Ils ont toujours fait ça, quand ils ont commencé, ils ont breveté la réduction de bruit. Tu vois ? Donc là ils brevettent le Dolby Atmos. Après il y a un truc que je trouve super intéressant c'est que le grand public connaît Dolby grâce aux salles de cinéma, donc ils identifient ça tout de suite à une qualité sonore, une spatialisation et à du son cinéma. Donc il y a un truc où c'est pas mal que ce soit eux parce que c'est réputé comme structure. Chez les gens du son comme dans le grand public je trouve que c'est plutôt une bonne chose. Après, ils sont venus ici évidemment certifier la cabine. On a eu pas mal de discussions avec Dolby France et c'est des gens ultra intéressants et passionnés. Donc c'est bien que ce soit centralisé ouais avec une structure qui gère. Après peut-être sur le côté monopole ça va peut-être changer. Il y a peut-être d'autres choses qui vont se passer dans l'avenir. Je sais que j'avais mixé des titres au format Sony et c'est quand même vraiment moins bien quoi. Enfin en termes de qualité sonore, d'utilisation du programme, de sérieux même professionnel du process, Dolby va quand même beaucoup plus loin.

Tu avais pu intégrer la suite de Sony dans cette configuration du studio ?

Ouais, après c'est principalement du casque. En tout cas moi je n'ai pas mixé les formations sur des enceintes je l'ai fait qu'au casque.

Du coup, question un peu d'ordre général, comment est-ce que tu vois la pérennité de cette technologie dans les années à venir ?

Il y a eu un gros rush après Apple (*l'arrivée du catalogue Atmos sur Apple Music*). Là ça s'est calmé, il y en a moins par contre ça va prendre du temps. C'est ce que disait Clearmountain, va y avoir ça dans les voitures, va y avoir ça dans les barres de son. Petit à petit en fait, tous les constructeurs vont commencer à vraiment s'y mettre, ça va se développer. Mais je pense qu'il y a un vrai intérêt et je pense que ça va marcher.

Et ce que tu penses qu'on se dirige vers un système où les petites cabines et studio seraient un peu écrasés ? Genre le monopole serait consacré aux plus gros studios ?

C'est exactement comme quand il y a eu la crise de l'industrie du disque où plus personne n'allait en studio. Là ça revient, il y a du travail pour tout le monde et ça va être le cas avec Dolby enfin avec ses formats là aussi. Après il faut qu'il y ait un côté sérieux des mixeurs. Ici quand on a fait certifier la cabine on a compris pourquoi. C'est super important d'avoir une cabine aux normes entre guillemets. Tu ne peux pas faire n'importe quoi et juste faire un mix Atmos en réduction binaural au casque ce n'est vraiment pas assez. J'espère que tous les labels et les maisons de disques ne vont pas à un moment se dire "merde, on n'a pas assez contrôlé. Il y a des gens qui nous ont envoyé des choses qui n'étaient finalement pas assez qualitatives". Donc il faut que chacun prenne ce format au sérieux à mon avis et qu'effectivement même si tu as une petite cabine qu'elle soit quand même assez bien équipée pour te prétendre mixeur Atmos. Je ne suis pas sûr qu'il y ait quelque chose de l'ordre du monopole. Après c'est aussi aux jeunes mixeurs de se rendre dans des cabines qui sont bien faites quoi. Enfin il y a un truc, tout est ouvert, je pense que ça s'est quand même vraiment démocratisé. Il ne faut pas hésiter à aller voir les gens des studios comme ce que tu fais toi tu vois, d'aller à la rencontre des grands ou petits studios et de discuter, de s'intéresser, de peut-être même ramener des projets à travailler. Il y a de plus en plus de studios qui se montent un peu partout. Il y a aussi un truc franchement purement commercial : ça fait vendre plein d'enceintes et plein de choses donc ouais c'est juste sur un côté financier. Ici quand on s'est équipé ça nous a coûté 15000 euros je crois. Donc ce n'est pas rien quand même. Ce n'est pas n'importe quel studio qui peut du jour au lendemain faire une dépense comme ça.

On parlait justement du fait que dans les perspectives d'avenir il faudrait que les artistes s'emparent de la technologie dans leur processus de création. Il y a beaucoup moins la présence de l'artiste au mix Atmos qu'au mix stéréo, par exemple. Qu'est-ce que tu penses de cette perspective ?

Ouais ça c'est exactement une des perspectives, ça va se développer clairement. Il y a plein de producteurs de musique qui sont intéressés par ces formats et je pense qu'il y a tellement de trucs à faire. Ça peut être tellement intéressant de penser le projet dès le début en Atmos et ensuite de le décliner. Je serai ravi de faire des projets comme ça et on va en développer je pense ici.

Par rapport aux perspectives par rapport au casque, les moteurs de rendu binauraux vont probablement évoluer et s'améliorer au fur et à mesure du temps. Du coup, la problématique c'est : les paramètres de rendu binaural qu'on affecte aujourd'hui, au vu des algorithmes de décodage, seront-ils toujours pertinents dans 10 ans ? Où est-ce qu'on aura un effet daté 2020 ?

C'est comme les mix dans les années 80 où il y a des reverb numériques terribles. Ça fait partie du truc. En fait c'est l'évolution du processus clairement. Ça ne me fait pas peur quoi franchement. Il faudra se reposer la question à ce moment.

Toi tu les gères comment du coup les paramètres de rendu binaural ?

A la spatialisation, je réfléchis ça dès le début du mix, et je sais que tel objet va être proche ou loin enfin je sais plus exactement les termes là. Et ça c'est assez cool, ça marche bien. Dans mes sessions de mix, je sais que de tel objet à tel objet ce sera toujours tel type de distance.

Est-ce que tu souhaites aborder des choses dont on n'a pas parlé ?

Il y a quand même un truc là, vu que je suis vraiment dans les deux mondes, il y a toujours cette problématique de quand tu fais un album après faut qu'il soit transcrit dans un concert. Avec L-ISA notamment de L-Acoustics et tout ça, ça va être super intéressant de sortir de la musique pour les événements ou pour les concerts déjà en Atmos. Il y a un truc où vraiment il faut commencer à réfléchir comme ça, avoir une réflexion. Il y a l'album, le processus de la sortie, mais il y a aussi le live et si dès le début on pense à ça, ça peut faire des concerts ultra intéressants. L'avenir est vraiment là. Comme l'industrie musicale a explosé en live, pour moi une des évolutions majeures c'est vraiment ça clairement. Voilà, tout est interconnecté là il faut développer ça dès le début, dès la prise clairement pour aller ensuite le diffuser de l'album jusqu'au live.

Annexe 8 : Mesures des niveaux des extraits utilisés dans le test perceptif

Item	Integrated	Range	True peak	Maximum short-term	Maximum momentary
AloiseSauvageAppleSpatialAudio.wav	-18.0 LUFS	10.7 LU	-inf	-15.6 LUFS	-14.4 LUFS
AloiseSauvageAtmosAmazon.wav	-18.0 LUFS	9.8 LU	-inf	-15.5 LUFS	-14.2 LUFS
AloiseSauvageStereo.wav	-18.0 LUFS	8.4 LU	-inf	-15.6 LUFS	-14.3 LUFS
AloiseSauvageStereoUpmixHalo.wav	-18.0 LUFS	10.9 LU	-inf	-15.4 LUFS	-13.9 LUFS
BjorkAppleSpatialAudio.wav	-18.0 LUFS	9.7 LU	-inf	-15.0 LUFS	-12.7 LUFS
BjorkAtmosAmazon.wav	-18.0 LUFS	9.1 LU	-inf	-14.1 LUFS	-12.7 LUFS
BjorkStereo.wav	-18.0 LUFS	8.5 LU	-inf	-13.7 LUFS	-10.8 LUFS
BjorkUpmixHalo.wav	-18.0 LUFS	8.2 LU	-inf	-14.7 LUFS	-12.7 LUFS
BoygeniusAppleSpatialAudio.wav	-18.0 LUFS	10.4 LU	-inf	-15.5 LUFS	-14.0 LUFS
BoygeniusAtmosAmazon.wav	-18.0 LUFS	10.5 LU	-inf	-16.1 LUFS	-15.1 LUFS
BoygeniusStereo.wav	-18.0 LUFS	7.5 LU	-inf	-16.2 LUFS	-15.3 LUFS
BoygeniusStereoUpmixHalo.wav	-18.0 LUFS	7.7 LU	-inf	-16.3 LUFS	-15.6 LUFS
DamsoAppleSpatialAudio.wav	-18.0 LUFS	13.8 LU	-inf	-16.2 LUFS	-12.6 LUFS
DamsoAtmosAmazon.wav	-18.0 LUFS	12.6 LU	-inf	-16.6 LUFS	-13.1 LUFS
DamsoStereo.wav	-18.0 LUFS	11.2 LU	-inf	-17.0 LUFS	-15.6 LUFS
DamsoStereoUpmixHalo.wav	-18.0 LUFS	11.5 LU	-inf	-17.2 LUFS	-15.4 LUFS
HarryStyleAppleSpatialAudio.wav	-18.0 LUFS	2.2 LU	-inf	-16.9 LUFS	-15.6 LUFS
HarryStyleAtmosAmazon.wav	-18.0 LUFS	2.3 LU	-inf	-16.8 LUFS	-15.8 LUFS
HarryStyleStereo.wav	-18.0 LUFS	1.9 LU	-inf	-17.2 LUFS	-15.9 LUFS
HarryStyleStereoUpmixHalo.wav	-18.0 LUFS	1.9 LU	-inf	-17.2 LUFS	-16.1 LUFS
ParamoreAppleSpatialAudio.wav	-18.0 LUFS	5.1 LU	-inf	-16.5 LUFS	-14.2 LUFS
ParamoreAtmosAmazon.wav	-18.0 LUFS	5.2 LU	-inf	-16.6 LUFS	-14.7 LUFS
ParamoreStereo.wav	-18.0 LUFS	6.8 LU	-inf	-15.9 LUFS	-14.3 LUFS
ParamoreUpmixHalo.wav	-18.0 LUFS	6.7 LU	-inf	-16.0 LUFS	-14.4 LUFS
StromaeAppleSpatialAudio.wav	-18.0 LUFS	12.3 LU	-inf	-12.6 LUFS	-10.7 LUFS
StromaeAtmosAmazon.wav	-18.0 LUFS	11.8 LU	-inf	-13.0 LUFS	-11.5 LUFS
StromaeStereo.wav	-18.0 LUFS	9.9 LU	-inf	-13.9 LUFS	-13.0 LUFS
StromaeStereoUpmixHalo.wav	-18.0 LUFS	8.8 LU	-inf	-13.7 LUFS	-12.5 LUFS
TearsForFearsAppleSpatialAudio.wav	-18.0 LUFS	2.3 LU	-inf	-17.1 LUFS	-15.1 LUFS
TearsForFearsStereo.wav	-18.0 LUFS	1.9 LU	-inf	-17.2 LUFS	-15.7 LUFS
TearsForFearsStereoUpmixHalo.wav	-18.0 LUFS	1.7 LU	-inf	-17.3 LUFS	-16.1 LUFS
TearsForFearsAtmosAmazon.wav	-18.0 LUFS	2.3 LU	-inf	-16.6 LUFS	-14.9 LUFS

Annexe 9 : Questionnaire perceptif

Déroulé du test :

Vous allez être amené-e à écouter et comparer 4 versions de différents extraits musicaux d'environ 1 minute. Chacune des 4 versions, nommées de 1 à 4, est assignée à un bouton d'écoute correspondant sur le contrôleur.

Pour chaque extrait il vous sera demandé de :

- Écouter l'ensemble des quatre versions (stimuli 1,2,3,4). Vous pouvez écouter les différentes versions autant de fois que souhaité.
- Classer votre préférence pour chaque stimulus sur une échelle de 0 à 5 (0 = je n'aime pas du tout, 2,5= pas de préférence, 5 = aime beaucoup). Plusieurs stimuli peuvent avoir la même note.
- Classer votre sensation d'immersion pour chaque stimulus sur une échelle de 0 à 5 (0 = pas immersif, 5 = très immersif). Plusieurs stimuli peuvent avoir la même note.
- Les deux critères sont notés de manières indépendantes. Exemple : une sensation importante d'immersion n'implique pas nécessairement une note haute sur l'échelle de préférence et réciproquement.
- Vous pouvez ensuite ajouter des commentaires pour un ou des stimuli afin de préciser votre classement si vous le souhaitez.
- Vous pouvez rejouer l'extrait autant de fois que vous voulez. Une fois que vous avez terminé la notation, vous pouvez passer à l'extrait suivant.
- Un entrainement pour vous présenter le fonctionnement du test vous sera présenté avant de démarrer l'évaluation. Merci de ne pas modifier le volume sonore pendant l'ensemble du test
- Il n'y a pas de mauvaises réponses, si vous ne percevez aucune différence entre les différentes versions vous pouvez l'indiquer !

Grille d'évaluation :

Evaluer la préférence pour chacun des 4 extraits

0 1 2 3 4 5

Evaluer la sensation d'immersion pour chacun des 4 extraits

0 1 2 3 4 5

Commentaires additionnels sur l'un ou plusieurs des extraits :

Informations supplémentaires :***Informations sur le/la participant-e***

- **Nom Prénom :**
- **Age :**
- **Etes vous (*raier les mentions inutiles*) :**
 - **professionnel-le de l'audio**
 - **étudiant-e en école de son**
 - **enseignant-e en école de son**
 - **autre (préciser) :**
- **Quel moyen de diffusion utilisez vous le plus pour écouter de la musique ?**

- **Quelle est votre fréquence d'écoute hebdomadaire (heures/semaine) et sur quel support ? (CD, Streaming...)**

- **Etes vous familier-e avec l'écoute de musique au format immersif ? (Dolby Atmos, Apple Spatial Audio, Sony 360RA...)**

Annexe 10 : Résultats de l'évaluation du critère préférence

EXTRAIT	SUJET	STEREO	AMAZON	APPLE	UPMIX
Aloïse Sauvage	1	2,5	4	3,5	2
	2	2	2	2	3
	3	2	1	3	1
	4	2,5	3,7	3	3,7
	5	2,5	2,5	2,5	2,5
	6	4	4	4	2
	7	4	3	5	2
	8	3	5	5	1
	9	2,5	4	3	3,5
	10	5	4	3	1
	11	3	3	2	2
	12	4	2	1	4
	13	3,5	5	4	2
	14	1,5	4	1	3,5
	15	4	2,5	4,5	2,5
	16	4,5	4,5	3,5	2,5
	17	3,5	3,5	3,5	3,5
	18	3	3	2,5	3
	19	4	3	2	1
	20	3	3	1	2
	21	2	3	4	3,5
	22	3,75	4	4	0,5
	23	2	4	2	4
	24	4	3	3,5	4
	25	3	3	3	3,5
	26	2,5	2,5	2,5	2,5
	27	4	3,5	3,5	3,75
	28	2	4	2	4
	29	4	4	3,5	4
Bjork	1	3	3,5	4,5	5
	2	4	2	2	3
	3	2,5	3,5	3,5	2,5
	4	4	3	3	4
	5	1	4	4	5
	6	1	3	5	2
	7	2	3	4	4
	8	0	5	5	0
	9	3,5	2,25	2,5	3
	10	4	1	2	4
	11	2	4	3	2
	12	4	2,5	2,5	1
	13	5	1	1	4,5
	14	3	2	3,5	4
	15	2,5	0,5	0,5	3,5
	16	4,5	3	3	4
	17	5	2	4	2

	18	3,5	2,5	2,5	4
	19	4	1	3	2
	20	3	4	5	4
	21	2	4	4	3
	22	2,5	1	1,75	4
	23	2	5	5	2,5
	24	5	2,5	2,5	4,5
	25	3,75	3,75	3,75	3,75
	26	3	3,75	3,5	3,25
	27	2,5	3,5	3,5	2,5
	28	1	1	3,5	4
	29	4	1,5	1,5	4,5
Boygenius	1	5	4	2	2
	2	3,5	2	2	2
	3	3,5	2,5	1	1
	4	1,7	2,7	3	3,2
	5	2	1	4	4
	6	1	1	4	2
	7	5	2	3	1
	8	5	4	1	5
	9	4	2	4	3
	10	3	4	2	3
	11	2	3	2	4,5
	12	4	1	2,5	4,5
	13	5	1,5	1,5	4,5
	14	5	2	3	4,5
	15	4	0,5	1,5	3,25
	16	4	2,5	2,5	4
	17	4,5	3	3	4
	18	2	1,5	3,5	3
	19	4	1	1	3
	20	4	1	1	4
	21	3	3,5	2,5	3
	22	1	0,5	3	4
	23	4	2	4,5	1
	24	4,5	2,5	2,5	4,5
	25	1,5	3,5	3,5	3
	26	4	3	3	3,5
	27	4	2	3	2
	28	5	2	4	3
	29	4	2	2	5
Damso	1	4,5	2	2,5	5
	2	1,5	1,5	1,5	3,5
	3	1,5	3	4	0,5
	4	2,8	3,7	3,9	2,5
	5	0	4	3	3
	6	3	1	1	4
	7	2	4	5	3
	8	5	5	5	5

	9	3	3,5	3,5	3
	10	2	2	4	1
	11	4	2	2,5	5
	12	4,5	2	2	5
	13	5	1,5	1,5	4
	14	4	2	1	4,5
	15	2	0,5	2	4
	16	2,5	3	4	2
	17	4	2,5	2	4
	18	1	3,5	3	4
	19	1	3	4	2
	20	5	3	1	4
	21	2,5	4	4	3
	22	0,5	2	2	4
	23	1	3	3	5
	24	1	1,5	1,5	3,5
	25	1,5	3	3	2
	26	2	4	3	2,5
	27	4	1,5	1,75	2,75
	28	3,5	2,5	2	4,5
	29	4,5	2,5	2	5
Harry Style	1	1	3,5	3	5
	2	2	1	3,5	1
	3	4	1	2	3
	4	3,5	2,5	2,5	3,5
	5	4	1	2	4
	6	1	4	3	2
	7	1	1	2	5
	8	2	4	5	2
	9	3	2,75	2	3,5
	10	3	4	1	2
	11	4	1,5	1,5	3,5
	12	5	0,5	0,5	4
	13	5	0,5	1,5	3,5
	14	4,5	1	1	4
	15	4,5	1	0,5	4
	16	3,5	1	0	3,5
	17	4	1,5	1,5	3
	18	4	3	3	3,5
	19	4	1	1	3
	20	5	1	0	4
	21	3,5	2	1,5	4
	22	0,5	1	4	1,5
	23	5	3,5	3	5
	24	3,5	2,5	2,5	4,5
	25	2	3	3	4
	26	2,5	4,5	4	2,5
	27	3	2,75	2,5	4
	28	4	0	1	4
	29	1,5	5	5	2,5

Paramore	1	3	1,5	2	1
	2	2,5	2,5	2,5	2,5
	3	3,5	2	1,5	2
	4	3	2,5	2,1	2,5
	5	2	1	3	3
	6	4	3	3	3
	7	4	2	2	4
	8	5	3	2	4
	9	3	3,5	3,5	4
	10	4	2,5	4	3
	11	2	3,5	3	2,5
	12	2,5	4	4	2,5
	13	1,5	4,5	4,5	1
	14	4,5	2	3	1
	15	3,5	2	1	1,5
	16	4	1,5	3,5	1,5
	17	4,5	3	3	4
	18	2	1,5	3,5	3
	19	4	1	1	3
	20	4	1	1	4
	21	3	3,5	2,5	3
	22	1	0,5	3	4
	23	4	2	4,5	1
	24	4,5	2,5	2,5	4,5
	25	1,5	3,5	3,5	3
	26	4	3	3	3,5
	27	4	2	3	2
	28	5	2	4	3
	29	4	2	2	5
Stromae	1	4	2	2,5	3
	2	3	1,5	1,5	3
	3	1	2,5	3,5	2,5
	4	3,5	2,8	3,2	3,2
	5	4	2	3	2
	6	1	2	4	3
	7	1	1	3	5
	8	4	5	5	5
	9	4	2,5	4	2,25
	10	1	2	4	4
	11	3,5	3,5	2,5	2,5
	12	2	4	5	2
	13	5	4	0,5	1
	14	2,75	2	3,5	4
	15	3	1,5	2	4
	16	4	4	2,5	1,5
	17	2,5	3,5	4,5	2,5
	18	4	3	3	4
	19	2	3	4	3
	20	3	1	1	4
	21	3,5	4	2,5	2

	22	1	0,5	4	3,25
	23	3,5	1,5	2	4
	24	3	3,5	4,5	4
	25	3	4	4	3,75
	26	3,5	3	2,75	4
	27	2,5	2,75	3	3,75
	28	4	2	5	3
	29	3	4	4,5	2,5
Tears for Fears	1	2	4,5	4	2
	2	1,5	1,5	3	1,5
	3	3,7	1,5	1,25	3,5
	4	2,5	1,8	1,8	1,8
	5	4	1	1	3
	6	3	3	3	3
	7	1	2	1	3
	8	5	4	5	5
	9	3,5	2	3,25	4
	10	2	4	5	4
	11	4,5	2,5	3,5	3,5
	12	4	2	2	3
	13	5	1	3	4
	14	5	2	1,5	4
	15	4	0,5	2,25	1
	16	4	2	3	3
	17	3,5	3	4	3
	18	4	3	4	2
	19	2	4	3	1
	20	4	5	3	2
	21	2	2	4	3
	22	3,5	4	0,5	1
	23	4	1	5	4
	24	3,5	2	4,5	2,5
	25	2,5	1,75	4,5	2,75
	26	4,5	2,5	3,5	3
	27	4	1,25	1,5	2
	28	3	2	1	5
	29	4,5	2,5	2,5	4

Annexe 11 : Résultats de l'évaluation du critère immersion

EXTRAIT	SUJET	STEREO	AMAZON	APPLE	UPMIX
Aloise Sauvage	1	2	3,5	3,5	1,5
	2	3	3	3	3
	3	2	1	3	1
	4	2,5	3,7	3,7	3,7
	5	2,5	2,5	2,5	2,5
	6	3	3	3	4
	7	5	2	4	4
	8	5	5	5	2
	9	4	3,25	3,5	2,75
	10	1	4	5	2
	11	4	1	1	3
	12	2,5	5	0	2,5
	13	5	4	3,5	1,5
	14	1	3,5	2,5	2,5
	15	4	1,5	3	0,5
	16	2,25	2,25	4	2,25
	17	3	3	3	3
	18	2,5	2,5	2,5	2,5
	19	2	3	4	1
	20	1	1	1	1
	21	2	2,5	3	4
	22	2	3	1,75	3,5
	23	5	4	2	3,5
	24	2,5	2,5	3,5	3,5
	25	3	3	3	3
	26	2,5	2,5	2,5	2,5
	27	3,5	3,25	3,75	4
	28	3	4	3	4
	29	4,5	3,5	4,5	3
Bjork	1	3,5	4	5	4,5
	2	3,5	2	2	3
	3	2,5	4,5	4,5	2,5
	4	3,5	2,2	2,8	3,5
	5	1	3	3	4
	6	2	3	5	2
	7	2	3	4	4
	8	0	5	5	0
	9	4,5	4	3	3,75
	10	4	2	3	2
	11	4	3	2	4
	12	5	3	3	0
	13	5	0,5	1,5	3,5
	14	2	2,5	3,5	4
	15	3,5	0,75	1,25	4
	16	5	0	2	4
	17	5	2,5	4	3

	18	4	1,5	1	3
	19	3	1	2	4
	20	2	4	5	5
	21	2	4	4	3
	22	3,25	0,75	1,5	4
	23	5	3,5	3,5	4
	24	4,5	4	3,5	5
	25	3	3	3	3
	26	4,5	2,5	2,5	3,75
	27	3	4	3,5	3
	28	4	3	3,5	3,5
	29	5	1,5	1,5	4,5
Boygenius	1	5	4	2	2
	2	2,5	2,25	1,75	2,75
	3	1,5	3,7	1	1
	4	1,5	1,5	2,3	2,6
	5	5	2	4	3
	6	4	1	3	2
	7	5	2	3	3
	8	5	5	5	5
	9	4,5	1,5	4,5	3
	10	1	4	2	2
	11	4	3	1	4
	12	4,5	2	1	3,5
	13	4	1,5	1,5	5
	14	4,5	3	4	3
	15	4	0	2	3
	16	2	4	3,5	2
	17	4,5	2	2,5	3,5
	18	3,5	2,5	3	3
	19	4	2	2	3
	20	5	0	1	4
	21	4	2,5	3,5	2
	22	4	3,5	3	1,75
	23	4,5	2,5	3,5	1
	24	4,5	2,5	3	4
	25	4,5	3	4	2
	26	5	2	3	4
	27	4	2	3	2
	28	4,5	4	2	2,5
	29	4,5	3	3	4,5
Damso	1	4,5	2	2,5	5
	2	1,5	1,5	1,5	3,5
	3	1	3,5	3,7	1
	4	2,8	2,8	2,8	2,2
	5	0	3	3	3
	6	2	1	1	4
	7	2	4	5	3
	8	5	4	4	5
	9	4	3	3	4

	10	2	4	3	1
	11	3,5	2	2	4,5
	12	4,5	2	2	5
	13	4	1,5	0,5	5
	14	2,5	3	3	4
	15	2,5	0,5	2	4
	16	4,5	2	1,5	4
	17	4	2,5	2	4
	18	1	4	3	4
	19	1	3	4	2
	20	5	2	1	4
	21	2	3,5	3,5	2,5
	22	4	1	1	4
	23	5	3,5	3	4
	24	3,5	1,5	2	4,5
	25	3,5	2,5	2,5	2,5
	26	2	4	3	2
	27	4	1,75	2	3,25
	28	4,5	1	2	4
	29	4,5	2	2	5
Harry Style	1	4,5	2,5	2	5
	2	3,5	0,5	0,5	2,5
	3	1,7	4	3	1,5
	4	3	2,5	2,5	2,8
	5	4	2	1	4
	6	5	2	1	4
	7	2	1	1	4
	8	5	2	2	5
	9	3,75	2,75	2,5	3,5
	10	5	3	3	2
	11	3,5	3	1,5	3
	12	5	0,5	0,5	4
	13	5	1	1	4
	14	5	2,5	0	3
	15	5	0,5	0,25	4,5
	16	3,5	4	3	3,5
	17	3,5	2	1,5	4
	18	4	3	3	3,5
	19	3	2	2	4
	20	5	0	0	5
	21	4	3	0,5	4
	22	4	1	2	4
	23	5	3,5	3	5
	24	4,5	2,5	1,5	3,5
	25	3	2	2	4
	26	5	3	2	4
	27	3	2	1,5	4
	28	5	2	1	4
	29	5	2	2	4,5
Paramore	1	3	1,5	2	1
	2	2,5	2,5	2,5	2,5

	3	1,7	3,7	1,5	3,5
	4	2	3,5	3	2,5
	5	4	1	3	3
	6	4	2	2	2
	7	4	2,5	4	2
	8	3	4	5	2
	9	4	2	2	3,5
	10	4	1,5	2	3
	11	4	2,5	2,5	3
	12	4	1	2	3,75
	13	1,5	5	4	0,5
	14	4	1	3,5	3
	15	4	2,5	0,5	2,75
	16	3	2	4	2
	17	4,5	0,5	3	2
	18	3	3	2	2
	19	4	2	2	2
	20	2	2	2	2
	21	3,5	2,5	4	4
	22	4	1	1	1
	23	1	4	1,5	3
	24	2,5	3	4	2
	25	4	1,5	1,5	4
	26	4	3	3	3
	27	4,5	2	2	3
	28	4	3,5	3	3
	29	4,5	2,5	2,5	2
Stromae	1	4	2	2,5	3
	2	3,5	1	1	2,5
	3	4	1,7	0,7	2,4
	4	2	1	3	4
	5	5	3	2	4
	6	4	2	2	3
	7	2	2	3	4
	8	5	3	3	3
	9	3,5	2,5	3	3,5
	10	2	1	5	3
	11	4	3	3	2,5
	12	4	2	2	5
	13	5	4	1	0,5
	14	4,5	3	2	4
	15	3,5	0,5	0,5	4,5
	16	4	4	2,5	1,5
	17	2,5	3	4,5	2
	18	3	4	4	3
	19	2	3	4	3
	20	3	0	0	4
	21	4,5	1,5	3,5	2,5
	22	4	1	2	3
	23	4,5	4	4,5	4
	24	4	3	3,5	4,5

	25	4	4,5	4	4,5
	26	3	5	4	2
	27	2	3	1	4
	28	4	1	2	4
	29	3,5	4	4	3,5
Tears for Fears	1	2	4	3,5	3,5
	2	3	1,5	1,5	1,5
	3	2,5	1,7	1,5	3,25
	4	3	4	1	2
	5	4	4	1	2
	6	4	2	3	3
	7	1	1	2	2
	8	5	4	5	5
	9	4,5	3	3	3,5
	10	1	4	3	2
	11	4	2,5	2,5	3,75
	12	5	2	1	4
	13	5	2	3	4
	14	5	3	2	4
	15	3	2	3,75	3,5
	16	4	1	1	3
	17	3	2	4	2,5
	18	3	2	2	3,5
	19	2	4	3	1
	20	5	5	1	3
	21	3	4	2,5	2
	22	4,5	3,5	2	2
	23	5	4	3	2,5
	24	3,5	4,5	2,5	2,5
	25	4	1,5	3,5	2
	26	4,5	2	3,5	3
	27	4,5	1	1,5	3,5
	28	3,5	2,5	2,5	4
	29	4,5	1,5	2	4

Annexe 12 : Résultats d'ANOVA pour le critère « préférence »

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	232	736,45	3,174353448	1,501536377		
AMAZON	232	593,2	2,556896552	1,360385132		
APPLE	232	656,75	2,830818966	1,436784175		
UPMIX	232	725,9	3,12887931	1,292344193		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	57,55772629	3	19,18590876	13,72615819	9,16331E-09	2,614536786
A l'intérieur des groupes	1291,532522	924	1,397762469			
Total	1349,090248	927				

Figure 32 : ANOVA critère « préférence » - Ensemble des 8 titres

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	91,25	3,146551724	0,82820197		
AMAZON	29	97,7	3,368965517	0,822931034		
APPLE	29	87	3	1,196428571		
UPMIX	29	77,45	2,670689655	1,193128079		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	7,464310345	3	2,488103448	2,463048302	0,066162406	2,685643451
A l'intérieur des groupes	113,1393103	112	1,010172414			
Total	120,6036207	115				

Figure 33 : ANOVA critère « préférence » - Aloïse Sauvage

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	87,25	3,00862069	1,689655172		
AMAZON	29	78,75	2,715517241	1,583589901		
APPLE	29	92,5	3,189655172	1,431496305		
UPMIX	29	95,5	3,293103448	1,424415025		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	5,573275862	3	1,857758621	1,212407384	0,308612186	2,685643451
A l'intérieur des groupes	171,6163793	112	1,532289101			
Total	177,1896552	115				

Figure 34 : ANOVA critère « préférence » – BJORK

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	103,2	3,55862069	1,603226601		
AMAZON	29	63,2	2,179310345	1,06455665		
APPLE	29	75,5	2,603448276	1,024630542		
UPMIX	29	94,45	3,256896552	1,381736453		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	33,88627155	3	11,29542385	8,90428805	2,42434E-05	2,685643451
A l'intérieur des groupes	142,0762069	112	1,268537562			
Total	175,9624784	115				

Figure 35 : ANOVA critère préférence – Boygenius

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	78,8	2,717241379	2,257906404		
AMAZON	29	76,7	2,644827586	1,118990148		
APPLE	29	78,65	2,712068966	1,373866995		
UPMIX	29	101,25	3,49137931	1,528940887		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	14,01465517	3	4,671551724	2,975650701	0,034684224	2,685643451
A l'intérieur des groupes	175,8317241	112	1,569926108			
Total	189,8463793	115				

Figure 36 : ANOVA critère « préférence » – Damso

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	93,5	3,224137931	1,814039409		
AMAZON	29	61	2,103448276	1,930880542		
APPLE	29	63	2,172413793	1,808497537		
UPMIX	29	99	3,413793103	1,054802956		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	41,04094828	3	13,68031609	8,280786762	5,04718E-05	2,685643451
A l'intérieur des groupes	185,0301724	112	1,652055111			
Total	226,0711207	115				

Figure 37 : ANOVA critère « préférence » – Harry Style

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	97,5	3,362068966	1,194581281		
AMAZON	29	68,5	2,362068966	0,944581281		
APPLE	29	81,1	2,796551724	0,96820197		
UPMIX	29	82	2,827586207	1,254926108		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	14,58646552	3	4,862155172	4,458350507	0,005356109	2,685643451
A l'intérieur des groupes	122,1441379	112	1,09057266			
Total	136,7306034	115				

Figure 38 : ANOVA critère « préférence » – Paramore

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	85,25	2,939655172	1,23953202		
AMAZON	29	78,05	2,69137931	1,271083744		
APPLE	29	94,45	3,256896552	1,390665025		
UPMIX	29	91,7	3,162068966	0,974224138		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	5,525237069	3	1,84174569	1,511019447	0,215575997	2,685643451
A l'intérieur des groupes	136,5141379	112	1,218876232			
Total	142,039375	115				

Figure 39 : ANOVA critère « préférence » – Stromae

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	99,7	3,437931034	1,193866995		
AMAZON	29	69,3	2,389655172	1,321853448		
APPLE	29	84,55	2,915517241	1,782875616		
UPMIX	29	84,55	2,915517241	1,224839901		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	15,93387931	3	5,311293103	3,846368921	0,011559194	2,685643451
A l'intérieur des groupes	154,6562069	112	1,38085899			
Total	170,5900862	115				

Figure 40 : ANOVA critère « préférence » – Tears For Fears

Annexe 13 : Résultats d'ANOVA pour le critère « immersion »

Analyse de variance: un facteur

RAPPORT DÉTAILLÉ				
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance
STEREO	232	815,95	3,517025862	1,432208865
AMAZON	232	590	2,543103448	1,339173011
APPLE	232	596	2,568965517	1,367366025
UPMIX	232	725,45	3,126939655	1,182420464

ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	153,8379957	3	51,2793319	38,54742296	1,78657E-23	2,614536786
A l'intérieur des groupes	1229,189892	924	1,330292091			
Total	1383,027888	927				

Figure 41 : ANOVA critère « immersion » – Ensemble des 8 titres

Analyse de variance: un facteur

RAPPORT DÉTAILLÉ				
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance
STEREO	29	85,25	2,939655172	1,444889163
ATMOS AMAZON	29	85,95	2,963793103	1,08533867
APPLE SPATIAL AUDIO	29	87,7	3,024137931	1,280289409
UPMIX ATMOS	29	77,2	2,662068966	1,055474138

ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	2,251465517	3	0,750488506	0,616925471	0,605449336	2,685643451
A l'intérieur des groupes	136,2477586	112	1,216497845			
Total	138,4992241	115				

Figure 42 : ANOVA critère « immersion » – Aloïse Sauvage

Analyse de variance: un facteur

RAPPORT DÉTAILLÉ				
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance
STEREO	29	99,25	3,422413793	1,745997537
ATMOS AMAZON	29	77,7	2,679310345	1,67080665
APPLE SPATIAL AUDIO	29	89,05	3,070689655	1,444913793
UPMIX ATMOS	29	97,5	3,362068966	1,422259852

ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	10,0325	3	3,344166667	2,128694121	0,10058129	2,685643451
A l'intérieur des groupes	175,9513793	112	1,570994458			
Total	185,9838793	115				

Figure 43 : ANOVA critère « immersion » – Bjork

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	114,5	3,948275862	1,291871921		
ATMOS AMAZON	29	71,95	2,481034483	1,365788177		
APPLE SPATIAL AUDIO	29	78,05	2,69137931	1,171083744		
UPMIX ATMOS	29	84,1	2,9	1,124821429		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	36,93706897	3	12,31235632	9,942217896	7,28276E-06	2,685643451
A l'intérieur des groupes	138,6998276	112	1,238391318			
Total	175,6368966	115				

Figure 44 : ANOVA critère « immersion » – Boygenius

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	90,3	3,113793103	2,050517241		
ATMOS AMAZON	29	72,05	2,484482759	1,112339901		
APPLE SPATIAL AUDIO	29	71,5	2,465517241	1,062339901		
UPMIX ATMOS	29	103,95	3,584482759	1,323054187		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	25,37706897	3	8,459022989	6,098514747	0,000699836	2,685643451
A l'intérieur des groupes	155,3510345	112	1,387062808			
Total	180,7281034	115				

Figure 45 : ANOVA critère « immersion » – Damso

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	118,45	4,084482759	0,951625616		
ATMOS AMAZON	29	61,75	2,129310345	1,074199507		
APPLE SPATIAL AUDIO	29	46,75	1,612068966	0,891009852		
UPMIX ATMOS	29	109,8	3,786206897	0,718374384		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	128,7905819	3	42,93019397	47,238208	8,295E-20	2,685643451
A l'intérieur des groupes	101,7858621	112	0,90880234			
Total	230,576444	115				

Figure 46 : ANOVA critère « immersion » – Harry Style

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	98,2	3,386206897	0,968374384		
ATMOS AMAZON	29	68,7	2,368965517	1,140788177		
APPLE SPATIAL AUDIO	29	75	2,586206897	1,108374384		
UPMIX ATMOS	29	73	2,517241379	0,754156404		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	18,15267241	3	6,050890805	6,0940161	0,000703708	2,685643451
A l'intérieur des groupes	111,2074138	112	0,992923337			
Total	129,3600862	115				

Figure 47 : ANOVA critère « immersion » – Paramore

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	104	3,586206897	0,89408867		
ATMOS AMAZON	29	72,7	2,506896552	1,70137931		
APPLE SPATIAL AUDIO	29	77,2	2,662068966	1,776009852		
UPMIX ATMOS	29	94,4	3,255172414	1,034704433		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	22,21612069	3	7,405373563	5,479188972	0,001501185	2,685643451
A l'intérieur des groupes	151,3731034	112	1,351545567			
Total	173,5892241	115				

Figure 48 : ANOVA critère « immersion » – Stromae

Analyse de variance: un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
STEREO	29	106	3,655172414	1,376847291		
ATMOS AMAZON	29	79,2	2,731034483	1,440788177		
APPLE SPATIAL AUDIO	29	70,75	2,439655172	1,069889163		
UPMIX ATMOS	29	85,5	2,948275862	0,885621921		
ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	23,35954741	3	7,786515805	6,525268579	0,000415362	2,685643451
A l'intérieur des groupes	133,6481034	112	1,193286638			
Total	157,0076509	115				

Figure 49 : ANOVA critère « immersion » – Tears For Fears