

École Nationale Supérieure Louis-Lumière

Mémoire de Master Spécialité Son – Promotion 2016

FLORENT DENIZOT

L'Approche Orientée Objet au cinéma:

ENTRE SURENCHÈRE TECHNOLOGIQUE ET OUTIL DE CRÉATION SONORE

DIRECTEUR INTERNE: MOHAMMED ELLIQ

DIRECTEUR EXTERNE : RAPHAËL SOHIER

RAPPORTEUR : ÉRIC URBAIN



École Nationale Supérieure Louis-Lumière

Mémoire de Master Spécialité Son – Promotion 2016

FLORENT DENIZOT

L'Approche Orientée Objet au cinéma:

ENTRE SURENCHÈRE TECHNOLOGIQUE ET OUTIL DE CRÉATION SONORE

DIRECTEUR INTERNE: MOHAMMED ELLIQ

DIRECTEUR EXTERNE : RAPHAËL SOHIER

RAPPORTEUR: ÉRIC URBAIN

REMERCIEMENTS

Je remercie le corps enseignant de l'ENS Louis Lumière pour m'avoir aidé à élaborer mon sujet de mémoire et à le mener à bout.

Je remercie également le personnel de cette même école, de nous avoir permis, à mes amis de promotion et moi-même, d'étudier dans de bonnes conditions ; ainsi que :

Mohammed Ellio

bien évidemment, pour avoir dirigé ce mémoire, avec tout le sérieux et la rigueur qu'on lui connaît.

Raphaël Sohier

pour avoir accepté de diriger ce mémoire et m'avoir aidé de le mener à bien par ses réflexions et par la mise à disposition de sa salle pour mes expérimentations

Jérôme Wiciak

aidé à la construire avec bienveillance.

Éric Urbain

pour sa lecture attentive en tant que rapporteur du mémoire.

Serge ARTHIUS

pour avoir répondu à toutes mes questions d'ordre technique et pour avoir finalisé l'installation de la salle de Raphaël Sohier.

Dominique SCHMIT

pour avoir prêté une oreille attentive et mis à notre du système.

Jean-Pierre Halbwachs

disposé à répondre à mes interrogations.

Laurent Stehlin

pour la pertinence de ses remarques et pour ses connaissances du cinéma numérique.

Selim Azzazı

pour son regard juste et la rigueur de son raisonnement, qui, sans aucun doute, m'a permis d'envisager ce sujet avec plus de recul.

Gérard LAMPS

qui, quelque part, est à l'origine de ma démarche et m'a pour m'avoir répondu avec sincérité et avoir éclairé ma réflexion au regard de ses années d'expérience.

Armelle Mahé

de m'avoir confié ses expériences et son positivisme durant nos entretiens.

Florent FAJOLE

pour m'avoir accordé son temps, sa patience et son dévouement- qui me furent d'une grande aide pour la constitution de ma réflexion.

Simon Bonnani

pour avoir mis à ma disposition son film et avoir disposition les logiciels nécessaires au fonctionnement entretenu un discours optimiste et engagé sur le cinéma de demain.

Ariane et Laurence Lucon

soucieux que je mène à bien mes recherches et toujours pour leur traduction de l'anglais au français, du français à l'anglais et de ma pensée vers le français.

Enfin je tiens à remercier ma famille, pour son soutien durant cette période et surtout pour m'avoir encouragé à effectuer ce que j'aime, en me permettant de finir mes études.

RÉSUMÉ

Désormais les technologies de spatialisation telles que VBAP, DBAP ou encore ambisonie, jusqu'à maintenant reléguées au stade de l'expérimentation, trouvent désormais leur place dans les nouveaux formats audio fondés sur l'approche orientée objet.

À chaque « nouveauté technologique » son appropriation artistique ; ces systèmes semblent prédestinés à proposer des expériences sensorielles de réalité augmentée, s'adressant directement au cinéma-spectacle de type Géode ou Imax. Or, la récente normalisation du mode objet laisse entrevoir son expansion future dans le domaine du cinéma et de la diffusion télévisuelle, comme le confirme le succès de films, tels que *Gravity*¹ ou *The Revenant*², qui utilisent cette méthode.

Dans ce nouveau paradigme où le spectacle cinématographique semble se réorienter vers une écriture spatialisée du son, il nous a paru important de comprendre les enjeux et les opportunités offertes par les nouveaux formats orientés objet, tout en questionnant leur légitimité dans la création cinématographique.

Ce document propose un état des lieux de l'approche objet; en questionnant son intégration dans les salles de cinéma d'un point de vue technique et technologique premièrement; pratique ensuite, afin d'envisager concrètement son insertion dans la chaine de postproduction; esthétique enfin pour tenter d'évaluer ses possibilités créatrices au regard d'une narration cinématographique commune.

Mots clefs : VBAP, DBAP, approche orientée objet, downmixage automatique, spatialisation, espace sonore, postproduction sonore, son au cinéma.

¹ Alfonso Cuarón, *Gravity*, 91 min, États-Unis, Royaume-Uni, sorti en France le 23 octobre 2013.

² Alejandro G. IÑÁRRITU, *The Revenant*, 156 min, États-Unis, sorti en France le 24 février 2016.

ABSTRACT

Spatial sound effects such as VBAP, DBAP or even Ambisonics, which had, up until now, been relegated to the ranks of experimental practice, have now found a place among the new audio techniques said to be "object" oriented.

Each "technological innovation" has its own corresponding artistic device; these systems seem bound to propose sensory experiments in augmented reality, appealing directly to cinema-as-spectacle venues such as La Géode or Imax. Therefore, the recent standardization of object as a mode, foreshadows its future development in the field of cinema and television broadcasting, as confirmed by recent movie successes using this method (*The Revenant*³, *Gravity*⁴).

In this new paradigm where cinematographic spectacle is reconfigured as a spacesound oriented script, we have wished to expose the stakes and opportunities offered by these new object-oriented techniques, whilst questioning their legitimacy in film design.

This document presents a situational analysis of the object-based approach. It questions its implementation in movie theatres, initially from a technical and technological point of view and subsequently from a practical perspective with regard to its place in the postproduction workflow. Lastly, it studies the creative impact on the film's narrative content.

Keywords: VBAP, DBAP, object-based audio approach, automatic downmix, spatialization, sound space, sound postproduction, cinema sound.

³ Alfonso Cuarón, *Gravity*, 91 min, États-Unis, Royaume-Uni, sorti en France le 23 octobre 2013.

⁴ Alejandro G. IÑARRITU, *The Revenant*, 156 min, États-Unis, sorti en France le 24 février 2016.

Table des matières

I	NTRODUCTION	5
1	ÉVOLUTION DU SON MULTICANAL DANS LE CONTEXTE CINÉMATOGRAPHIC	OUE
	CTUEL	
	1.1 Modification d'écoute chez le spectateur : de la monophonie au mixage orienté objet	11
	1.1.1 De la Monophonie au Multicanal, une histoire aux multiples facettes	11
	1.1.2 Insertion du 5.1 et ses dérivés	18
	1.1.3 Nouvelle façon d'appréhender le son spatial : le mixage orienté objet	23
	1.2 APPARITION DU CONCEPT DE MIXAGE ORIENTÉ OBJET	29
	1.2.1 Les bases du mixage orienté objet	29
	1.2.2 Le Dolby Atmos	36
	1.2.3 Le DTS MDA : Mixage en kit	39
	1.2.4 Le Format Européen ADM	43
	1.3 La quête de l'immersion entre nouveauté et résurgence	44
	1.3.1 Une volonté de plus de crédulité ?	44
	1.3.2 L'image 3D ou le mythe du cinéma total ?	48
	1.3.3 Enjeux du son 3D et de l'image 3D	51
	1.4 CONCLUSIONS ET ORIENTATION DES MIXAGES ACTUELS	58
2	EXPÉRIMENTATION DE L'APPROCHE OBJET	65
	2.1 EXPLOITATION DE LA MATIÈRE SONORE DANS LA CRÉATION CINÉMATOGRAPHIQUE	65
	2.1.1 Intégration des ambiances en mode objet : Pour quel film ?	65
	2.1.2 Les sons effets et leur intégration.	71
	2.1.3 Limites, possibilités et ouvertures de la création en mode objet	77
	2.2 Particularités de la postproduction des formats orientés objet	85
	2.2.1 Intégration du mode objet dans la chaine de travail	85
	2 2 2 En phase de montage son	86

	2.2.3 En phase de mixage	88
	2.2.4 Vers une refonte des métiers	89
2	3 Conclusion et tendances à venir	90
3 P	'ARTIE PRATIQUE: ANALYSES PERCEPTIVES DU MIXAGE ORIENTÉ OBJET	95
3	1 Contexte d'étude	95
	3.1.1 Réflexions préliminaires.	95
	3.1.2 Principaux axes d'études	96
	3.1.3 Problématiques liées à la compatibilité descendante :	97
3	2 Protocole du test perceptif	100
	3.2.1 Présentation générale du test.	100
	3.2.2 Définition des paramètres d'études	101
	3.2.3 Le protocole	104
	3.2.4 Choix d'un public test	105
3	3 MISE EN PLACE DU TEST	106
	3.3.1 Préparation des séquences	106
	3.3.2 Choix d'un corpus de séquences	107
	3.3.3 Limites du test	116
3	4 Analyse et interprétation	117
	3.4.1 Analyse	117
	3.4.2 Interprétations.	124
	3.4.3 Récapitulatif	129
Co	NCLUSION	131
Вів	LIOGRAPHIE	137
Fil	MOGRAPHIE	143
Ani	NEXES	145
A	_ Description du langage BIFS :	.146
В	_ Installation electroacoustique du Dolby Atmos :	.147
C		

1. Nouvelle configuration pour les enceintes surround?149	
3. Une nouvelle exploitation de la dynamique sonore?157	
D_ MISE EN ESPACE DES OBJETS SONORES : PRINCIPES PHYSIQUES	
1. Présentation161	
2. VBAP	
3. DBAP	
4. Dual Balance Panning171	
5. Ambisonie	
6. VBAP, DBAP, HOA : des parallèles intéressants175	
E_ OUTILS DE MISE EN ESPACE DÉVELOPPÉS PAR DOLBY	
1. Le RMU :	
2. Le Dolby Local Renderer : mode d'emploi et utilisation179	
3. Le Dolby Atmos Panner	
F_ COMPATIBILITÉ DES MIXAGES, DOWNMIXAGE AUTOMATIQUE	
1. Fonctionnement général des processeurs de diffusion orientée objet192	
2. Fonctionnement du Dowmixeur Dolby Atmos :195	
G_ Déscription du schéma de principe du décodeur MPEG H Audio 3D :197	
H_ Descriptif des films utilisés	
1. Les Medouny, 2016, 16'	
2. Bamako Saint Afrique, 2013, 14'	
3. Les Saisons, 2016, 1 h 37'	

Tables des Figures

•	Figure 1.1 : Photogramme d'une pellicule comportant respectivement de gauche à droite, le format numérique SDDS, le format Dolby Digital et enfin le Dolby Surround SR18
•	Figure 1.2 : Configuration Auro 3D 11.1 dans une salle de cinéma.
•	Figure 1.3 : Schéma de composition du système de diffusion 22.2. Source www.nhk.or.jp22
•	Figure 1.4 : Configuration Imm Sound 23.1. Les chiffres et flèches en rouge désignent respectivement les canaux de la couche horizontale ainsi que les enceintes qu'ils alimentent25
•	Figure 1.5 : Configuration des haut-parleurs du système Dolby Atmos. En rouge, l'emplacement des haut-parleurs préexistant aux normes 5.1 et 7.1. En vert, les haut-parleurs optionnels
•	Figure 1.6 : Illustration de la différence entre le modèle égocentrique à gauche et allocentrique à droite. Source SMPTE
•	Figure 1.7 : Schéma comparatif des différentes possibilités de flux audio-numériques depuis l'arrivée des formats orientés objet. Source ITU-R BS.2159-4
•	Figure 1.8 : Schéma de principe du décodeur MPEG H Audio 3D
•	Figure 1.9 : Modification des chaines A et B de traitement du signal audio au cinéma pour des formats audio-immersifs. Source SMPTE
•	Figure 1.10 : Schéma de principe de Dolby Atmos. Source www.dolby.com38
•	Figure 1.11 : Schéma de principe d'utilisation des beds et des objets. Source www.dolby.com39
•	Figure 1.12 : Différentes configurations électroacoustiques du DTS MDA
•	Figure 1.13 : Description des objets-audio dans le Format DTS MDA. Source MDA White Paper42
•	Figure 1.14: Worflow d'une chaine de postproduction DTS MDA. Source MDA White Paper42
•	Figure 2.1 : Photogramme de la séquence d'introduction des Medouny. Saïd marche de dos vers le passage sous le pont
•	Figure 2.2 : Photogramme de la séquence d'introduction des Medouny. Saïd passe sous le petit pont
•	Figure 2.3 : Principe de mise en espace et de fabrication d'une acoustique en fonction de la modélisation de salle choisie
•	Figure 2.4 : Capture d'écran de la fenêtre d'édition du logiciel Audiostage. La fenêtre d'édition 3D View propose une représentation graphique de la modélisation de l'espace dans lequel évolue le spectateur
•	Figure 2.5 : Illustration du principe de création d'images de synthèse par synthèse particulaire. En remplaçant la caméra virtuelle par un microphone virtuel, le parallèle peut être fait pour les applications audio
•	Figure 3.1 : Les Medouny, Simon Bonanni, 2016, Photogramme. Saïd rentre dans la salle alors que Silvio entraine Malik

•	Figure 3.2 : Les Medouny, Simon Bonanni, 2016, Photogramme. Les trois frères sont en voiture. Malik à droite, Silvio conduit et Saïd se trouve à l'arrière	
•	Figure 3.3 : Bamako Saint Afrique, Frédéric Jolfre, 2013, Photogramme	112
•	Figure 3.4 : Les Saisons, Jacques Perrin et Jacques Cluzaud, 2016, Photogramme	113
•	Figure 3.5 : Les Saisons, Jacques Perrin et Jacques Cluzaud, 2016, Photogramme	115
•	Figure 3.6 : Récapitulatif des résultats de différences significatives	118
•	Figure 3.7 : Moyennes des résultats obtenus pour la variable indépendante immersion	119
•	Figure 3.8 : Moyennes des résultats obtenus pour l'immersion en fonction des séquences	120
•	Figure 3.9 : Moyennes des résultats obtenus pour la variable indépendante lisibilité du son	121
•	Figure 3.10 : Moyennes des résultats obtenus pour la lisibilité du son en fonction des séquences.	121
•	Figure 3.11 : Moyennes des résultats obtenus pour la variable indépendante qualité de timbre	122
•	Figure 3.12 : Moyennes des résultats obtenus pour la qualité du timbre en fonction des séquence	es 122
•	Figure 3.13 : Moyennes des résultats obtenus pour l'adéquation à l'image en fonction des séquences	123
•	Figure 3.14 : Tableau du taux de corrélation c en fonction de la séquence	124
•	Figure C.1: Schéma d'une zone d'écoute centrale, à gauche une configuration discrètes et à dro une configuration où les enceintes acoustiques sont appairées	
•	Figure C. 2 : Découpage de la salle en zones et de régions, deux configurations possibles	151
•	Figure C.3 : Orientation verticale des enceintes arrière.	153
•	Figure C.4: Positionnement des enceintes acoustiques du plafond	154
•	Figure C.5 : Positionnement longitudinal des enceintes au plafond. Les figures du milieu et de droite représentent les deux configurations possibles par région	154
•	Figure C.6 : Orientation de l'axe acoustique pour les haut-parleurs du plafond	155
•	Figure C.7 : Tableau des niveaux de calibration en fonction des systèmes de diffusion	158
•	Figure C.8: Représentation des différentes technologies mises en jeu dans les dispositifs objet. I a) VBAP deux dimensions, en b) dual-balance et en c) distance-based	
•	Figure D.9 : Représentation du Triangle actif dans l'utilisation du VBAP. Source Ville Pulkki	164
•	Figure D.10 : Placement de la source virtuelle dans un système de coordonnées cartésiennes	167
•	Figure D.11: Etalement d'une source sonore virtuelle par aiguillage de plusieurs directions activant un trio de haut-parleurs	169
•	Figure D.12 : Schéma de principe d'une chaine d'encodage et décodage ambisonique	175
•	Figure E.13 : Schéma d'intégration du Local Renderer dans une configuration où ProTools contles métadonnées objets	
•	Figure E.14 : Différentes configurations possibles du logiciel Local Renderer	181
•	Figure E.15 : Configuration des pistes bed et objets ainsi que leurs bus associés	184

Figure E.16 : Exemple de chemin du signal audio et des métadonnées	186
• Figure E.17 : Représentation du chemin du signal des métadonnées de spatialisation propriétair aux consoles et au Dolby Atmos. Source : Authoring for Dolby Atmos © Sound Manual	
• Figure E.18 : Configuration la plus simplifiée des échanges entre la console de mixage, le RMU la station d'édition ProTools. Source : Authoring for Dolby Atmos © Sound Manual	
• Figure E.19 : Fenêtre du plugin Dolby Atmos Panner. Source : Authoring for Dolby Atmos © So Manual	ound 189
• Figure E.20 : Fenêtre du plugin, Dolby Atmos Panner avec l'outil Size en action	190
Figure E.21 : Fenêtre du plugin pour une piste objet stéréophonique.	191
• Figure F.22 : Fenêtre du paramétrage du rendu en fonction du type de source	195
• Figure H.23: Les Medouny, Simon Bonanni, 2016, Photogramme	198
Figure H.24 : Bamako Saint Afrique, Frédéric Jolfre, 2013, Photogramme	199
• Figure H.25 : Les Saisons, Jacques Perrin et Jacques Cluzaud, 2016, Photogramme	200

Introduction

« Afin d'être représenté, le réel doit être connu, et la connaissance est toujours en soi une forme de représentation [...] la « réalité » que chaque nouvelle technologie s'engage à représenter est en grande partie définie par des systèmes représentationnels préexistants.⁵ »

Dans l'histoire des arts, le cinéma ne semble pas avoir été épargné des modèles de représentation préexistant à sa naissance. Tout comme la photographie s'est inspirée de la peinture, le cinéma sonore est issu de plusieurs domaines : la radio, le théâtre, la musique orchestrale, la phonographie et le spectacle vivant. Ce qui explique en partie pourquoi, dans les débuts du cinéma, la notion de théâtralité était encore très présente. D'autre part, le cinéma muet était tout de même accompagné de bonimenteurs et de pianiste, proposant non seulement une interprétation vocale et musicale mais aussi un lien entre le spectacle visuel et la salle de projection.

Lorsque les premières expérimentations sonores sont apparues, c'est dans l'héritage et le contexte théâtral que les techniques de diffusion se sont développées, rejoignant rapidement la notion de « scène sonore » que ces modèles préexistant impliquaient. La volonté de créer un espace sonore s'accordant avec la « réalité » est très vite apparue comme la ligne directrice à adopter dans le spectacle audiovisuel. Le développement de la diffusion sonore au cinéma n'est pas motivé uniquement par les progrès technologiques, mais est aussi dû à une demande artistique et économique de la part de l'industrie cinématographique, modifiant culturellement l'esthétique des films ainsi que les expériences sensorielles des spectateurs.

⁵ Rick Altman (dir.), Sound theory, sound practice, New York, Routledge, coll. AFI film readers, 1992, p. 46. Traduit, par nos soins, de l'anglais: « In order to be represented, the real must be known, and knowledge is always already a form of representation [...] the "reality" which each new technology sets out to represent is in large part defined by preexistent representational systems ».

Ce qui était clairement assumé comme artifice visible se dissimule petit à petit dans les coulisses des salles de projection et très vite, les haut-parleurs derrière l'écran prennent la place des bonimenteurs et de la fosse d'orchestre. Dans un but de transparence parfaite et afin de proposer un spectacle visuel et sonore total, réalisateurs, techniciens et sociétés de l'industrie cinématographique s'adonnent au rêve de corréler le son et l'image. Une pluralité de systèmes se succède avant d'arriver à un dispositif, somme de toutes les expériences et modèles passés, mondialement répandu aujourd'hui : le 5.1. Ce procédé de diffusion, considéré alors comme « système du futur » a fait l'objet de grands débats à son apparition. Des critiques universitaires émettent leurs avis controversés et ne voient pas l'avancée technologique de la spatialisation comme un nouveau moyen pour penser le cinéma. Il a donc fallu un certain temps aux réalisateurs et aux professionnels pour s'affranchir du caractère gadget initialement attribué au système 5.1 et utiliser pleinement ses possibilités créatrices.

Ces mêmes questions semblent se reposer avec la récente apparition de « nouveaux » formats dits orientés objet qui, au premier abord, semblent plus convenir à des programmes de réalité augmentée de type Géodes qu'à des fins cinématographiques. Ainsi l'association d'une image dite à réalité virtuelle augmentée et d'un espace sonore éclaté dit « 3D » semble être actuellement et logiquement la meilleure application possible. Néanmoins, c'est pourtant dans un but de questionner la validité du mixage objet dans la narration cinématographique majoritaire, à savoir la 2D, que nous allons mener cette étude. Ces nouveaux formats, dont les récentes expériences cinématographiques ont été remarquées, proposent une utilisation moins conventionnelle du son spatialisé, s'éloignant doucement des normes de représentation pratiques et esthétiques appliquées jusqu'ici.

Ces nouvelles manières d'envisager le son spatial sont bien entendu mises en application par des technologies de spatialisation. Ces technologies, parmi lesquelles citons le VBAP et son dérivé, le DBAP, ainsi que l'ambisonie, bien qu'elles soient établies depuis la fin des années 90, ont mis du temps à s'intégrer dans le domaine de la restitution cinématographique. Ces principes de description du champ sonore en éléments mathématiques fondent les bases du mixage orienté objet et renouvellent la notion de spatialisation actuelle qui est orientée canal. En effet au lieu d'enregistrer les conséquences de la source sonore au travers des outils de spatialisation, le mixage objet enregistre la position de cette source dans un espace virtuel afin de le reproduire dans différents espaces réels. Le mixage orienté objet ouvre la possibilité d'une conception d'un autre espace sonore où le contenu est rendu de manière différente selon les configurations disponibles, répondant en partie aux problématiques de transportabilité bien connues des mixeurs et permettant donc potentiellement une nouvelle approche du son spatial.

Ces technologies tentent aujourd'hui de s'intégrer dans une logique de postproduction cinématographique. Les formats Dolby Atmos, DTS MDA ou encore ADM voient le jour mais ne sont pas encore totalement contextualisés dans une chaine de production cinématographique. Dans le contexte de notre recherche et le temps imparti à sa réalisation, il nous est apparu important voire fondamental de nous axer sur un système en particulier, ne pouvant prétendre à une étude complète de tous les formats orientés objet. Logiquement et au regard de la faible implantation (en France) desdits formats, nous nous sommes dirigés vers la firme américaine Dolby et son système de diffusion 3D, le Dolby Atmos, pour le moment le plus répandu.

Notre questionnement préalable à la construction de cette recherche est simple : Si l'avenir du spectacle cinématographique se réoriente dans une écriture spatialisée du son, jusqu'à quel point les formats orientés objet peuvent-ils répondre aux exigences d'une conception sonore filmique en évolution tout en étant compatibles avec les formats antérieurs? Afin de dégager au mieux les enjeux portés par ces formats, nous essaierons à travers cette étude de proposer une

approche globale des systèmes orientés objet : en tentant de répondre à des questions d'ordre technique au préalable pour comprendre ce qui fait leur particularité par rapport aux systèmes orientés canaux ; pratique ensuite afin d'envisager leur intégration dans la chaine de postproduction sonore cinématographique ; esthétique enfin, pour tenter d'évaluer les possibilités et effets des nouveaux systèmes au regard de la narration cinématographique.

Ce mémoire fait donc constat de nos recherches, interrogations et suppositions face à ce sujet trop vaste pour être étudié dans tous ses détails. Ce document, à la fois théorique et pratique, tente de mettre en exergue les potentialités du mixage objet et d'en proposer une approche générale, fruit de nos expérimentations.

Nous proposons donc premièrement au lecteur une approche historique de la spatialisation du son au cinéma afin de mieux envisager l'apparition des nouveaux formats orientés objet de manière culturelle et esthétique. Ce bref rappel nous permettra de questionner légitimement l'éternel retour du son spatialisé dans la narration cinématographique, de manière à discerner les principales caractéristiques des formats orientés objets émergents et d'en dégager leurs particularités. De la même manière nous étudierons la configuration électroacoustique des salles proposée par Dolby dans le but de comprendre quels changements ont été effectués par rapport au système de référence 5.1 et ce que cela implique concrètement dans la diffusion sonore.

Ensuite nous proposerons une démarche pratique de l'utilisation du mode objet dans la création cinématographique. Nous y ferons état de nos expérimentations, menées dans la salle de montage au format Dolby Atmos de Raphaël Sohier, sur différentes séquences de films en 2D. Ces expérimentations s'articuleront autour de comparaisons avec le système 5.1 pour mieux déceler les avantages et limites des deux dispositifs. Notre prise de position mixte oscillant entre le poste de monteur son et celui de mixeur, proposera une expérience subjective de notre confrontation

aux méthodes de spatialisation objet. Nous tenterons aussi et autant que faire se peut de pointer les éléments de spatialisation qui semblent complexes à envisager dans un downmixage. Enfin cette partie sera l'occasion d'un bref retour sur nos expériences pratiques, et ce qui selon nous reste à améliorer, permettant ainsi de proposer aux lecteurs une réflexion sur l'intégration du processus de création en mode objet dans la chaine de postproduction.

En dernier lieu, nous confronterons le fruit de notre travail et de nos observations à une réalité plus objective au travers de tests perceptifs. Ce retour sur expérience, permettant de relativiser nos hypothèses subjectives, donnera lieu à l'évaluation des principaux atouts esthétiques présupposés de l'utilisation du mode objet, en le comparant au système 5.1, premièrement par son downmixage automatique puis par une version 5.1 réalisée manuellement. L'analyse de ces tests permettra de dégager, nous l'espérons, des tendances permettant d'ouvrir la voie à de futures recherches sur la concordance supposée de ces systèmes à une narration cinématographique plus globale.

Si le corps de ce mémoire peut parfois paraître dense, cela s'explique par une volonté de proposer un document le plus complet possible, s'adressant à un public large et décrivant avec précision les éléments qu'il étudie. Il témoigne également d'un processus de recherche, avec ses hésitations et ses tâtonnements. Nous avons pris soin de ménager des rappels entre les parties, afin que le lecteur puisse s'affranchir d'une lecture linéaire et naviguer aisément dans le texte selon ses points d'intérêt.

1 ÉVOLUTION DU SON MULTICANAL DANS LE CONTEXTE CINÉMATOGRAPHIQUE ACTUEL

1.1 MODIFICATION D'ÉCOUTE CHEZ LE SPECTATEUR : DE LA MONOPHONIE AU MIXAGE ORIENTÉ OBJET

1.1.1 DE LA MONOPHONIE AU MULTICANAL, UNE HISTOIRE AUX MULTIPLES FACETTES

« le réel ne peut jamais être présenté, seule la représentation peut être représentée.⁶ »

Cette citation de Rick Altman au sujet de l'espace sonore semble aller à l'encontre de plus d'un siècle de recherche de la retranscription de la réalité dans l'image et le son au cinéma. Pourtant les avancées technologiques dans le domaine de la diffusion sonore ont toujours été marquées par une volonté de recréer une réalité sonore, et ce, dès l'apparition du célèbre *The Jazz Singer*⁷ de Alan Crossland en 1927. Adopté comme repère historique de l'arrivée du son au cinéma, le Vitasound (son procédé sonore associé) s'orienta rapidement dans la recherche et le développement de la diffusion multicanal, visant à reproduire une réalité sonore induite. Dans les années 30, les supports optiques d'enregistrement étant limités à un canal, la diffusion sur plusieurs haut-parleurs s'effectuait par l'intermédiaire d'un opérateur, qui, par l'activation d'une bascule, envoyait le signal monophonique sur un des haut-parleurs qui lui étaient mis à disposition. Ce dispositif permettait dans les débuts du cinéma sonore d'envoyer le son des dialogues derrière l'écran et le son de l'orchestre à partir de la fosse. Ceci dans le but de recréer la spatialisation

⁶ Rick Altman (dir.), Sound theory, sound practice, op. cit., p. 45. Traduit, par nos soins, de l'anglais : « The real can never be presented, representation alone can be represented »

⁷ Alan Crosland, *The Jazz Singer (Le Chanteur de Jazz)*, 88 min, États-Unis, sorti en France le 4 octobre 1929, 1927.

sonore connue des salles de cinéma. La tâche incombant à l'opérateur étant fastidieuse, les techniciens s'accordèrent sur le fait qu'il était tout à fait acceptable d'avoir conjointement les dialogues et la musique d'orchestre diffusés en un seul et même point derrière l'écran. Cette anecdote du début du cinéma parlant, entraine par son échec les prémisses du mixage orienté canal⁸, alors qu'il constitue sans le savoir les fondements du mixage orienté objet.

Cette expérience a mis l'accent sur un point, si le son pouvait être localisé sur l'écran et dans la fosse, il pouvait de la même manière, être localisé en accord avec l'évolution spatiale des éléments visuels à l'écran. Néanmoins, le retour à une période monophonique, marqua pendant un temps l'abandon d'un système multienceinte. Rapidement, les innovations techniques à l'image et les bandes sons Hollywoodiennes se complexifiant, le son spatialisé revint au cœur des débats entre techniciens du son. Dans cette optique, de nombreux essais de systèmes visant à reproduire la réalité sonore furent réalisés. Du côté français ce fut Napoléon⁹ d'Abel Gance qui instaura avec André Debrie la notion de perspective sonore. Lors de la projection du film le 11 mai 1935 dans la salle de la Paramount, plus de 32 haut-parleurs étaient dissimulés dans la salle en plus de ceux placés derrière l'écran. Ce système apparut en opposition totale à l'enceinte centrale installée depuis peu dans les salles. « Cette sonorisation se révéla bouleversante lorsque le film fut présenté, pendant son exclusivité parisienne, avec cette fameuse « perspective sonore » enfin dévoilée au public. Celui-ci, littéralement enveloppé de sons, se trouvait projeté au cœur de l'action et devenait acteur à son tour [...] la rue en révolution prenait possession de la salle où les clameurs retentissaient de partout ; la tempête rugissait et déferlait avec rage sur elle et chaque spectateur se trouvait emporté, comme Bonaparte, dans le tourbillon des éléments déchaînés¹⁰. ».

⁸ Le mixage orienté canal ou son homonyme anglais *channel-based* implique qu'un son soit attribué à un canal. Ce canal en question se voit attribuer une ou plusieurs enceintes selon sa fonction et son placement dans le système de diffusion sonore employé.

⁹ Abel Gance, Napoléon, 330 min, France, sorti le 7 avril 1927, 1935.

¹⁰ Roger ICART, Abel gance ou le promethée foudroyé, Lausanne, L'âge D'homme, coll. Histoire Théorie

Ce système, malgré son grand succès et son engouement vis-à-vis du public, ne put s'adapter à d'autres films et persister dans les salles obscures. Effectivement plusieurs éléments rendaient incompatible le système de perspective sonore à l'exploitation dans les salles. Premièrement pour des soucis de ressources humaines, car chaque point de diffusion sonore était un phonographe activé par un opérateur à différents moments du film ce qui nécessitait un grand nombre d'opérateurs et une grande connaissance du film. D'autre part, les supports de stockage des sons étaient onéreux et n'avaient pas de grande capacité. Enfin l'installation sonore et visuelle que nécessitait la projection de ce film n'allait pas dans le sens de la logique d'exploitation cinématographique.

Peu de temps après le succès parisien d'Abel Gance apparaît le système Fantasound inventé par William Garity alors technicien dans la compagnie Walt Disney. Ce système de diffusion est basé sur trois canaux destinés aux haut-parleurs derrière l'écran et deux canaux pour alimenter des haut-parleurs placés dans la salle. Notons la ressemblance avec l'actuel système 5.1 aujourd'hui standardisé. Le procédé Fantasound fut élaboré pour les besoins du film *Fantasia* réalisé par Leopold Stokowski et sorti en 1940. Ce système fut un échec à cause du nombre démesuré de haut-parleurs qu'il nécessitait; selon les salles trente à quatre-vingts haut-parleurs étaient nécessaires au bon fonctionnement du procédé Fantasound. En dépit des efforts fournis par les créateurs de *Fantasia* pour définir un « environnement sonore¹¹ » complet, la version sonore immersive ne sera projetée que dans six salles de cinéma aux États-Unis (alors le pays d'origine). Le film est projeté dans les autres salles dans le standard monophonique. Bien que le film ait été pensé pour être diffusé dans le système Fantasound, il est évident que pour des raisons économiques et de logique d'implantation du système, le film a été plus largement

Cinéma, 1984, p. 248.

¹¹ Robert E. CARR et R. M. HAYES, Wide screen movies: a history and filmography of wide gauge filmmaking, Jefferson, N.C, McFarland & Co, 1988, p. 248.

regardé et écouté dans des salles n'ayant pas le dispositif requis. Par conséquent la version sonore tronquée du film a été de plus grande utilité que la version complète.

À la suite de l'élaboration de ces nombreux systèmes multicanal commence la pérennisation et la vulgarisation du système stéréophonique. Cette implantation est accélérée non pas par l'industrie cinématographique mais par l'industrie du disque. L'enregistrement musical effectué sur disques microsillon à deux canaux est accompagné au fur et à mesure par les platines vinyles. Un système vient de naître, une industrie de se créer. Fortement encouragées par une nouvelle façon de consommer la musique, les platines viennent alors s'installer dans les foyers. L'intégration de systèmes de diffusion chez le particulier va devenir un acteur économique fondamental pour le développement et la standardisation de nouveaux systèmes. Ce succès commercial s'accompagne aussi d'une autre exploitation de la stéréophonie en salle de cinéma.

La stéréophonie est très vite idéalisée; elle devient une convention culturelle et esthétique, plaçant ce système au rang d'anthologie et lui conférant un caractère presque sacré. L'étude historique de la stéréophonie comme un phénomène esthétique a conduit à étudier le développement des conventions de représentation et des systèmes de médiation. L'étude et la compréhension de l'acoustique subjective (alors entamée au XVIII° siècle par Giovanni Battista Venturi dès 1796) prend un second souffle. C'est à partir de ce type de recherche que recommença la perpétuelle quête d'un système de diffusion qui s'accommoderait et se justifierait dans la manière de satisfaire notre capacité (psychoacoustique) auditive de perception. Cette période s'accompagne de recherches sur l'acoustique et la psychoacoustique. Apparait alors la notion de son comme phénomène spatial; il semble donc naturel d'écouter en stéréophonie. La monophonie est désormais présentée dans la littérature audiophile comme une écoute inadaptée et non naturelle face aux nécessités psychoacoustiques qu'entraine l'écoute humaine. La notion de *Staging* est intro-

duite par Paul Théberge, Hyle Devine et Tom Everrett comme étant « une notion qui peut être prise pour indiquer à la fois la possibilité d'un référent externe (mais un qui n'est pas, en soi, pleinement synonyme de « réel ») et le processus associé à sa représentation technologique. 12 ». En d'autres termes, cette notion recouvre un état d'esprit où l'on cherchait à tout prix à reconstituer, via la stéréophonie, une scène sonore réaliste, correspondant aux expériences que l'on pourrait avoir au théâtre, à l'opéra, etc. Cette notion s'accorde particulièrement bien avec les conventions de projection en salle de cinéma, fortement liées à un héritage théâtral avec des spectateurs assis et orientés face à l'écran; le son apparaît alors comme une scène sonore qui proviendrait de l'écran. Cela continue à être vrai, bien que les nouveaux systèmes de diffusion évoluent vers une augmentation radicale du nombre de haut-parleurs arrière comme nous le rappellent les auteurs :

« Aussi, alors que l'écran cinématographique n'est pas une « scène », il est cependant situé dans un contexte théâtral : à l'ère de la télévision, d'internet et de la visualisation mobile, la représentation théâtrale demeure un composant essentiel dans la production et la distribution de films. Dans la pratique conventionnelle de la représentation cinématographique, la configuration des places assises pour les spectateurs face à l'écran et [...]à grande échelle ceci exige que la projection du son dans le cinéma préserve la spécificité de celui d'une scène de théâtre. 13 »

¹² Paul Théberge, Kyle Devine et Tom Everrett (dirs.), Living stereo: histories and cultures of multichannel sound, New York, Bloomsbury Academic, 2015, p. 5. Traduit, par nos soins, de l'anglais: « a notion that can be taken to indicate both the possibility of an external referent (but one that is not, in itself, wholly synonymous with the « real ») and the process associated with its technological representation ».

¹³ Ibid., p. 12. Traduit, par nos soins, de l'anglais : « So while the cinematic screen is not a « stage », it is nevertheless situated in a theatrical context : in an age of television, internet and mobile viewing, theatrical exhibition is still a core component in film production and distribution. Within conventional cinematic exhibition practice, the configuration of audience seating forces a front-facing orientation to the screen and [...]to a large degree this requires that sound projection in the cinema retain the character of a sound stage »

Entre 1946 et 1954, le cinéma connait un déclin dû à l'arrivée de la télévision dans les foyers. Le nombre de spectateurs chute de moitié et l'industrie cinématographique semble activement rechercher un moyen de solliciter l'envie d'aller au cinéma. Entre 1950 et 1970, plusieurs formats de diffusion vont voir le jour sans pour autant persister dans le temps. Parmi ces systèmes, retenons tout d'abord le procédé Cinerama de Todd AO (Cinerama et Todd-AO (contraction de Michael Todd, nom du concepteur du procédé, et d'American Optical, la société d'optique qui exploite le procédé) désignent deux procédés différents mais précurseurs du format 7.1) apparu en 1952 qui propose à la sortie du film *This is Cinerama*¹⁴ la projection d'une image extra large (146° d'ouverture) avec l'adjonction de sept pistes analogiques magnétiques (LLcCRcRLsRs). En réponse, la 20th Century Fox présente en 1953 La Tunique¹⁵ d'Henri Koster en Cinémascope. Ce procédé d'anamorphose inventé en 1926 par Henri Chrétien permet à partir d'une pellicule 35 mm désanamorphosée d'avoir un ratio d'image de 2,35 :1. 4 pistes magnétiques sont adjointes à l'image pour former un LCRS. La caractéristique commune à ces dispositifs est l'élargissement considérable du champ de vision du spectateur. Généralement autour de 60°, cette grande ouverture visuelle sur l'écran implique le mouvement des yeux et de la tête du spectateur. Les ingénieurs de ces systèmes audiovisuels ont considéré judicieux de pouvoir coordonner point visuel et point sonore. Ainsi, durant cette période, bien des systèmes disposent de 5 haut-parleurs derrière l'écran, permettant par conséquent de couvrir toute la largeur de l'écran. En 1976, ces différents procédés vont laisser la place au Dolby Stéréo connu aussi sous le nom de Dolby Surround, qui viendra uniformiser et standardiser la diffusion du son au cinéma. La firme américaine Dolby, connue alors pour ses réducteurs de bruit, invente un système matriciel permettant à partir de deux pistes optiques nommées respectivement Lt Rt, la diffusion d'un mixage LRS. Ce système

¹⁴ Merian C. Cooper, Gunther von Fritsch, Ernest B. Schoedsack et Michael Todd Jr, *This Is Cinerama*, 115 min, États-Unis, 1955.

¹⁵ Henry Koster, *The Robe (La Tunique)*, 135 min, États-Unis, sorti en France le 4 décembre 1953.

innovant permet de solutionner notamment le problème du stockage des pistes audio qui affecte les systèmes multicanaux précédemment cités. En effet, seules deux pistes sons peuvent être écrites en marge du photogramme. Ce système permet donc, à partir d'un nombre de pistes réduit, de bénéficier de plusieurs canaux de restitution. En 1982, le Dolby Surround Pro Logic apparaît pour compléter le décodage matriciel du Dolby Surround. Le décodeur Dolby Pro Logic vient émuler une voie centrale avant, répondant ainsi à la demande de séparation des voies de droite et de gauche avec une enceinte centrale pour les dialogues. En octobre 1987, à la réunion de la SMPTE, Tomlinson Holman fait une proposition de normalisation du système 5.1 qui sans le savoir va permettre de standardiser l'écoute cinéma jusqu'à aujourd'hui. Malgré tout Mark Kerins nous indique qu'il n'en était pas aussi sûr le jour de la rencontre.

« Aujourd'hui, le son digital « 5.1 » [...] est répandu dans les foyers et presque omniprésent dans les salles de cinéma. Mais à la fin des années 80, lorsque Holman a fait sa proposition à la SMPTE, personne n'était sûr de la forme que prendrait le son digital cinématographique et du succès qu'il rencontrerait. L'industrie du 7^e art, après tout, a vécu une longue histoire d'inventions de technologies du son qui ont juste été vite écartées. ¹⁶. »

En 1989, les premières bases d'un format 5.1 pour le cinéma sont définies par un groupe d'étude de la SMPTE. Ce groupe définit un format numérique avec les données disposées sur le film, entre les perforations (voir *Figure 1.1*) sur un seul côté de la pellicule.

¹⁶ Mark Kerins, Beyond Dolby (stereo): cinema in the digital sound age, Bloomington, Indiana University Press, 2011, p. 19. Traduit, par nos soins, de l'anglais: « Today, digital ''5.1'' sound [...] is commonplace is home and nearly ubiquitous in theatres. But in the late 1980 s, when Holman made his proposal to the Society of Motion Picture and Television Engineers (SMPTE), no one was sure what form cinematic digital [surround] sound would take or how successful it would be. The movie industry, after all, had a long history of introducing new sound technologies only to quickly discard them. »

La pérennisation du 5.1 fut bien sûr encouragée par son utilisation commerciale en 1992 avec l'arrivée du Dolby Digital. Ce procédé permet d'obtenir jusqu'à six canaux numériques totalement discrets. Néanmoins apparaît de nouveau la notion de compatibilité descendante qui, outre la stratégie commerciale qu'elle subodore, implique que le format 5.1, n'étant pas encore le système de référence, doit pouvoir s'adapter aux différentes configurations des salles de cinéma. Pour des raisons de simplicité de mise en œuvre, de coût modéré et de réalité de production, le 5.1 discret était alors matricé automatiquement en Lt Rt et imprimé optiquement sur la pellicule.

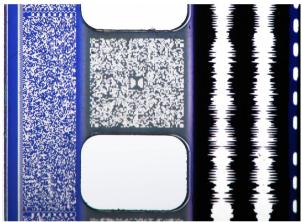


Figure 1.1 : Photogramme d'une pellicule comportant respectivement de gauche à droite, le format numérique SDDS, le format Dolby Digital et enfin le Dolby Surround SR

1.1.2 Insertion du 5.1 et ses dérivés

Comme beaucoup d'autres technologies, le processus de création du 5.1 fut parsemé de succès et d'échecs, de progrès et de régressions qui écartent une approche seulement historique et technologique du développement du système. Le montage son analogique, fastidieux dans son processus d'élaboration, ne permettait qu'une édition limitée de plusieurs sons simultanément. Le savoir faire du créateur passait essentiellement par le choix du son.

L'apparition de machines de montage numérique intégrant un système de stockage direct to disk tel que le Akai DD 1500 permet le montage simultané de 8 pistes audionumériques. Ces machines peuvent également être reliées entre elles afin d'augmenter encore le nombre de pistes pouvant être jouées simultanément. Bien que ce nombre paraisse ridicule au regard des centaines de pistes gérées actuellement par les stations de montage, cela augmente alors grandement les possibilités créatives des monteurs son. Cette station de montage multipistes n'est pas anodine dans l'apparition du système 5.1. En effet, la numérisation du son augmente le nombre de pistes discrètes tandis que le système 5.1 demande encore plus de canaux. La pérennisation de ce système n'aurait pas pu se faire sans la création de stations de montage pouvant l'alimenter. Cette logique d'alimenter les canaux d'un système est désormais admise et dans le monde professionnel elle est entrée dans les mœurs comme le modèle basé canal (channel-based en anglais).

En 1993, DTS (Digital Theater System) et Sony se joignent à cette évolution numérique en proposant le DTS 5.1 et le SDDS 7.1. Entre 1993 et 2006, DTS et son concurrent Dolby, ne cessent de faire évoluer le 5.1 en y ajoutant des canaux selon leurs méthodes de matriçage et d'encodage propres. Les années 2000 sont marquées pour Dolby par une réédition de leur système Pro Logic en Dolby Surround, Pro Logic II, Pro Logic IIx, Pro Logic IIz. Ces systèmes à matriçage permettent à partir d'un programme stéréophonique de reconstituer des canaux audio supplémentaires à ceux déjà préexistant, par exemple le Dolby Pro Logic IIz crée deux canaux surround en élévation à partir d'un format discret 7.1.

Bien que plusieurs compagnies aient développé des systèmes répondant aux normes fixées par la SMPTE, c'est le Dolby Digital qui apparaît comme le plus attractif aux yeux des producteurs et des distributeurs, et ceci en grande partie grâce à son procédé de rétrocompatibilité.

Ces dix dernières années témoignent d'un remaniement du mode de consommation de l'œuvre audiovisuelle. Les grandes avancées dans le domaine des nanotechnologies ont totalement redéfini les médiums d'accès à l'objet cinématographique. La démocratisation d'appareils mobiles (téléphones, ordinateurs portables, tablettes, etc.), combinée avec les nouvelles capacités d'échange de données d'internet, donne accès à des flux audio et vidéo un peu partout et n'importe quand dans le monde. Comme nous l'avons remarqué, la vulgarisation de la TV dans les années 50 a constitué une forte concurrence pour le cinéma, mais sans parvenir à le détrôner et cette conjoncture a même constitué un vecteur d'évolution dans son expression et dans sa forme. La TV elle, bien que très dépendante du contenu cinématographique dans ses débuts, s'éloigna rapidement des salles de projection pour créer un contenu propre au médium et à sa manière de le consommer. Il apparaît aujourd'hui, avec l'amélioration des installations visuelles et audio chez le particulier, que la consommation du cinéma ne se fait à nouveau plus en majorité dans les salles obscures. La démocratisation des supports DVD, ou récemment du Blu-Ray disc permet une qualité de restitution du film honorable voire supérieure (pour le Blu-Ray) au standard de diffusion en salles. Bien que leur fréquentation soit en légère hausse entre 2013 et 2016¹⁷, il est évident que les salles de cinéma ne sont plus l'unique lieu de consommation du film. Le cinéma se joue directement chez soi, ou ailleurs, et ce sur des médiums allant du téléviseur équipé ou non de Home Theater à l'ordinateur personnel. De nos jours, très rares sont les personnes ne visionnant des films qu'au cinéma. L'œuvre cinématographique a maintenant plusieurs vies et son existence est conditionnée par l'industrie et la société de consommation dont elle dépend.

Il est difficile dans le cadre de ce mémoire d'évaluer les tendances à venir, néanmoins notons celle des industriels qui se dirigent à nouveau vers un renouvellement technologique stratégique. On observe alors l'émergence de la vidéo haute-

¹⁷ Propos recueilli sur le site internet du CNC, onglet statistiques, rubriques fréquentations des salles.

définition, la résurgence de la 3D accompagnée de l'augmentation de la cadence image HFR (*High Frame Rate*) et de sa dynamique. En contraste à ces innovations technologiques de l'image, le son reste attaché au format développé dans les années 1980. La réaction est tardive chez les designers de systèmes audio. D'autant plus que le passage au cinéma numérique relance le marché par les nouvelles technologies et ravive l'intérêt pour la spatialisation du son.

Face à cette demande, plusieurs fabricants proposent leurs systèmes en prenant pour base le standard multicanal 5.1 et en s'appuyant sur les capacités grandissantes des processeurs. Autant dire que, sur le papier, l'évolution n'est pas flagrante vu que les propositions gravitent autour de l'ajout de canaux. Il est assez aisé voire tentant d'augmenter quantitativement le nombre de haut-parleurs, afin d'améliorer qualitativement l'immersion du spectateur. Chacun de ces systèmes présente des intérêts, notamment sur le rendu qui, grâce à un foisonnement de haut-parleurs, permet d'avoir des sources sonores plus stables et précises. Cependant, ces évolutions restent basées sur le principe de stéréophonie et ne proposent pas d'alternative aux défauts inhérents à ce type de technologie. Plusieurs designers ont proposé leurs innovations de procédés de restitution. Citons en premier lieu Tomlinson Holman, à l'origine de la dénomination précédente de 5.1, qui propose avec le 10.2 un dérivé de celui-ci. Ce principe de restitution comporte deux haut-parleurs Wide Left et Wide Right ajoutés à côté des enceintes L et R, une paire de haut-parleurs en hauteur (45° par rapport à l'axe de l'écran), une paire de renforts de basse surround et une enceinte surround centrale arrière.

La compagnie allemande Barco propose elle, l'Auro 3D comme nouveau format immersif. Ce système consiste à rajouter des sources sonores indépendantes au système de diffusion classique installé dans les salles de cinéma. Dans une installation Auro 3D, chaque haut-parleur issu d'un système de diffusion 5.1 traditionnel trouvera son homologue en vertical, ainsi le système de base double dans le sens de

la hauteur. Ce système présentant des variantes allant du 9.1 au 13.1 propose une dimension verticale en rajoutant à ses canaux en hauteur des sources issues du plafond.



Figure 1.2 : Configuration Auro 3D 11.1 dans une salle de cinéma.

Au Japon, un format très similaire et proposant aussi une restitution verticale des sources sonores prend la dénomination de NHK 22.2. Premièrement dévoilé durant le salon annuel de l'AES en 2004, cette structure développée pour accompagner le nouveau standard de diffusion cinéma 8 K Super Hi-Vision, est composée de trois niveaux dont nous invitons le lecteur à prendre connaissance dans la Configuration Imm Sound 23.1. Les chiffres et flèches en rouge désignent respectivement les canaux de la couche horizontale ainsi que les enceintes qu'ils alimentent. Le

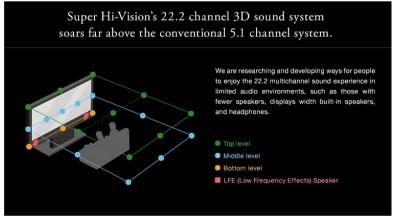


Figure 1.3 : Schéma de composition du système de diffusion 22.2. Source www.nhk.or.jp

principe reste le même pour la couche supérieure, représentée en vert, et celle inférieure représentée en jaune et orange (*Figure 1.3*).

1.1.3 Nouvelle façon d'appréhender le son spatial : le mixage orienté objet

Dans l'approche orientée objet, le mixage des différentes sources sonores du film est soumis à un processeur de diffusion qui viendra calculer, en fonction de la configuration de la salle, le signal qu'il doit envoyer à chaque haut-parleur. Différents systèmes en utilisent le principe. Les technologies employées pour l'Imm Sound et le Dolby Atmos sont décrites en Annexe D.

Outre atlantique, Karlheinz Brandenburg, connu aussi pour sa contribution à la création de la norme de réduction de débit (compression) MP3, propose le système Iosono. Ce système, à mi-chemin entre les techniques de sonorisation du spectacle vivant et du cinéma, utilise la technologie WFS (*Wave Field Synthesis*), ou de synthèse par front d'onde en français, afin de reproduire le champ sonore. Cette technique développée par A.J. Berkhout, de D. de Vries et P. Vogel¹⁸, s'appuie sur le principe physique de Huygens, qui suggère la décomposition du champ sonore en une multitude de sources ponctuelles. Les haut-parleurs multiples, agissant comme des sources ponctuelles permettent la restitution du champ sonore par synthèse du front d'onde. Notons tout de même que la WFS nécessite un grand nombre d'enceintes pour avoir une reproduction du champ sonore acceptable. Dans sa version actuelle, le système allemand Iosono se base sur le rendu de 64 sources sur 128 canaux au maximum¹⁹.

¹⁸ A. J. Berkhout, D. de Vries et P. Vogel, *Acoustic control by wave field synthesis*, The Journal of the Acoustical Society of America, 1 mai 1993, vol. 93, no 5, pp. 2764-2778.

¹⁹ Bergame PÉRIAUX, Jean-Luc OHL, Patrick THÉVENOT et Florian CAMERER, Le son multicanal: de la production à la diffusion du son 5.1, 3D et binaural, Paris, Dunod: INA, 2015, p. 355.

Notons l'existence en France, d'un système concurrent développé par Sonic Emotion, dont l'originalité et les solutions qu'il propose par rapport aux systèmes précédemment cités, ont fait l'objet d'un mémoire à L'ENS Louis-Lumière²⁰. Etienne Corteel, directeur scientifique de la société Sonic Emotion, propose dans son système de WFS adapté au cinéma une version allégée du nombre de haut-par-leurs par rapport à ses concurrents. Rémi Carreau et Thibaut Macquart nous rappellent que la WFS ne nécessite pas « une organisation équivalente entre la production et la restitution du contenu²¹ ». En effet, la WFS se distingue des formats dérivés du 5.1 par une utilisation du mixage orienté objet. Ce procédé, dont nous ferons plus tard l'étude approfondie, se construit à l'opposé du mixage orienté canal qui, dès la phase de production, nécessite une réflexion sur la configuration de diffusion du film (5.1, 7.1, etc.). Dans les technologies orientées objet, le mixage des différentes sources sonores du film est soumis à un processeur de diffusion qui viendra calculer, en fonction de la configuration de la salle, le signal qu'il doit envoyer à chaque haut-parleur.

Imm Sound

Du côté américain, plus particulièrement chez Dolby, la recherche d'un nouveau système a été plus tardive. Selon Paul Théberge, Hyle Devine et Tom Everrett²², le point de bascule se situe en 2008 où AMC Theaters, l'une des plus grandes chaînes de salles de cinéma des Etats Unis, demande à Dolby Labs de penser le futur format multicanal. AMC, par ce biais, recherche à transposer au son ce qui a été fait avec l'image 3D, c'est-à-dire, tenter de relancer la vente de tickets. La réponse de Dolby est le Dolby Surround 7.1, qui comme cité plus haut, se base sur la structure existante du 5.1 en séparant les surround en quatre zones : gauche, arrière

²⁰ Rémi Carreau et Thibaut Macquart, *Utilisation de la technologie WFS dans la création sonore cinématographique : posibilités et limites*, Mémoire de Master, ENS Louis-Lumière, 2015.

²¹ Ibid., p. 67.

²² Paul Théberge, Kyle Devine et Tom Everrett (dirs.), Living stereo, op. cit., p. 230.

gauche, droit, arrière droit. Le premier film projeté dans ce format est *Toy Story 3*²³, et bien que ce procédé ne soit pas révolutionnaire, il engage néanmoins un processus de recherche dans les Laboratoires Dolby. En 2012, Dolby officialise le rachat de la compagnie Imm Sound et la sortie de son nouveau format le Dolby Atmos.

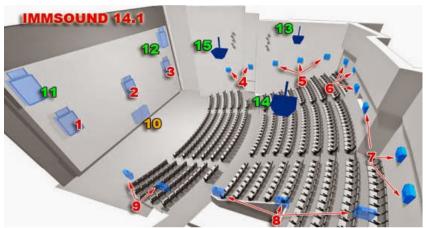


Figure 1.4 : Configuration Imm Sound 23.1. Les chiffres et flèches en rouge désignent respectivement les canaux de la couche horizontale ainsi que les enceintes qu'ils alimentent.

Etant donné l'implication du système Imm Sound²⁴ dans la création du Dolby Atmos, et au regard de l'étude approfondie que nous développerons au sein de ce mémoire, nous vous proposons un survol non exhaustif du système de diffusion espagnol afin d'introduire ensuite le Dolby Atmos. Bien entendu, le rachat d'Imm Sound par Dolby ne permet pas de rentrer dans les détails techniques des algorithmes de calcul du format audio orienté objet mais aussi de l'ambisonie d'ordre supérieur et des calculs basés sur la perception spatiale du son, que l'Imm Sound est censé utiliser dans son système ; ces caractéristiques techniques étant en effet sous la propriété intellectuelle détenue par Dolby. Néanmoins, nous proposerons des hypothèses des diverses techniques mises à contribution dans le Dolby Atmos.

²³ Lee Unkrich, Toy Story 3, 103 min, États-Unis, sorti en France le 14 juillet 2010.

²⁴ Toni Mateos, Vicente López et Alex Artigas, *Immersive 3D sound for cinema*, Documentation de présentation technique Immsound, 2012.

Le procédé Imm Sound, dans sa version la plus complète en 23.1, permet une diffusion sonore avec dimension verticale selon 3 couronnes²⁵ d'enceintes schématisées dans la *Figure 1.4*.

Outre l'extension zénithale des sources, ce système se démarque de ceux cités plus haut non seulement par son augmentation du nombre de voies latérales et arrières mais surtout par l'ajout du processeur de diffusion Imm 3DSP qui vient calculer le signal audio qu'on lui envoie en fonction de la configuration de restitution dont il dispose. À la différence des configurations orientées canal où le nombre de hautparleurs arrière dépend de la salle, l'installation Imm Sound garde un nombre de haut-parleurs fixe. Notons qu'il existe une version 14.1 pour les salles de plus petite dimension. Ces emplacements étant renseignés dans le processeur de diffusion Imm 3DSP, celui-ci permet un placement des objets sonores censé être plus précis que celui d'un système orienté canal classique. L'imm Sound et l'approche mixage objet qu'il utilise proposent de se dégager de systématisme où un canal surround alimente une rangée de haut-parleurs. Ici le processeur de diffusion va rendre la position de la source sonore en fonction de la position des haut-parleurs dans la salle. Remarquons néanmoins, que la notion d'orienté canal n'est pour autant éloignée étant donné que le canal 8 dans la Figure 1.4 alimente parfois jusqu'à trois haut-parleurs.

Au-delà du gain en précision commercialement repris dans son dossier de présentation²⁶, le dispositif Imm Sound a la particularité d'être compatible avec les formats 5.1 et 7.1. Ceci via un calcul du processeur et par, nous l'avons vu, une configuration électroacoustique similaire sur l'axe horizontal de l'écran. Imm Sound propose ainsi de solutionner le problème des downmixages de plus en plus récurrent dans le domaine professionnel. Le concepteur espagnol propose, à partir

²⁵ Bergame Périaux, Jean-Luc Ohl, Patrick Thévenot et Florian Camerer, *Le son multicanal*, *op. cit.*, p. 40.

²⁶ Toni Mateos, Vicente López et Alex Artigas, « Immersive 3D sound for cinema », op. cit.

d'un mixage effectué sous leur système, de laisser le processeur calculer les différents formats de diffusion 14.1, 7.1, 5.1 et stéréo. Ce principe de déclinaison automatique de downmixage fut amorcé chez les mixeurs par l'utilisation de logiciels permettant notamment le passage du 5.1 à la stéréo. Il sera repris par Dolby sous le slogan commercial « Créez une fois, optimisez partout²⁷. », fidèle à leur philosophie de rétrocompatibilité amorcée, comme vu précédemment, dès l'évolution de leur système de restitution. Cette philosophie ayant pour but d'effectuer un passage progressif d'un système ancien à sa version récente, sans pour autant discriminer les salles ayant un dispositif non à jour. Au-delà des apparences, il paraît évident qu'il s'agit aussi d'une stratégie commerciale afin de pouvoir faire évoluer leur système tout en conservant l'embargo du marché.

Dolby Atmos

En 2012 Dolby Labs propose avec le Dolby Atmos, un nouveau format de mixage, d'encodage et de restitution sonore dit 3D. Par l'utilisation de ce terme, Dolby introduit une dimension verticale de restitution sonore. Au premier abord, le Dolby Atmos semble proposer une version complétée du système Dolby Surround 7.1. En effet, comme le suggère la *Figure 1.5*, on retrouve les trois voies LCR installées derrière l'écran et un réseau de haut-parleurs surround répartis en 4 zones (latéral, gauche, arrière gauche, arrière droit, latéral droit) issues du 7.1. Interviennent alors les réseaux d'enceintes acoustiques complémentaires du Dolby Atmos décrite dans la *Figure 1.5* et approfondie en Annexe B.

²⁷ DOLBY LABORATORIES INC., *Un son de prochaine génération pour le cinéma*, Documentation de présentation technique, 2012, Consultable sur www.dolby.com, p. 6.

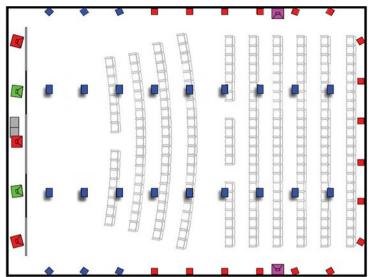


Figure 1.5 : Configuration des haut-parleurs du système Dolby Atmos. En rouge, l'emplacement des haut-parleurs préexistant aux normes 5.1 et 7.1. En vert, les haut-parleurs optionnels.

Néanmoins, sous ces aspects de 7.1 amélioré, la notion de zone centrale d'écoute se substitue à celle du *sweet spot* propre aux dispositifs orientés canal. Cette zone centrale d'écoute va modifier l'emplacement des enceintes surround notamment dans l'alignement de leur axe acoustique. Mais la particularité de ce système ne semble pas résider seulement dans l'ajout d'enceintes suspendues. Comme la WFS et le format proposé par Imm Sound, le Dolby Atmos utilise le modèle orienté objet, qui permet entre autres une configuration plus libre du nombre de haut-parleurs. Le document de présentation de Dolby²⁸ laisse entrevoir que la nouveauté se situe aussi dans la redéfinition des caractéristiques des haut-parleurs utilisés, permettant une augmentation notable du niveau sonore supportable par une enceinte.

²⁸ DOLBY LABORATORIES INC., « Un son de prochaine génération pour le cinéma », op. cit.

1.2 APPARITION DU CONCEPT DE MIXAGE ORIENTÉ OBJET

1.2.1 LES BASES DU MIXAGE ORIENTÉ OBJET

Historique:

Les premières expériences de son multicanal commencent chez les exploitants aux alentours de 1927, qui essayent d'inventer des systèmes pouvant envoyer différents sons dans différentes enceintes. Leur intention se différencie de celle de Bell Labs, qui, à la même période, élabore un système multipiste, dont chaque piste est orientée vers une enceinte. Les exploitants tentant de reproduire un univers sonore théâtral cohérent, ont équipé leurs cinémas d'enceintes derrière l'écran mais aussi au niveau de ce qui s'apparente à la fosse d'orchestre. Ainsi, l'opérateur a alors pour tâche d'envoyer les dialogues et les sons narratifs derrière l'écran et la musique d'orchestre en fosse. Cette opération est réalisée à partir d'une seule et unique piste. Bien entendu, la difficulté de manipulation pour l'opérateur n'a pas permis au système de perdurer.

Néanmoins, cet épisode anecdotique pourrait constituer en quelque sorte la base du mixage orienté objet. D'autres facteurs ont par la suite accéléré sa création. Pardelà les années, l'industrie cinématographique, les exploitants et les industriels du son au cinéma ont promu petit à petit un nombre important de canaux s'accompagnant d'une installation matérielle de plus en plus lourde, coûteuse et imposante. Cela commence avec le bien connu 5.1 qui, par sa configuration électro-acoustique dans les salles mène au 7.1 qui n'en est qu'une simple amélioration. D'autres formats surenchérissant le nombre imposant de canaux voient le jour comme le NHK 22.2. Le résultat de ce combat technologique de la part des industriels entraine pour les techniciens du son un nombre encore plus important de canaux à

gérer et du travail en plus pour adapter leur ouvrage à chacun de ces nouveaux formats. Bien que les différentes firmes du son au cinéma se soient relancées dans une course au nombre de canaux, il semble que la solution se trouve à l'opposé de ce que nous connaissons aujourd'hui.

Présentation:

• Channel-based:

Il y a actuellement trois manières et techniques pour envisager la spatialisation dans le domaine de la diffusion cinématographique. Tout d'abord le modèle orienté canal²⁹, dont nous avons décrit précédemment les principaux fondements. Dans ce modèle, le nombre de canaux employés détermine la configuration de son système de restitution. Le signal sonore est donc conçu pour une enceinte placée à un endroit précis dans la salle de cinéma. Le modèle orienté canal est le plus commun des modèles grâce à sa simplicité d'utilisation et sa transparence de fonctionnement. Cela semble fonctionner honorablement lorsqu'on travaille la matière sonore en ayant connaissance du type de diffusion (stéréo, 5.1, 7.1) comme le suggère la *Figure 1.7*. Par contre, il arrive que pour un type de diffusion donné, le nombre de haut-parleurs alimentés par un canal varie. C'est le cas notamment dans les salles de cinéma. Dans ce cas-là, les informations spatiales qui étaient satisfaisantes en auditorium semblent se distordre en salle de cinéma. Pour autant, le modèle orienté canal paraît avoir fait ses preuves pour des programmes où la spatialisation du contenu n'est pas un élément essentiel.

²⁹ Les anglo-saxons utilisent le terme de *speaker-feed* qui résume assez justement la manipulation sonore des techniciens audiovisuels quant à l'utilisation de tels systèmes.

• Scene-based:

Il existe un deuxième concept principalement développé par Michael Gerzon qui est celui de l'holophonie. Celui-ci se base sur la description et la décomposition du champ acoustique en harmoniques. Cette décomposition du champ acoustique s'effectue principalement à l'aide d'harmoniques sphériques. Ces harmoniques vont de 1 à 5 pour l'ordre le plus élevé. Cette conception de champ acoustique s'éloigne de la première qui ramenait le son à une source acoustique. Ce système est connu sous le nom de HOA pour *High Order Ambisonic*. L'HOA est une approche alternative pour capturer et rendre un champ sonore complexe en deux ou trois dimensions. La manipulation et la diffusion d'un tel signal requiert une simple matrice qui calcule le signal selon la configuration électroacoustique employée. En effet, cette technique encode le son en différentes composantes spatiales suivant les axes x, y et z d'un repère cartésien. Les signaux d'un format ambisonique n'alimentent pas directement des haut-parleurs mais contiennent tout de même l'information directionnelle d'un champ sonore.

• *Object-based*:

Le modèle orienté objet³⁰ a été proposé par TUD (*Technology University of Delft*); ses bases en Europe ont été amenées par le projet Carrouso³¹. Ce projet s'appui sur la norme MPEG-4 BIFS (Cf. annexe A) utilisée aujourd'hui pour représenter les métadonnées audio. Ces métadonnées contiennent des informations spatiales géométriques qui accompagnent l'information audio brute. Le format orienté objet décrit donc un champ spatial en termes d'événements audio joués à des positions spécifiques de la salle. En ce sens il diffère de la conception classique des salles de cinéma où le son est spatialisé par rapport au spectateur (système égocentrique), et propose un système allocentrique basé sur la spatialisation du son dans l'auditorium

³⁰ Object-based description en anglais.

³¹ Sandra Brix, Thomas Sporer et Jan Plogsties, « CARROUSO-An European approach to 3D-audio », Convention Paper 5314, Presented at the 110th AES Convention, Amsterdam, 2001.

(voir *Figure 1.6*). L'approche orientée objet s'appuie sur trois points caractéristiques :

- Position de la source sonore : coordonnées spatiales (cartésiennes et/ou polaires)
- Caractéristiques de la source sonore : taille (étalement), directivité.
- Paramètres de la salle de diffusion : taille de la salle, réverbération, etc.

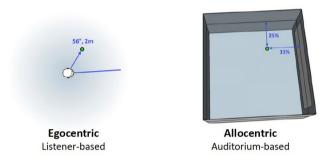


Figure 1.6 : Illustration de la différence entre le modèle égocentrique à gauche et allocentrique à droite. Source SMPTE

Il s'agit donc d'un modèle adaptable qui ne fait référence à aucune position de haut-parleur particulière; en ce sens ce modèle est censé s'adapter à plusieurs configurations correspondant à des systèmes de diffusion basés sur le modèle orienté objet (WFS, Dolby Atmos, DTS MDA). Ce système adaptatif offre³² à l'utilisateur une certaine souplesse quant au choix de son installation audio. En effet, le système choisi par l'utilisateur tentera par des algorithmes de restitution de rendre l'information audio en concordance avec les informations spatiales qui lui sont associées. Le modèle orienté objet inclut des données de taille et de localisation qui permettent de restituer avec plus d'exactitude un contenu audio. Le contenu audio est représenté comme des objets audio contenant le contenu audio pur, accompagné

³² Ajoutons tout de même que selon la configuration audio que l'on utilise, le rendu du programme peut énormément changer si on réduit le nombre de points source au strict minimum.

de métadonnées décrivant la position de l'objet audio et ses propriétés (directivité, taille, etc.) en temps réel.

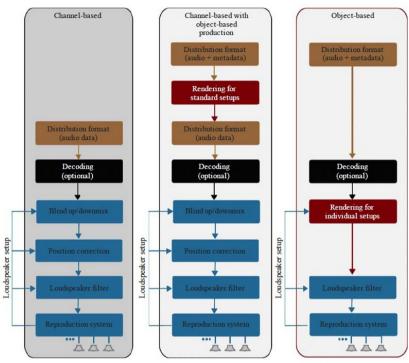


Figure 1.7 : Schéma comparatif des différentes possibilités de flux audionumériques depuis l'arrivée des formats orientés objet. Source ITU-R BS.2159-4

Volonté de compatibilité entre formats :

De nos jours, la quasi-totalité des productions audio mène à un mixage au format multicanal. Selon les formats utilisés, la compatibilité se fait complexe. On trouve néanmoins des solutions logicielles permettant de *downmixer* ou de *upmixer* des contenus d'un format à un autre. Très tôt dans l'histoire de la diffusion au cinéma, se pose la question de la compatibilité d'un contenu audio spatialisé avec les différents systèmes de diffusion installés dans les salles. Ce problème a été solutionné industriellement par le développement de systèmes d'encodage et de décodage propriétaires propres à chaque firme telles que Dolby, DTS, SDDS.

Aujourd'hui, ce qui était avant le lieu de combats industriels entre les grandes firmes semble devenir un argument d'avenir. L'approche objet est sujet de recherches et d'intégration dans des normes techniques telles que la norme MPEG4, MPEG7 ou encore le MPEG-H 3D. Cette dernière norme propose un système de codage et de décodage permettant une configuration électroacoustique flexible chez l'utilisateur. Les formats audio 3D fleurissant et les installations domestiques étant très hétérogènes, la norme MPEG-H Audio 3D propose de rendre un univers audio 3D sous des environnements moins sophistiqués que ce que nécessitent les formats qu'elle encode. La *Figure 1.8* présente le schéma de principe du décodeur MPEG-H Audio 3D illustrant les principales parties du système décrite en Annexe G.

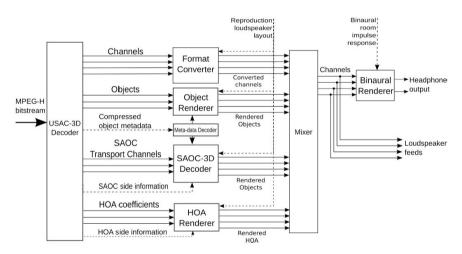


Figure 1.8 : Schéma de principe du décodeur MPEG H Audio 3D.

On retrouve cette approche dans la sonorisation WFS dans le spectacle vivant, et depuis peu dans la diffusion cinématographique dans des procédés propriétaires tels que Dolby Atmos, DTS MDA (Multi Dimensional Audio), Iosono ou Sonic Emotion. Dans l'approche orientée objet, chaque source est définie par une piste et son information spatiale peut alors être positionnée dans l'espace dans n'importe quel format multicanal de diffusion.

Transportabilité des mixages :

L'idée est que les techniciens du son ne se voient plus contraints d'élaborer une bande son pour un canal en particulier ni pour un format. Chaque piste audio ou du moins chaque son composant le flux sonore d'un programme audiovisuel serait accompagné d'informations adressant le son à une position physique relative à la salle de cinéma dans laquelle le flux sera diffusé. En ce sens, cette façon de penser s'éloigne totalement des règles établies durant presque un siècle de cinéma parlant. De la même manière il n'y a plus un nombre de pistes fixe en relation avec le format auquel il est destiné (le format 5.1 est par exemple composé de 6 pistes audio) mais un certain³³ nombre de pistes audio qui viendront se lire en temps réel, en fonction de la configuration des salles et selon un placement effectué à posteriori par le mixeur. L'association de la piste audio et son flux d'informations combinées, appelées métadonnées, vient créer le terme d'objet.

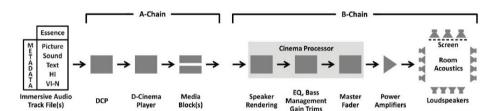


Figure 1.9 : Modification des chaines A et B de traitement du signal audio au cinéma pour des formats audio-immersifs. Source SMPTE.

Le terme d'objet sonore, qui selon la firme Dolby peut être considéré « comme un groupe d'éléments sonores qui partagent la même localisation physique dans l'auditorium³⁴ », peut être statique ou en mouvement. Son positionnement est assuré par des métadonnées qui renseignent la position de la source sonore à un instant t. Cette infrastructure, ayant de fortes ressemblances avec le moteur audio des jeux vidéo, a pour but de modeler le signal audio en fonction des configurations électroacoustiques des salles de cinéma (voir *Figure 1.9*). Ainsi, la tâche complexe

³³ Le nombre de pistes peut varier selon les formats établis par les différentes firmes industrielles.

³⁴ DOLBY LABORATORIES INC., « Un son de prochaine génération pour le cinéma », op. cit.

d'homogénéisation et d'optimisation de la bande son, afin qu'elle soit retranscrite au mieux dans n'importe quelle salle de cinéma, qui incombe habituellement au technicien du son (en l'occurrence, le mixeur) échoit, dans un système orienté objet, à un ordinateur qui vient théoriquement décoder les intentions du mixeur et les retranscrit en fonction de la salle dans laquelle le film se joue. Ainsi, comme le souligne la firme DTS dans son communiqué³⁵, l'intérêt du mixage orienté objet réside dans le fait qu'il permet de se libérer du système de diffusion cible (ou du moins de le rendre plus flexible) tout en conservant une fidélité du contenu audio spatial.

1.2.2 LE DOLBY ATMOS

« l'empreinte actuelle du produit et de la technologie de Dolby, a permis de cibler avec précision les exigences de la prochaine génération du son cinématographique numérique, dans des domaines couvrant aussi bien la conception et l'édition du son que le réenregistrement, le matriçage, le packaging, la distribution et la diffusion dans les salles³⁶. »

Nous tenterons dans cette partie d'informer le lecteur sur les principaux traits technologiques du système Dolby Atmos. Le lecteur remarquera que plusieurs points technologiques sont repris dans en Annexe. Par conséquent un descriptif de la configuration électroacoustique est effectué en Annexe B et en Annexe C (normes de diffusion). L'Annexe D propose un récapitulatif des technologies mis en œuvre dans la spatialisation orientée objet propre au Dolby Atmos. Enfin l'Annexe E proposera une étude des différents outils développé par Dolby dont nous nous servirons par la suite de ce mémoire.

³⁵ DTS INC., MDA Multi-Dimensional Audio, a 3D Audio Solution for Professional Applications, Documentation de présentation technique, 2015, nº 1.1.0.0.

³⁶ DOLBY LABORATORIES INC., « Un son de prochaine génération pour le cinéma », op. cit., p. 5.

La dernière innovation technique de la firme Dolby, le Dolby Atmos, commercialisé à partir de 2012, vise à donner une nouvelle raison au spectateur d'aller au cinéma en recyclant une nouvelle fois le progrès immersif que le système permet. Cette démarche industrielle s'inscrit dans la lignée des projections 4 K, IMAX ou encore projection en 3D. Tout d'abord timide dans son implantation européenne, le système commence à se répandre en France (une cinquantaine de salles sont équipées en France selon Dolby) alors qu'il est bien implanté à l'étranger notamment en Chine et aux États-Unis. Les premiers films notables sortis en salle dans ce format sont Brave³⁷ en 2012 suivi de Chasing Mavericks³⁸ en 2012 également. La première production française à s'aventurer dans ce format est Taken 2³⁹. Un an plus tard, Gravity⁴⁰ est un des premiers films à retenir notre attention quant à son utilisation du système Dolby Atmos. Entre 2015 et 2016, 64 films sont mixés dans ce format dont *The Revenant*⁴¹ qui propose une utilisation subtile du dispositif, laissant entrevoir un avenir bien plus prometteur que ce que l'on pourrait penser de prime abord. Ce système, créé en collaboration avec la compagnie Espagnole Imm Sound, propose une diffusion dite de 3D audio immersive qui se détacherait d'un mixage figé que les systèmes de diffusion classiques 5.1 ou 7.1 nécessitent.

³⁷ Mark Andrews, Brenda Chapman et Steve Purcell, *Brave (Rebelle)*, 93 min, États-Unis, sorti en France le 1er août 2012.

³⁸ Michael APTED et Curtis HANSON, *Chasing Mavericks*, 116 min, États-Unis, sorti en France le 28 novembre 2012.

³⁹ Olivier Megaton, Taken 2, 92 min, France, sorti le 3 octobre 2012.

⁴⁰ Alfonso Cuarón, Gravity, op. cit.

⁴¹ Alejandro G. IÑÁRRITU, The Revenant, 156 min, États-Unis, sorti en France le 24 février 2016.

Description Générale :



Figure 1.10 : Schéma de principe de Dolby Atmos. Source www.dolby.com

Ce système multicanal met au maximum à disposition 64 points de diffusion pour les salles de cinéma les plus grandes. En effet, le nombre de haut-parleurs varie en fonction de la taille de la salle ce qui permet d'adapter la configuration selon ses besoins réels. Ces 64 haut-parleurs sont répartis dans la salle suivant un cahier des charges strict élaboré par Dolby.

Un total de 128 pistes audio s'offre aux techniciens du son pour fournir le système de diffusion. Ces 128 pistes audio doivent être assimilées au concept d'objets sonores. Ces pistes sont gérées par un processeur appelé RMU (*Rendering Master Unit*), qui vient adresser les sons aux différentes enceintes, selon les informations données par le mixeur. C'est aussi ce processeur qui, selon la configuration de la salle, viendra calculer de la meilleure manière la restitution du son en accord avec les intentions du mixeur.

Néanmoins, Dolby ne s'affranchit pas du mixage orienté canal car, dans les 128 pistes, 10 d'entre elles sont attribuées à un *bed*⁴² 9.1, terme utilisé par Dolby pour désigner un élément sonore statique ou évoluant lentement. Dans leur document de présentation technique⁴³, certains éléments de la bande son (son ambiant et réverbération) nécessitent d'être diffusés largement, par des plateaux de haut-parleurs. Ces

⁴² Bed dont la traduction française lit fait sens lorsqu'on fait référence au système de diffusion orienté canal 7.1 ou 5.1. En effet, le bed 9.1 de chez Dolby correspond à un système de diffusion 7.1 auquel il est adjoint 2 canaux supplémentaires qui viennent alimenter les rampes de haut-parleurs zénithaux pour former un 9.1.

⁴³ DOLBY LABORATORIES INC., « Un son de prochaine génération pour le cinéma », op. cit.

beds permettent donc de bénéficier des avantages de l'orienté canal et sa rapidité d'exploitation. Ainsi, 118 pistes audio sont attribuées aux objets sonores et 10 au bed comme le résume la *Figure 1.4*. Ces objets sonores, une fois spatialisés, peuvent être diffusés sur n'importe lequel des points d'émission. Les objets sonores peuvent être regroupés à deux ou à plusieurs en fonction de l'effet recherché.



Figure 1.11 : Schéma de principe d'utilisation des beds et des objets. Source www.dolby.com

Ce système propose donc une ambivalence, permettant à la fois de rester sur un mode de diffusion orienté canal, afin de construire un environnement ambiant et de venir le préciser par des sons ponctuels grâce au mode orienté objet (voir *Figure 1.11*).

1.2.3 LE DTS MDA: MIXAGE EN KIT

Pour rester dans la course infernale des systèmes de diffusion audio 3D, DTS lance en avril 2015 son nouveau format : le DTS MDA. Ce système tente de se démarquer de ses concurrents par le codage *open source* de ses objets audio et la gratuité de sa licence. Dans la même démarche, MDA a été proposé aux comités de standardisation comme format audio immersif alternatif aux systèmes propriétaires fermés comme celui proposé par Dolby. Cette proposition aboutit en avril 2015 à une standardisation du flux audio-immersif auprès de l'organisme ETSI dans son communiqué⁴⁴.

⁴⁴ ETSI, MDA; Object-Based Audio Immersive Sound Metadata and Bitstream, ETSI TS 103 223, 2015, vol. 1.1.1.

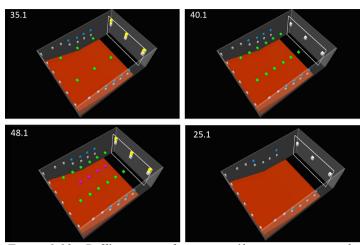


Figure 1.12 : Différentes configurations électroacoustiques du DTS MDA.

Lors de la présentation du DTS MDA au comité de standardisation SMPTE, DR. Ton Kalker (de chez DTS Inc.) dévoile quatre propositions de configuration (Cf. *Figure 1.12*) DTS MDA pour les salles de cinéma. Ces différentes configurations ont été élaborées après étude et généralisation des salles de cinéma en allant de la plus grande (configuration 48.1) à la plus petite (configuration 25.1). Les versions du DTS MDA avec élévation (35.1, 40.1 et 48.1) sont très proches de ce que propose Dolby avec le Dolby Atmos.

Comme ses concurrents, DTS propose avec MDA, un format qui libère les créateurs de bandes son de la préoccupation au sujet de la transportabilité de leur travail sous plusieurs environnements. Grace à leur séparation des éléments audio purs et des informations spatiales, la firme prétend que les intentions créatrices seront préservées et rediffusées de la façon la plus optimale selon la configuration audio. Ce format permettrait une configuration électroacoustique quasi libre en termes de placement et du nombre de haut-parleurs utilisés, et ceci dans des configurations particulières (*Home Cinema*) ou publiques (cinéma). Le DTS MDA s'impose comme un nouveau format d'enregistrement, de transmission et d'algorithme de restitution Audio 3D. Ce format a pour stratégie d'implantation une

approche moins verrouillée que les autres systèmes, notamment en ce qui concerne les installations.

Fonctionnement général :

Bien que les arguments de DTS restent à priori les mêmes que Iosono ou encore Dolby, leur fonctionnement diffère légèrement.

Les objets-audio sont décrits dynamiquement comme des points (étendus ou ponctuels) source positionnés dans l'espace à une position spatiale précise échantillonnée dans le temps. Le nombre d'objets audio n'est pas limité. Ces objets sont encapsulés en PCM pour le flux audio et en MDA bitstream pour la position spatiale. Ce dernier conteneur renseigne aussi le type de contenu audio (dialogue, musique, effet, etc.) et d'autres propriétés comme le gain (voir Figure 1.13). Ces objets audio identifiés, l'utilisateur a un total contrôle de son contenu ; il peut ainsi gérer le niveau sonore du contenu, désactiver la piste, etc. Bien que l'ouverture du contenu audio à l'utilisateur puisse sembler nuisible à l'œuvre cinématographique, il peut permettre néanmoins des traitements audio par catégorie d'objet, ce qui peut trouver un réel intérêt pour les configurations domestiques. Par traitement audio, nous entendons la norme Loudness qui a pour but le contrôle précis des niveaux et de la dynamique audio des programmes destinés (pour le moment) à l'utilisation domestique. Cette tendance à donner accès à l'œuvre audiovisuelle nous semble préoccupante dans le domaine du cinéma où les logiciels de création de DCP permettent déjà, avec la norme MPEG-H 3D, un contrôle des niveaux sonores par objet.

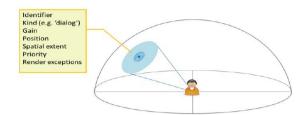


Figure 1.13: Description des objets-audio dans le Format DTS MDA. Source MDA White Paper.

<u>Intégration du DTS MDA dans la chaine de post-production cinématogra-phique :</u>

Le DST MDA comporte trois parties bien distinctes :

- La partie programme qui est constituée du plugin de panoramique *MDA Creator* et du *Stand Alone MDA Player*. C'est ici que vont être créées les différentes données du MDA.
- La partie flux de données qui se compose de l'audio, des métadonnées et du MDA bitstream.
- La partie traitement qui procède à l'analyse et à la diffusion des sons à partir des informations reçues.

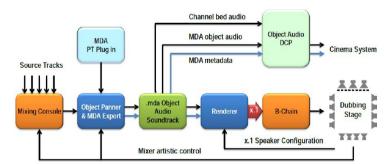


Figure 1.14: Worflow d'une chaine de postproduction DTS MDA. Source MDA White Paper.

1.2.4 LE FORMAT EUROPÉEN ADM

Souvent méconnu de la communauté audio, parce qu'européen, le format ADM⁴⁵ (*Audio Definition Model*) est un format ouvert de description d'informations audio 3 D (*channel based*, *object based* et *scene based* (HOA, Cf. Annexe D.5.). À la différence des Américains, les Européens ont opté pour l'unification et la création d'un format audio orienté objet unique. Ce projet fut élaboré en collaboration avec la BBC, France TV, Radio France, Fraunhofer et l'UER (Union Européenne de Radio-Télévision). Ce format se propose comme standard ouvert et totalement gratuit. Dans cette optique, la transparence est à l'honneur, toutes les informations sont accessibles sur le site internet de l'UER⁴⁶ que nous invitons notre lecteur à consulter pour plus d'informations sur ce format. Depuis juillet 2015, le comité de standardisation ITU a standardisé le format ADM et postule pour unifier les technologies de DTS, Dolby et Auro-3D.

Notons que dans cette même démarche l'agence nationale de la recherche finance actuellement le projet Edison 3D qui vise à élaborer des outils et des solutions à la production audio 3D. De ce projet émane bien sûr l'envie d'élaborer un consensus aux différents formats 3D existants : « le projet Edison 3D se propose de réaliser des avancées scientifiques et des outils qui permettront aux acteurs de l'industrie audiovisuelle d'être aussi indépendants que possible des multiples formats à venir. »⁴⁷. Ce projet de recherche, très intéressant dans notre étude, ne pourra pas, hélas, nous aider dans l'élaboration du mémoire, sa parution étant en effet programmée courant été 2016.

⁴⁵ Operating Eurovision and Auroradio EBU, *Audio Definition Model*; *Metadata Specification*, TECH 3364, 2014, no 1.0.

⁴⁶ https://tech.ebu.ch/docs/tech/tech3364.pdf

⁴⁷ Citation issue de l'article de présentation du projet de recherche Edison 3D consultable sur le site : http://www.agence-nationale-recherche.fr/? Projet=ANR-13-CORD-0008.

1.3 LA QUÊTE DE L'IMMERSION ENTRE NOUVEAUTÉ ET RÉSURGENCE

1.3.1 Une volonté de plus de crédulité ?

« Dans la période classique monophonique, l'aimantation spatiale facilite les choses pour légitimer l'immobilité latérale du son : qu'une voiture entre par la gauche ou par la droite, que des personnages traversent l'écran, et qu'en même temps les sons ne fassent pas réellement de même, ne gêne plus personne. Or, rien ne va plus si ces déplacements et ces localisations imaginaires, avec la multiplication des pistes sonores et des haut-parleurs indépendants, deviennent réels, illustrant cette loi du cinéma : plus on est réaliste au niveau perceptif et plus on approche des limites de la croyance ou plutôt de ce que les Anglo-Saxons appellent très bien « the voluntary suspension of disbelief », la suspension volontaire de l'incrédulité⁴⁸. »

Cette notion de suspension volontaire de l'incrédulité, largement répandue dans les premiers films de science-fiction, peut être matérialisée par un seuil où le spectateur entre dans l'univers et l'histoire du film en acceptant les incohérences entre le réel et la narration du film. La connivence du spectateur passe par l'unité cinématographique de toutes les compétences que concentre le film (réalisation, scénario, jeu d'acteurs, tournage, montage, etc.). Ces domaines de compétence étant de plus en plus connus du spectateur, ce qui relevait de la magie du cinéma est reconnu comme relevant de la prouesse technique et le seuil de suspension de l'incrédulité semble se rehausser. Selon nous, les attentes spectatorielles évoluent tant du point vue de l'image que du son, et dans un flagrant parallèle à l'incrédulité de Saint-Thomas, le spectateur veut y croire et a besoin de plus en plus d'indices pour se rattacher à l'histoire. La technologie mise en œuvre lors de la création d'un film ne cesse d'évoluer dans ce sens ; s'amorce alors un combat difficile entre les tech-

⁴⁸ Michel Chion, *Un art sonore le cinema : histoire, esthétique, poétique*, Paris, Cahiers du cinema, coll. *Cahiers du cinéma essais*, 2003, p. 121.

niques d'hyper-réalisme mises en œuvre pour tenter de satisfaire les attentes des spectateurs et le spectateur qui voit sans cesse ces limites repoussées, réévaluant sans cesse son seuil de suspension volontaire de l'incrédulité. Cet hypothétique système de rétroaction en boucle infinie doit cependant être nuancé, car la base de la connivence entre le spectateur et l'histoire racontée reste du domaine de la mise en scène et du scénario et ne doit pas se substituer à la complaisance de belles images et d'un beau son.

Pour autant, l'évolution des systèmes de restitution passe, nous l'avons remarqué, par l'ajout systématique de haut-parleurs. Néanmoins, il est important de notifier à notre lecteur, que tous « les concepteurs qui ont pour rêve ultime de recréer la vie devraient considérer premièrement comment nos yeux et nos oreilles perçoivent la réalité⁴⁹ ». Il est vrai que de nombreuses études ont été menées dans les domaines psycho-visuel et psychoacoustique afin de comprendre comment fonctionnent notre vision et notre audition. Ceci dans le but de recréer artificiellement des sensations réalistes de vision et d'écoute. Pour autant, les recherches dans le domaine de la perception humaine n'ont pas eu le même engouement, surtout lorsque les modalités de vision et d'écoute dans une salle de cinéma ne sont pas les mêmes que dans la vraie vie. En effet, comme le précisent Gérard Pelé et Laurent Millot « tout système qui crée des sources sonores virtuelles et possède en même temps des sources réelles, dispersées et de caractéristiques déterminées, est susceptible d'engendrer des conflits de la perception spatiale⁵⁰ ». D'autre part, en un siècle de cinéma sonore et de divers dispositifs sonores, le cinéma demeure un spectacle frontal, où l'histoire racontée par le réalisateur se détourne rarement de l'écran.

⁴⁹ Ion Allen, « Matching the sound to the Picture », Dolby Technical Papers, Reprinted by permission from « The Proceedings of the AES 9th International Conference : Televison Sound Today and Tomorrow », San Francisco, 1991, p. 1. Traduit de l'anglais par nos soins.

⁵⁰ Laurent MILLOT et Gérard Pelé, « La sensation d'espace dans le spectacle audiovisuel », Rémi Adjiman et Bruno Cailler (dirs.), in *Une architecture du son*, Paris, coll. *Cahiers de Champs Visuels n° 1 & 2*, 2005, p. 284.

Néanmoins, comme le suggère Claude Baiblé, un spectateur dans un dispositif de « géode sonore⁵¹ » peut être charmé par l'immersion « bruissante et enveloppante⁵² ». L'auteur, en traitant du dispositif Dolby Atmos, nous met en garde sur les conditions d'écoutes recouvrant quasiment tout le champ auditif. Celui-ci semble divisé en 2 parties :

- La zone frontale, restreinte quasiment à l'angle du champ de vision, principalement et volontairement écoutée par le spectateur.
- La zone hors champ ou d'ambiance, « entendue mais inécoutée⁵³ » qui vient s'ajouter ou compléter la zone frontale.

Une des conséquences à risques de cette division de la scénographie sonore serait que l'écoute du spectateur soit attirée voire happée « en de multiples azimuts porteurs de sens⁵⁴ » alors que le spectacle cinématographique reste frontal. L'auteur rejoint ensuite Ion Allen à propos de l'adéquation du découpage visuel et sonore. Le cinéma est un art du découpage, qui d'un point de vue visuel propose différents axes et valeurs de plan d'une même scène, provoquant des changements de localisation tranchés et totalement acceptés par le spectateur. Néanmoins, l'idée que le son vienne coller à une réalité sonore induite par l'image et son découpage est paradoxale. L'espace scénique proposé par la succession de plans n'est pas en adéquation avec l'espace réel filmé, marquer les changements d'espace au son signifierait aller à l'encontre de l'espace scénique. D'autre part des changements spatiaux brusques introduiraient une discontinuité temporelle allant à l'encontre de la continuité narrative du film. Claude Baiblé, en ouverture de son article, use de la rhétorique au sujet de l'apparition du Dolby Atmos:

⁵¹ Claude BAIBLÉ, *Atmosphère, Atmosphère, est-ce que j'ai une gueule d'atmosphère*?, Dossier sur la spatialisation du son, Lettre CST 145, 2013, p. 22.

⁵² *Ibid*.

⁵³ Ibid.

⁵⁴ *Ibid*.

« Que peut donc faire le son spatialisé sur des plans toujours centrés, quoique diversement orientés? Peut-on dissocier aussi facilement le point de regard du point d'écoute, sans entrainer des réactions d'orientation quasi-automatiques, réactions ici sans objet, puisque l'écran rectangulaire reste frontalement posé? [...] Que faire des sons latéraux, verticaux, ou dorsaux, lorsqu'ils sont actifs dans la mise en scène? S'agit-il alors de construire seulement une ambiance molle, sans transitoires susceptibles d'éveiller les neurones phasiques de l'attention réflexe? [...] De susciter ou d'éviter les détournements d'attention? De se limiter aux plans séquences, en écartant le découpage classique en plans brefs? À l'évidence les problèmes déjà rencontrés en 5.1 se trouvent élargis avec le nouveau dispositif. 55 »

Bien que ces questions soient légitimes dans leur approche, elles nous semblent peu contextualisées. Il est vrai que certains films proposent des bandes sons simplistes dans leur adéquation spatiale avec l'image, illustrant bien souvent les problèmes cités ci-dessus et le caractère gadget qu'un tel système peut revêtir. Cet aspect gadget tentant trop souvent de se substituer à la mise en scène a longtemps accompagné la réputation du 5.1 et semble aujourd'hui, par lien de parenté, caractériser le Dolby Atmos. Cependant comme Mark Kerins a pu le remarquer concernant l'esthétique du 5.1⁵⁶, une fois cette période de découverte du système révolue, réalisateurs et techniciens (du son) proposent une bande son plus réfléchie, suivant des codes propres à un style. L'utilisation de la zone arrière décontextualisée de la technologie qui l'emploie, semble s'adapter et étayer l'histoire du film. Dans son mémoire de fin d'études⁵⁷, Selim Azzazi résume la problématique de sa réflexion sur le son multicanal : « l'enjeu n'est pas de coller à une réalité sonore induite mais à une narration par des sons ». Cette idée totalement désengagée d'une quelconque technique de restitution, vient répondre selon nous, aux interrogations menées par

⁵⁵ Ibid.

⁵⁶ Mark Kerins, Beyond Dolby (stereo), op. cit., p. 84.

⁵⁷ Sélim Azzazi, *Réflexion sur l'apport de la diffusion multicanal dans l'écriture cinématographique*, Mémoire de fin d'études, ENS Louis-Lumière, 1999.

Claude Baiblé⁵⁸ et étudiées par Michel Chion⁵⁹. Bien entendu, les questions perceptives évoquées par Claude Baiblé sont connues de la plupart des mixeurs et monteurs son et sont même à l'origine de codes et pratiques de mixage à part entière.

À ce sujet nous invitons le lecteur à consulter le mémoire de fin d'études *La diffusion multicanal au cinéma*: *nécessité d'une écriture visuelle et sonore adaptées*⁶⁰. De plus, bien que l'étude soit menée sur le 5.1, elle constitue une approche valable pour les systèmes orientés objet émergents. Par ailleurs, l'étude perceptive du système Atmos dépasse le cadre de notre étude et d'autre part nécessite un recul aujourd'hui inexistant vis-à-vis de sa récente arrivée sur le marché. Néanmoins, le mémoire de fin d'études de Nicolas Moreau, *L'orientation réflexe dans le cinéma multicanal*⁶¹, nous paraît être une base d'étude solide pour une première approche des problèmes de réception des sons au cinéma.

1.3.2 L'image 3D ou le mythe du cinéma total?

« Les textes abondent [...] où les inventeurs n'évoquent rien moins que ce cinéma intégral donnant la complète illusion de la vie, dont nous sommes encore loin aujourd'hui⁶² »

⁵⁸ Claude BAIBLÉ, *Problèmes de spatialisation en son multicanal*, Espaces pluriels, images et sons, septembre 2004, Cahier Louis-Lumière, nº 2, 09/2004.

⁵⁹ Michel CHION, Un art sonore le cinema, op. cit.

⁶⁰ Christophe VINGTRINIER, La diffusion multicanal au cinéma : nécessité d'une écriture visuelle et sonore adaptées, Mémoire de fin d'études, ENS Louis-Lumière, 1997.

⁶¹ Nicolas Moreau, *L'orientation réflexe dans le cinéma multicanal*, Mémoire de fin d'études, ENS Louis-Lumière, 1999.

⁶² André BAZIN, Le mythe du cinéma total et les origines du cinématographe, Paris, Critique n°6, 1946, p. 555.

Lors de la rédaction de ce mémoire, nos études sur les avancées technologiques et l'esthétique de réalité qu'elles tentent de procurer nous ont amené à croiser le célèbre mythe du cinéma total⁶³ d'André Bazin. Celui-ci soutient cette conception idéaliste selon laquelle il y aurait au commencement de toute aventure cinématographique, la volonté de saisir le réel – ou de cloner le réel – et de le restituer dans sa complexité la plus totale. Le cinéma poursuivrait une quête radicale, en cherchant constamment à être à l'image de la vie et la copie fidèle de la nature : « le mythe directeur de l'invention du cinéma est [...] celui du réalisme intégral, d'une recréation du monde à son image⁶⁴ ». Dans l'idéologie d'André Bazin, le futur du cinéma tend vers une captation et une restitution hyper-réalistes globalisantes. Pour lui l'invention du cinéma est avant tout placée sous l'autorité pleine et entière de ce principe de réalisme intégral qualifiant le cinéma d'art total. Néanmoins, comme le souligne Martin Barnier et Kira Kitsopanidou⁶⁵, André Bazin après le visionnage d'essais cinématographiques stéréoscopiques ne semblait pas totalement convaincu de l'évolution de la forme cinématographique vers le relief, bien que les caractéristiques de reproduction du réel pouvaient justifier son discours sur le cinéma total. Cependant, tandis que le principal initiateur de cette pensée semble questionner maintenant les médiums qui la proposent, d'autres théoriciens comme René Barjavel continuent d'en faire l'étude. Dans Cinéma total: Essai sur les formes futures du cinéma⁶⁶, l'auteur souligne que le cinéma approche, avec le relief et l'immatérialité physique de l'image qu'il engage, de sa représentation la plus aboutie, la plus totale.

⁶³ André BAZIN, Qu'est-ce que le cinéma ?, Paris, Éd. du Cerf, coll. Septième art, n° 60, 1985, p. 19.

⁶⁴ Ibid.

⁶⁵ Martin Barnier et Kira Kitsopanidou, *Le cinéma 3-D : histoire, économie, technique, esthétique*, Paris, Armand Colin, 2015, p. 17.

⁶⁶ René Barjavel, Cinéma total: Essai sur les formes futures du cinéma, Paris, Denoël, coll. Romans Français, 1944, p. 51-52.

L'histoire de la stéréoscopie au cinéma ne montre pas pour autant le même engagement que Barjavel, mais s'approche plus de celle de son prédécesseur André Bazin. Comme l'explique avec précision Martin Barnier et Kira Kitsopanidou⁶⁷, l'histoire du cinéma en relief – et ce dès l'invention de la stéréoscopie fixe – est constituée de hauts et de bas, de périodes d'apparitions et de disparitions respectivement associées à des moments de grâce et des moments de désenchantement. Suivant cette représentation, les avis semblent se partager face à la nouvelle vague - ou retour éternel - de cinéma en relief que l'on connaît aujourd'hui. Certains soutiennent qu'elle ne se distingue pas des vagues précédentes, qu'il s'agit d'un gadget essentiellement commercial et éphémère que l'industrie et l'économie cinématographiques ne cessent de remettre au goût du jour. Sylvain Grain fait lui preuve d'optimisme face à cette technologie qui est « le fruit d'un regard réaliste sur l'histoire et le contexte actuel du cinéma 3D en tant que médium de création⁶⁸ ». Cet optimisme est d'autant plus renforcé que la maturité technologique⁶⁹ – souvent remise en cause dans l'histoire du cinéma en 3D – actuelle permet de distinguer cette nouvelle vague des précédentes. D'autre part, le cinéma 3D a pu se populariser récemment avec la standardisation de la copie de distribution numérique ou DCP (Digital Cinema Package), élaborée en juillet 2005 par DCI (Digital Cinema Initiatives) LLC, qui a permis de faciliter la distribution du film en relief dans les salles équipées d'un seul projecteur numérique. Cette projection numérique, maintenant majoritaire, permet de proposer des projections dites 3D que l'on regarde à l'aide de lunettes individuelles. En ce sens, le numérique permet au cinéma en relief de se détacher de la simple curiosité.

⁶⁷ Martin Barnier et Kira Kitsopanidou, Le cinéma 3-D, op. cit., p. 5.

⁶⁸ Ibid.

⁶⁹ Dans maturité technologique, nous entendons le passage en 2006 du cinéma analogique au cinéma numérique. Ce qui a permis de résoudre un problème technique du cinéma 3D en pellicule. En effet, ce dernier étant projeté à l'aide de deux projecteurs, son synchronisme était souvent complexe voire hasardeux. Ce qui a souvent pénalisé la pérennisation du cinéma 3D en raison des maux de têtes qu'il procurait à son spectateur. Avec le numérique le cinéma en relief s'effectue à partir d'un seul projecteur, écartant tout problème de synchronisme.

À l'image de l'âge d'or du cinéma en relief américain, le récent regain d'intérêt du cinéma 3D fait preuve d'une production filmique assez importante et de qualité variable, mais, à la différence des années 50, le mouvement actuel se démarque par la diversité des productions qu'il propose. Notons tout de même que le cinéma 3D subit, tout comme le son 3D, les conséquences des films qu'il propose à savoir un catalogue de films d'animation ou de grand spectacle. Ceci entrainant allègrement le cinéma d'auteur et indépendant à refuser de s'essayer aux images stéréoscopiques. Selon nous le cinéma en relief, à condition qu'il ne meure pas à nouveau, gagnerait à être exploré par tous les styles de cinéma.

1.3.3 Enjeux du son 3D et de l'image 3D

« Le « réalisme » des matières de l'expression cinématographique n'est que le résultat d'un très grand nombre de conventions et de règles, conventions et règles qui varient selon les époques et les cultures. Faut-il rappeler que le cinéma n'a pas toujours été sonore, n'a pas toujours été en couleurs, et que même lorsqu'il le fut, le réalisme des sons ou des couleurs s'est singulièrement modifié au fil des années[.] Or à chaque étape [...], le cinéma n'a cessé d'être jugé réaliste. Le réalisme apparaît alors comme un gain de réalité [...] par rapport à un mode antérieur de représentation. Mais ce gain est infiniment reconductible, du fait des innovations techniques, mais aussi parce que la réalité, elle, n'est jamais atteinte. 70 ».

En vue des différentes tentatives d'introduction du relief dans le cinéma et de son récent assaut dans les années 2000, il est constructif de savoir ce qui distingue cette vague des précédentes, afin d'établir ce qui est de l'ordre de la nouveauté et ce qui relève de la continuité. Réponse qui au premier abord, se situe dans la cohabitation

⁷⁰ Jacques Aumont (dir.), *Esthétique du film*, 3. éd., s.l., Armand Colin, coll. *Armand Colin Cinéma*, 2008, p. 95.

de deux systèmes images et son, qui proposent des caractéristiques similaires et évoluent parallèlement (ou en redondance) dans les effets qu'ils mettent en œuvre.

Il est évident que cette partie, découlant directement de la précédente, parle de l'alliance du son dit 3D et de l'image dite 3D. De manière toute aussi évidente, Victor Gambier évoque déjà dans son mémoire de fin d'étude « *Nouvelles approches pour la bande sonore d'un film en relief*⁷¹ » la nécessité de faire évoluer le système 5.1 ou de créer un système sonore permettant de rendre compatible les effets sonores (ici de jaillissement) avec les effets que le cinéma en relief offre. Ces effets 3D sont recensés au nombre de cinq par Martin Barnier et Kira Kitsopanidou⁷²:

- 1. L'immersion provoquée par l'image en relief
- 2. L'effet de jaillissement de l'écran
- 3. L'attraction ou l'aspiration de la vision
- 4. La profondeur de champ remaniée et le gain que cela implique sur la perspective
- 5. L'effet de volume.

Ces effets, bien que séparés et classés, interagissent et sont plus ou moins tous liés entre eux afin d'améliorer l'immersion du spectateur. En ce qui concerne l'immersion, nous l'avons vu, l'image tend, avec l'élargissement des écrans et du relief, à confondre la réalité du film avec celle du spectateur. Comme introduit plus haut et en 1.3.1, le son cherche pour éviter la redondance, depuis plusieurs années, à atteindre ce même degré de confusion. C'est donc dans la lignée de procédés immersifs tel que le Cinerama, l'Imax ou le 5.1, que cinéma 3D et son 3D semblent

⁷¹ Victor Gambier, *Nouvelles approches pour la bande sonore d'un film en relief*, Mémoire de fin d'études, ENS Louis-Lumière, 2010.

⁷² Martin Barnier et Kira Kitsopanidou, Le cinéma 3-D, op. cit., p. 3.

renouveler la culture du cinéma immersif. L'accomplissement ultime est actuellement proposé par le son 3D, proposant, sur le papier, une expérience sonore et
sensorielle totale, pour une immersion accrue au sein du film. Cette totale
immersion passe par la notion de sphère sonore marquant l'éternel retour des effets
englobants. Ces systèmes de diffusion 3D visent à arroser le public d'un ou
plusieurs sons, diffusés par une ou plusieurs enceintes, afin de reproduire l'effet
réaliste dans le film de la provenance de sons sur un axe horizontal et vertical.
Nous ne réengagerons pas le débat entamé au sujet du réalisme de reconstitution au
cinéma en 1.3.1; cette problématique bien essentielle à l'utilisation de ces systèmes
englobants semble totalement déborder le cadre de notre étude vis-à-vis du temps
qui nous est imparti. Néanmoins, citons *Styles filmiques*. 2 : Les réalismes : Cassavetes, Forman, Kiarostami, Loach, Pialat⁷³ comme ouvrage de référence à ce sujet.

L'immersion et la réalité dont *Gravity*⁷⁴ a pu faire preuve, sont en partie dues aux effets spéciaux améliorant les sensations réalistes chez le spectateur, qui semble en demander toujours plus. Ceci n'a pas toujours été le cas, comme le remarque Ariel Rogers dans *Cinematic Appeals*: *The Experience of New Movie Technologies*⁷⁵. L'auteur nous explique que l'ancienne 3D se mettait souvent en abîme par la profusion des sorties d'objets de l'écran, provoquant un mouvement de recul chez le spectateur. D'après l'auteur, la tendance inverse est remarquable avec la nouvelle 3D et son usage plus maitrisé et subtil des effets de jaillissement. Nous pensons, comme Martin Barnier, que ceci est rendu possible grâce à la concomitance du son 3D, comme le propose Le Dolby Atmos, et de l'image 3D.

⁷³ Frank Curot, Styles filmiques. 2: Les réalismes: Cassavetes, Forman, Kiarostami, Loach, Pialat, Paris, Lettres Modernes Minard, coll. Études cinématographiques, n° 69, 2004, 291 p.

⁷⁴ Alfonso Cuarón, Gravity, op. cit.

⁷⁵ Ariel Rogers, *Cinematic appeals: the experience of new movie technologies*, New York, Columbia University Press, coll. *Film and culture*, 2013, 330 p.

Actuellement, certains professionnels avancent même que cette adéquation améliore grandement la profondeur dans l'image. On aurait l'impression d'être dans le film. Cette notion de mettre le spectateur à la place de l'acteur, (re) devient un sujet de préoccupation du cinéma actuel. En effet, notons qu'il ne s'agit pas d'une nouveauté mais bien d'une caractéristique de la narration cinématographique :

« système de représentation iconique, le dispositif scénique propre au cinéma et les phénomènes d'identification primaires et secondaires [accentués par le dispositif de représentation] font que le spectateur se trouve inclus dans la scène représentée, et qu'il devient ainsi en quelque sorte partie prenante de la situation à laquelle il assiste⁷⁶ ».

L'alliance de l'image et du son dans le cinéma 3D tend à s'inscrire dans la mouvance de la réalité augmentée, dans une sorte de cinéma inédit où tout reste à définir. Cependant, cet engagement du spectateur n'est pas propre au cinéma 3D, comme le suggère *Strange Days*⁷⁷ de Katryn Bigelow où la réalité augmentée, est à la fois thème et forme du film. L'engagement du spectateur au cinéma s'inscrit en parallèle de la mise en scène théâtrale où les comédiens évoluent sur scène et dans la salle, rompant ainsi la ségrégation spatiale entre l'espace réel (l'espace du public) et la réalité diégétique (l'espace de la mise en scène). L'identification du spectateur au film par le biais de l'acteur semble de plus en plus être soutenue par l'immersion du spectateur à la place de l'acteur, en s'affranchissant de l'objectivité présupposée et en assumant pleinement la part subjective due à l'immersion.

Cette immersion du public passe aussi par la mise en abyme du système de restitution en lui-même. Le cinéma et son public semblent renouer avec le côté spectaculaire et pseudo-réaliste dont il est issu. Ainsi, comme dans les différentes tenta-

⁷⁶ Jacques Aumont (dir.), Esthétique du film, op. cit., p. 107.

⁷⁷ Kathryn Bigelow, Strange Days, 145 min, États-Unis, sorti en France le 7 février 1996, 1995.

tives du cinéma de Méliès, le spectacle ne s'effectue plus uniquement frontalement, la salle prend vie. Bien que Claude Baiblé nous mette en garde face à cette prise de conscience de la salle⁷⁸ ou du dispositif de restitution par le spectateur, il semble que certains films usent du nouveau rapport à la représentation que cela introduit. L'immersion dans le monde du film passe par des liens, des ponts entre les éléments diégétiques du film et la salle de représentation. Cette mise en abyme est remarquable à chaque époque marquant l'arrivée d'un nouveau système de représentation où l'on observe des exemples de films présentant les effets propres à ces systèmes.

Cet effet de pont entre la salle et l'univers du film passe souvent par l'effet de volume que procurent le son et l'image 3D. Dans *Gravity*⁷⁹, ces effets sont matérialisés par des objets flottant et se heurtant en apesanteur dans la navette spatiale, des sons ponctuels dans leur localisation viennent matérialiser les objets visibles et non visibles. Ce déplacement anarchique des sons et des images vient faire baigner le spectateur au sein des éléments visuels et sonores l'environnant, par le biais d'une profondeur de champ améliorée à l'image et d'un étalement des plans sonores, possible grâce à un plus grand nombre d'enceintes adressables discrètement au sein de la salle. Cette profondeur de champ est, selon nous, rendue possible par la couronne d'enceintes surround allant jusqu'à l'écran et par les deux rangées de haut-parleurs suspendus au plafond. Au premier abord, il semblerait que la volonté soit de recréer des champs réverbérés dans toute la salle. De cette manière, l'idée de plan sonore peut apparaître. Le gain en réalité, et donc en crédibilité, apporté par la profondeur de champ et par les plans sonores permet ainsi d'accentuer la participation du spectateur au sein du film. Notons aussi que ces effets vont dans le sens du spectaculaire, qui, selon Caroline Renouard, est une caractéristique du cinéma, au même titre que l'aspect réaliste. Ce à quoi Andrew Dudley ajoute que le

⁷⁸ Claude BAIBLÉ, « Problèmes de spatialisation en son multicanal », op. cit., p. 58-60.

⁷⁹ Alfonso Cuarón, Gravity, op. cit.

spectateur en a conscience et que l'espace créé à l'image et aux sons constitue un monde de référence, qu'il semble accepter. Ceci modifie d'ailleurs les attentes des spectateurs afin de justifier le surcoût de leur place de cinéma spécial (3D ou Dolby Atmos). Le spectateur veut voir et entendre ce pourquoi il paye. Les limites de ce raisonnement amènent à penser que les spectateurs ne vont pas au cinéma pour l'histoire mais pour un dispositif de restitution, ce qui est remarquable par la présence du nom de ce système sur l'affiche du film. Les réalisateurs de cette nouvelle vague en ont conscience et réalisent souvent un clin d'œil technique au système de diffusion afin de se libérer de cette attente du spectateur. Ceci passe souvent par le jaillissement sonore et visuel d'un élément du film, à l'image de la publicité Haribo avant le début du film, où une vague de bonbons s'écroule sur nous, suscitant chez le jeune public l'effet kinesthésique d'attraper les bonbons. Cet effet d'immensité, de globalité amenée visuellement est repris au son par les capacités grandissantes de diffusion. Une notion d'amplitude est alors créée. Dans le cinéma 3D, tout n'est pas en relief, il arrive que certains plans soient quasiment équivalents à ceux de la 2D. Tout comme au son 3D, il arrive parfois que le son soit en monophonique diffusé centralement. L'atout du Dolby Atmos, selon certains mixeurs, serait comme en 5.1, une plus grande amplitude entre la monophonie et le son multicanal, permettant des effets d'attraction, d'aspiration et de contraste entre des scènes fortement spatialisées et d'autres centrées, quasiment monophoniques. Dans la version 5.1 de *Mommy*⁸⁰ de Xavier Dolan, l'effet est assez visible. Ce film évoque la vie compliquée d'une mère et son fils, celui-ci ayant des troubles hyperactifs rendant leur relation difficile. Formellement, le film débute en format 1:1 carré avec un son monophonique et ce n'est qu'au moment où le fils semble vivre un instant de liberté que celui-ci tire littéralement les bords de l'écran pour l'élargir et atteindre le format 1 :85. Cette transition est marquée au son par un passage de la monophonie à la stéréophonie (LCR) et, à certains moments, par l'utilisation des

⁸⁰ Xavier Dolan, Mommy, 139 min, Canada, sorti en France le 8 octobre 2014.

enceintes arrière. L'exemple que *Mommy*⁸¹ propose en 1:85 et en 5.1 peut être transposable à la 3D et au Dolby Atmos dans *Gravity*⁸². Les formats de représentation semblent plus que jamais impliqués dans la mise en scène et l'histoire, et le développement de ces systèmes s'accorde à augmenter l'ampleur de leurs effets.

Tous les effets de correspondance images et sons proposés ci-dessus peuvent souvent passer pour des effets de foire. Ces critiques régulièrement formulées sont connues des réalisateurs actuels qui utilisent ces effets en toute connaissance de cause, ayant bien conscience de leurs limites nocives. Il semblerait donc que ces réalisateurs ne se cantonnent plus à créer une réalité en adéquation avec notre réel ce qui, selon Charles Barr est perdu d'avance :

« Le problème est de concevoir une manière d'envelopper chaque individu du public d'un monde visuel total, de la même façon qu'il est possible de l'entourer d'un monde sonore absolu. Une fiction radiophonique peut donner une représentation satisfaisante et intégrale de ce que l'on peut entendre dans la réalité. Nos perceptions visuelles sont plus complexes que nos perceptions sonores et sont plus étroitement liés avec nos autres sens dans notre expérience quotidienne. Nous pouvons donc imaginer une réalité visuelle totale en lisant un livre ou en écoutant un morceau, mais même dans un Circlorama — même dans ce type de cinéma nous restons au centre de notre propre réalité — les gens à côté de nous, le sol sous nos pieds, l'espace entre l'écran et nous... par conséquent, il est impossible de nous « soumettre » entièrement à un cinéma total comme nous pouvons le faire en fiction radiophonique. 83 »

⁸¹ *Ibid*.

⁸² Alfonso Cuarón, Gravity, op. cit.

⁸³ Charles Barr, CinemaScope: Before and After, Film Quarterly, 1963, vol. 16, nº 4, p. 24. Traduit, par nos soins de l'anglais: « The problem is to devise some way of surrounding each individual in the audience with a total visual world, in the same way that it's possible to surround him with a total aural one. A radio play can give a satisfactory total representation of what we could in reality. Our visual perceptions are more complex than our aural ones, and are more closely bound up with the other factors in our experience. We can imagine a total visual reality, in reading a book or hearing a play, but even in a Circlorama – type cinema we are still at the center of our own reality – the people next to us, the ground beneath us, the space between us and the screen... so it's impossible to ''submit'' ourselves

La bonne utilisation de ce type de système rejoint l'idée du *could* (pouvoir) et du *should* (devoir) de Walter Murch au sujet de l'utilisation des enceintes arrière dans un film. En d'autres termes, les systèmes son et image 3D, si leur maniement est mesuré, améliorent les sensations éprouvées par un spectateur par rapport à un film en 2D et en 5.1. Reste à savoir si l'alliance entre une image plane 2D et un système Dolby Atmos propose les mêmes modalités spectactorielles.

1.4 CONCLUSIONS ET ORIENTATION DES MIXAGES ACTUELS

Nous avons vu durant l'étude succincte des différents systèmes, orienté canal et orienté objet, aujourd'hui proposés, que les principales évolutions s'axent sur une augmentation conséquente du nombre de haut-parleurs surround sur le plan horizontal et vertical. Les diverses propositions de configurations proposées par les différents fabricants de systèmes ne semblent pas pour autant s'éloigner radicalement des configurations 5.1 et 7.1 déjà répandues. La différence principale se situe dans la technologie de diffusion qu'ils emploient, WFS pour Iosono et Sonic Emotion et VBAP (Annexe D.2.) pour DTS MDA, Dolby Atmos et ADM. Ces différentes technologies permettent de se libérer de la notion de canal audio, chaque objet sonore est traité et décrit séparément selon ses paramètres de spatialisation sans se préoccuper de son format de restitution.

Actuellement ces installations et configurations semblent s'intégrer, à petite échelle, chez quelques exploitants ayant les moyens de renouveler leur dispositif sonore. D'un point de vue commercial ces systèmes apparaissent à un tournant technologique fort où la dématérialisation du support cinématographique entraine une autre manière de consommer l'objet audiovisuel. Ces nouvelles apparitions technologiques répondent ainsi en partie à une nécessité pour l'industrie du cinéma de se renouveler afin de redonner envie d'aller dans les salles en proposant aux spectateurs des expériences qu'ils ne peuvent pas avoir chez eux dans leurs salons.

Une fois de plus, les réalités commerciale et cinématographique s'engagent dans une quête de réalisme à travers des dispositifs de représentation proposant un spectacle audiovisuel encore plus immersif. Notre approche rapide et condensée permet d'ores et déjà d'affirmer que cette quête de l'immersif est une caractéristique fondamentale du cinéma et bien que les constructeurs l'utilisent comme argument commercial de choix, elle a toujours fait partie du système de représentation cinématographique. L'évolution de la diffusion sonore s'est souvent fixée comme objectif de s'approcher d'une écoute en adéquation avec notre perception réelle. Cette recherche de réel fait partie intégrante du système de représentation iconique du cinéma et a pour but d'inscrire le spectateur dans la scène à laquelle il assiste.

Dans le domaine de l'image, l'approche « perceptivo-cognitive »⁸⁴ proposée par Jacques Aumont semble aller dans ce sens. Cette théorie suppose entre autres que le spectateur, conscient de ce qu'il regarde, se comporte comme un partenaire actif de l'image, émotionnellement et cognitivement. Cet engagement du spectateur dans la représentation passe aussi par la notion de « double réalité des images »⁸⁵ définie par l'auteur comme étant la capacité du spectateur à substituer à son vécu la réalité visuelle (et sonore) qu'on lui présente en analogie avec sa propre connaissance du réel. L'auteur ajoute au sujet de l'image que « nous sommes conscients que cette image est plate et la voyons comme telle, mais nous y voyons aussi un fragment d'espace à trois dimensions »⁸⁶. Selon nous, ces théories facilement compréhensibles d'un point de vue de l'image, sont à notre connaissance, bien plus complexes à envisager au niveau du son.

⁸⁴ Jacques Aumont, L'image, Paris, A. Colin, 2011, p. 80-87.

⁸⁵ Ibid., p. 39.

⁸⁶ *Ibid*.

À ce que Jacques Aumont avance, Jean-Pierre Oudart avait déjà proposé une distinction intéressante entre l'effet de réel et l'effet de réalité. L'auteur définit l'effet de réalité comme étant liée au système de représentation (perceptif) à la différence de l'effet de réel qui quant à lui tient à l'inscription du spectateur au sein de ce système. À quoi Jacques Aumont ajoute que « cette intrusion du spectateur fait qu'il ne perçoit plus les éléments de la représentation comme tels, mais qu'il les perçoit comme étant les choses elles-mêmes »⁸⁷.

Cette problématique de l'inscription du spectateur au cœur du système de représentation semble aujourd'hui renouvelée, requestionnée et remise en question par l'arrivée de systèmes d'autant plus immersifs, proposés en partie par les formats orientés objet. La difficulté de mise en œuvre d'un espace sonore cohérent à l'image essaie de se solutionner en partie par des systèmes proposant une spatialisation accrue et plus ponctuelle. À ce sujet, Laurent Jullier⁸⁸ propose déjà, à propos du 5.1, une réflexion intéressante, bien que très axée sur l'aspect commercial des constructeurs de systèmes. L'auteur propose une vision très pessimiste de ce qu'il définit comme « une véritable idéologie du son numériques multipistes, participant d'un déplacement topologique de l'action filmique. »89 Il définit comme idéologie les arguments qu'utilisent les commerciaux pour vendre leurs systèmes aux ingénieurs du son, exploitants et par conséquent spectateurs. Ces arguments gravitent autour des concepts d'immersion et de dynamique sonore qu'il définit respectivement comme « le spectateur poisson⁹⁰ » et « le bombardement (le spectateur-cible)⁹¹ ». Bien que nous n'adhérions pas entièrement à son discours véhément, dénonçant trop facilement les travers des industriels du cinéma, nous retenons tout

⁸⁷ Jacques Aumont (dir.), *Esthétique du film*, 3. éd., s.l., Armand Colin, coll. *Armand Colin Cinéma*, 2008, p. 107.

⁸⁸ Laurent Jullier, « ASK FOR MORE! Le discours quantitatif chez les géants du son numérique », Thierry Miller et Laurent Jullier (dirs.), in *Analyse et réception des sons au cinéma*, Paris, L'Harmattan, coll. *Champs visuels*, 2007, pp. 289-297.

⁸⁹ Ibid., p. 289-290.

⁹⁰ Ibid., p. 290.

⁹¹ Ibid., p. 291.

de même de son propos le constat d'un changement visible dans notre manière de consommer le spectacle audiovisuel.

« Alors que les vingt premières années du cinéma ont été, pour reprendre la terminologie de Noël Burch, le temps du passage de la confrontation exhibitionniste à l'absorption diégétique, la tendance est susceptible de s'inverser doucement, avec une forme nouvelle de confrontation exhibitionniste oscillant entre l'exemplification technique (écoutez ce superbe système!... the audience is listening THX sound...) et la transformation de l'auditoire en cibles résonnantes⁹² ».

à quoi il ajoute,

« combinée à l'obsession quantitative, pareille confusion à propos du lieu du film mène à cette croyance infondée selon laquelle le nombre d'effets de manche sonores détermine mathématiquement l'effet affectif sur le spectateur⁹³ »

Cependant bien que ces « révélations » soient en partie vérifiables nous pensons que le public, les ingénieurs du son, les exploitants, etc. ne sont pas dupes. En ce sens, nous pensons que la description faite par l'auteur est ciblée à une catégorie de films et une catégorie d'exploitants qui utilisent ces systèmes de manière inadaptée. Laurent Jullier oublie d'autre part que quand bien même ces systèmes sont commercialisés ainsi, l'utilisation faite par les professionnels et les réalisateurs peut en être toute autre. De toute évidence, ce type de discours dédaignant les avancées technologiques, s'éloigne majoritairement de la pratique qui peut en être faite dans le cinéma.

L'auteur omet que cette notion d'immersion relève surtout d'une question de pratique, et que l'histoire de la diffusion sonore a toujours été dans cette optique, avec la volonté de recréer une scène sonore, comme nous l'avons vu avec la notion

⁹² Ibid., p. 296.

⁹³ *Ibid*.

de *staging*. Le point de concordance actuel entre tous les principes de restitution est qu'ils tendent à reformuler ces propositions immersives en par le biais de nouvelles technologies de spatialisation. Ces nouvelles propositions immersives sont aussi la conséquence d'une évolution des formats images qui proposent des projections sur des écrans de plus en plus grands marquant un nouveau type d'expérience spectatorielle.

« Quelque chose de basique a changé dans l'expérience audiovisuelle qui redéfini la relation établie entre le spectateur et l'écran, qui, maintenant, semble s'introduire dans l'espace du public, notamment à l'aide de la bande son, qui renforce cette extension de l'image en sortant même des limites de l'écran à l'aide d'enceintes stratégiquement positionnées sur les côtés au plafond et sur le mur au fond de la salle, pour entourer le spectateur de son. »⁹⁴

Néanmoins, les discours commerciaux à l'apparition de ces nouveaux systèmes ne semblent pas différents de ceux effectués dans le passé. Mais il demeure du ressort des réalisateurs et des ingénieurs du son de s'en approprier l'utilisation. L'argumentaire de Laurent Jullier peut être conforté dans le sens où l'utilisation de ces formats orientés objet s'adresse à des films spécifiques, représentatifs d'une part de la production cinématographique globale. Si les films majoritairement produits dans ces formats sont des films d'action (ou à effets) à gros budget (*Blockbuster*), nous pensons que leurs possibilités principales s'adaptent mieux à des films d'auteur. Ces dispositifs dits de prochaine génération s'inscrivent pour le moment en marge de la production cinématographique principale, proposant leurs technologies à des films grand public comme *Superman vs Batman*⁹⁵ ou *Bilbo le Hobbit*⁹⁶, dans lesquels l'histoire sans intérêt pousse les techniciens à dissimuler l'absence

⁹⁴ John Belton, Widescreen Cinema, Harvard, Harvard University Press., 1992, p. 187.

⁹⁵ Zack SNYDER, *Batman v Superman: Dawn of Justice*, 151 min, États-Unis, sorti en France le 23 mars 2016.

⁹⁶ Peter Jackson, *The Hobbit : The Desolation of Smaug (Le Hobbit : La désolation de Smaug)*, 161 min, États-Unis, sorti en France le 11 décembre 2013.

d'enjeux narratifs en utilisant le caractère spectaculaire du dispositif et en s'éloignant d'un processus de narration. Dans ce sens, les systèmes orientés objet, se voient affublés d'un aspect gadget, tout comme le 5.1 dans ses débuts.

Or il nous semble que la technologie objet permet, lorsque nécessaire, de prolonger la quête de la réalité au cinéma et non dans les effets spectaculaires. En ce sens nous pensons que les formats orientés objet se destinent plus à une production filmique d'auteur qu'aux films à effets auxquels on les rattache encore trop souvent.

Pour conclure, au vu des démarches commerciales, des productions ciblées et du public intéressé, il ne semble pas évident à l'heure actuelle d'envisager les dispositifs orientés objet comme des systèmes de prochaine génération. D'autant plus qu'il est encore moins certain qu'ils arrivent à se pérenniser comme le 5.1 a pu le faire (alors que son parcours chaotique est fortement semblable). De plus, même si ils tentent de s'imposer, le 5.1 restera pour le moment le format le plus regardé, en dépit du fait que les formats objet proposent de solutionner la problématique de compatibilité descendante par une automatisation du mixage vers plusieurs autres versions. Bien que l'avenir des formats orientés objet ne soit pas encore tracé, l'automatisation du travail de mixage vers différents autres formats répond à la logique de consommation de l'objet audiovisuel actuel. En effet, la multiplicité des formats de diffusion pour un film (LtRt, 5.1, Auro 3D, Dolby Atmos, DTS MDA, etc.) ne peut pas être intégrée manuellement au sein de la postproduction cinématographique. L'automatisation semble de prime abord répondre aux contraintes actuelles de postproduction filmique.

Enfin, ce qui nous semble primordial, est d'entrevoir comment les ingénieurs du son vont utiliser ces systèmes proposant *in fine* d'améliorer la sensation de réalité. C'est en essayant de se rapprocher le plus possible d'une écoute réaliste par la

multiplication du nombre de pistes et de haut-parleurs que l'on approche des limites de la croyance et de l'illusion cinématographique chez le spectateur.

« Tout comme le démontre Benjamin Wright [...], alors que les fabricants comme Dolby vantent leur système « surround » sur la base de la possibilité de placer le son avec une précision croissante dans la salle de cinéma pour immerger l'auditoire au centre d'un espace acoustique à trois dimensions, les ingénieurs du son sont tenus de respecter les idéologies concernant l'intelligibilité des dialogues, la cohérence narrative et diégétique, ils subissent les contraintes d'attacher les sons au cadre visuel de crainte de distraire le public du cadre essentiel de la représentation. Ainsi, les rangées d'enceintes frontales, Gauche-Centre-Droit — en outre, la scène sonore (sound stage) stéréophonique frontale — sont toujours privilégiées dans le mixage du film, sans tenir compte du nombre total de canaux audio disponibles à la spatialisation des sons dans le mixage. 97 »

⁹⁷ Paul Théberge, Kyle Devine et Tom Everrett (dirs.), Living stereo, op. cit., p. 12-13. Traduit, par nos soins, de l'anglais: « As Benjamin Wright argues [...], while manufacturers such as Dolby promote their surround systems on the basis of their ability to place sound with increasing accuracy within the cinema and to immerse the audience within the three dimensional acoustic space, sound engineers are bound by ideologies of speech intelligibility, narrative and diegetic coherence, and the need to ties sounds to the visual frame for fear of distracting audiences from the audiences from the essential space of representation. Thus the frontal, Left-Center-Right array of speakers — in effect, the stereo sound stage — is always priviliged in film mixing, regardless of the total number of sound channel employed in the mix »

2 Expérimentation de l'approche objet

2.1 EXPLOITATION DE LA MATIÈRE SONORE DANS LA CRÉATION CINÉMATOGRAPHIQUE

Afin de poursuivre notre recherche sur le downmixage automatique des systèmes de diffusion orientés objet, il nous est apparu nécessaire de s'interroger sur les enjeux esthétiques de ce qu'offre l'approche objet dans la chaine de post-production sonore au cinéma. Nous nous sommes concentrés sur deux éléments de la bande son, à savoir les ambiances et les effets. De plus, dans le temps imparti à notre étude, nous avons choisi, bien que ce paramètre nous semble intéressant (Cf. Annexe C), de ne pas étudier les effets du gain en dynamique sonore du système. D'autre part, nous conseillons la lecture de l'Annexe E.3. pour un descriptif détaillé du logiciel de mis en espace *Dolby Atmos Panner*.

2.1.1 INTÉGRATION DES AMBIANCES EN MODE OBJET : POUR QUEL FILM ?

Durant l'étude théorique du système de diffusion orienté objet, nous pensions déjà à quelques idées de mise en pratique. La mise en espace précise et ponctuelle des sons nous a amené à remanier la conception des ambiances au sein d'un film. Dans un dispositif orienté canal, l'ambiance d'une scène peut être issue de plusieurs sons sommés électriquement afin d'alimenter un ou plusieurs canaux. Notre expérimentation consiste à effectuer un mélange acoustique de ces différentes sources. Ainsi, chaque source sonore n'est plus pensée comme un son unique alimentant un canal diffusant le son souhaité mais comme des objets sonores évoluant indépendamment les uns des autres. La perception de ce mélange sonore s'effectue grâce à la spatialisation des sources en mode objet et non électriquement au sein d'un réseau de plusieurs enceintes issues du même canal. L'hypothèse de rendu nous semble

intéressante au regard de la méthode classique d'élaboration de l'ambiance d'une scène sonore. Premièrement, la spatialisation des différents objets sonores permet de diminuer l'inter-masquage des sources, problématique connue des monteurs son et des mixeurs habituellement solutionnée par la suppression d'une source au profit d'une autre. La spatialisation permet une meilleure fusion de ces objets sonores puisque perçus au point d'écoute comme la somme acoustique de plusieurs objets sonores diffusés par des enceintes distinctes. Selon nous, cette logique d'étalement spatial des sources peut être perçue comme unique. Lors des montages de séquences de films de la partie pratique de mémoire, nous avons même été agréablement surpris de l'intégration de ces différents objets avec la mise en scène.

La conception des ambiances en termes d'objets sonores semble s'intégrer plus facilement à l'image, estompant les conflits scénographiques évoqués en 1.3. En effet, la mise en espace des objets sonores peut s'effectuer dans toute la salle, au plus près de l'écran comme au plus loin. Dans bien des cas, les ambiances placées au plus près de l'écran dans le premier tiers de la salle procurent des sensations immersives sans pour autant détourner l'attention du spectateur de l'écran.

Effectivement, la spatialisation de ces ambiances, tout en restant globalement frontale, semble éclaircir le contenu sonore situé derrière l'écran. Le résultat obtenu semble procurer une sensation d'aspiration vers l'écran sensiblement plus juste par rapport à certains types de mises en scènes immersives. Ce type de sensation est perceptible par la hiérarchisation des plans sonores, exacerbée par la précision de diffusion que procure la spatialisation. Le nombre de haut-parleurs adressables de manière autonome, permet de créer de la profondeur dans la combinaison des éléments sonores. Cependant, cette profondeur se crée à partir de la salle et non derrière l'écran comme la logique le voudrait. Nous pensons que la profondeur des plans sonores passe essentiellement par l'existence d'un hors champ au regard de la

mise en scène. Ce hors champ semble exister de manière plus subtile et intégrée que s'il était diffusé par une rangée de haut-parleurs.

De plus, la diffusion verticale de certaines ambiances renforce la sensation d'espace et de profondeur. Le mélange acoustique entre des ambiances horizontales et zénithales procure une sensation d'aération qui rend les plans sonores moins confus et étriqués. Bien sûr, comme nous le rappelle Aymeric Dupas⁹⁸, notre mauvaise perception de localisation en élévation permet un placement plus souple des ambiances sonores sans pour autant être susceptible de retenir l'attention du spectateur. D'autre part, le contenu sonore des ambiances utilisées joue beaucoup sur la sensation d'élévation, ainsi, des sons placés sur le plan horizontal peuvent être perçus en élévation. Ce qui nous amène à penser que l'intérêt de la spatialisation des ambiances en élévation ne se trouve pas dans la cohérence en localisation mais plutôt dans la sensation d'un espace sonore uniforme. D'autre part, lorsque les ambiances se retrouvent en mouvement, l'intérêt ne se situe pas au point de vue d'une réalité spatiale mais plutôt comme sensation d'espace dans lequel évoluent les personnages. Ce qui va dans le sens d'une mise en scène immersive où l'on recherche souvent à confondre le point de vue subjectif du personnage avec celui du spectateur comme l'a expérimenté par exemple le film Strange Days⁹⁹. Non seulement le spectateur se retrouve en immersion d'un point de vue spatial mais aussi du point de vue sensoriel. En effet, le caractère englobant de tels systèmes met en avant l'interaction sensorielle avec le spectateur dont l'état psychologique, émotionnel et moteur dépend du traitement de l'univers immersif qui lui est proposé.

⁹⁸ Aymeric Dupas, *Multicanal au cinéma : un canal zénithal ?*, Mémoire de fin d'études, ENS Louis-Lumière, 2007.

⁹⁹ Kathryn Bigelow, Strange Days, op. cit.

Ce glissement a déjà été initié selon nous par des films où la caméra tend de plus en plus à être assumée en tant que telle. Comme si le cadreur et sa caméra constituaient le seul médium entre la scène jouée et la salle de cinéma.

Cet effet n'est pas propre à une époque, puisque certains films – La dame du lac¹⁰⁰ en particulier – et certains mouvements cinématographiques – La Nouvelle Vague. Le Néoréalisme – peuvent en faire l'exemple. Néanmoins, il nous apparaît que l'effet est de plus en plus utilisé. L'exemple le plus connu se situe dans la séquence d'introduction de Il faut sauver le soldat Ryan¹⁰¹. Mais des films comme Elephant¹⁰², The tree of life¹⁰³, Les fils de l'homme¹⁰⁴, Enter the Void¹⁰⁵ ou encore The Revenant¹⁰⁶ sont des exemples de formes cinématographiques où l'existence de la caméra joue un rôle fondamental dans la narration. L'expérience proposée dans les salles de cinéma tend vers une forme de réalité augmentée où le facteur sensoriel est pris en compte, se rapprochant du marché florissant des jeux vidéo à vision subjective ou encore FPS (First Person Shooter) où interaction visuelle et sonore font partie prenante de l'évolution de l'utilisateur dans le récit. Dans Bad Motherfucker¹⁰⁷, clip vidéo réalisé par Ilya Naishuller, le champ visuel du spectateur est confondu avec celui du personnage principal pendant toute la vidéo à la différence de *Strange Days*¹⁰⁸ où la camera personnage ne prend vie qu'à certains moments.

¹⁰⁰ Robert Montgomery, La dame du lac, 105 min, États-Unis, sorti en France le 14 avril 1948.

¹⁰¹ Steven Spielberg, *Il faut sauver le soldat Ryan*, 169 min, États-Unis, sorti en France le 30 septembre 1998.

¹⁰² Gus Van Sant, Elephant, 81 min, États-Unis, sorti en France le 22 octobre 2003.

¹⁰³ Terrence Malick, *The Tree of Life*, 139 min, États-Unis, sorti en France le 17 mai 2011.

¹⁰⁴ Alfonso Cuarón, *Children of Men (Les fils de l'homme)*, 109 min, États-Unis, Royaume-Uni, sorti en France le 18 octobre 2006.

¹⁰⁵ Gaspar Noé, Enter the Void, 161 min, sorti le 5 mai 2010.

¹⁰⁶ Alejandro G. IÑÁRRITU, The Revenant, op. cit.

¹⁰⁷ Bad Mother fucker de Ilya Naishuller, clip vidéo pour le groupe Biting Elbows sorti sur la chaine de vidéo gratuite youtube en 2013. Notons que les clips vidéo ont souvent fait l'œuvre de tentative de caméra subjective comme ce fut le cas en 1997 avec le clip Smatch My Bitch Up du groupe musical The Prodigy.

¹⁰⁸ Kathryn Bigelow, Strange Days, op. cit.

Le succès de ce clip vidéo a permis au réalisateur de sortir un long métrage, Hardcore Henry¹⁰⁹, basé sur le même procédé immersif à l'image et s'essayant cette fois à l'écriture sonore multicanal proposée par le mixage orienté objet. Ce dernier semble symboliser les limites de l'utilisation de l'immersion dans la forme cinématographique. L'univers sonore y est saisissant. Mais l'identification au personnage est de plus en plus floue et le spectateur a l'impression d'être le personnage. Or, l'identification est rompue lorsque le personnage principal – vous – parle à d'autres personnages. Ceci expliquant pourquoi les films ou les séquences de films à vision plus ou moins subjective évitent toute forme de dialogue du personnage principal. D'autre part, la bande son très immersive, se rapprochant de l'interaction visuelle et sonore du jeu vidéo, ne semble pas en accord avec nos modalités d'écoute cinématographique. Effectivement, il semblerait que le gain en précision, en réalisme et en immersion, fruit d'une écriture sonore orientée objet, tente de recréer un univers sonore réel. Ceci passe par une bande son riche en éléments sonores dont la spatialisation accrue et le niveau global sont fabriqués à l'image de ce que l'on peut vivre dans le réel. Cependant, nous ne pensons pas que le spectateur accepte un tel conditionnement. En ce sens, Hardcore Henry¹¹⁰, de par sa narration visuelle et sonore axée sur l'impression de réalité augmentée, s'éloigne d'une forme cinématographique où le spectateur, bien que plongé dans la narration du film, entretient une certaine distance de par sa forme et son contenu.

Néanmoins, bien au-delà des dérives cinématographiques de ce que le mode objet peut induire, il semble avoir un intérêt pour un cinéma plus conventionnel. En effet, le mode objet permet de renouer avec certains systèmes de prise de son multicanal. Dans le cadre de la postproduction cinématographique, la majeure partie des sons utilisés provient de systèmes monophoniques ou au plus stéréophoniques. Les prises de sons multi-microphoniques ne sont pour autant inexistantes, mais leur

¹⁰⁹ Ilya NAISHULLER, *Hardcore Henry*, 96 min, États-Unis, sorti en France le 13 avril 2016. 110 *Ibid*.

rendu spatial et spectral en salle de cinéma reste souvent moins intéressant qu'une ambiance issue d'une prise de son monophonique et stéréophonique. Effectivement, la plupart du temps, les prises de sons d'ambiances multi-microphoniques ont un rendu intéressant en écoute discrète, où chaque haut-parleur est alimenté par le signal d'un seul microphone. Cependant, ce n'est plus le cas lorsque les enceintes, alimentées par un même canal, s'additionnent, ce qui est le cas de toutes les salles de cinéma. Cette utilisation ne trouve donc que peu de finalité au regard des systèmes de diffusion orientés canal. La solution semble être amenée par les systèmes de mixage orientés objet qui permettent d'orienter le signal audio vers un seul haut-parleur (Cf. Annexe E.3.), autorisant à conserver les paramètres de spatialisation effectués lors de la prise. La restitution devient plus fidèle et l'enveloppement parfois plus réaliste qu'une ambiance fabriquée par plusieurs éléments stéréophoniques ou monophoniques. Nous pensons notamment que cette immersion peut ne pas être perçue ou ressentie directement par les spectateurs. Dans plusieurs cas, les ambiances multicanal, comme la quadriphonie, s'intègrent particulièrement bien à l'image. L'amélioration de la réponse en fréquences des enceintes arrière propose une meilleure cohérence entre les sons d'écran et les sons provenant de la salle. Il semblerait que l'ambiance recentre plus notre attention sur l'image là où sur des systèmes classiques l'attention était amenée par une utilisation frontale du système. Néanmoins, il reste à étudier la validité de l'utilisation des prises de son multi-microphoniques par rapport de leur zone d'écoute optimale. En mode objet, lorsque le nombre de haut-parleurs utilisés pour la diffusion coïncide avec le nombre de microphones utilisés pour la captation, la restitution de l'espace sonore est meilleure lorsque le spectateur est placé à un endroit bien déterminé (restriction de la zone d'écoute). En réduisant le nombre de haut-parleurs sollicités pour obtenir un espace cohérent, le mode objet s'oriente dans un principe de diffusion où la place du spectateur demeure encore plus importante que dans les principes de diffusion orientée canal.

2.1.2 Les sons effets et leur intégration

Tout comme dans le cas des ambiances, l'intérêt du mode objet dans la création sonore tend à une utilisation des effets sur une zone restreinte d'enceintes. Notons tout de même que la précision de diffusion en mode objet peut porter préjudice en ce qui concerne l'orientation réflexe chez le spectateur. Dans ce sens, les dispositifs orientés objet nécessitent les mêmes précautions d'utilisation par rapport à la spatialisation. Bien que l'étude approfondie de ce sujet sorte du domaine de notre recherche, nous renvoyons notre lecteur aux mémoires de fin d'étude de Nicolas Moreau¹¹¹ et de Christophe Vingtrinier¹¹².

Au fur et à mesure de nos essais, nous avons remarqué que le mode objet peut intégrer des effets plus naturellement au sein de la bande son. La précision en déplacement dynamique au regard des critères d'orientation réflexe de l'attention conduit à une utilisation des enceintes du premier tiers de la salle au détriment des deux tiers restants. Néanmoins, le déplacement dynamique de la source entre l'écran et le premier tiers de la salle semble s'effectuer de manière plus fluide et moins dérangeante. L'apport supplémentaire du mode orienté objet par rapport au mode orienté canal classique est de pouvoir sortir très légèrement de l'écran certains éléments ponctuels de la bande son. Cette sortie de la diffusion en façade permet de produire des effets d'intégration acoustique intéressants avec la salle. En effet, là où la réverbération artificielle, le délai, le filtrage, etc. étaient utilisés pour tenter de décoller¹¹³ le son de l'écran, le mode objet permet d'arriver plus naturellement au même type d'effet. Étonnamment, le placement d'un objet sonore en combinant l'utilisation des paramètres X, Y, Z et *Size* décrit en 3. produit des effets d'intégration acoustique intéressants.

¹¹¹ Nicolas Moreau, L'orientation réflexe dans le cinéma multicanal, op. cit.

¹¹² Christophe VINGTRINIER, La diffusion multicanal au cinéma : nécessité d'une écriture visuelle et sonore adaptées, op. cit.

¹¹³ Ce terme est souvent utilisé pour caractériser une diffusion sonore plus large et moins frontale.

Dans *Les Saisons*¹¹⁴, une séquence met en scène une biche, qui, étrangement agitée, joue dans l'eau, attirant le regard intrigué d'un oiseau perché sur un arbre. L'enjeu sonore de la scène se retrouve confronté au découpage qui présente une alternance de plans de la biche et de l'oiseau en train de l'observer. Sur les plans sur l'oiseau, la biche doit être placée sur un autre plan sonore, afin de correspondre à la mise en scène. L'effet acoustique a été obtenu en déplaçant l'objet sonore vers la salle et en élévation. La diffusion de l'objet fut élargie (avec le paramètre *Size*) à plusieurs enceintes pour estomper sa précision.

Il semblerait donc que l'intégration des effets sonores ne passe plus exclusivement par des traitements du son en lui-même, mais aussi par la spatialisation de celui-ci au sein de la salle. Néanmoins, l'intégration des effets par des outils plus traditionnels est toujours possible voire plus intéressante. Durant nos expériences, nous avons essayé de placer des réverbérations artificielles dans les haut-parleurs zénithaux totalement décorrélés spatialement de leur source généralement frontale¹¹⁵. Il s'avère que la sensation d'espace perçue paraît plus naturelle que si nous utilisons une réverbération multicanal sur le plan horizontal. L'emploi systématique de réverbérations sur le plan horizontal et sur le plan vertical n'est donc pas nécessaire; au contraire, l'effet naturel s'en trouve d'autant plus estompé. Tout en étant en dehors de l'image, les réverbérations artificielles au plafond semblent s'intégrer dans un processus englobant sans pour autant devenir gênantes. Ainsi, il devient possible de simuler des acoustiques de salles de manière plus réaliste. Ceci fonctionne particulièrement bien sur les changements d'espace, de l'intérieur vers l'extérieur et vice versa. Tout comme le 5.1 permet une amplitude de diffusion allant de la monophonie au 5.1, les formats orientés objet avec un plafond de hautparleurs élargissent les possibilités de spatialisation en élévation. Il est désormais

¹¹⁴ Jacques Perrin et Jacques Cluzaud, Les Saisons, op. cit.

¹¹⁵ La source frontale est elle aussi intégrée dans la bande son du film par une réverbération artificielle.

possible de jouer sur des acoustiques sur différents plans, augmentant la sensation de relief.

Dans la séquence d'introduction des *Medouny*, le personnage de Saïd évolue dans différents espaces. La caméra suit son évolution de dos en plan séquence. Le personnage, qui marche d'un pas pressé, part d'un parking, passe sous un petit pont et s'avance vers un hangar dans lequel il finira par rentrer. Sur cette séquence d'introduction, dont le filmage adopte un point de vue très similaire à celui de Saïd (Cf. Figure 2.2), la spatialisation ponctuelle du mode objet nous a permis de suivre l'emplacement des sources sonores en fonction de l'axe de la caméra. Ainsi, à l'aide de réverbérations et de délais, nous avons simulé la sensation d'éloignement d'une cloche d'église, en rapport avec l'éloignement de la caméra de sa position initiale. La cloche est envoyée dans trois réverbérations. La première réverbération (stéréophonique) est utilisée afin d'agir directement sur le rapport champ direct/champ diffus de la cloche et nous permet de simuler l'éloignement. Cette réverbération suit alors le positionnement (dynamique) de la source initiale. Une autre réverbération (stéréophonique), combinée avec des délais, est utilisée pour créer les réflexions de la cloche sur les différentes parois que le personnage va croiser durant son évolution. Cette réverbération est placée dans les deux canaux zénithaux permettant une diffusion stéréophonique homogène au plafond. Enfin, une troisième réverbération multicanal (5.1) est utilisée sur le plan horizontal afin de réinjecter un champ diffus essentiellement frontal pour rendre la diffusion cohérente sur tous les plans. Ce principe a permis d'intégrer le son de cloche de manière beaucoup moins artificielle et beaucoup plus cohérente au regard de la mise en scène et des enjeux que suggère le plan séquence.



Figure 2.1 : Photogramme de la séquence d'introduction des Medouny. Saïd marche de dos vers le passage sous le pont.

De la même manière, lorsque Saïd va arriver sous le petit pont, la sensation de relief et de plan sonore est amenée en intégrant le direct à l'acoustique du passage bétonné. En effet, dans le direct, Saïd traine sa chaussure sur le gravier, provoquant ainsi l'excitation des flotteurs échos, des réverbérations et d'autres phénomènes acoustiques perceptibles dans ce type de lieux. Le réalisateur, attaché au son du direct, désire augmenter ce changement d'acoustique fort. Dans cette optique, le direct est envoyé dans des réverbérations (suivant le même dispositif cité plus haut) et dans plusieurs délais (spatialisés en mode objet) pour simuler l'excitation acoustique du lieu. L'effet acoustique, convainquant d'un point de vue réaliste, accentue le processus de filmage immersif de la caméra-personnage.

Nous constatons qu'avec très peu d'éléments, il est aisé de rendre un espace cohérent sans recourir à des couches et des couches de sons additionnels. Néanmoins, comme nous l'avons vu, la création en mode objet n'écarte pas l'utilisation du mode orienté canal ; au contraire, elle permet parfois de la compléter simplement. En effet, il arrive qu'une simple réverbération orientée canal puisse permettre d'arriver au même rendu qu'une réverbération spatialisée en objet.



Figure 2.2 : Photogramme de la séquence d'introduction des Medouny. Saïd passe sous le petit pont.

Enfin, pour résumer, notre utilisation du dispositif s'articule plus sur une écoute centrée sur le spectateur que frontalement sur l'écran ; ce qui semble rentrer dans une logique de filmage de caméra spectateur. Cette logique de filmage assume totalement l'omniprésence de la caméra et de l'artifice cinématographique. La magie du cinéma est comme rompue; le spectateur a conscience du processus d'élaboration filmique ; le spectateur assiste plutôt à une performance audiovisuelle captée par la caméra et les microphones. Dans cette fausse transparence, tout est fait pour que le spectateur ait la sensation d'être avec les comédiens sur la scène. Dans ce processus, la caméra s'intègre à la narration et est sans cesse mise en abyme. C'est le cas dans *The Revenant* où le personnage principal, à l'article de la mort, est filmé en très courte focale et en très gros plan. Alors que le froid et ses blessures semblent confirmer son destin funèbre, sa respiration vient peu à peu fabriquer de la buée sur l'objectif de la caméra, matérialisant à la fois son combat contre la mort et la plausibilité de la scène. Le spectateur assiste à une performance du comédien en train d'exercer son métier. Il l'observe, l'analyse et lorsque le spectateur semble le plus distant face à la représentation apparaît un autre élément,

la buée, matérialisant le fait que tout ceci n'est que du cinéma mais en même temps il fait réellement très froid et les tremblements de Leonardo Dicaprio ne sont probablement pas tous simulés pour la mise en scène.

Dans cette interaction entre la caméra, la mise en scène et les personnages, le son, lui, n'est pas autant remis en question que la caméra. Le spectateur a conscience que la prise de son est effectuée en même temps que l'image mais il ne la dissocie pas. D'ailleurs, dans cette logique de captation du moment, tous les sons ajoutés en postproduction doivent concorder avec l'esthétique de Found Footage, où l'image et le son inscrits sur la pellicule proviennent du lieu du tournage. Les films répondant à ce genre cinématographique n'utilisent bien souvent pas le son dans sa forme la plus cohérente à la logique de captation (à savoir la monophonie). Il est intéressant de constater que la bande son exploite pleinement l'immersion par le son spatialisé et l'existence d'un hors champs sonore. C'est lui qui, par mille artifices, cette fois-ci méconnus du spectateur, va rendre le décor plus réaliste, c'est même lui qui va justifier le décor (pourtant réel) du film. Cette notion va de pair avec la notion de scène sonore, déjà introduite par Paul Théberge avec le concept de Staging (Cf. 1.1.1). Dans cette optique, le son lui, doit fournir une sorte de transparence avec l'image qu'il ponctue. Le mode orienté objet semble plus que tout répondre à cette problématique en s'axant plus sur l'écoute d'un spectateur caméra que sur une écoute centrée sur l'image. On ne bouge plus la source, mais on déplace le spectateur et son point de vue avec des trajectoires automatisées. Ce qui est important dans la création en mode objet est le réalisme induit par le processus filmique. Le son doit lui, coller au réalisme acoustique. Cela passe parfois par des bandes sonores épurées, mais comportant une mise en espace accrue des sons. Nous constatons qu'avec très peu d'éléments, il est alors possible de construire l'univers sonore de cette caméra spectateur.

Néanmoins, la question demeure du côté de la narration, est-ce que nous ne passons pas plus de temps à construire un espace plausible qu'une narration par des sons ? Est-ce que même si le processus filmique tend à rendre le son immersif ou en accord avec une réalité acoustique, nous ne passons pas à côté de l'essentiel du travail sonore ? Ou est-ce que le nouveau type de cinéma tend à une utilisation cantonnée à travailler sur des phénomènes perceptifs immersifs ?

2.1.3 Limites, possibilités et ouvertures de la création en mode objet

Durant nos expérimentations, nous nous sommes rendu compte que les outils de traitement spatial du son ne sont pas optimisés pour un mode de travail en orienté objet. Ces outils, émanant d'une logique de traitement orienté canal, ne permettent pas une gestion simple et efficace lorsqu'ils sont utilisés pour des sons orientés objet. L'exemple le plus frappant est pour nous l'utilisation des réverbérations et des délais, principaux logiciels de traitement et outils de création sonore. Appliquer une réverbération en adéquation avec le son spatialisé en mode objet devient complexe. En premier lieu, la logique d'exploitation¹¹⁶ des réverbérations s'applique difficilement au système orienté objet. Pour éviter toute dichotomie entre le son direct et le son réverbéré, la réverbération doit suivre la spatialisation du son direct, impliquant que la réverbération soit aussi en mode objet sur une autre piste. Deux problèmes s'opposent à ce mode de fonctionnement :

• Le processeur de spatialisation des objets introduit un retard de calcul sur les sons spatialisés. Ce retard est de 2 images et demie. Ce retard n'est pas perçu par le logiciel d'édition ProTools, qui ne peut pas le compenser. Le fonctionnement de réverbération par envoi d'auxiliaire ne fonctionne pas

¹¹⁶ Le processus de réverbération d'une source sonore se décompose en trois parties. Une piste est dédiée au son d'origine (non traité) et une autre est dédiée au son traité par la réverbération. L'envoi du son d'origine est réalisé par des envois effectués par des pistes auxiliaires. Il est alors possible de manipuler l'influence du son « direct » par rapport au son « réverbéré ».

non plus. Le délai introduit par l'envoi en piste auxiliaire ne peut également pas être compensé par ProTools. Par conséquent un pré-délai de 1,25 images s'introduit dans le processus de réverbération. Ce qui rend impossible l'utilisation de réverbérations courtes.

• Le suivi de spatialisation de la réverbération par rapport à la source s'effectue de manière indépendante. Il n'y a pas de moyens pour effectuer un lien entre les deux modules de spatialisation utilisés par la piste de son direct et la piste de réverbération.

D'autre part, il n'existe pas encore de réverbération prévue pour les *beds*. L'essentiel des plugiciels de réverbération actuels proposent au plus de couvrir 8 canaux, laissant de côté les canaux zénithaux. Seul le développeur *Exponential Audio*, semble proposer une alternative intéressante dans sa suite de plugiciels de réverbération 117. L'option 3D Link, propose de créer des liens entre plusieurs réverbérations pour ainsi arriver au nombre de canaux escompté. Non seulement chaque réverbération réagit à ses propres entrées, mais il introduit aussi les entrées des autres réverbérations qui peuvent être retardées, diffuses et filtrées tout comme dans un espace réel. Ces liens permettent de verrouiller les paramètres de telle sorte qu'ils agissent comme une réverbération unifiée, tout en conservant la possibilité de contrôler les réverbérations de manière séparée lorsque nécessaire.

L'inadaptation est d'autant plus flagrante lorsque la comparaison est effectuée avec le mode orienté canal où des plugiciels, tels que le *Spat v3*¹¹⁸ de l'IRCAM, proposent une utilisation plus flexible et moins complexe de la mise en espace des sons. Le logiciel *Spat* permet de spatialiser des sources en prenant en compte

¹¹⁷ La société *Exponential Audio* propose une suite de plugiciels de réverbération composés de *Phoenix Verb Surround*

¹¹⁸ Le *Spat v3* est un plugiciel de spatialisation de sources sonore créé par la compagnie Flux. Il permet entre autres de recréer des acoustiques à partir de sources monophoniques ou stéréophoniques. Le plugiciel possède 16 entrées et peut alimenter des systèmes jusqu'au 8.0.

plusieurs paramètres acoustiques¹¹⁹ axés sur un modèle perceptif avancé. Mais il implémente aussi déjà des lois de spatialisation similaires à celles employées dans les systèmes de spatialisation orientée objet (VBAP, DBAP, ambisonique). Néanmoins, bien que le *Spat v3* implémente un bon nombre des paramètres que nous avons tenté d'effectuer manuellement lors de nos essais de création en mode objet, le plugiciel n'est pas encore utilisable sous un univers orienté objet propriétaire comme le DTS MDA, Dolby Atmos, etc.

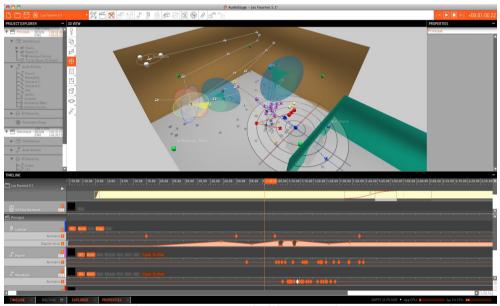


Figure 2.3 : Principe de mise en espace et de fabrication d'une acoustique en fonction de la modélisation de salle choisie.

Il semblerait, pour le moment, qu'un gouffre existe entre les possibilités créatrices rendues possibles par les dispositifs orientés objet et les outils permettant de les formaliser. Des recherches sur le rendu d'un environnement sonore complexe par un moteur audio ont été menées par Nicolas Tsingos, Emmanuel Gallo et George

¹¹⁹ Le son est spatialisé en prenant en compte la variation de l'intensité sonore en fonction de la distance, la perte des aigus par éloignement, la directivité de la source, la présence de la source et son orientation par rapport à l'auditeur, etc. Le *Spat v3* permet aussi de prendre en compte l'acoustique d'une source dans un espace grâce aux effets de réflexion primaire et de réverbération.

Drettakis¹²⁰. Leurs écrits proposent un moteur audio permettant de placer une source sonore dans un environnement ou encore dans une architecture particulière.

De cette recherche et à l'heure actuelle, seuls quelques développeurs incorporent à leurs logiciels une modélisation de l'environnement. C'est le cas de Longcat Technologies, société française spécialisée dans le développement de technologies audio 3D, qui propose avec son logiciel de synthèse *Audiostage*, une ergonomie et des outils adaptés aux nouvelles manières d'élaborer la matière sonore induites par le mode objet (Cf. *Figure 2.4*).

De toute évidence, le logiciel *Audiostage* incorpore dans sa conception les nouveaux codes de création sonore. En effet, nos expériences menées en 2.1.2 nous ont amené à recréer un espace sonore cohérent en adéquation avec le déplacement de la caméra dans le décor du film. Nous avons avancé le concept de *caméra spectateur* (le spectateur est mis à la place de la caméra et suit ses déplacements dans le décor, comme s'il était lui-même en train de s'y déplacer), nous constatons que l'on peut appliquer cette même notion d'un point de vue sonore à travers le *microphone spectateur*.

La prise de son directe témoigne déjà en partie de l'espace sonore de ce décor. Cependant, cette prise de son étant le plus souvent monophonique, elle ne permet pas le rendu spatial du décor sonore. Le montage son, à travers le concept de microphone spectateur, cherche à récréer la dimension spatiale de ce décor. Il est possible de répondre à cette forme de mise en scène dans le monde de l'orienté canal, mais ce processus de création est plus abouti dans une approche objet. Cependant, nous venons de voir que sa mise en place est très complexe.

¹²⁰ Nicolas TSINGOS, Emmanuel GALLO et George DRETTAKIS, « Perceptual audio rendering of complex virtual environments », REVES/INRIA Sophia-Antipolis, 2004.

Le logiciel *Audiostage* propose de recréer spatialement ce décor simplement. Tout d'abord, en le modélisant de manière architecturale puis dans un second temps en composant un environnement sonore au sein de cet espace. Le décor sonore est rendu à partir d'un point d'écoute assimilé à un microphone virtuel. À la manière d'un perchman dans une prise de son directe, ce microphone virtuel va pouvoir se mouvoir dans un espace qui, cette fois, est virtuel.

Ainsi, ce décor sonore modélisé et « capté » par le microphone virtuel, est alors rendu (réellement) dans la salle de cinéma, plaçant le spectateur au sein du nouvel espace.

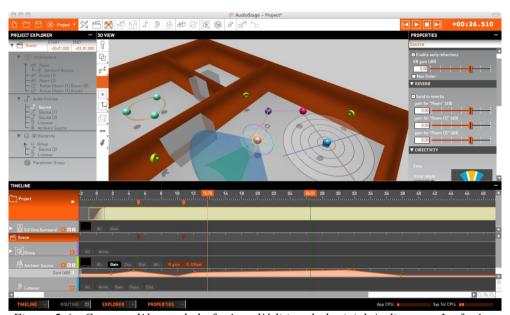


Figure 2.4 : Capture d'écran de la fenêtre d'édition du logiciel Audiostage. La fenêtre d'édition 3D View propose une représentation graphique de la modélisation de l'espace dans lequel évolue le spectateur.

En bref, l'intérêt de ce type de modélisation est de permettre aux ingénieurs du son une utilisation complète des possibilités de l'approche objet, sans pour autant avoir à en assumer les lourds processus techniques de mise en place. La démarche créatrice est alors résumée au choix des sources et à l'architecture dans laquelle nous voulons les mettre en espace.

Pour le moment, ce logiciel fonctionne avec les systèmes de diffusion orientée objet proposés par DTS et Barco. Il ne semble pas exister d'adaptation pour le système Dolby Atmos ou encore le système ouvert européen ADM. De plus, les objets sonores manipulés dans le logiciel n'ont pas de correspondance avec ceux utilisés par les dispositifs orientés objet. La spatialisation s'effectue donc en interne du logiciel en fonction de la configuration électroacoustique dont l'utilisateur dispose. Dans ce sens l'utilisation du logiciel n'implémente pas la transportabilité proposée par les processeurs orientés objet.

Cette application permet de déplacer l'écoute du spectateur dans une modélisation de l'environnement. En effet, ce logiciel permet de modéliser un espace en y insérant des éléments architecturaux (murs, portes, etc.) afin de simuler une réalité acoustique la plus en accord possible avec la scène visuelle. Le point d'écoute du spectateur peut alors évoluer au sein de la scène sonore animée, la cohérence acoustique étant calculée en temps réel en fonction du positionnement du spectateur virtuel dans la modélisation de la salle (Cf. *Figure 2.3*).

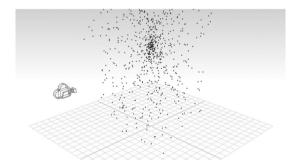


Figure 2.5 : Illustration du principe de création d'images de synthèse par synthèse particulaire. En remplaçant la caméra virtuelle par un microphone virtuel, le parallèle peut être fait pour les applications audio.

L'utilisation de microphones virtuels s'intègre récemment dans les études de Charles Verron et George Drettakis dans *Procedural audio Modeling for Particle-Based Environnemental Effects*¹²¹ et de Nuno Fonseca dans son étude *3D Particle Systems for Audio Applications*¹²². Ces recherches s'inspirent de la création d'images de synthèse, dont la conception est basée sur des systèmes de particules ¹²³ animées dans un espace 3D virtuel dont une caméra, elle aussi virtuelle, vient capter une image en 2D (Cf. *Figure 2.5*).

Appliqué au monde de l'audio, Nuno Fonseca décrit un système de rendu audio permettant de capter cet espace virtuel sonore global créé par plusieurs particules évoluant dans le temps. Ces particules sont des points dans l'espace qui ont une durée de vie, qui peuvent se déplacer ou rester statiques et interagir entre eux. L'utilisateur ne contrôle pas chaque particule individuellement, mais il peut agir sur les paramètres globaux du logiciel comme le nombre de particules en jeu, leur agitation aléatoire, etc. Ce rendu s'effectue à l'aide d'un ou de plusieurs microphones virtuels répartis dans l'espace qui viennent capter la scène sonore créée. Chaque microphone virtuel délivre un flux audio dans le but d'alimenter les systèmes de haut-parleurs. Il nous semble intéressant que dans de futurs travaux, de tels moteurs de rendu audio puissent fonctionner directement en parallèle avec les processeurs de diffusion orientée objet. Dans ce cas, les particules peuvent être directement assimilées à des objets sonores permettant de bénéficier de la transportabilité et de la spatialisation proposées dans le mixage objet.

¹²¹ Charles Verron et George Drettakis, « Procedural audio modeling for particle-based environmental effects », Convention Paper 8764, Presented at the 133rd AES Convention, San Francisco, 2012.

¹²² Nuno Fonseca, « 3D Particle Systems for audio Applications », DAFx-13, proc. of the 16th Int. Conference on Digital Audio Effects, Maynooth, 2013.

¹²³ Un système particulaire est un dispositif stochastique fait de plusieurs petites particules qui évoluent dans le temps. Ce concept est utilisé dans le domaine de l'image de synthèse pour obtenir plusieurs effets visuels comme la pluie, le feu, la fumée, la poussière, etc.

Une des applications les plus évidentes se situe dans la création sonore (synthèse de sons). En effet, plusieurs particules peuvent être utilisées simultanément pour reconstituer un événement sonore complexe. Nous pensons aussi qu'à partir d'une banque son d'éléments ponctuels, les systèmes particulaires peuvent reconstituer une ambiance complexe en traitant chaque particule de manière indépendante (en niveau, en timbre, en durée, etc.); Comme cela a été expérimenté par Charles Verron et Georges Drettakis qui ont essayé de synthétiser (virtuellement) une ambiance de pluie, de feu, etc.

Il apparaît que les recherches et développements commencent à entrevoir les possibilités qu'ouvrent les formats orientés objet. Néanmoins, il est encore tôt pour entrevoir quelles directions esthétiques vont naître à la prise en main du mode objet par le corps professionnel. Au cours de notre étude nous avons tenté de proposer une approche de création sonore fusionnelle avec ce que nous avons appelé la *caméra spectateur*. Ce concept nous a amené à suggérer la notion de *microphone spectateur*. La création sonore semble s'orienter vers une quête de réalisme acoustique où le choix des sons s'accompagne d'une spatialisation qui devient de plus en plus complexe afin d'approcher la sensation d'assister (de l'intérieur) à une scène sonore. Nous avons vu que pour le moment, les outils permettant d'atteindre ces objectifs ne sont pas encore totalement adaptés à cette logique de création. Cependant, il nous semble que les chercheurs et les développeurs empruntent le chemin de la synthèse de cette scène sonore par différents moyens, notamment par la recréation d'un espace sonore virtuel lui-même capté par des microphones fictifs.

Ceci permettant de gérer plusieurs paramètres de prise de son afin d'obtenir le résultat le plus en adéquation possible avec le processus filmique. En revanche, nous constatons que ces processus de spatialisation se basent tous plus ou moins sur la monophonie dirigée. En outre, ces dispositifs tentent de s'approcher du rendu des prises de sons multicanal comme le HOA. Bien que les paramètres gérés par les

logiciels de simulation d'espace ne puissent pas être modifiables dans le cadre d'une prise de son, une étude comparative et approfondie reste à mener entre les deux dispositifs. En effet nous pensons que les systèmes de prise de son HOA peuvent avoir une utilité dans les systèmes orientés objet alors que leur utilisation est rare dans des dispositifs orientés canal. Selon nous, deux conceptions de la création sonore voient le jour, l'une orientée vers la synthèse d'espaces sonores virtuels et l'autre orientée vers la captation in situ d'espaces sonores réels.

2.2 Particularités de la postproduction des formats orientés objet

2.2.1 Intégration du mode objet dans la chaîne de travail

Durant la confection des séquences de notre partie pratique, nous nous sommes confrontés au côté organisationnel que la création sonore en mode objet implique dans la chaine de postproduction. Cette confrontation nous semble d'autant plus légitime que les avis et les pratiques divergent, constat représentatif d'un nouvel outil dont les codes et l'organisation dans la chaine de travail restent à découvrir. De plus il apparaît que la spatialisation des éléments sonores semble influencer la façon d'appréhender la création, ce qui selon nous fait bouger les frontières des différents corps de métier.

Actuellement, parmi les films réalisés dans des dispositifs orientés objet plusieurs approches ont été tentées. D'une manière générale, la création en mode objet débute soit durant le montage son soit pendant le mixage. Notons cependant que le terme mixage orienté objet est le plus couramment utilisé, soulignant ainsi la phase dans laquelle est envisagé le passage au mode objet. Or, nous proposons d'employer le terme d'approche objet, qui lui souligne l'importance de penser ce processus de création de manière plus globale.

2.2.2 EN PHASE DE MONTAGE SON

Dans la chaine de post-production audio, chaque acteur de l'élaboration de la bande son travaille dans des lieux différents où les configurations sont variables. Ceci est principalement applicable au secteur du montage son qui s'effectue dans des salles de taille moyenne dans un environnement 5.1 avec des écoutes de proximité. Nous l'avons vu, le mode orienté objet tend à rendre transportable le travail effectué dans différents lieux d'écoute, ce qui lui permet d'être réalisé au stade du montage son.

À l'heure actuelle deux manières sont référencées pour l'élaboration des montages sons en vue d'un mixage objet. La première consiste à travailler sous un univers orienté canal, avec des dispositifs de diffusion dits classiques, comme le 5.1, et le passage en mode objet s'effectue en phase de mixage. Il n'y a aucune possibilité de mise en espace d'objets sonores. De ce fait les différentes possibilités créatrices des dispositifs objets ne sont pas prises en compte et le montage son semble s'orienter vers les codes de création propres à l'utilisation des formats multicanal classiques. En outre, si le montage son prévoit des sons pour le mode objet, la réflexion est quelque peu biaisée puisque pensée et écoutée sous un univers orienté canal. Ce processus est révélateur d'une réalité industrielle. Très peu de producteurs acceptent de post-produire leurs films dans des formats dits de prochaines générations, et, quand certains l'envisagent, le format principalement diffusé restera un standard répandu dans toutes les salles.

Nous nous questionnons alors sur l'importance narrative de ces sons puisque de toute évidence ils sont généralement prévus et pensés lors d'un dernier passage de montage son additionnel. En effet, ce mode de fonctionnement implique que le montage son soit divisé en deux phases. La première vise à délivrer un travail dans un format classique en accord avec les intentions du réalisateur. La deuxième elle, semble totalement dévolue à créer du contenu pour le nouveau système de diffusion et non plus pour le film. Cette approche peu convaincante et totalement désengagée

du processus narratif filmique ne peut que créer des bandes son objet à caractère gadget puisque l'essentiel de la narration passe par le format orienté canal principal.

Une deuxième manière de concevoir la matière sonore préconise d'appréhender la bande son en mode objet dès le montage son. C'est d'ailleurs le mode de fonctionnement que nous avons adopté lors de la confection de certaines séquences de notre partie pratique. Ce processus permet de prévisualiser le rendu du film en salle, ce qui influence totalement la réflexion et le travail de recherche de sons. D'autant plus que les codes d'utilisation esthétique semblent refondés et nécessitent un apprentissage stylistique. En outre, cette manière de concevoir la matière sonore nous semble la plus adaptée pour créer un contenu sonore en adéquation avec les intentions scénographiques du réalisateur.

Par ailleurs une création sonore fortement liée au travail de spatialisation estompe les frontières de la répartition des tâches en fonction des métiers. En effet, il apparaît que cette tâche, communément rattachée au travail du mixeur, tend à se préparer en phase de montage son. Selon nous, deux facteurs propres au mode objet encouragent ce fonctionnement. Premièrement, la transportabilité du mode objet permet de conserver la spatialisation effectuée en phase de montage vers les salles de mixages, ce qui n'est pas forcément le cas lorsque l'on travaille sous un environnement orienté canal. Deuxièmement parce que la lourdeur technologique et technique nécessite du temps de préparation et que ce processus ne semble pas (ou plus) s'adapter au temps accordé en phase de mixage.

2.2.3 EN PHASE DE MIXAGE

Dans certains cas, il arrive même que l'écriture scénographique en mode objet intervienne en phase terminale de mixage, révélant les indécisions et hésitations de la production à effectuer le film dans un format Dolby Atmos, DTS MDA, etc. Cette décision tardive implique que la postproduction du film s'effectue à partir d'un montage son orienté canal classique 5.1 ou au mieux 7.1. Bien entendu, ces types de montages n'excluent pas l'utilisation de formats objets, mais soulignent qu'ils ne sont pas exploités à leur juste titre et accentuent leur utilisation gadget comme nous l'avons introduit précédemment dans le cadre du montage son. Cependant, et dans bien des cas, une recherche de scénographie sonore est entamée entre le réalisateur, le monteur son et le mixeur. Dans cette optique, le montage son évolue durant le mixage en s'imprégnant des différents retours sur le vif des autres corps de métier. Cette manière de travailler, souvent associée à une conception américaine, ne semble pas inintéressante étant donné la qualité des bandes sons produites. Effectivement, la scénographie sonore du fameux Gravity¹²⁴, cité à maintes reprises dans ce mémoire, fut essentiellement conçue durant le mixage¹²⁵. Néanmoins, comme l'argumente Benjamin Wright¹²⁶, cette prise en main a été rendue possible par le soutien financier de Dolby qui a financé les jours de mixage additionnels pour le mixage orienté objet. Par ce simple constat, nous questionnons la possibilité d'effectuer le passage des sons en mode objet en phase de mixage. Il paraît donc évident qu'une décision du passage en mode objet prise en amont de la phase de mixage (donc dès la phase de montage) est plus judicieuse, tant d'un point de vue artistique que d'un point de vue économique.

¹²⁴ Alfonso Cuarón, Gravity, op. cit.

¹²⁵ Le mixage a été réalisé en partie par le mixeur Skip Lievsay. Une première version 7.1 a été réalisée, puis dans un second temps un mixage objet en Dolby Atmos a été entrepris.

¹²⁶ Paul Théberge, Kyle Devine et Tom Everrett (dirs.), Living stereo, op. cit., p. 238.

Le film *Les Saisons*¹²⁷ offre une alternative intéressante puisqu'il effectue tout le travail préparatoire propre aux formats orientés objets en phase de montage son. La spatialisation est alors effectuée en amont du mixage. En effet, ceci est rendu possible car les informations de spatialisation passent par les métadonnées, la communication entre la salle de montage et l'auditorium s'effectue de manière transparente via une simple session informatique. D'autre part, les retouches peuvent s'effectuer de manière très simple. Selon Gérard Lamps¹²⁸, le mixage orienté objet permet de décloisonner les frontières existantes entre le montage son et le mixage. Celui-ci explique que cela lui permet de se (re) centrer sur des aspects de mixage purs, comme par exemple le niveau des éléments entre eux et la focalisation sonore.

2.2.4 Vers une refonte des métiers

De toute évidence les formats orientés objets proposent une réévaluation des idéologies et des pratiques de la conception de la bande son d'un film. En effet, nous constatons que les propositions de bandes son effectuées en mode objet tentent de faire bouger les croyances, les convictions et les philosophies qui semblent scléroser la création sonore. Selon Howard Becker « Les conventions artistiques orientent les décisions qui doivent être prises au regard du travail produit dans une catégorie artistique donnée, bien qu'une convention particulière puisse être révisée en fonction d'un projet donné¹²⁹ ». Dans ce sens, les conventions régulent les paramètres formels d'un travail effectué par les artistes en leur fournissant un modèle permettant une plus grande cohésion artistique collective.

¹²⁷ Jacques Perrin et Jacques Cluzaud, Les Saisons, op. cit.

¹²⁸ Propos recueillis lors d'un entretien téléphonique le 4 décembre 2015.

¹²⁹ Howard S. Becker, *Art As Collective Action*, American Sociological Review, 1974, vol. 39, nº 6, p. 770-771. Traduction par nos soins à partir de l'anglais « artistic conventions cover all the decisions that must be made with respect to works produced in a given art world, even though a particular convention may be revised for a given work ».

Or, il apparaît que les conventions de mixage et de montage son ont très peu changé depuis les neuf décennies de cinéma sonore, bien que chaque innovation technique ait tenté de les remodeler sans grand succès. Néanmoins certains professionnels¹³⁰, apparemment moins conservateurs, y voient une manière de concevoir la bande son d'un film de façon déconventionnée, proposant à l'extrême de penser la bande son en mode objet et d'utiliser la version automatique dépendante 5.1 ou 7.1. Des avis convergent pour dire que le format objet a un caractère émancipateur et que les versions descendantes automatiques dépassent ce qui aurait pu être fait de manière native.

2.3 CONCLUSION ET TENDANCES À VENIR

De notre point de vue, certaines formes cinématographiques encore minoritaires semblent s'orienter vers un cinéma où l'aspect sensoriel s'intègre de plus en plus dans la manière de concevoir un film. Cette forme de cinéma ne se cantonne pas aux films grand spectacle, où les effets sensationnels, souvent utilisés à outrance, s'ancrent plus dans une logique de saturation sensorielle que de narration sonore; mais aussi à des films dont le sujet semble moins axé sur le spectaculaire que sur le dramatique comme *Les 33*¹³¹ ou encore sous la forme de comédie dramatique avec *Eddie the Eagle*¹³².

Bien que les concepteurs de formats audio orientés objet espèrent redéfinir le côté technique, pratique et expérimental du cinéma audio, leurs réussites à s'imposer comme solution viable et innovante dépend largement de la communauté audio, celle-ci étant responsable du contenu sonore des films. Or, cette communauté artistique est constituée de professionnels ayant des pratiques conventionnées induisant

¹³⁰ C'est le cas de Sreejesh, mixeur Indien qui a récemment mis en ligne ses expériences de mixage réalisées sous divers formats dont le format orienté objet proposé par Dolby. Ses mémoires sont consultables sous forme d'article sur le site internet https://film-mixing.com/.

¹³¹ Patricia RIGGEN, The 33, 127 min, Chili, Etat-Unis, sorti en France le 2 mars 2016, 2015.

¹³² Dexter Fletcher, *Eddie the Eagle*, 106 min, États-Unis, Royaume-Uni, Allemagne, sorti en France le 4 mai 2016.

leur identité et leurs compétences spécifiques, ancrées dans des années de pratique. On peut donc supposer que la nouvelle approche objet mettra du temps à s'intégrer aux mœurs, en dépit des efforts des fabricants pour imposer leurs systèmes.

Le travail créatif résultant n'en est pas moins dépendant des conventions et des techniques partagées privilégiant la clarté, l'intelligibilité et la narration des sons. Ce travail, sur les formats orientés canal, est pleinement contrôlé, le mixage que livre le mixeur constitue le produit artistique fini, écartant la possibilité à un programme informatique d'en modifier ses composantes. Avec les formats orientés objet, le processus et le contrôle artistique sont plus flous. Il apparaît que l'écriture finale que constitue le mixage est beaucoup moins figée. La logique générale de travail s'en trouve modifiée et les pré-mixages rendus obsolètes. Les formats orientés objet poussent à ne plus enregistrer la conséquence des sources mais les sources elles-mêmes. Avec les processeurs de rendu numérique, le travail de mixage (non figé) est prolongé par la machine, ce qui peut laisser perplexes les professionnels sur le rendu de leur travail par la machine. Dans ce même courant de pensée, ces formats encouragent les mixeurs à automatiser une partie de leur travail pour former les différentes versions de leur labeur. Par conséquent, les processeurs de rendu qui délivrent automatiquement une version 5.1 et 7.1 d'un mixage orienté objet original, lancent un défi aux mixeurs et aux monteurs dans leur manière de créer et de conserver une identité artistique. Remarquons que ce phénomène n'est pas spécifique à la rétrocompatibilité des formats audio mais que l'on peut faire le rapprochement avec les filtres photographiques sur les réseaux sociaux, les filtres acoustiques sur nos baladeurs, etc. Dans cette vague d'uniformisation artistique, il convient à chacun de mener sa propre réflexion.

Pour le moment, la prochaine génération de formats audio semble proposer des possibilités artistiques intéressantes et innovantes. Néanmoins, ils requièrent un auditeur et des praticiens pour en étudier et assimiler la conception et l'esthétisme.

Pour le moment nous pensons que leurs possibilités créatrices restent limitées par des contraintes économiques et humaines, principalement gouvernées par les conventions et les différentes idéologies de pratique.

Cependant, il est possible d'affirmer que de tels mixages nécessitent un temps supplémentaire pour la conception du film. Premièrement, au niveau de la recherche qu'ils impliquent en phase de montage son où, là aussi, l'utilisation esthétique d'un tel système reste à approfondir. Deuxièmement, en phase terminale de mixage, où bien des partis pris de mise en scène sont choisis en très peu de temps, parfois au préjudice des possibilités qu'offrent ces nouveaux dispositifs. *En Solitaire*¹³³ de Christophe Offenstein semble en être un exemple flagrant. En effet, pour des raisons de temps, le film a été envisagé au montage son dans un format 7.1. Ce n'est qu'en fin de mixage (les deux derniers jours), qu'une version Dolby Atmos fut entreprise. Ceci est symptomatique d'un système qui tente de s'imposer, mais qui pour des contraintes extérieures se trouve relayé en fin de production et de manière facultative. Dans cette mouvance de travail et de manque de temps, la compatibilité descendante est effectuée automatiquement. Ceci peut néanmoins fonctionner comme dans le cas d'une mise en scène où le mixage orienté objet est peu utilisé.

D'autre part, il existe une production cinématographique dans la logique scénaristique de laquelle l'utilisation du mixage orienté objet semble s'intégrer pleinement. C'est le cas du film *Gravity*¹³⁴ déjà cité à plusieurs reprises dans notre étude ou encore du film *Les Saisons*¹³⁵ de Jacques Perrin et de Jacques Cluzaud. La production de ces films a la particularité d'avoir pensé en amont aux différentes possibilités techniques et esthétiques que proposent les systèmes de mixage orientés objet. En l'occurrence, des tests de montage son et de mixage orienté objet

¹³³ Christophe Offenstein, En solitaire, 101 min, France, sorti en France le 6 novembre 2013.

¹³⁴ Alfonso Cuarón, Gravity, op. cit.

¹³⁵ Jacques Perrin et Jacques Cluzaud, Les Saisons, 97 min, sorti en France le 27 janvier 2016.

ont été entrepris en phase de montage image. Cette démarche semble pour nous la plus logique et la plus complète puisqu'elle réunit et convoque les différents acteurs de la post-production du film. Les possibilités scénographiques augmentées permises doivent être préparées au moins en phase de montage son afin de proposer une mise en scène sonore réfléchie et subtile pour ne pas tomber dans l'effet de foire symptomatique de l'apparition de nouveaux dispositifs de diffusion. Des films comme *The Revenant* ¹³⁶, Sicario ¹³⁷ ou encore American Sniper ¹³⁸ semblent renouveler leur mise en scène en connaissance des potentialités du son spatialisé.

¹³⁶ Alejandro G. IÑÁRRITU, The Revenant, op. cit.

¹³⁷ Denis VILLENEUVE, Sicario, 121 min, États-Unis, sorti en France le 7 octobre 2015.

¹³⁸ Clint Eastwood, American Sniper, 133 min, États-Unis, sorti en France le 18 février 2015.

3 PARTIE PRATIQUE: ANALYSES PERCEPTIVES DU MIXAGE ORIENTÉ OBJET

3.1 Contexte d'étude

3.1.1 RÉFLEXIONS PRÉLIMINAIRES

À la base de notre mémoire de recherche, nous avions décidé de ne pas réaliser de tests perceptifs, pensant que leur mise en œuvre et leurs débouchés n'étaient pas en accord avec notre approche d'étude préliminaire des systèmes orientés objet. Néanmoins, lorsque nous en avons dégagé les quelques approches d'utilisation, il est apparu nécessaire d'en proposer une étude objective, ne serait-ce que pour relativiser nos conclusions préalables avancées en 2. Notre intention ici est de pouvoir conclure ou du moins ouvrir notre recherche au domaine principal auquel les systèmes de diffusion se destinent : proposer une expérience cinématographique au spectateur. Nous avons donc décidé de soumettre nos expérimentations à un public de futurs professionnels du son (les étudiants de la section son de l'Ecole Nationale Supérieure Louis-Lumière) pour des raisons que nous développerons en 3.2.4.

Cependant, durant notre vaste étude sur les formats objets, il nous est apparu primordial d'orienter nos tests perceptifs sur les critères que le mixage orienté objet prétend améliorer. Il est fondamental pour nous de prendre du recul face aux caractéristiques immersives et réalistes que les principaux formats objets émergents proposent. Bien que nous ne prétendions pas dégager l'avenir d'un nouveau langage sonore cinématographique, nous avons vu dans notre partie précédente que l'esthétique du mixage tend à se centrer sur l'écoute et sur l'expérience sensorielle

de l'auditeur. Il est pour nous primordial d'observer s'il existe une relation entre expérience sensorielle et narration cinématographique.

De manière logique nous nous sommes appuyés et inspirés des expériences menées par Thibaut Macquart et Rémi Carreau dans leur étude de l'utilisation de la WFS dans la création sonore cinématographique. Les deux auteurs proposaient d'évaluer l'apport de la WFS et de vérifier si « la construction d'un espace sonore différent via l'utilisation d'un système qui, par nature, propose une autre approche d'illusion perceptive, influence notre perception d'un objet audiovisuel ¹³⁹ ». Leur évaluation passait par une étude comparative avec le système 5.1 pour deux raisons. Premièrement, pour faciliter l'évaluation de l'apport de la WFS en effectuant une comparaison avec un système de référence connu. Enfin « il s'agissait bien d'évaluer la WFS dans un cadre audiovisuel, domaine dans lequel le 5.1 est largement implanté, maitrisé et semble, malgré tout, bien dimensionné pour répondre aux problématiques perceptives du son à l'image ». Aussi notre étude intègre un mixage 5.1 classique comme élément de référence mais, à la différence de l'étude citée, nous mettons en garde notre lecteur sur le caractère conventionnel, pratique et idéologique qu'il présuppose.

3.1.2 Principaux axes d'études

Nous avons vu, au regard de notre partie précédente, que l'esthétique du mixage tend à se centrer sur l'écoute et sur l'expérience sensorielle de l'auditeur et il est pour nous important de savoir si cette démarche est valable dans le cadre de la narration cinématographique. De plus, nous voulions savoir si cette esthétique est corrélée à l'apparition de formats nouvelle génération, comme certains réalisateurs l'avancent¹⁴⁰. D'autre part, cette expérience spectatorielle (totale) n'est commu-

¹³⁹ Rémi Carreau et Thibaut Macquart, Utilisation de la technologie WFS dans la création sonore cinématographique : posibilités et limites, op. cit., p. 206.

¹⁴⁰ Nous faisons référence à une interview vidéo du réalisateur Alfonso Cuaron qui explique que son film

niquée que par des systèmes orientés objet encore minoritairement installés dans les salles. En ce sens, la version d'exploitation du film la plus en accord avec les intentions sonores artistique n'est vue que par un faible pourcentage du public. Pour répondre à cette problématique, les processeurs de diffusion objet ont prévu une compatibilité descendante (Cf. Annexe F) et proposent des mixages 5.1 et 7.1 automatiques.

Bien entendu, l'enjeu artistique que représentent les downmixages automatiques devait faire partie de notre investigation. En effet, il nous était fondamental d'évaluer si les intentions artistiques étaient perçues de la même manière que dans le mixage original. Dans ce test nous avons donc tenté d'amener des éléments de réponse afin d'apprécier si le downmixage automatique est un processus viable au regard de la production artistique.

Des systèmes comme le Dolby Atmos tentent par le mode de fonctionnement orienté objet de répondre à la question de la transportabilité. Néanmoins, cet aspect technique permettant de contrôler la transportabilité de la spatialisation des sources passe à travers un procédé de matriçage dont les technologies sont gardées secrètes. Ce grand flou technologique dans la chaine du son peut devenir effrayant au moment où la généralisation du DCP et du 5.1 discret permettait un contrôle de la chaine audio par les techniciens du son.

3.1.3 Problématiques liées à la compatibilité descendante :

Lors de la prise en main du dispositif de spatialisation décrit en 2, plusieurs paramètres nous ont semblé intéressants à retenir : la pleine bande arrière, le faible taux de distorsion et la grande dynamique sonore. Néanmoins, tous ces critères peuvent devenir problématiques lorsqu'ils sont transposés sur des systèmes

n'aurait pas eu le même impact immersif s'il n'avait pas été conçu en Dolby Atmos. Interview issue de https://www.youtube.com/watch?v=a-PsjshThNM et consultée le mercredi 11 mai.

antérieurs. Le système 5.1, système le plus répandu dans les salles, est de qualité très hétérogène et les caractéristiques techniques des dispositifs de diffusion (hautparleurs, amplificateurs, etc.) se détériorent au fil des années et des projections successives. Les downmixeurs automatiques ne prennent pas en compte les capacités réduites des systèmes antérieurs par rapport aux spécifications préconisées par les nouveaux. En effet, le contenu basses fréquences des signaux spatialisés dans la salle par des enceintes à large bande se retrouve diffusé par un réseau d'enceintes d'ambiances, généralement à la réponse en fréquences réduite et au taux de distorsion élevé.

Les sons dont le contenu spectral est large, diffusés de manière homogène et précise dans toute la salle, se retrouvent diffusés par une rampe de haut-parleurs émettant le même signal. D'une part, la précision spatiale devient floue et étalée dans toute la salle mais son contenu spectral va être altéré. Cela peut aller de la suppression d'une partie du contenu (par filtrages ou interférences destructives, etc.) à l'augmentation en niveau du contenu (par interférences constructives, etc.). Ce qui est le cas des ambiances diffusées en plusieurs points objet ayant une forte corrélation. Le contenu hautes fréquences se retrouve déformé par le phénomène de filtrage en peigne et les basses fréquences se retrouvent augmentées en niveau. Le contenu sonore arrière se retrouve profondément dénaturé et les sensations d'enveloppement et d'immersion, produites par un mixage objet, se retrouvent diminuées voire à contre sens dans un mixage automatique orienté canal. En outre, en s'éloignant des intentions premières du réalisateur, ce contenu sonore peut même devenir problématique au regard des conflits scénographiques sonores.

Notre étude sur la compatibilité descendante proposée par les différents formats orientés objet nous amène aussi à nous interroger sur l'état de la diffusion sonore dans les salles de cinéma standards. Il existe encore une très grande disparité de qualité de restitution dans les salles. Les petits exploitants n'ont pas les moyens de

mettre à jour leur installation vieillissante. Les grands multiplexes, qui investissent dans leur installation, n'ont plus le personnel qualifié pour la maintenance et la bonne utilisation de ces systèmes. Les projectionnistes sont remplacés par des serveurs supervisant les projections à des kilomètres de la salle de cinéma.

D'un autre côté, comme décrit en Annexe C, des recommandations de diffusion sonore tendent à améliorer les conditions d'écoute en salle. Ces recommandations sont censées pousser les exploitants vers une homogénéisation de leurs salles, permettant ainsi une meilleure transportabilité du contenu artistique produit en amont. Monteurs et mixeurs ont une connaissance globale de l'état actuel de la diffusion au cinéma et tentent d'adapter leur travail pour le retrouver dans une pluralité de salles de cinéma. Ce processus bride souvent les potentialités artistiques d'un film et conduit à une utilisation restreinte des systèmes existants.

Le processus entamé par les industriels de la diffusion au cinéma avec les systèmes de prochaine génération vise à améliorer cette qualité de diffusion. Leur système de licence et de contrôle des salles proposant leur dispositif garantit une certaine qualité. Leur système économique a la vertu d'homogénéiser un certain nombre de salles. Ainsi, le récent regain d'énergie des différents constructeurs semble aller dans le sens des professionnels du cinéma. Selon nous, certains films¹⁴¹ récents illustrent ce nouveau souffle et semblent se libérer des contraintes, entraînant les films dans une logique d'amélioration.

¹⁴¹ Nous pensons notamment à *The Revenant* d'Alejandro González Iñárritu qui propose des tentatives de mise en espace des voix, bien que cette tentative ait déjà été réalisée sans grand succès avec le système 5.1. La voix normalement diffusée par l'enceinte centrale au cinéma, se trouve spatialisée sur toute la surface de l'écran et une partie de la salle. Pari audacieux, selon Florent Lavallée, qui par sécurité opte pour une diffusion de la voix au centre. Selon lui, l'enceinte centrale est une des enceintes dont le fonctionnement est indispensable, puisqu'elle contient le message principal du film. Ainsi il sera plus difficile pour des exploitants et des spectateurs de ne pas s'apercevoir d'un dysfonctionnement de celleci.

Ces films, dont le format d'origine n'est proposé que par une minorité de salles, seront diffusés dans leur version descendante. D'autre part, quel parti pris adopter lorsque les dispositifs de diffusion qualifiés de prochaine génération apportent une plus-value à la mise en scène du film, sachant que la version descendante sera diffusée plus largement? Il semblerait que deux courants se dégagent. Le premier, partisan de la quête artistique, n'effectue aucun compromis¹⁴² en mettant de côté les formats descendants. Le deuxième effectue des compromis aux dépens des effets de mise en scène sonore possibles. Ces réflexions ne sont pas propres aux nouveaux formats actuels mais semblent, à la différence de leurs prédécesseurs, exacerbées par les possibilités de spatialisation accrue qu'ils proposent. Les enjeux semblent d'autant plus problématiques que dans bien des cas, les versions descendantes sont fabriquées automatiquement par voie logicielle (décrit en Annexe F). Ceci étant symptomatique du temps de postproduction court, rendant le mixage manuel de versions descendantes impossible, relayant à la machine ce que l'humain devrait théoriquement faire. Aujourd'hui, un fossé technologique existe entre les formats de diffusion de prochaine génération et les systèmes classiques. La compatibilité descendante ne semble pas présenter les mêmes enjeux qu'auparavant.

3.2 PROTOCOLE DU TEST PERCEPTIF

3.2.1 Présentation générale du test

Le test a été soumis aux étudiants de la section son de l'école nationale Louis-Lumière. Les tests ont été effectués entre le mercredi 4 et le dimanche 8 mai 2016, dans la salle de montage son équipée d'un système Dolby Atmos 15.1 de Raphaël Sohier. Par ailleurs le lecteur trouvera un descriptif de la configuration des sessions

¹⁴² Le parallèle peut être fait dans le domaine de l'image, avec le cas du film *Les Huits Salopards* de Quentin Tarantino. Le film a été projeté dans son format d'origine (70 mm) dans une seule salle en France.

ProTools à l'utilisation du Dolby Atmos en Annexe E.1.. Les sessions tests avaient une durée de trente à quarante minutes.

Chaque sujet passait le test de manière individuelle, idéalement placé au *sweet spot* afin d'écarter de nos résultats toute incohérence de jugement liée à un mauvais placement de l'auditeur.

3.2.2 DÉFINITION DES PARAMÈTRES D'ÉTUDES

Ce type de projection test a un but esthétique, culturel et technique. En premier lieu, nos interrogations porteront sur la cohérence sonore et visuelle de l'univers sonore installé dans les séquences, sur sa vraisemblance (acoustique et spatiale) et sur l'immersion qu'il est censé provoquer.

Hypothèses d'études :

Hypothèse 1 : La progression des systèmes de diffusion cinématographique vers un univers de mixage orienté objet repousse les limites perceptives d'écoute. Des systèmes proposant un grand nombre de points objets sur une dimension verticale et horizontale peuvent améliorer l'impression de réel, d'immersion et d'implication narrative chez le spectateur.

Hypothèse 2 : Cette précision et l'amplitude de spatialisation permettent d'intégrer au film un aspect sensationnel et une immersion plus importante chez le spectateur. Ceci augmente l'amplitude immersive et la crédibilité en allant de la monophonie au 9.4 + objets sonores. Ce large choix de possibilités permet de produire des mixages différents de ceux envisagés dans un système 5.1 traditionnel.

Hypothèse 3 : L'absence de réflexion humaine dans le processus automatique de downmixage va entrainer des aberrations quant au rendu des effets créés en mode objet, qui ne peuvent pas être supportés par une configuration 5.1. Ces aberrations vont causer des problèmes d'adéquation à l'image.

Etablissement des variables :

Variables intersujet:

Tous les sujets du test sont soumis aux mêmes séquences mais dans un agencement différent. La projection de ces séquences s'effectuera dans le même lieu et dans les mêmes conditions.

Variables intrasujet:

Durant notre étude, nous nous sommes aperçu que la sensation d'immersion passait sensiblement par l'amélioration de la bande passante (dans le bas du spectre) des enceintes arrière en configuration objet. Les sons ainsi spatialisés proposent un rendu plus convainquant et naturel. Bien que le résultat puisse paraître évident, nous avons donc pris la décision d'étudier l'influence du *bass management* des enceintes arrière sur l'immersion. Par conséquent, des projections tests ont eu lieu de manière équitable sans *bass management* et d'autres avec.

<u>Critères de jugement : (variables indépendantes)</u>

Les sujets du test devront à l'issue des séquences apprécier leur sensation d'immersion durant la scène. Ce critère décrit, selon nous, l'aptitude du sujet à s'imprégner de la mise en scène et de l'histoire qui lui est racontée. Ce terme se distingue de son sens commun rattaché à l'enveloppement sonore du spectateur et

rejoint l'idée selon laquelle il devient partie prenante de la situation à laquelle il assiste 143 ».

D'autre part nous demandons au sujet d'établir une note sur ce que nous appelons la lisibilité du son. Celle-ci décrit selon nous la capacité du sujet à percevoir un élément sonore de la bande son et à l'assimiler à la narration audiovisuelle. Nous qualifions de lisible une scène dont on parvient à discerner les différents éléments sans que ceux-ci ne rentrent en concurrence les uns avec les autres, rendant la scène sonore floue. Cette lisibilité passe par différentes pratiques d'intégration des éléments sonores, d'un point de vue spatial et acoustique. Comme vu précédemment (Cf. 2.1), il nous semble que la lisibilité peut être améliorée par le mixage objet. La lisibilité passe donc par la transparence du message sonore en accord avec l'image et fait que le spectateur se retrouve inclus dans la scène, c'est un paramètre qui selon nous est corrélé à l'immersion de celui-ci.

De plus, il nous semblait intéressant de soumettre à l'étude perceptive des candidats le critère de qualité du timbre. Dans l'absolu, cette notion fait référence à la qualité du son reproduit, par rapport à sa source d'origine. Ici, le sujet ne pourra évidemment pas avoir de référent. Par conséquent, il sera amené à faire une étude comparative en prenant comme référent le mixage qui lui paraitra proposer le meilleur rendu de la matière sonore. D'autre part il paraissait intéressant pour nous d'évaluer si la différence de bande passante des enceintes arrière entre les différents formats était perceptible et pouvait être associée à d'autres critères étudiés ici. Enfin, il nous était important de confirmer ou d'infirmer ce que nous avons soutenu dans notre troisième partie : le mixage-objet permettrait un meilleur démasquage des sources impliquant une amélioration du timbre.

¹⁴³ Jacques Aumont (dir.), Esthétique du film, op. cit., p. 95.

Dernièrement, il nous est apparu indispensable de juger de l'adéquation de la bande sonore au regard du spectacle visuel. Pour nous cet aspect est primordial puisqu'il détermine, au fond, une des raisons de notre étude. L'apparition de nouveaux systèmes peut proposer une écriture sonore intéressante tout comme des abus et des utilisations erronées allant dans le sens contraire du respect de la mise en scène. L'histoire du son au cinéma (Cf. 1.1) recèle un bon nombre d'échecs et de répétitions. Neuf décennies de cinéma sonore, parsemées de dispositifs toujours plus enveloppants, ont finalement proposé un spectacle frontal.

3.2.3 LE PROTOCOLE

Le test se composait de cinq séquences images proposant chacune un mixage objet, son downmixage automatique et un mixage 5.1 classique réalisé manuellement. Au total, quinze séquences ont été soumises à l'évaluation par les sujets tests. Le sujet n'avait aucun moyen formel de distinguer l'origine du mixage qu'il écoutait. En ce sens notre test perceptif s'inscrit dans un processus d'étude absolue. Il nous semblait important que le sujet ne puisse pas comparer les différents mixages d'une même séquence au risque d'orienter son jugement en fonction du type de mixage de référence. De cette manière notre étude proposera d'étayer nos jugements sur notre comparaison relative entre les différents systèmes.

L'intégralité des quinze séquences est soumise à un premier visionnage dans un ordre aléatoire afin que le sujet s'habitue à celles-ci et à leurs mixages associés. Cette mise en condition nous paraissait nécessaire pour que le sujet puisse s'habituer aux différents systèmes qui lui étaient proposés et prenne connaissance des différences perceptives que les différents mixages proposent. De cette manière le sujet peut préalablement apprécier l'échelle de notation et commencer à l'appréhender.

Dans un second temps, le sujet rentrait en phase d'analyse des séquences, il lui était alors demandé d'évaluer celles-ci selon les quatre critères définis en 3.2.2. Afin qu'aucun quiproquo de sens ne soit effectué, un rappel de la signification des critères est effectué oralement avant chaque début de test. Chaque critère était évalué par l'attribution d'une note s'échelonnant de 0 à 10. L'évaluation s'effectue par le biais d'un formulaire papier. La phase d'analyse se décomposait en trois sessions matérialisant les trois mixages possibles que le sujet pouvait rencontrer. Les trois sessions étaient composées de 5 séquences. En ce sens l'ensemble des séquences issues d'une même session ne comportait qu'un seul type de mixage parmi les trois décrits plus haut.

Afin de minimiser l'effet d'ordre de projection des séquences et des mixages à étudier, l'ordre des sessions et des séquences les composant était choisi aléatoirement et ceci pour chaque sujet. Le sujet n'avait le droit de regarder qu'une seule fois les séquences qui lui étaient proposées dans la première phase et la deuxième phase. La main était laissée au sujet afin de qu'il puisse naviguer dans la session ProTools à son propre rythme, le temps de réponse entre chaque question était donc laissé libre. Le sujet naviguait dans les séquences grâce à l'appel de numéro de marqueur sous le logiciel d'édition ProTools. Aux quinze séquences étaient attribué un numéro entre 1 et 15. Le sujet n'avait aucun moyen de savoir à quelle séquence et mixage associé correspondait le numéro de marqueur. Ce numéro de marqueur à appeler était indiqué à chaque séquence de chaque session.

3.2.4 CHOIX D'UN PUBLIC TEST

L'établissement de nos critères d'évaluation nous a conduit à proposer ce test à un public initié au domaine du son, censé avoir reçu l'enseignement d'un langage commun et une connaissance des concepts mis en jeu dans celui-ci. Néanmoins

pour des raisons de faisabilité, le test n'a été soumis qu'aux étudiants de la section son de l'Ecole Nationale Supérieure Louis-Lumière.

D'autre part, nous notifions à notre lecteur que dans notre note d'intention ce test était initialement prévu pour un public d'experts, mais des soucis logistiques nous ont amené à revoir nos ambitions.

3.3 MISE EN PLACE DU TEST

3.3.1 Préparation des séquences

Au vu du temps qui nous était imparti, nous n'avons pas effectué tous les montages des différentes séquences, bien que cela aurait été préférable pour minimiser les variations esthétiques introduites par les différents monteurs. Nous avons décidé de prendre en compte ce paramètre dans notre étude, tout en essayant d'homogénéiser autant que faire se peut les différents extraits en phase de mixage, réalisé cette foisci par nos soins et ce dans les trois formats (Dolby Atmos, Downmixage automatique 5.1 et 5.1).

Pour aller plus en détails, les stimuli de notre test ont été élaborés de trois manières différentes. Deux extraits sont issus du film *Les Medouny*¹⁴⁴, dont la prise de son directe et le montage ont été effectués par nos soins. Pour l'extrait de la séquence 3, nous sommes partis d'un montage son préexistant (effectué sous un environnement multicanal) afin de réaliser le mixage en mode objet et dans les formats dérivés. Enfin, les deux dernières séquences, issues du film *Les Saisons*¹⁴⁵, ont la particularité d'avoir été montées en mode objet par un professionnel. Les différents mixages ont ensuite été effectués par nos soins.

¹⁴⁴ Simon Bonanni, Les Medouny, 16 min, France, sorti courant février 2017.

¹⁴⁵ Jacques Perrin et Jacques Cluzaud, Les Saisons, op. cit.

3.3.2 Choix d'un corpus de séquences

Lors du choix du corpus nous nous sommes demandé s'il n'était pas préférable de tourner nous-mêmes des séquences afin de mettre en exergue certains aspects discutés en 2. Néanmoins, une telle démarche nous écarte du domaine d'application de production filmique dans lequel nous essayons d'ancrer notre étude. Nous avons donc choisi d'effectuer nos tests à partir de séquences déjà réalisées, dans un souci de nous confronter au processus réel d'élaboration du son cinématographique.

Dans une première phase de recherche nous avons sélectionné six films, dont nous avons extrait plusieurs séquences pour lesquelles il nous paraissait intéressant d'envisager un traitement en mode objet. Ce corpus initial constitua notre première approche pratique. A l'issue d'une première recherche nous avons sélectionné quatorze séquences d'une durée moyenne de cinquante secondes. Bien entendu lorsque nous avons pris la décision d'effectuer un test perceptif, il est apparu évident qu'il n'était pas possible de soumettre autant de séquences à une évaluation. Nous nous sommes fixés un nombre de cinq séquences à sélectionner, ceci permettant de raccourcir la durée du test afin de minimiser la fatigue des sujets.

Des critères fixes nous ont aidé à arriver au bon nombre de séquences. Premièrement, la qualité du montage son a été jugée au regard de la mise en scène et a permis d'éliminer des séquences qui présentaient des défauts d'écriture sonore risquant d'altérer le jugement de nos sujets. Deuxièmement, dans le cadre d'un test court, il nous était primordial de sélectionner des séquences mettant en exergue les critères que nous cherchions à étudier. Enfin, nous avons préféré proposer des séquences dont la durée était légèrement plus longue afin de laisser le temps d'en saisir les principaux enjeux narratifs. Nous sommes parvenus à une sélection de cinq séquences, mais au prix d'une moins grande diversité dans les genres cinématographiques.

Trois films ont servi à élaborer ces cinq séquences. Les deux premiers extraits sont issus du film *Les Medouny*¹⁴⁶ réalisé dans le cadre de la partie pratique du mémoire sur le plan séquence de Simon Bonanni. Le troisième extrait provient du film *Bamako-Saint-Afrique*¹⁴⁷ réalisé par Fédéric Jolfre, utilisé comme exercice de mixage durant notre enseignement à l'école. Enfin les deux derniers extraits proviennent du film *Les Saisons*¹⁴⁸ de Jacques Perrin et Jacques Cluzaud sorti le 27 janvier 2016. Une description plus détaillée est en annexe H.

Nous vous proposons maintenant de prendre connaissance du descriptif des séquences.

<u>Séquence 1 : 1'15</u>

Au petit matin, à l'heure dorée, Saïd va à la salle de boxe afin de s'entrainer. L'extrait a été choisi car totalement tourné en plan séquence et en caméra épaule. L'extrait démarre en suivant Saïd de dos, marchant vers la salle tout en fumant son « joint ». Il s'arrête pour le terminer en cachette avant de rentrer dans la salle d'entrainement, où Silvio entraine Malik dans la plus grande rigueur depuis un moment déjà. La caméra abandonne Saïd afin de tourner autour de Silvio et Malik. Lorsque Malik semble arriver au maximum des efforts qu'il peut fournir, la caméra se réoriente rapidement sur Saïd en fond de salle, en train d'enfiler ses gants. Celuici échange un regard furtif avec son grand frère Silvio avant de commencer à boxer le sac de frappe qui se trouve devant lui. La caméra bascule vers Silvio, apparemment perplexe du comportement qu'adopte son frère.

Cette séquence a été choisie pour plusieurs raisons ; la première, la plus importante pour nous, est sa forme visuelle immersive déjà évoquée en 2.1. Le plan séquence permet une immersion plus accrue du spectateur par son procédé filmique même, à

¹⁴⁶ Simon Bonanni, Les Medouny, op. cit.

¹⁴⁷ Frédéric Jolfre, Bamako Saint Afrique, 14 min, France, sorti le 3 novembre 2013.

¹⁴⁸ Jacques Perrin et Jacques Cluzaud, Les Saisons, op. cit.

propos duquel nous avons vu qu'un parallèle pouvait être établi dans le domaine du son à travers le concept de *microphone-spectateur*. D'autre part cet extrait est intéressant spatialement, puisque la caméra évolue dans différents espaces, que nous avons essayés de rendre à travers un travail sur les différents espaces sonores effectué dès le montage son.



Figure 3.1 : Les Medouny, Simon Bonanni, 2016, Photogramme. Saïd rentre dans la salle alors que Silvio entraine Malik.

De plus il nous paraissait intéressant que Saïd, le benjamin de la famille, soit omniprésent même lorsqu'il quitte le champ visuel de la caméra. Dans la version du mixage objet nous avons donc pris la décision de faire des rappels sonores hors champ de sa présence. Ces appels hors champs semblent, de notre point de vue, proposer un apport à la narration puisqu'ils suggèrent l'existence ininterrompue des trois personnages et prédisposent le spectateur à considérer ces trois personnages comme une unité au sein de la séquence. Cette vie du personnage hors champ est rendue possible par les caractéristiques de spatialisation précises dont les subtilités ont été évoquées en 2.1. En outre, lorsque le personnage de Saïd n'est plus dans le champ de la caméra, la spatialisation de sa présence sonore ponctue l'interaction entre la caméra, Malik et Silvio. À la fin de l'extrait le panoramique partant de Saïd

pour revenir vers Silvio est mis en exergue par un panoramique sonore inverse plaçant Saïd dans le hors champs droit de l'écran, soulignant la distance qui sépare les deux personnages. Enfin cette spatialisation accrue semble se justifier dans le caractère intrusif que Saïd entretient par rapport à ses deux frères, soutenant l'idée qu'il n'est pas le bienvenu.

Pour conclure, cet extrait par bien des aspects présente des notions d'interactivités entre la mise en scène, la caméra et la narration. Cette séquence propose d'évaluer si ces interactions sont corrélées voire mises en exergue par une forme d'interactivité sonore mise en place par le mixage objet.

<u>Séquence 2 : 1'05</u>



Figure 3.2 : Les Medouny, Simon Bonanni, 2016, Photogramme. Les trois frères sont en voiture. Malik à droite, Silvio conduit et Saïd se trouve à l'arrière.

Silvio, Malik et Saïd se retrouvent ensemble en voiture. La caméra amorce son mouvement depuis l'arrière, nous regardons Saïd perdu dans ses pensées. Au fur et à mesure nous passons à l'avant de la voiture. Vraisemblablement Malik est perplexe et Silvio, le regard froid et vague, semble se contenter de conduire la

voiture. Personne ne parle, seul le vrombissement du véhicule est perceptible. Dans un plan cut nous passons dans un intérieur appartement, apparemment chez le père des trois frères. Seuls les bruits de couverts résonnent dans la pièce. Cette fois-ci les trois jeunes boxeurs font comme si de rien n'était pour tenter de cacher leur embarras de se retrouver ainsi réunis devant l'autorité paternelle, jusqu'au moment où Karim, le père, se rendant compte de la situation, tente maladroitement de briser la glace.

Nous avons choisi cette séquence pour étudier le paramètre immersif au regard d'un changement d'espace, cette fois-ci découpé. De plus le passage de la voiture vers l'intérieur appartement procure selon nous un changement d'espace brusque mettant en exergue le mutisme des personnages révélateur des non-dits et conflits familiaux. La simplicité de la mise en scène nous semble inversement proportionnelle à l'effet produit. Ce pourquoi nous avons décidé d'effectuer une construction sonore épurée, se basant sur des éléments simples. Une simple ambiance de voiture a donc été utilisée, et la scène de repas ne comporte que le son direct, diffusé au centre. Nous avons donc effectué une prise de son avec un dispositif de croix IRT dans la voiture du tournage. Dans sa diffusion objet la prise de son est rendue par quatre haut-parleurs discrets, ce qui produit selon nous une ambiance globale très subtilement immersive. En effet la qualité de timbre et la bonne concordance entre le système de prise de son et le dispositif de diffusion semblent être un atout. Dans sa version downmixée l'ambiance est diffusée par dix haut-parleurs, rendant l'ambiance de voiture plus confuse, moins précise, moins immersive. Dans sa version 5.1 nous avons pris la décision d'effectuer une diffusion plus frontale, car nous n'appréciions pas le rendu des ambiances de voiture dans les arrières.

En outre, le but de cette séquence est d'étudier l'effet immersif du passage d'une séquence où le spectateur est immergé dans la voiture avec les personnages à une autre totalement frontale et monophonique. De cette manière, nous essayerons

d'évaluer si l'immersion du spectateur est corrélée à la répartition spatiale des ambiances.

<u>Séquence 3 :</u> 0'55



Figure 3.3 : Bamako Saint Afrique, Frédéric Jolfre, 2013, Photogramme.

Djibril, en train d'effectuer un pilotage musclé de son 4x4, ne semble guère satisfait des prestations de son copilote, perdu entre ses cartes et sa feuille de route. Les deux compagnons retrouvent in extremis leur chemin, mais dans son empressement, l'équipage manque de percuter une petite fillette en train de manger des fraises des bois sur le bas-côté de la route.

Nous avons sélectionné cette séquence pour son caractère ultra découpé, se rapprochant de la logique d'une séquence d'action où la saturation sensorielle visuelle est corrélée à l'ultra découpage spatial du son. En ce sens cet extrait constitue l'exemple type des conflits d'ordre spatiaux évoqués largement en 1.3 et mis en exergue ici par un montage image très découpé. Cet extrait constitue selon nous un bon exemple cinématographique où des conflits d'ordre scénographique peuvent survenir entre la projection frontale de l'image et la répartition spatiale des sons dans la salle. Ces conflits, nous l'avons vu, ont conduit les professionnels à l'élaboration de codes et de pratiques conventionnés, dans le but de raisonner l'utilisation des systèmes multicanaux et de ne pas en faire un usage outrancier. Néanmoins, sans prétendre remettre en question le bien-fondé de ces codes, qui entre autre, résolvent dans une certaine mesure les problèmes perceptifs cités, nous voulions déterminer si le mixage objet était assujetti aux mêmes contraintes perceptives.

En outre, l'apparition de nouveaux systèmes coïncidant souvent avec une utilisation gadget des sons arrière, nous voulions confronter les systèmes orientés objet aux mêmes contraintes afin de déterminer si les conflits d'ordre audio-visuel étaient augmentés ou diminués par un système de prime abord encore plus discriminant.

<u>Séquence 4 : 1'15</u>



Figure 3.4: Les Saisons, Jacques Perrin et Jacques Cluzaud, 2016, Photogramme.

L'expression commune « le calme avant la tempête » résume assez justement l'extrait proposé. Le spectateur suit un groupe de loirs jusqu'à l'intérieur d'un tronc d'arbre leur servant de refuge. Les rongeurs s'attroupent entre eux chaleureusement, mais leurs mouvements erratiques semblent signifier que l'heure n'est pas au repos. Et pour cause, les arbres s'agitent, grincent, remuent dans le chaos le plus total. Une tempête s'abat sur la forêt et les arbres la constituant luttent pour préserver leur posture centenaire. Alors que le vent s'estompe peu à peu, la caméra surprend un groupe de grues effectuant un vol migratoire vers des terres plus propices à leur bien-être.

Cet extrait a été sélectionné pour son approche des différents plans sonores. La grande profondeur de champ que propose le découpage axe la mise en scène dans une recherche de différents plans visuels. Le montage son, dans le but de renforcer cette profondeur, utilise plusieurs éléments sonores ponctuels (grincements, craquements, vent dans les feuilles) afin de créer un premier plan sonore très proche du

spectateur. Ces effets sont généralement suivis d'un effet de jaillissement proposant une contiguïté sonore, vers la salle, du premier plan visuel. Le dernier plan à l'image est lui matérialisé par des ambiances de forêts (mouvements de feuilles très diffus, mouvements atmosphériques).

Notre motivation première ici est d'évaluer la possibilité d'amener de la profondeur en proposant un premier plan sonore, localisé dans la salle. Deuxièmement, cet extrait au caractère immersif dans son décor, son sujet et son filmage proposent une bande son très immersive où la spatialisation exacerbe les paramètres d'enveloppement du spectateur. Un retour de l'expérience spectatorielle s'impose.

<u>Séquence 5 : 1'00</u>



Figure 3.5: Les Saisons, Jacques Perrin et Jacques Cluzaud, 2016, Photogramme.

Un ours, étendu nonchalamment sur une branche d'arbre, ne semble pas se soucier de la pluie qui s'abat sur la forêt. Alors que les gouttes se font de plus en plus grosses, les animaux les plus vulnérables tentent de s'abriter rapidement tandis que d'autres font leur réapparition.

Cet extrait fut sélectionné pour son utilisation de l'élévation comme paramètre susceptible d'améliorer la sensation d'immersion chez le spectateur. Néanmoins, il nous apparaît logique de soumettre à l'aveugle aux sujets tests les différents mixages pour savoir si la spatialisation d'éléments sonores en accord avec leur provenance dans la vie réelle peut présenter un intérêt. Cette étude nous semblait d'autant plus justifiée qu'il a été démontré que nous percevons instinctivement certains sons¹⁴⁹ en élévation.

3.3.3 Limites du test

À l'origine de notre mémoire, nous avons voulu proposer une étude large, technique, pratique et esthétique. Cette étude, certainement trop ambitieuse, à ouvert sur des tests perceptifs, qui proposent eux aussi un domaine d'étude peu restreint. Par conséquent, nous avons conscience que les sujets tests avaient beaucoup de critères à évaluer en même temps. Associée aux variables dépendants du sujet (fatigue, utilisation de l'échelle, compréhension de la sémantique), la difficulté de jugement se retrouve exacerbée dans notre étude.

De plus, les extraits testés n'ont pas un processus de fabrication similaire. Autant la diversité de provenance des sessions de montage son était un de nos critères de sélection initiale (Cf. 3.3.1), autant elle est aussi une des limites de notre protocole. Une confection identique des séquences aurait été préférable pour éviter aux sujets de juger l'esthétique du montage son. Par ailleurs, les extraits ont été mixés par nos soins, et nous avons pleinement conscience de notre expérience réduite qui diffé-

¹⁴⁹ En effet, lorsque des sons issus d'objets sonores réels évoluant en hauteur (les oiseux, la pluie, les feuilles, etc.) sont retranscrits à travers un enregistrement, ils sont généralement perçus en élévation par l'auditeur. Outre l'aspect culturel dans la perception spatiale de l'élévation, Jens Peter Blauert (J. BLAUERT, Spatial Hearing: The Psychophysics of Human Sound Localization, MIT Press, 1997.)introduit que des paramètres fréquentiels peuvent rentrer en compte dans la perception des sons en élévation. Cette théorie est aussi soutenue par Robert A. Bulter et Richard A. Humanski (Robert A. Butler et Richard A. Humanski, Localization of sound in the vertical plane with and without high-frequency spectral cues, Perception & Psychophysics, 1992, vol. 51, n° 2, pp. 182-186.).

renciera sans aucun doute nos travaux de ceux réalisés par des professionnels aguerris. De plus le mixage étant effectué par une seule personne, nous ne pouvons intégrer objectivement la pluralité stylistique des mixages qu'offre le spectacle audiovisuel.

D'autre part, nous avons préféré restreindre le nombre d'extraits à analyser, au détriment d'une plus grande pluralité de séquences et de genres cinématographiques. Or nous voulions, en premier lieu, adapter notre étude à la forme cinématographique de la manière la plus large possible, mais ceci relève d'un idéal incompatible avec l'étude que nous menons. Nous ne pouvons pas adresser notre étude à l'ensemble des genres du spectacle cinématographique.

Chaque séquence a été choisie en vue de l'étude d'un critère précis. Toutefois, pour ne pas influencer le sujet test, et donc altérer son jugement, nous avons procédé à l'évaluation des quatre critères sur l'ensemble des séquences.

Bien que l'immersion et l'adéquation à l'image soient des paramètres indépendants des stimuli présentés, en revanche la qualité du timbre ou encore la lisibilité sont, elles, ciblées sur certains extraits. Les résultats étant moyennés sur l'ensemble des séquences, l'analyse précise par séquence risque d'être noyée dans du bruit de mesure.

D'autre part remarquons que si notre étude se destine au spectacle cinématographique, le choix du public soumis aux tests n'est pas assimilable à un public naïf et étranger au processus de production cinématographique. Il serait indispensable de poursuivre notre étude en l'élargissant à un public moins spécifique.

3.4 Analyse et interprétation

3.4.1 Analyse

Vingt-deux personnes ont assisté au test perceptif. Quatre variables indépendantes ont été établies (Cf. 3.2.2). La présence du *bass management* arrière constitue notre seule variable dépendante inter-sujets (B), dix personnes ont passé le test sans *bass managment* et douze avec. Les variables séquences (S), mixages (M), constituent nos variables intra-sujets. Ce protocole nous permet d'effectuer une analyse de variance (ANOVA). Pour minimiser l'influence de l'utilisation d'échelle de notation, les résultats ont été normalisés.

L'analyse statistique calcule le seuil de significativité (p) pour les facteurs intrasujets et inter-sujets, permettant de savoir si leurs influences sur les variables indépendantes (immersion, lisibilité du son, qualité du timbre, adéquation du son à l'image) sont significatives. En d'autres termes, une valeur est dite significative lorsque p est inférieur à 0,05. Elle indique que la différence perçue par le sujet ne relève pas du hasard. Néanmoins, l'ANOVA ne nous permet pas d'affirmer clairement une préférence, mais plutôt de dégager des tendances.

IMMERSION		LISIBILI	LISIBILITE DU SON		QUALITE DU TIMBRE		ADEQUATION A L'IMAGE	
Variables	Р	Variables	Р		Variables	Р	Variables	Р
S	0,010	S	0,469		S	0,004	S	0,000
M	0,000	M	0,340		M	0,015	М	0,992
S*M	0,343	S*M	0,578		S*M	0,568	S*M	0,242
S*B	0,001	S*B	0,193		S*B	0,332	S*B	0,242
M*B	0,021	M*B	0,641		M*B	0,001	M*B	0,093

Figure 3.6 : Récapitulatif des résultats de différences significatives.

De plus, pour faciliter la lecture et éviter les lourdeurs répétitives, nous proposons d'abréger les différents mixages proposés. Ainsi le mixage orienté objet trouve comme raccourci m. ob, le downmixage automatique 5.1 s'écrira dm et le 5.1 manuel prendra la dénomination 5.1.

Immersion:

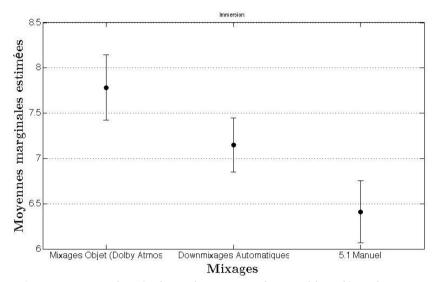


Figure 3.7 : Moyennes des résultats obtenus pour la variable indépendante immersion.

Pour la variable immersion l'analyse révèle que quatre facteurs (S, M, S*B et M*B) sont significatifs. Concrètement, la *Figure 3.7* démontre que l'ensemble des mixages objet et des dm ont une influence sur le critère d'immersion par rapport au 5.1, et ce de manière globale quelle que soit la séquence étudiée. En ce sens nous ne pouvons avancer avec certitude qu'il existe des différences entre les mixages par séquence, même si les moyennes des résultats obtenus par séquence laissent entrevoir des différences intéressantes.

D'autre part notons que l'analyse ne trouve pas de différence significative entre le m.ob et le dm sur le critère d'immersion. Par contre notons qu'il existe une différence significative entre les variables S*B et S*M, en outre une différence d'immersion a été nettement perçue entre les séquences avec et sans bass management et ce en fonction des mixages et des séquences. De plus il semblerait qu'une différence quasi significative (p=0,065) se dessine entre le m.ob et le 5.1 à la séquence 4 et la séquence 5. De plus la séquence 2 présente des différences significatives visibles entre le m.ob et le 5.1 (p=0,000) et entre le m.ob et le dm (p=0,005).

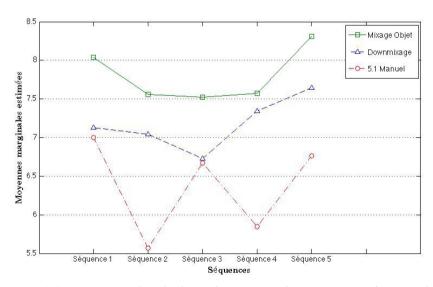


Figure 3.8 : Moyennes des résultats obtenus pour l'immersion en fonction des séquences.

Lisibilité du son :

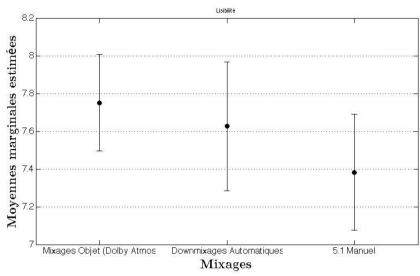


Figure 3.9 : Moyennes des résultats obtenus pour la variable indépendante lisibilité du son.

Aucune différence significative n'a été décelée. Il semblerait qu'aucune différence de lisibilité n'ait été appréciée en fonction des séquences et des mixages.

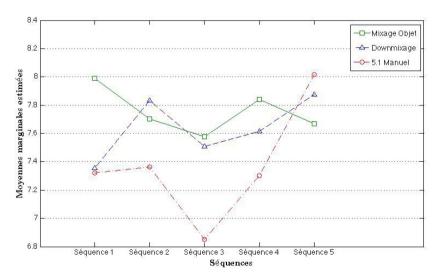


Figure 3.10 : Moyennes des résultats obtenus pour la lisibilité du son en fonction des séquences

Qualité du timbre :

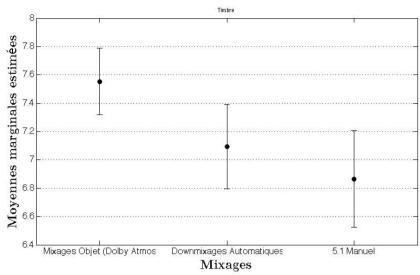


Figure 3.11 : Moyennes des résultats obtenus pour la variable indépendante qualité de timbre.

La variable S est significative, ce qui indique qu'il y a une influence globale de l'évaluation du critère de qualité de timbre en fonction de la séquence. Cependant il n'y a pas d'influence sur les mixages et ce, que le *bass managment* soit actif ou non.

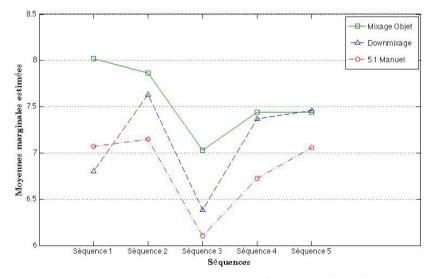


Figure 3.12 : Moyennes des résultats obtenus pour la qualité du timbre en fonction des séquences

La qualité du timbre semble être légèrement plus appréciée pour les mixages objet par rapport aux mixages effectués en 5.1 (Cf. Figure 3.12). De plus aucune différence de timbre ne semble être marquante entre le mixage objet et son downmixage ainsi que le 5.1 manuel vis-à-vis du 5.1 automatique. Néanmoins remarquons que plusieurs différences significatives sont perçues entre les différents mixages de la séquence 1. En effet (p=0,020) entre le mixage objet et entre le mixage 5.1 et (p=0,048) entre le mixage objet et le downmixage automatique.

Adéquation du son à l'image :

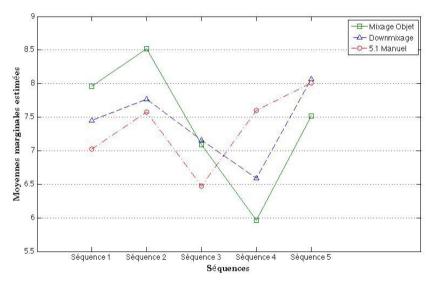


Figure 3.13 : Moyennes des résultats obtenus pour l'adéquation à l'image en fonction des séquences.

Le facteur S est significatif, ce qui induit qu'il y a une influence globale de la séquence sur l'adéquation du son à l'image. Par contre, aucune différence n'est perceptible en fonction des mixages proposés. Cependant il s'avère qu'une différence significative existe entre le mixage objet et le mixage 5.1 de la séquence 4. Il semblerait que l'adéquation à l'image soit mieux pour un mixage 5.1 manuel qu'en mixage objet.

Enfin il existerait une corrélation (c=0,463) entre les facteurs d'immersion et d'adéquation à l'image, l'immersion participe à l'adéquation à l'image mais la corrélation est modérée, il doit y avoir un ou plusieurs autres facteurs qui y participent. Notons aussi que la corrélation est forte pour la séquence 1.

	Séquence 1	Séquence 2	Séquence 3	Séquence 4	Séquence 5
Corrélation	0,732	0,436	0,578	0,317	0,363
Corrélation moyenne			0,463		
Corrélation moyenne	1 1	/1	0,463		

Figure 3.14 : Tableau du taux de corrélation c en fonction de la séquence.

3.4.2 Interprétations

D'emblée les résultats statistiques permettent de détecter des différences de perception entre les différents mixages proposés. Or, pour en fournir une analyse plus pertinente, il est important de mettre en regard les résultats des tests et notre travail d'élaboration des extraits. Ceci permettra entre autres de dégager des tendances que nous avons conjecturées en 2.1 et 2.2. Enfin, nous essaierons d'étayer notre propos en insérant les remarques que nous nous sommes faites lors du mixage des séquences, et celles faites par les personnes ayant participé au test.

Immersion:

Les résultats les plus concluants et intéressants de notre étude sont sur le jugement de l'immersion. Une des lignes directrices de notre mémoire était d'étudier si la débauche de techniques associée aux nouveaux systèmes orientés objet avait un intérêt quantifiable auprès des spectateurs. Nous avons supposé tout au long de notre mémoire que les systèmes actuels proposaient déjà des possibilités immersives convenables et plus que suffisantes. Quelque part, nous cherchions à remettre en cause la logique de recherche de systèmes encore plus immersifs pour encore plus de réalité. Nous nous attendions à ce qu'il n'y ait pas de différence flagrante

entre les systèmes 5.1 et orientés objet. Cependant nos analyses contestent nos hypothèses.

L'étude révèle que les sujets ont une sensation d'immersion plus grande sur les mixages objet et leurs downmixages associés. Cependant nous rappelons que notre définition de l'immersion s'éloigne de celle établie par les fabricants de systèmes. D'ailleurs, certaines personnes ont remarqué que les m.ob sont souvent perçus plus naturellement que les dm et le 5.1. Bien que les analyses des résultats ne permettent pas d'affirmer des corrélations par séquence, la première séquence du film *Les Medouny*¹⁵⁰ explique, selon nous, en partie cet effet de naturel. Plusieurs commentaires ont été faits sur la spatialisation du clocher dans le m.ob, où il a été jugé comme « aérien » ou encore « moins faux que dans le 5.1 ». Cette séquence d'introduction semble en effet bien se prêter à une sensation acoustique large où des éléments ponctuels (et pourtant très localisés) s'intègrent en profondeur. Des remarques similaires ont été faites sur la sensation d'espace lors de l'entré de Saïd dans la salle de boxe. Les ambiances mouvantes semblent nous absorber dans cette salle où pourtant les enjeux sémantiques paraissent clairement prendre le dessus.

La deuxième séquence du film *Les Medouny*¹⁵¹ présente des résultats très convainquants qui dépassent nos attentes. En effet, les différences de mixage dans les différents formats nous semblaient faibles et les effets mis en avant nous semblaient très subtils. L'effet immersif produit par l'ambiance de voiture en croix IRT nous semblait finement perceptible et nous avons mixé cette séquence dans cette optique. En effet les arrières, même peu sollicités, permettaient une sensation immersive très intéressante. Néanmoins, lorsque nous avons écouté le dm, l'espace décrit nous semblait plus confus et moins immersif notamment au niveau du timbre altéré à cause de la diffusion. Cependant, l'analyse des résultats montre que le dm propose sensiblement les mêmes sensations d'immersion. Il apparaît que notre

¹⁵⁰ Simon Bonanni, Les Medouny, op. cit.

¹⁵¹ *Ibid*.

jugement comparatif entre les deux mixages a été biaisé par un jugement du respect du timbre alors mieux rendu sur le m.ob notamment dû à l'extension basse fréquence des ambiances arrière par le *bass management*. Pour ces mêmes raisons, nous avions mixé la séquence 5.1 beaucoup plus frontalement. Logiquement l'immersion ressentie nous paraissait moindre mais le mixage beaucoup moins confus.

Immersion et adéquation à l'image :

L'indifférence entre le m.ob et le dm se maintient d'autant plus sur les séquences où la spatialisation est beaucoup plus marquée et enveloppante. C'est le cas de la séquence 4 ; pour laquelle le résultat a été le plus significatif. En effet lors du mixage nous avions été surpris de la quantité de sons montés, l'univers nous semblait très réaliste et en concomitance avec l'image. Nous sommes restés assez proche de ce qui nous a été communiqué, dans le but d'expérimenter jusqu'où le spectateur pouvait se retrouver comme cible résonnante des sons l'entourant. Globalement, nos tests suggèrent que le m.ob est beaucoup plus immersif que le 5.1. Néanmoins cette tendance semble être inversement proportionnelle en ce qui concerne l'adéquation à l'image. Nous remarquons que le 5.1 a été pour cette séquence plus apprécié que le m.ob et le dm. De plus nous avons vu que les différences d'adéquation à l'image sont corrélées au type de séquence. Nous présumons que l'immersion peut effectivement être perçue est appréciée par le public, cependant cela peut aussi aller à l'encontre de l'adéquation à l'image pour la séquence 4, ce qui n'est pas en soi une découverte pour notre lecteur. Par contre, l'image de cette séquence nous paraissait très propice à l'immersion, de fait nous pensions que les résultats seraient plus modérés. Notamment puisque nous suivions de très près des animaux terrestres et aériens évoluant dans une forêt en pleine tempête. Nous pensons que ce point de vue marqué et très différent de celui de

l'être humain pourrait être la cause de cette inadéquation du son à l'image. Nous remarquons que le montage son très évolué nous amène à éprouver une sensation hyper réaliste. Or, le spectateur se retrouve à proximité d'animaux auxquels il tente de s'identifier; cependant, ne partageant pas le même mécanisme perceptif, le spectateur se trouve très vite détaché de ce processus immersif. Nous nous sommes fait cette réflexion lors du mixage de la version 5.1, plusieurs sources ont été supprimées pour rendre la scène plus dans une logique de tableau. Le spectateur s'en trouvait moins impliqué du point de vue des sens et pouvait s'adonner à un visionnage plus divertissant que saisissant.

Sur *Les Medouny*¹⁵², nous constatons au contraire que la tendance semble s'inverser : l'adéquation à l'image a été jugée meilleure pour le m.ob, même si les différences ne sont pas toujours significatives. Nous pensons que c'est en partie dû à la présence humaine au sein du film qui influence nettement notre perception de l'immersion sonore.

Lisibilité du son :

Nos hypothèses de départ ne peuvent pas être affirmées. En effet, aucune différence de lisibilité n'est discernable entre les différents paramètres. Il nous paraissait pourtant flagrant lors des mixages que le m.ob proposait une bonne lisibilité qui sur certaines séquences améliorait la sensation d'immersion et d'adéquation à l'image. Or, il semblerait que la lisibilité ne soit pas un critère qui influence la perception dans le cadre de notre test. Ce constat permet cependant de montrer que la lisibilité ne dépend pas du mixage proposé. Aussi, à chaque mixage, la lisibilité du son semble être acceptable. Ce qui, avec du recul nous paraît évident, car c'est un élément primordial que l'on cherche toujours à privilégier. Cette recherche de lisibilité passe souvent par la supression de sons, au détriment du rendu de l'espace.

152 *Ibid*.

Ainsi, sans accroître dans l'absolu cette lisibilité, le m.ob permet de conserver les sons impliqués dans la construction de l'espace sonore. Cela ouvre alors peut être sur de nouvelles habitudes de mixage en ce qui concerne la zone frontale.

Le mixage relève principalement du choix des sons (présents dans le montage) au regard de la mise en scène. La sensibilité est donc le maître mot de cette tâche. Néanmoins, nous pensons que ces choix sont influencés par des critères perceptifs, dont la lisibilité des sources sonores fait partie. Lorsque nous avons effectué le mixage de la séquence 1 en 5.1 nous avons opté pour une diffusion frontale des ambiances, cette décision a été influencée par les critères de qualité de timbre et de lisibilité du son. La lisibilité des sources spatialisées dans les arrières étant confuse, nous avons préféré ne conserver que l'ambiance principale (clocher, voiture au loin, fond d'air) avant pour une meilleure lisibilité de la scène.

Qualité du timbre :

D'un point de vue général, sur l'ensemble des séquences, il y a une préférence du m.ob en ce qui concerne la qualité du timbre. Bien que l'étude statistique ANOVA décèle globalement une influence du timbre sur la perception des séquences, cette différence significative entre les mixages est plus remarquable sur la séquence 1. Nous supposons que cela est dû à la nature de cette séquence, notamment du point de vue de l'espace et du passage de l'extérieur vers l'intérieur. En 2.1, nous avons déjà vu le travail effectué sur les acoustiques et les ambiances afin de reconstituer un espace cohérent en fonction de la mise en scène. Il semble que nous puissions également attribuer la réussite immersive de cette séquence à la qualité du timbre.

En ce qui concerne l'extérieur, les effets de réverbérations et de délais effectués en mixage objet nous semblaient nettement altérés dans le dm, et cela a été confirmé par le ressenti de deux personnes qui ont jugé que le timbre du clocher était dégradé, mettant en avant une acoustique grossière. Cela peut s'expliquer par le fait que deux réverbérations différentes, sur deux plans, horizontal et vertical, se retrouvent mêlées dans un seul plan lors du dm. Nous pensons de plus que cet effet a été mieux retranscrit sur le 5.1 fait manuellement, néanmoins la différence n'étant pas significative nous ne pouvons l'affirmer.

Dans l'ambiance intérieure, le m.ob séparait clairement les ambiances de ventilation, placées au plafond, des autres spatialisations horizontales. La subtilité du traitement spatial des ambiances en m.ob semble donc avoir été bien perçue, et nous supposons que cela est corrélé avec l'utilisation du *bass management* qui permet de mieux les définir. Il semble en revanche que le dm ait eu plusieurs conséquences sur la perception de ces ventilations. D'abord, un effet de masquage fréquentiel les a rendues plus floues, tandis que leur placement sur le même plan horizontal que les autres n'a plus permis leur identification particulière.

3.4.3 RÉCAPITULATIF

A l'issue de ces tests, nous remarquons que la question des conventions et des influences artistiques et pratiques déjà soulevée dans la conclusion du 2.3 est toujours bien présente. Il nous paraît évident, après analyse et interprétation, que notre manière de concevoir une bande son est souvent dépendante du système dans lequel on l'appréhende. Sans remettre en cause le 5.1, qui au fond répond parfaitement à certains besoins de mise en scène sonore, nos tests révèlent que dans d'autres cas il n'est pas forcément la solution la mieux adaptée aux volontés artistiques et scénaristiques.

Les résultas souvent observables sur les séquences 1, 2, 4 et 5 amènent à supposer que la création sonore en mode objet s'appréhende dès le montage son. Il semblerait également que, mal exploité, ce système puisse aboutir à des phéno-

mènes d'inadéquation à l'image, comme nous avons pu l'observer pour la séquence 4. Nous avions volontairement pris le parti d'accentuer les effets immersifs car la narration semblait s'y prêter. Cela a fonctionné, l'immersion a été jugée des plus probantes sur cette séquence. Néanmoins, nous avons également observé que cela a conduit à une moins bonne adéquation à l'image. Avec le recul, nous pensons que les effets immersifs, poussés à leur extrême, ont abouti chez le spectateur à une saturation sensorielle inappropriée à la mise en scène. Cet écueil, classique, est celui d'un excès d'enthousiasme face à la surenchère des possibilités de spatialisation offertes par un tel système. Il convient donc, malgré l'engouement suscité par la découverte de nouvelles techniques, de garder à l'esprit que l'enjeu de la création sonore est de servir la mise en scène sans se mettre en avant ostensiblement.

D'autant plus que le mixage objet semble tirer son épingle du jeu plus sur des effets subtils, qui ne sont pas forcément perçus précisément mais ressentis globalement par l'ensemble des spectateurs. Cependant, ce n'est pas le mixage objet en luimême qui propose un rendu plus intéressant du point de vue de l'immersion, puisque, globalement, le downmixage automatique, bien que présentant selon nous des défauts intrinsèques dus à la manière dont il est obtenu, semble souvent présenter les mêmes qualités immersives et les mêmes tendances d'adéquation à l'image que le mixage objet. Ainsi, il s'agirait avant tout d'une autre manière d'appréhender l'immersion, permise par le mixage objet, mais qui pourrait se retrouver également dans un système downmixé 5.1. Notre scepticisme initial concernant le downmixage automatique, que nous supposions moins satisfaisant qu'un mixage 5.1 manuel, a donc été mis en défaut par nos tests au cours desquels il est apparu finalement plus efficace, au moins dans le cas de la recherche de l'immersion.

CONCLUSION

Dans une éternelle recherche de davantage de réalisme, les avancées technologiques du son au cinéma se sont focalisées sur une restitution la plus fidèle possible de la spatialisation. Ce processus a abouti récemment à l'émergence de systèmes de diffusion orientés objet proposant, selon les constructeurs, une expérience sensorielle « comme dans la vraie vie ». La normalisation de ce format laissant supposer son expansion future, il nous a paru judicieux de le contextualiser dans une production cinématographique plus étendue que celle du grand spectacle à laquelle il est pour l'instant cantonné. Nous voulions donc à l'origine, étudier ses potentiels apports aussi bien d'un point de vue technique, pratique, qu'esthétique.

Nous faisons aujourd'hui le constat d'un renouvellement des questionnements soulevés par les avancées technologiques. L'histoire de la spatialisation commence avec la stéréophonie puis plus récemment avec le 5.1 qui a fait l'objet de nombreuses recherches psychoacoustiques et esthétiques aboutissant à une bonne connaissance de ses potentialités et limites. Ce mémoire, en se proposant d'étudier globalement un système de diffusion permettant, selon ses promoteurs, d'améliorer encore la spatialisation, s'inscrivait donc d'emblée dans la lignée de ces recherches. Il s'agissait entre autres de voir si le mixage objet était soumis aux mêmes limites que le 5.1 ou s'il ne permettait pas de les dépasser pour offrir de nouvelles potentialités. Depuis le début nous avons conscience que la spatialisation ne doit pas être quelque chose d'essentiel à la bonne intelligibilité d'un film. Toutes les expériences de son spatialisé, et ce jusqu'au 5.1, ont privilégié une diffusion monophonique des éléments les plus importants (dialogues) pour assurer leur restitution dans toutes les configurations de diffusion. Néanmoins, nous avons remarqué que certains films conçus sous un univers objet, majoritairement vus dans des installations 5.1 dans

leur version downmixée automatiquement, semblaient garder l'attrait propre à une nouvelle approche de la spatialisation. La question de la compatibilité descendante est alors apparue prépondérante, les systèmes de diffusion orientés objet étant encore loin d'être les plus courants dans les salles de cinéma.

Il s'agissait donc également de comprendre comment rendre la spatialisation sur des systèmes descendants, et si ce protocole n'entrainait pas une perte qualitative trop importante et des aberrations d'ordre spatial pouvant aller à l'encontre de la mise en scène. Pour comprendre le downmixage automatique, nous avons donc dû préalablement étudier en profondeur le fonctionnement du mixage objet. Nous avons procédé à une étude succincte de tous les systèmes proposant actuellement un mixage objet, qui nous sont apparus relativement similaires, mis à part quelques variations dans la flexibilité des configurations. D'aspect extérieur, ils ressemblent à une version étendue du 7.1, augmentée d'une dimension verticale. En revanche, leur principe de fonctionnement, étudié en profondeur, diffère totalement de tout système orienté canal : chacune des multiples enceintes est adressable de façon discrète. Cette étude nous a permis de formuler des hypothèses d'utilisation quant à la meilleure façon de tirer parti de ces nouveaux outils.

Une étude approfondie des principes mêmes du mixage objet nous a permis de dégager les aptitudes supposées de ce nouveau système, à savoir un gain qualitatif potentiel par rapport au 5.1. Parmi les différentes qualités apparues, le gain en intermasquage, le dégagement de la zone frontale de l'écran ainsi que l'amélioration supposée du timbre semblent être des éléments inhérents au mixage objet lui permettant en théorie de dépasser certaines limites de conception attachées au fonctionnement orienté canal (phénomènes de masquage, de filtrage, etc. provoqués par la diffusion d'un même son par plusieurs enceintes). Néanmoins, le mixage objet ne peut s'affranchir totalement des contraintes dues au cadre audiovisuel dans lequel il s'inscrit. S'il apparaît qu'une utilisation raisonnée

et subtile de ce nouveau système permet de repousser les frontières de ce que l'on pensait possible dans la recréation de l'espace sonore, en revanche l'écueil de la surenchère d'effets est encore plus grand que pour le 5.1. On observe donc généralement une dichotomie entre l'optimisme débordant des concepteurs vantant les qualités de leurs systèmes, et le scepticisme accru des professionnels et des théoriciens qui se méfient d'une spatialisation encore plus poussée amenant souvent à un effet gadget plutôt qu'à un réel intérêt narratif.

Comme il était difficile de faire la distinction entre, d'une part, ce qui semblait pouvoir relever de l'argument de vente et d'autre part, ce qui apparaissait comme un rejet lié plus à des habitudes de pratique et des conventions qu'à une réelle expérience du système, nous avons procédé à notre propre expérimentation en cherchant à mettre en pratique les innovations rendues possibles par le mixage objet. Il nous est vite apparu que ce système ne se limitait pas aux usages grossiers décriés par ses détracteurs, et qui semblent plus provenir des films eux-mêmes que du dispositif utilisé. Ainsi, contrairement à ce qui est le plus souvent employé dans les productions actuelles, le gain immersif ne semble pas se déployer dans un amas de sons arrivant de toute part mais plutôt dans la précision du placement de quelques sources ponctuelles choisies. Nous avons pu recréer de manière satisfaisante, à notre point de vue, des espaces sonores cohérents voire réalistes grâce au placement subtil de quelques objets sonores, et ce pour différents types de mise en scène, du documentaire animalier au film intimiste en passant par le film à effets. Durant le mixage, en cherchant à dégager la zone frontale (derrière l'écran) afin de la rendre moins confuse, nous avons été naturellement amenés à utiliser les enceintes du premier tiers de la salle, permettant l'étalement des plans sonores et créant ainsi de la profondeur dans l'axe de l'écran. En termes de ressenti, il nous semblait que notre rapport au spectacle audiovisuel était moins distant, la scène sonore décrite à l'écran plus lisible, l'espace sonore recréé dans le premier tiers de la salle plus cohérent tout en s'inscrivant dans l'image.

Néanmoins, nous nous sommes heurtés à certaines limites. D'un point de vue pratique, les outils classiques de création sonore ne sont pas encore adaptés au mixage objet, ce qui rend malaisé le travail sous cet environnement. De même, les pratiques ne sont pas encore harmonisées ; or nous avons testé différentes configurations où l'approche objet avait été décidée soit en phase de mixage, soit dès la phase de montage. L'importance d'utiliser une matière sonore adéquate nous est apparue cruciale, nécessitant sans doute une réflexion préalable au tournage permettant d'optimiser la phase de postproduction pour ne pas tomber dans une utilisation gadget où l'on rajoute en fin de chaîne des sons qui n'ont pas été prévus dès le montage. Il nous semble ainsi essentiel pour obtenir des résultats satisfaisants de revoir nos pratiques afin d'appréhender le mode objet comme un outil d'expression et non comme une simple technique.

Afin de vérifier la validité de nos hypothèses personnelles concernant la définition de l'espace sonore agencé par l'utilisation du mixage objet, nous avons mis en place des tests visant à confirmer et compléter nos théories. La finalité était de savoir si ce genre de système s'intègre bien dans le spectacle audiovisuel. Pour ce faire, nous avons soumis à un public test trois versions de mixage différentes, un mixage objet, son downmixage automatique et une version 5.1 manuelle élaborée par nos soins selon les pratiques habituelles. Le but était de comparer qualitativement ces trois systèmes de diffusion selon différents critères clefs de la création sonore, à savoir la sensation d'immersion, la lisibilité du son, la qualité du timbre et l'adéquation à l'image. Si le gain immersif du mixage objet a bien été perçu et généralement apprécié sur l'ensemble des séquences proposées, des résultats similaires ont été obtenus pour le downmixage automatique, au sujet duquel nous étions tout d'abord sceptiques. En ce qui concerne la qualité du timbre, la tendance générale montre une préférence globale pour le mixage objet, suivi de son downmixage. Ainsi, ces résultats semblent suggérer que ce n'est pas la diffusion en mode objet en elle-même qui est à l'origine de ce gain globalement reconnu, mais qu'il proviendrait plutôt d'une approche du mode objet dont les répercussions peuvent se ressentir également dans son downmixage. La notable différence observée entre le mixage objet et son downmixage d'une part, et la version classique 5.1 d'autre part, nous conduit à penser que, bien plus que le système de diffusion, c'est l'approche même du mixage objet qui permet sa réussite en ouvrant sur d'autres possibilités, en affranchissant le mixage de pratiques et d'esthétiques admises liées au 5.1.

Cependant, certaines limites sont apparues. La lisibilité du son, que nous supposions améliorée par la description précise de l'espace sonore induite par le mixage objet, n'a finalement pas abouti à des différences significatives entre les trois mixages. De même, au sujet de l'adéquation à l'image, les apports n'ont pas été perçus de manière flagrante. En revanche, un cas d'inadéquation à l'image a montré que le mixage objet et son downmixage sont encore plus susceptibles de rompre le contrat audiovisuel que le mixage 5.1. Il convient aussi de rappeler que les séquences ont été choisies en fonction d'une scénographie particulière, impliquant la spatialisation de manière accrue au sein de la narration. Les résultats les plus probants ont par ailleurs été observés sur une forme à part entière, à savoir le plan séquence, et nous devons donc pondérer notre analyse qui s'est cantonnée à un domaine d'étude particulier dans lequel la spatialisation a une importance plus marquée. Ainsi, si nos tests nous permettent de conclure à une bonne intégration du mixage objet dans une certaine forme du spectacle audiovisuel, nous ne sommes pas à l'heure actuelle en mesure de pouvoir affirmer son efficacité dans tous les domaines. Nous nous heurtons ici aux limites de nos tests qui ne permettent pas une étude exhaustive des potentialités du mixage objet. Pour en explorer un éventuel apport narratif, il serait nécessaire de poursuivre nos expérimentations au sein d'un récit complet et pas en étudiant de simples séquences extraites de leur contexte filmique.

Ainsi, à l'issue de ce mémoire, nous comprenons que le mixage objet semble présenter un réel apport du point de vue de l'immersion et de la création d'espaces sonores, sans se cantonner à un genre de film grand public adepte d'effets. Il convient de rappeler que même dans ce genre de production, le mixage objet, pensé en amont et utilisé à bon escient, peut s'affranchir de sa mauvaise réputation de gadget, et remplir tout à fait son rôle en apportant cohérence spatiale et immersion. D'un point de vue plus global, nos tests semblent prouver, à travers le succès du downmixage automatique, que ce n'est pas seulement le système de diffusion orienté objet qui fait la réussite du procédé, mais que l'on peut retrouver également certaines de ses qualités dans les versions descendantes. Il apparaît donc que cet apport provient autant d'une approche du mixage objet que du système en luimême, approche qui permet de s'affranchir de conventions et d'habitudes de pratique induites par des années d'expérience du mixage orienté canal. Néanmoins, en proposant un renouvellement de nos approches de la spatialisation, le mixage objet semble effleurer les limites de ce qui peut être fait dans le paradigme du cinéma.

Il convient donc, pour exploiter pleinement ce nouvel outil, d'adapter les pratiques et l'approche du processus créateur, afin de repenser l'espace sonore et visuel en fonction de ces nouveaux paramètres. De plus, nous nous devons de rester vigilants face à l'automatisation grandissante des processus de création audiovisuelle. Si, de prime abord, cette tendance répond à une conjoncture économique où le foisonnement des formats de diffusion rend impossible le travail manuel, elle pourrait, si l'on n'y prend pas garde, remettre en cause l'Humain au sein du processus artistique intrinsèque du cinéma. Il peut être tentant de rester attaché au système orienté canal, qui garantit l'intégrité de l'œuvre à sa sortie de l'auditorium; mais il est sans doute préférable d'accepter les avancées technologiques, afin de mieux les comprendre pour pouvoir en tirer profit sans qu'elles remettent en cause notre liberté artistique.

BIBLIOGRAPHIE

OUVRAGE:

ALTMAN Rick (dir.), *Sound theory, sound practice*, New York, Routledge, coll. *AFI film readers*, 1992, 291 p.

Aumont Jacques (dir.), *Esthétique du film*, 3. éd., s.l., Armand Colin, coll. *Armand Colin Cinéma*, 2008, 244 p.

BARJAVEL René, Cinéma total : Essai sur les formes futures du cinéma, Paris, Denoël, coll. Romans Français, 1944.

BARNIER Martin et KITSOPANIDOU Kira, *Le cinéma 3-D : histoire, économie, technique, esthétique*, Paris, Armand Colin, 2015, 188 p.

BARR Charles, *CinemaScope : Before and After*, Film Quarterly, 1963, vol. 16, no 4, pp. 4-24.

BAZIN André, *Qu'est-ce que le cinéma*?, Paris, Éd. du Cerf, coll. *Septième art*, n° 60, 1985, 7-372 p.

BECKER Howard S., *Art As Collective Action*, American Sociological Review, 1974, vol. 39, n° 6, pp. 767-776.

BLAUERT J., Spatial Hearing: The Psychophysics of Human Sound Localization, MIT Press, 1997.

CHION Michel, *Un art sonore le cinéma : histoire, esthétique, poétique*, Paris, Cahiers du cinéma, coll. *Cahiers du cinéma essais*, 2003, 478 p.

Curot Frank, Styles filmiques. 2 : Les réalismes : Cassavetes, Forman, Kiarostami, Loach, Pialat, Paris, Lettres Modernes Minard, coll. Études cinématographiques, n° 69, 2004, 291 p.

ICART Roger, *Abel gance ou le promethée foudroyé*, Lausanne, L'âge D'homme, coll. *Histoire Théorie Cinéma*, 1984, 7-489 p.

KERINS Mark, *Beyond Dolby (stereo): cinema in the digital sound age*, Bloomington, Indiana University Press, 2011, 378 p.

PÉRIAUX Bergame, OHL Jean-Luc, Thévenot Patrick et Camerer Florian, *Le son multicanal : de la production à la diffusion du son 5.1, 3D et binaural*, Paris, Dunod : INA, 2015.

ROGERS Ariel, *Cinematic appeals : the experience of new movie technologies*, New York, Columbia University Press, coll. *Film and culture*, 2013, 330 p.

OUVRAGE COLLECTIF:

CARR Robert E. ET HAYES R. M., Wide screen movies: a history and filmography of wide gauge filmmaking, Jefferson, N.C, McFarland & Co, 1988, 5-502 p.

MILLOT Laurent et PELÉ Gérard, « La sensation d'espace dans le spectacle audiovisuel », Rémi ADJIMAN et Bruno CAILLER (dirs.), in *Une architecture du son*, Paris, coll. *Cahiers de Champs Visuels n° 1 & 2*, 2005, pp. 7-384.

Théberge Paul, Devine Kyle et Everrett Tom (dirs.), *Living stereo : histories and cultures of multichannel sound*, New York, Bloomsbury Academic, 2015, 289 p.

REVUE:

BAIBLÉ Claude, *Problèmes de spatialisation en son multicanal*, Espaces pluriels, images et sons, septembre 2004, Cahier Louis-Lumière, nº 2, 09/2004.

BENNETT John C., BARKER Keith et EDEKO Frederick O., *A New Approach to the Assessment of Stereophonic Sound System Performance*, J. Audio Eng. Soc, 1985, vol. 33, nº 5, pp. 314-321.

BERKHOUT A. J., VRIES D. de et VOGEL P., *Acoustic control by wave field synthesis*, The Journal of the Acoustical Society of America, 1 mai 1993, vol. 93, n° 5, pp. 2764-2778.

BUTLER Robert A. et Humanski Richard A., *Localization of sound in the vertical plane with and without high-frequency spectral cues*, Perception & Psychophysics, 1992, vol. 51, no 2, pp. 182-186.

GERZON Michael A. et WORAM John M., *Blumlein Stereo Microphone Technique and Author's Reply*, J. Audio Eng. Soc, 1976, vol. 24, no 1, pp. 36, 38.

HERRE Jürgen, HILPERT Johannes, KUNTZ Achim et PLOGSTIES Jan, *MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding*, J. Audio Eng. Soc, 2015, vol. 62, nº 12, pp. 821-830.

Pulkki Ville, *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*, J. Audio Eng. Soc, 1997, vol. 45, n° 6, pp. 456-466.

ARTICLE DE CONVENTION (COMMUNICATION À CONGRÈS):

Breebaart Jeroen, Engdegard Jonas, Falch Cornelia, Hellmuth Oliver, Hilpert Johannes, Hoelzer Andreas, Koppens Jeroen, Oomen Werner, Resch Barbara, Schuljers Erik et others, « Spatial audio object coding (SAOC)-The upcoming MPEG standard on parametric object based audio coding », Convention Paper 7377, Presented at the 124th Convention, Amsterdam, 2008.

BRIX Sandra, SPORER Thomas et PLOGSTIES Jan, « CARROUSO-An European approach to 3D-audio », Convention Paper 5314, Presented at the 110th AES Convention, Amsterdam, 2001.

Fonseca Nuno, « 3D Particle Systems for audio Applications », DAFx-13, proc. of the 16th Int. Conference on Digital Audio Effects, Maynooth, 2013.

PULKKI Ville, « Uniform spreading of amplitude panned virtual sources », Laboratory of Acoustics and Audio Signal Processing Helsinki University of Technology, proc. of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, 1999.

Pulkki Ville, Karjalainen Matti et Välimäki Vesa, « Localization, Coloration, and Enhancement of Amplitude-Panned Virtual Sources », Audio Engineering Society Conference : 16th International Conference : Spatial Sound Reproduction, 1999.

TSINGOS Nicolas, Q. ROBINSON Charles et A.C CRUM Poppy, « Evaluation of panning algorithms for theatrical applications », Dolby Laboratories Inc., Rapport de la Conférence ICSA, Erlangen, 2014.

VERRON Charles et Drettakis George, « Procedural audio modeling for particle-based environmental effects », Convention Paper 8764, Presented at the 133rd AES Convention, San Francisco, 2012.

ARTICLE:

BAIBLÉ Claude, *Atmosphère, Atmosphère, est-ce que j'ai une gueule d'atmosphère ?*, Dossier sur la spatialisation du son, Lettre CST 145, 2013, p. 22.

LOSSIUS Trond, BALTAZAR Pascal et HOGUE Théo DE LA, *DBAP-distance-based amplitude panning*, Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2009.

MÉMOIRE ET THÈSE:

AZZAZI Sélim, *Réflexion sur l'apport de la diffusion multicanal dans l'écriture cinématographique*, Mémoire de fin d'études, ENS Louis-Lumière, 1999.

CARREAU Rémi et MACQUART Thibaut, *Utilisation de la technologie WFS dans la création sonore cinématographique : posibilités et limites*, Mémoire de Master, ENS Louis-Lumière, 2015.

DUPAS Aymeric, *Multicanal au cinéma : un canal zénithal ?*, Mémoire de fin d'études, ENS Louis-Lumière, 2007.

GAMBIER Victor, *Nouvelles approches pour la bande sonore d'un film en relief*, Mémoire de fin d'études, ENS Louis-Lumière, 2010.

MOREAU Nicolas, *L'orientation réflexe dans le cinéma multicanal*, Mémoire de fin d'études, ENS Louis-Lumière, 1999.

PÉREZ-LÓPEZ A., Real-Time 3D Audio Spatialization Tools for Interactive Performance, Mémoire de Master, Université de Pompeu Fabra, Barcelone, 2014.

SALAÜN BERTET Stéphanie, Formats audio 3D hiérarchiques: caractérisation objective et perspective ces systèmes Ambisonics d'ordres supérieurs, Thèse de doctorat en Acoustique, Institut National des Sciences Appliquées, Villeurbanne, 2009.

VINGTRINIER Christophe, *La diffusion multicanal au cinéma : nécessité d'une écriture visuelle et sonore adaptées*, Mémoire de fin d'études, ENS Louis-Lumière, 1997.

DOCUMENT DE PRÉSENTATION TECHNIQUE:

ALLEN Ion, « Matching the sound to the Picture », Dolby Technical Papers, Reprinted by permission from « The Proceedings of the AES 9th International Conference : Televison Sound Today and Tomorrow », San Francisco, 1991.

DOLBY LABORATORIES INC., What's new in Dolby Atmos mixing tools v1.6, Documentation de spécification technique, 2016.

DOLBY LABORATORIES INC., *Dolby Atmos Specifications : issue 3*, Documentation de spécification technique, 2015, Consultable sur www.dolby.com.

DOLBY LABORATORIES INC., *Dolby*© *AC-4*: *Audio Delivery for Next-Generation Entertainment Services*, Document de normalisation, 2015.

DOLBY LABORATORIES INC., *Un son de prochaine génération pour le cinéma*, Documentation de présentation technique, 2012, Consultable sur www.dolby.com.

DTS INC., MDA Multi-Dimensional Audio, a 3D Audio Solution for Professional Applications, Documentation de présentation technique, 2015, nº 1.1.0.0.

EBU Operating Eurovision and Auroradio, *Audio Definition Model*; *Metadata Specification*, TECH 3364, 2014, no 1.0.

ETSI, MDA; Object-Based Audio Immersive Sound Metadata and Bitstream, ETSI TS 103 223, 2015, vol. 1.1.1.

ETSI, Digital Audio Compression (AC-4) Standard Part 2: Immersive and personalized audio, ETSI TS 103 190-2, 2015, vol. 1.1.1.

MATEOS Toni, LÓPEZ Vicente et ARTIGAS Alex, *Immersive 3D sound for cinema*, Documentation de présentation technique Immsound, 2012.

Q. Robinson Charles, Tsingos Nicolas et Metha Sripal, *Scalable Format and Tools to extend the possibilities of Cinema Audio*, The SMPTE Motion Imaging J., 2012, vol. 121, nº 8, pp. 63-74.

TSINGOS Nicolas, GALLO Emmanuel et DRETTAKIS George, « Perceptual audio rendering of complex virtual environments », REVES/INRIA Sophia-Antipolis, 2004.

FILMOGRAPHIE

Andrews Mark, Chapman Brenda et Purcell Steve, Brave (Rebelle), 93 min, États-Unis, sorti en France le 1er août 2012.

APTED Michael et **HANSON Curtis**, *Chasing Mavericks*, 116 min, États-Unis, sorti en France le 28 novembre 2012.

BIGELOW Kathryn, *Strange Days*, 145 min, États-Unis, sorti en France le 7 février 1996, 1995.

BONANNI Simon, Les Medouny, 16 min, France, sorti courant février 2017.

COOPER Merian C., FRITSCH Gunther von, SCHOEDSACK Ernest B. et JR Michael Todd, *This Is Cinerama*, 115 min, États-Unis, 1955.

CROSLAND Alan, *The Jazz Singer (Le Chanteur de Jazz)*, 88 min, États-Unis, sorti en France le 4 octobre 1929, 1927.

Cuarón Alfonso, Children of Men (Les fils de l'homme), 109 min, États-Unis, Royaume-Uni, sorti en France le 18 octobre 2006.

Cuarón Alfonso, *Gravity*, 91 min, États-Unis, Royaume-Uni, sorti en France le 23 octobre 2013.

DOLAN Xavier, *Mommy*, 139 min, Canada, sorti en France le 8 octobre 2014.

Eastwood Clint, *American Sniper*, 133 min, États-Unis, sorti en France le 18 février 2015.

FLETCHER Dexter, *Eddie the Eagle*, 106 min, États-Unis, Royaume-Uni, Allemagne, sorti en France le 4 mai 2016.

GANCE Abel, Napoléon, 330 min, France, sorti le 7 avril 1927, 1935.

IÑÁRRITU **Alejandro G.**, *The Revenant*, 156 min, États-Unis, sorti en France le 24 février 2016.

JACKSON Peter, *The Hobbit : The Desolation of Smaug (Le Hobbit : La désolation de Smaug)*, États-Unis, sorti en France le 11 décembre 2013.

Jolfre Frédéric, Bamako Saint Afrique, 14 min, France, sorti le 3 novembre 2013.

Koster Henry, *The Robe (La Tunique)*, 135 min, États-Unis, sorti en France le 4 décembre 1953.

MALICK Terrence, *The Tree of Life*, 139 min, États-Unis, sorti en France le 17 mai 2011.

MEGATON Olivier, Taken 2, 92 min, France, sorti le 3 octobre 2012.

Montgomery Robert, La dame du lac, 105 min, États-Unis, sorti en France le 14 avril 1948.

NAISHULLER Ilya, *Hardcore Henry*, 96 min, États-Unis, sorti en France le 13 avril 2016.

Noé Gaspar, Enter the Void, 161 min, sorti le 5 mai 2010.

OFFENSTEIN **Christophe**, *En solitaire*, 101 min, France, sorti en France le 6 novembre 2013.

PERRIN Jacques et CLUZAUD Jacques, *Les Saisons*, 97 min, sorti en France le 27 janvier 2016.

RIGGEN Patricia, *The 33*, 127 min, Chili, Etat-Unis, sorti en France le 2 mars 2016, 2015.

Sant Gus Van, Elephant, 81 min, États-Unis, sorti en France le 22 octobre 2003.

SPIELBERG Steven, *Il faut sauver le soldat Ryan*, 169 min, États-Unis, sorti en France le 30 septembre 1998.

SNYDER Zack, *Batman v Superman : Dawn of Justice*, 151 min, États-Unis, sorti en France le 23 mars 2016.

UNKRICH Lee, Toy Story 3, 103 min, États-Unis, sorti en France le 14 juillet 2010.

VILLENEUVE Denis, Sicario, 121 min, États-Unis, sorti en France le 7 octobre 2015.

ANNEXES

A DESCRIPTION DU LANGAGE BIFS:

Inspiré du langage VRML (Virtual Reality Modelling Language), qui ne couvre que les représentations en 3D, la norme MPEG-4 BIFS (Binary Format for Scenes description), qui couvre les représentations en 3D et en 2D, décrit une scène audiovisuelle comme étant un ensemble de contenus audiovisuels individualisés (images fixes, objets vidéo, objets audio, textes, graphismes, etc.), appelés objets média primitifs, organisés de façon hiérarchique et liés par des relations spatiales et/ou temporelles. Le groupement d'objets média primitifs permet de former des objets média plus complexes et plus sophistiqués appelés objets média composites. Les objets média primitifs peuvent être des contenus audiovisuels naturels ou synthétiques. Chaque objet média primitif possède une dimension spatiale (forme, texture, taille, orientation, mouvement, etc., pour un objet média visuel, hauteur, timbre, espace, etc., pour un objet média auditif) et une dimension temporelle (début d'apparition ou date de naissance, durée de présence ou de vie, etc.).

Le fichier binaire BIFS contient des informations sur l'organisation, le positionnement dans l'espace et dans le temps, l'animation, les valeurs d'attributs des
objets média au sein de la scène audiovisuelle, la possibilité d'action sur les
attributs d'un objet média pour lui appliquer des transformations susceptibles de
modifier son apparence géométrique et/ou acoustique, la possibilité de regrouper
plusieurs objets média primitifs pour former un objet média composé, les
mécanismes d'interactivité entre l'utilisateur et la scène audiovisuelle, etc. La
description de la scène audiovisuelle peut évoluer en utilisant des mises à jour. Ce
fichier est adjoint aux objets média. Les objets média et ce fichier sont ensuite
encapsulés dans un même fichier dont l'extension est mp4.

B Installation électroacoustique du Dolby Atmos:

Déscription générale :

En partant de configurations connues dans l'ensemble des salles de cinéma, le Dolby Atmos se base sur une structure classique 5.1 voire 7.1. Ainsi, Dolby part de la configuration suivante :

- 3 voies derrière l'écran Gauche, Centre, Droit (avec voies intermédiaires Gauche et Droite pour les écrans de plus de 12 m¹⁵³). L'axe acoustique des enceintes est positionné aux 2/3 de la hauteur de l'image. Les enceintes gauche et droite, situées au milieu des bords imposés par le format 1,85 et le format Cinémascope, sont orientées vers un point situé aux 2/3 de la salle en partant de l'écran.
- deux rampes d'enceintes respectivement situées à gauche et à droite constituent les enceintes surround Ls et Rs. Ces rampes d'enceintes appelées par abus de langage enceintes ambiance commencent au 1/3 de la salle en partant de l'écran. Les enceintes sont écartées d'environ 2 à 3 mètres maximum. Elles sont orientées verticalement de manière à viser le spectateur assis à l'autre extrémité du rang qui leur fait face.
- deux rampes d'enceintes arrière situées à gauche et à droite.
- une voie dédiée au caisson de renfort des basses (LFE) ; ce canal est situé généralement en dessous de l'écran sur une dalle solidaire de celle de la salle ; il est préférable que le caisson de renfort des basses ne soit pas centré

¹⁵³ Cette notion de haut-parleurs optionnels pour les écrans de plus de 12 m de largeur n'est pas une amélioration proposée avec l'arrivée du Dolby Atmos. Ces voies additionnelles ont été proposées pour la première fois avec le procédé Fantasound en 1940 et reprises ensuite dans toutes les grandes salles où les dispositifs multicanal de base ne pouvaient pas fournir une couverture sonore acceptable.

par rapport à la largeur de la salle car cela produit généralement des problèmes acoustiques (phénomènes d'ondes stationnaires).

À cette configuration quelque peu classique viennent s'ajouter :

- deux rampes de haut-parleurs suspendus au plafond, couvrant toute la longueur de la salle.
- des haut-parleurs qui viennent faire la liaison entre les haut-parleurs frontaux (derrière l'écran) et les premiers haut-parleurs surround, venant fermer le réseau d'enceintes d'ambiance.
- des haut-parleurs dédiés au renfort des basses à l'arrière de la salle utilisés par le bass management.

L'évolution se situe du côté de l'élargissement de la bande passante des enceintes arrière. En effet pour permettre une diffusion homogène du son en façade et dans la salle, les enceintes d'ambiance doivent être capables de descendre jusqu'à 50 Hz. En deçà de cette valeur, le *bass management* et les renforts de basses arrière prendront le relais.

C REDÉFINITION DES NORMES DE DIFFUSION SONORE DANS LE SYSTÈME DE DIFFUSION DOLBY ATMOS :

Dans la version 3 des spécifications¹⁵⁴ que doit respecter chaque salle équipée en Dolby Atmos, Dolby repense la conception du système de restitution sonore tout en s'efforçant de conserver la compatibilité avec les systèmes de diffusion tels que le 5.1 et le 7.1. Dans cette dernière version des spécifications, la firme américaine met l'accent sur une refonte des performances électroacoustiques nécessaires dans une salle de cinéma.

1. Nouvelle configuration pour les enceintes surround ?

Bref rappel des normes de diffusion en salles de cinéma :

Le système de diffusion Dolby Atmos se base sur les normes de diffusion relatives au 5.1 ou au 7.1. Ces normes répondent aux exigences des règles de diffusion cinéma préconisées par l'ISO (*International Organization for Standardization*) et par la SMPTE (*Society of Motion Picture & Television Engineers*) telles que ISO 2969/1987 (E) et SMPTE ST202-2010.

Ces normes prévoient un nombre précis d'enceintes surround par côté. Les hautparleurs dédiés à un canal sont alimentés par le même signal et par conséquent diffusent le même son. Dans le système Dolby Atmos, initialement conçu, chaque haut-parleur est alimenté par un signal unique qui lui est attribué. L'installation technique devient ainsi encombrante et coûteuse pour les exploitants. Dans sa première version, ce type de système nécessite un canal d'amplification par point de diffusion. La configuration maximale, qui compte un total de 64 haut-parleurs,

¹⁵⁴ DOLBY LABORATORIES INC., *Dolby Atmos Specifications : issue 3*, Documentation de spécification technique, 2015, Consultable sur www.dolby.com.

exige, quant à elle, 32 amplificateurs stéréophoniques. En vue des caractéristiques techniques requises par Dolby, seuls des investissements importants peuvent permettre d'acquérir une telle installation; ce qui constitue un handicap majeur à l'expansion de cette technique. Conscients de cet enjeu, les ingénieurs de Dolby ont donc revu à la baisse leur cahier des charges. Néanmoins, bien que cette dernière révision ne semble pas affecter notablement la qualité de la diffusion en salle, elle pose, cependant, des limites à la validité des modèles physiques qui régissent le Dolby Atmos.

Définition d'un point référence d'écoute et d'une zone centrale d'écoute :

La configuration préconisée par le Dolby Atmos nécessite, au préalable, de définir quelques termes. Le premier étant le point de référence d'écoute qui se situe sur la ligne centrale de l'écran aux 2/3 de la salle de cinéma. Il s'agit de l'axe acoustique des enceintes d'écran comme le décrit la *Figure C.1*. Ce point central d'écoute détermine le centre d'une zone centrale d'écoute définie sous la forme d'un rectangle dont la profondeur équivaut au tiers de la distance du premier au dernier fauteuil et dont la largeur équivaut à un tiers de la salle. Comme nous le verrons

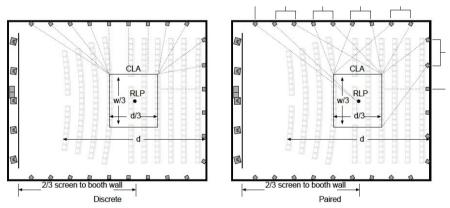


Figure C.1: Schéma d'une zone d'écoute centrale, à gauche une configuration discrètes et à droite une configuration où les enceintes acoustiques sont appairées.

plus tard, cette zone centrale d'écoute permet de déterminer l'orientation horizontale des enceintes latérales.

Configuration du Dolby Atmos sur le plan horizontal, définition des zones et des régions :

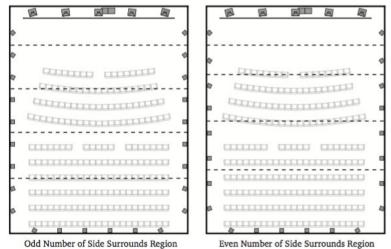


Figure C. 2 : Découpage de la salle en zones et de régions, deux configurations possibles.

Dans la dernière version du cahier des charges¹⁵⁵, les haut-parleurs surround sont sectorisés par zones et régions. Dans une salle de cinéma, il y a cinq zones : le côté gauche, le côté droit, l'arrière, le plafond gauche et le plafond droit. Les zones des côtés et du plafond sont quant à elles divisées en régions comme le suggère la *Figure C. 2*.

Ces régions sont divisées de sorte que chacune d'entre elles contienne deux paires de haut-parleurs gauche et droite, soit quatre haut-parleurs (deux de chaque côté de la salle). La région la plus proche de l'écran forme une exception car elle ne contient qu'une seule paire de haut-parleurs (gauche et droite, comme le présente la partie gauche de la *Figure C. 2*. Cette exception peut néanmoins donner naissance à une deuxième particularité. En effet, pour une salle comportant un grand nombre de

¹⁵⁵ Ibid.

haut-parleurs situés sur le côté, la deuxième région la plus proche contient elle aussi une seule paire de haut-parleurs (gauche et droite).

Ces enceintes acoustiques placées au plus près de l'écran doivent être alignées sur l'axe acoustique des enceintes d'écran. Les enceintes des murs latéraux arrières, situées dans la dernière région, doivent être positionnées aux 2/3 de la hauteur du mur. La droite formée par les enceintes avant et arrière détermine l'axe sur lequel les enceintes de côté doivent être placées joignant ainsi les deux haut-parleurs. Notons qu'il existe des tolérances sur le placement de ces enceintes sur cette droite. Les enceintes sont séparées de deux à trois mètres. Les enceintes se situant en dehors de la zone centrale d'écoute doivent avoir une orientation angulaire précise. La partie droite de la *Figure C.1* montre que, pour tout point situé sur le coin de la zone centrale, les axes acoustiques allant de ce point vers les enceintes du canal d'ambiance latéral ne doivent pas avoir entre elles un angle de plus de 30°.

La partition de la salle en zones et régions vise à établir des correspondances entre les haut-parleurs d'une même région. Le positionnement des enceintes acoustiques latérales peut s'effectuer de manière plus précise. Mais ces régions permettent aussi de regrouper plusieurs haut-parleurs sous forme de paires. Dans la dernière version de son cahier des charges¹⁵⁶, le constructeur propose d'appairer les haut-parleurs appartenant à une même zone. Les haut-parleurs sont alimentés par le même signal audio ; ainsi, un canal d'amplification peut être utilisé pour attaquer les deux enceintes appairées. Ceci permet une installation du système plus flexible pour les exploitants, mais rend, selon nous, la méthode de spatialisation décrite en Annexe D caduque, le principe de spatialisation étant plus performant sur des points sources discrets.

¹⁵⁶ *Ibid*.

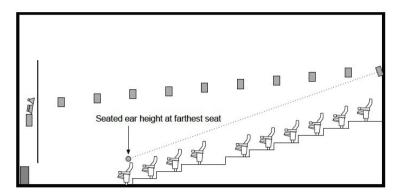


Figure C.3 : Orientation verticale des enceintes arrière.

L'orientation horizontale est donc impactée par cet appariement des enceintes par régions et s'organise comme précisé dans la partie droite de la *Figure C.1*. À l'inverse, l'orientation sur l'axe vertical est beaucoup plus simple et sans exception. Chaque haut-parleur surround doit être orienté de sorte que son axe acoustique vise le siège le plus éloigné de lui (Cf. *Figure C.3*).

Configuration des enceintes au plafond :

Une fois n'est pas coutume, le placement des enceintes de plafond est conditionné par un positionnement précis et fait lui aussi l'objet d'exceptions selon la configuration de la salle (taille, hauteur, nombre de places, etc.).

Concernant le placement longitudinal, les haut-parleurs zénithaux sont décomposés eux aussi en régions. Les haut-parleurs du plafond sont disposés en deux rangées couvrant la salle dans toute sa longueur. Ces deux rangées forment deux lignes parallèles positionnées de part et d'autre à équidistance entre l'enceinte centrale et les enceintes situées à droite et à gauche de l'écran. Ce qui correspond au positionnement des enceintes inter-gauche et inter-droite d'écran (Cf. *Figure C.4*).

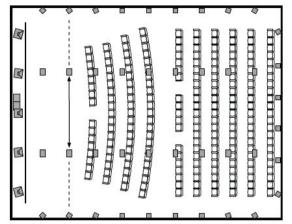


Figure C.4: Positionnement des enceintes acoustiques du plafond.

Cette figure montre aussi la configuration optimale des enceintes zénithales, à savoir le même nombre de couples de haut-parleurs suspendus au plafond que sur les murs latéraux de la même région. Ces enceintes zénithales sont donc placées sur la même ligne que les enceintes de côté de la même région. Notons que pour des salles dont le plafond est haut, l'espace entre les deux rangées de haut-parleurs peut s'agrandir selon des calculs d'angle que nous invitons à découvrir dans le document des spécifications techniques¹⁵⁷.

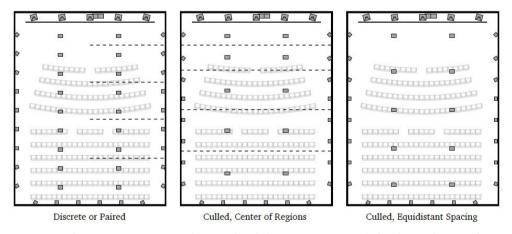


Figure C.5 : Positionnement longitudinal des enceintes au plafond. Les figures du milieu et de droite représentent les deux configurations possibles par région.

¹⁵⁷ Ibid.

Les haut-parleurs zénithaux peuvent eux aussi être appairés s'ils appartiennent à la même zone et à la même région. Les conditions sont les mêmes que sur le plan horizontal, à savoir l'appairage des haut-parleurs peut s'effectuer après la première zone d'enceintes discrètes la plus proche de l'écran. À la différence du plan horizontal, le nombre d'enceintes peut aussi varier. Les zones présentant quatre enceintes peuvent être réduites à deux suivant les deux configurations décrites dans la *Figure C.5*.

L'orientation de l'axe acoustique s'effectue selon un point situé à la demi-distance entre le point formé par l'axe central de la salle et l'axe équivalent au positionnement du haut-parleur (*Figure C.6*).

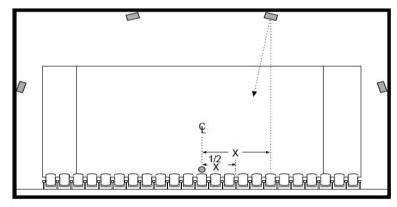


Figure C.6: Orientation de l'axe acoustique pour les haut-parleurs du plafond.

En dehors de la zone centrale d'écoute, les haut-parleurs sont soumis aux mêmes ajustements angulaires qu'au point de vue horizontal. Les haut-parleurs doivent, par conséquent, être orientés vers le bord de la zone centrale leur correspondant. Chaque angle formé par deux haut-parleurs adjacents ne doit pas dépasser 30°. D'autre part, si les enceintes sont appairées par région, celles-ci doivent avoir un angle comme indiqué dans la *Figure D.12*.

2. ÉLARGISSEMENT DU SPECTRE DES ENCEINTES SURROUND?

Dès les débuts de sa commercialisation, le Dolby Atmos avance que son système d'enceintes surround est pleine bande (*full-range*). Par conséquent, l'amélioration fréquentielle de la diffusion propose aux techniciens une nouvelle approche de l'utilisation des arrières. En effet, cette nouveauté laisse alors supposer que les sons pleine bande, préparés et mixés par les techniciens ingénieurs du son, ne seront plus altérés par la diffusion en salle. Et pourtant, le Dolby Atmos est censé respecter les normes ISO 2969 :1987 et SMPTE ST 202 :20210. Au point de mesure, la réponse en fréquence des enceintes surround devra suivre la Courbe ISO X.

Rappelons que les systèmes 5.1 et 7.1, régis par ces mêmes normes, proposent déjà un élargissement spectral des enceintes surround en comparaison avec le Dolby Surround dont la bande passante est de 100 Hz à 7 kHz. Cette bande passante réduite des écoutes surround persiste jusque dans les années 2000 où le Dolby Surround Pro Logic II propose son élargissement pour des effets arrière suivant encore une fois la courbe ISO X.

Cependant, la démarche d'élargissement spectral des surround prend source dans la volonté d'homogénéiser les réponses en fréquences des enceintes arrière avec la façade. D'autant plus que technologiquement parlant, les enceintes surround ne prenaient pas en charge les mêmes niveaux et les mêmes distorsions de fréquences pleine gamme que les canaux d'écran. Cet aspect technique aurait eu, selon Dolby, un impact direct sur la possibilité chez les mixeurs d'effectuer des spatialisations sonores de l'écran vers la salle. Ce constat aurait conduit les exploitants de salles de cinéma à ne pas mettre à niveau leur équipement surround avec des haut-parleurs de meilleure qualité créant le dilemme de la poule et de l'œuf qui a freiné l'adoption généralisée d'installations de qualité supérieure. En imposant un cahier

des charges aux exploitants, le système Dolby Atmos semble améliorer la qualité des équipements électroacoustiques installés, minimisant ainsi le problème relatif à la diffusion des bandes son chez certains exploitants par des systèmes de piètre qualité.

L'amélioration qui peut être notable est le renfort des basses des enceintes surround par *bass management*¹⁵⁸. En effet, bien que cette technique ne s'applique pas exclusivement à ce système, ce dernier tente de l'améliorer. Dans les configurations 5.1 et 7.1, le LFE est situé au niveau de l'écran. Le nombre de caissons de graves varie selon la taille des salles. Dans le nouveau système, Dolby propose deux caissons de graves à l'avant et deux renforts de graves à l'arrière (gauche et droite) pour les salles proposant un nombre de places assises supérieur à 500, et plus de sept hautparleurs par côté où la fréquence de transition (*cross over*) du *bass management* utilisé est supérieure à 80 Hz. Cela permet dans les grandes salles d'avoir un meilleur rendu spectral en basses fréquences là où un déséquilibre était notable entre l'avant et l'arrière. Néanmoins, la puissance et le côté omnidirectionnel des basses fréquences en acoustique ne permettent pas de justifier de notre point de vue l'addition d'une paire de subwoofers à l'arrière.

Au-delà des avances commerciales, le système d'enceintes surround préconisé par Dolby Atmos ne semble pas, de prime abord, proposer une avancée significative dans la diffusion sonore. Cependant, des progrès sont à noter sur le contrôle de la chaine de diffusion électroacoustique, ce qui permettrait une spatialisation des sources plus homogène et fluide entre l'écran et la salle.

¹⁵⁸ Le bass management consiste à envoyer dans le caisson des graves les fréquences basses contenues dans le signal audio envoyé aux enceintes surround. La fréquence de transition se situe entre 50 Hz et 120 Hz avec un filtrage électronique à 24 dB/octave (filtre d'orde 4). En outre, ce protocole permet à la fois de protéger les enceintes surround de la diffusion d'un contenu fréquentiel qui les endommagerait et d'améliorer le rendu sonore des basses fréquences dans les sources sonores situées dans les enceintes surround. Ce système de couplage doit être réalisé par des professionnels qui viendront mettre en phase les différentes enceintes surround avec le LFE et corriger préférentiellement le rendu en fonction de l'acoustique de la salle. Le bass management est alors contrôlé par un processeur de diffusion préalablement paramétré. En ce sens, l'utilisateur n'intervient pas à l'utilisation de ce système.

3. Une nouvelle exploitation de la dynamique sonore?

Le gain en qualité des équipements électroacoustiques se trouve essentiellement dans la redéfinition des niveaux de diffusion par les enceintes arrière. Le réglage des niveaux relatif au système Atmos ne diffère que très peu des normes électroacoustiques connues pour les installations 7.1 et 5.1 et récapitulées dans la *Figure C.7*. Néanmoins, les quelques différences notables représentent un tournant dans le monde de la diffusion cinématographique.

DCP	Ecran Gauche	Ecran Centre	Ecran Droit	Ambiance Gauche	Ambiance Droite	Ambiance arrière Gauche	Ambiance arrière Droite	Renfort de basses
5.1	85 dB(C)	85 dB(C)	85 dB(C)	82 dB(C)	82 dB(C)			85 dB(C)
7.1	85 dB(C)	85 dB(C)	85 dB(C)	85 dB(C)	85 dB(C)	85 dB(C)	85 dB(C)	85 dB(C)

Figure C.7 : Tableau des niveaux de calibration en fonction des systèmes de diffusion.

Dolby propose dans ses recommandations qu'au point de mesure (le *sweet spot*), chaque haut-parleur de façade soit capable de délivrer un niveau efficace (RMS) maximal de 105 dB_{SPL}. Ce niveau, bien qu'il soit à première vue élevé, est conforme aux recommandations de la norme SMPTE RP 200:2010, selon laquelle un signal de référence (bruit rose) à -20 dB_{FS} doit correspondre à un niveau de 85 dB_C pour chaque enceinte de la façade. Soit pour un niveau électrique du signal maximal (0 dB_{FS}), le niveau théoriquement diffusé par le canal frontal doit être 105 dB_C. Cependant, aucune mesure ouvrant sur une normalisation ne s'effectue à un tel niveau électrique. Le respect de la qualité de restitution correspondant à des niveaux électriques au-dessus du niveau de référence est de la responsabilité des exploitants.

Qui plus est, des systèmes de diffusion mal dimensionnés pourraient engendrer une distorsion harmonique voire provoquer l'endommagement du haut-parleur. Ainsi, en fixant la référence de 105 dB_{SPL}, comme niveau maximal délivré par chaque

haut-parleur frontal au point d'écoute, le système s'assure de la bonne reproduction du signal audio en façade. Cependant, aucune valeur de référence du taux de distorsion harmonique n'est communiquée ouvrant ainsi le débat sur le contrôle de la qualité des haut-parleurs.

Même s'il paraît complexe d'affirmer que le système de diffusion connaît une amélioration en façade, la déduction semble plus aisée en ce qui concerne les enceintes arrière. En effet, chaque enceinte située dans les surround doit être capable de restituer au point d'écoute un niveau moyen de 99 dB_{SPL}. Les enceintes arrière, situées dans la région la plus proche de l'écran, doivent être capable de restituer un niveau de 105 dB_{SPL} au point d'écoute. Les niveaux sonores de diffusion en salle de cinéma semblent s'approcher de plus en plus de ceux de la sonorisation. Une étude de la perception des niveaux sonores chez le spectateur dans une salle de cinéma nous paraît constituer un sujet d'étude intéressant et d'actualité. Face à de telles capacités de reproduction sonore, la question du niveau global des bandes sons dans les films semble se poser à l'heure où certains mixeurs sur-mixent les films dans l'espoir de retrouver un niveau sonore cohérent en salle de projection.

Au-delà de l'aspect physiquement dangereux que le système Dolby Atmos peut donc produire, il permet également d'effectuer des fluctuations de dynamique très importantes. Comme ce fût le cas dans *Gravity*¹⁵⁹, où un crescendo de sons oppressants vient trouver son *climax* dans un silence total. Nous pensons que les importants niveaux sonores que ce système rend possible sont plus intéressants quand ils sont associés à des niveaux plus bas. En l'occurrence, cette réflexion a déjà été menée au passage du cinéma au son numérique, où l'amélioration de la dynamique des supports d'enregistrement repoussa l'amplitude des niveaux en salles de projection. Néanmoins, il est important de souligner qu'auparavant, le report de

¹⁵⁹ Alfonso Cuarón, Gravity, op. cit.

bande son sur support magnétique ou optique jouait aussi le rôle de garde-fou dans le sens où la dynamique réduite ne permettait pas une large excursion des niveaux sonores.

Bien que cette amélioration technique permette de repousser les limitations dans les domaines artistiques, plusieurs plaintes dues au niveau sonore important de certains films ont été rapportées par des exploitants et des spectateurs. Pour répondre à ces problèmes, certains pays comme la Belgique et la Suisse commencent à réfléchir à une normalisation de la diffusion sonore au cinéma à l'image de ce qu'a fait la norme EBU-R128 pour la Télévision. Cette normalisation des mixages cinéma préviendrait d'une sur-exploitation du dispositif électroacoustique des salles. Cette normalisation, n'étant pas en vigueur en France, interroge tout de même les exploitants et les différents organismes de l'audiovisuel.

En conclusion, nous pensons que le système Dolby Atmos constitue au premier abord un outil offrant une meilleure dynamique par l'augmentation possible des niveaux sonores. Cette augmentation de niveau peut s'effectuer en répartissant l'énergie sonore dans toute la salle. De plus, il redéfinit les normes de diffusion en utilisant des systèmes qualitatifs dans une acoustique qualitative.

D MISE EN ESPACE DES OBJETS SONORES : PRINCIPES PHYSIQUES

1. Présentation

Cette partie, qui vise à étudier les différents principes mis en jeu dans la technologie employée par le système Dolby Atmos, a été complexe à entreprendre du fait de la politique interne de Dolby, qui, pour des raisons les concernant, ne communique pas énormément sur les technologies employées par son système. De plus, elle permettra de comprendre les moyens technologiques qui rendent possible la spatialisation en mode objet, à savoir le placement libre de sources sonores virtuelles dans un espace en trois dimensions.

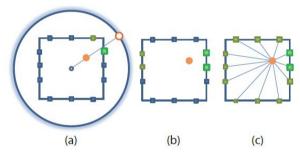


Figure C.8: Représentation des différentes technologies mises en jeu dans les dispositifs objet. En a) VBAP deux dimensions, en b) dual-balance et en c) distance-based.

Nous ne prétendons pas par la suite mettre à jour les ressources techniques du Dolby Atmos, néanmoins nous tenterons d'expliquer à partir de technologies connues les fondements que le système Dolby Atmos fournit. Les différentes technologies décrites par la suite sont issues des recherches ¹⁶⁰ et tests perceptifs ¹⁶¹ menés par Dolby. Ces éléments technologiques qui semblent répondre aux diffé-

¹⁶⁰ Charles Q. Robinson, Nicolas Tsingos et Sripal Metha, *Scalable Format and Tools to extend the possibilities of Cinema Audio*, The SMPTE Motion Imaging J., 2012, vol. 121, n° 8, pp. 63-74.

¹⁶¹ Nicolas TSINGOS, Charles Q. ROBINSON et Poppy A.C CRUM, « Evaluation of panning algorithms for theatrical applications », Dolby Laboratories Inc., Rapport de la Conférence ICSA, Erlangen, 2014.

rents enjeux du système Dolby Atmos, nous ont été plus ou moins confirmés par la firme lors d'échanges par courriel.

2. VBAP

Il existe plusieurs principes de panoramique multicanal, un des plus répandus est une extension des lois de panoramique d'amplitude utilisées dans le format stéréophonique. Cette loi d'amplitude à -3 dB s'applique par paire d'enceintes adjacentes *Pairwise Amplitude Panning* de manière à panoramiquer une source sonore avec une puissance constante d'un canal à l'autre. En stéréophonie, pour déterminer la relation entre le positionnement virtuel d'une source et l'amplitude proportionnelle nécessaire aux enceintes, Alan Blumlein a été le premier à développer une méthode connue sous le nom de « loi des sinus »¹⁶². Cette formule a été modifiée est complétée plus tard par J. C. Bennett sous une forme similaire appelée « loi des tangentes »¹⁶³.

Lorsqu'on étend ce principe de panoramique à un nombre d'enceintes supérieur à deux, on fait appel au VBAP (*Vector-Based Amplitude Panning*). Cette technologie, développée par Ville Pulkki¹⁶⁴, offre la possibilité de placer une source sonore virtuelle avec précision dans un système de haut-parleurs non prédéfini par le procédé. Cette approche permet l'utilisation d'un nombre illimité de haut-parleurs placés de manière arbitraire dans un univers en 2 ou 3 dimensions autour de l'auditeur. En utilisant le VBAP il est possible de créer des enregistrements pour une large palette de configurations de haut-parleurs et, en outre, de réaliser des

¹⁶² Michael A. Gerzon et John M. Woram, *Blumlein Stereo Microphone Technique and Author's Reply*, J. Audio Eng. Soc, 1976, vol. 24, no 1, pp. 36, 38.

¹⁶³ John C. Bennett, Keith Barker et Frederick O. Edeko, A New Approach to the Assessment of Stereophonic Sound System Performance, J. Audio Eng. Soc, 1985, vol. 33, no 5, pp. 314-321.

¹⁶⁴ Ville Pulkki, *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*, J. Audio Eng. Soc, 1997, vol. 45, n° 6, pp. 456-466.

enregistrements indépendamment des placements des haut-parleurs. Les haut-parleurs doivent néanmoins être équidistants de l'auditeur.

Le VBAP réutilise le principe du *panning* d'amplitude où l'équation utilisée pour le calcul des facteurs de gain est basée sur le concept de vecteurs et propose une reformulation de la loi des tangentes de J. C. Bennet. Cette méthode permet donc d'utiliser aussi le VBAP pour une utilisation stéréophonique, c'est-à-dire avec deux enceintes (2D). Cependant, cette reformulation innove particulièrement pour la spatialisation tridimensionnelle, où, contrairement à la loi originale des tangentes, le modèle peut être généralisé pour un trio de haut-parleurs.

La configuration stéréophonique classique peut être augmentée par un troisième haut-parleur placé à la même distance de l'auditeur que les autres. Les trois haut-parleurs forment alors un triangle appelé triangle actif (*Figure D.9*).

Comme dans les autres méthodes de spatialisation (à l'exception de la WFS), le fonctionnement optimal du VBAP nécessite que les distances entre chacune des enceintes et la position de l'auditeur soient similaires, c'est pourquoi une disposition circulaire du système de diffusion est préconisée. La source sonore virtuelle apparaît alors à l'intérieur d'un triangle formé par des haut-parleurs, ces groupes d'enceintes (formés de paires en 2D et de trios en 3D) sont appelés des bases et, lorsqu'un signal est présent sur n'importe quel élément d'une base, le calcul des facteurs de gain est effectué pour tous les membres de ce groupe. En fonction de la configuration, chaque enceinte peut appartenir à plusieurs bases, soit un maximum de deux bases pour la version 2D, et un nombre illimité pour la version 3D. La direction de la source virtuelle est dépendante de l'amplitude de diffusion de cette même source par les trois haut-parleurs. Cette relation s'opère à travers des gains appliqués à chaque haut-parleur. Pour les mêmes raisons qu'en stéréophonie, le déplacement de la source sonore doit s'effectuer à puissance constante, ce qui amène la relation (1):

$$g_1^2 + g_2^2 + g_3^2 = C (1)$$

Où g_1, g_2, g_3 sont les gains unitaires appliquées aux enceintes d'un triangle actif et C une constante.

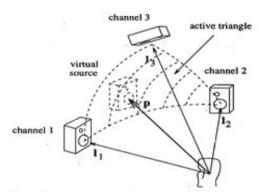


Figure D.9: Représentation du Triangle actif dans l'utilisation du VBAP. Source Ville Pulkki

La direction d'un son virtuel est donnée par la direction d'un vecteur qui est la somme de trois vecteurs en direction des trois haut-parleurs, dont le module est proportionnel à leurs gains. Dans l'optique de placer virtuellement une source sonore, l'algorithme du VBAP détermine premièrement quels sont les trois haut-parleurs les plus proches de l'emplacement virtuel de la source.

Ensuite, le gain appliqué à chaque haut-parleur est calculé par une combinaison linéaire en relation avec le placement des haut-parleurs selon un repère de coordonnées cartésiennes. Ici, x,y,z correspondent aux axes du système de coordonnées cartésiennes et w représente l'information monophonique du signal à spatialiser. Les gains appliqués aux différents haut-parleurs sont calculés de la manière suivante :

$$g_x = \cos(\theta)\cos(\gamma) \tag{2}$$

$$g_{y} = \sin(\theta)\cos(\gamma) \tag{3}$$

$$g_z = \sin(\gamma)$$
 (4)

$$g_{w} = 0.707$$
 (5)

Où γ est l'angle relatif à l'élévation et où θ l'angle relatif à l'azimut représenté dans la *Figure D.10*. Le calcul du facteur de gain est déterminé de façon similaire par un passage des coordonnées sphériques aux coordonnées cartésiennes. Notons de plus que pour trois haut-parleurs positionnés de manière orthogonale, les facteurs de gains sont équivalents à la valeur absolue des facteurs de gain calculée pour un système ambisonique (14).

La position des haut-parleurs face à l'auditeur forme un vecteur unitaire en trois dimensions. L'origine du repère cartésien décrit dans la *Figure D.10* et le point défini par l'enceinte 1 définissent le vecteur unitaire suivant $l_1 = [l_{11} \ l_{12} \ l_{13}]^T$. Les vecteurs unitaires l_1, l_2, l_3 définissent respectivement la direction des haut-parleurs 1, 2 et 3. Comme décrit dans la *Figure D.9*, la direction du son virtuel est définie sous forme d'un vecteur unitaire à trois dimensions $p = [p_1 \ p_2 \ p_3]^T$. L'expression de la source virtuelle comme combinaison linéaire des trois vecteurs unitaires l_1, l_2 et l_3 des facteurs de gain g_1, g_2 et g_3 peut s'exprimer sous la forme suivante :

$$p = g_1 l_1 + g_2 l_2 + g_3 l_3$$

$$p^{T} = g L_{123} \text{ avec } g = [g_1 g_2 g_3] \text{ et } L_{123} = [l_1 l_2 l_3]$$
(6)

Le vecteur *g* correspondant au vecteur de gain de la source virtuelle à placer peut être calculé selon la matrice suivante :

$$g = p^{T} L_{123}^{-1} = [p_{1} \ p_{2} \ p_{3}] \cdot \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix}^{-1}$$

$$(7)$$

 L_{123}^{-1} existe puisque la matrice inverse L_{123} définit l'espace tridimensionnel des haut-parleurs l_1 , l_2 et l_3 . Le facteur de gain doit être normalisé par l'équation (8) afin de répondre aux notions de puissances constantes pour un déplacement de source sonore développé en (1).

$$g^{scaled} = \frac{\sqrt{Cg}}{\sqrt{(g_1^2 + g_2^2 + g_3^2)}}$$
 (8)

À noter que lorsque trois haut-parleurs sont placés orthogonalement, le facteur de gain calculé avec le VBAP est équivalent à la valeur absolue du facteur de gain dans un système ambisonique tridimensionnel. En ce sens la technologie mise en œuvre dans le VBAP peut permettre des parallèles intéressants dans les systèmes où on l'emploie. En d'autres termes, nous ne pouvons pas affirmer que le Dolby Atmos utilise une combinaison avec ce procédé. Néanmoins son prédécesseur Imm Sound, avait affirmé l'utilisation de techniques ambisoniques dans sa description technique et il ne serait pas surprenant de le retrouver en œuvre dans le système de Dolby.

Le système VBAP tridimensionnel peut être appliqué, comme évoqué précédemment, à des systèmes comprenant plus de trois haut-parleurs placés dans l'espace verticalement et horizontalement. L'avantage du VBAP est que le placement de ces haut-parleurs ne nécessite pas un positionnement précis. Cependant, comme tout autre système de panoramique d'amplitude, la source virtuelle ne pourra pas être positionnée en dehors de la base active. Par conséquent, l'erreur maximale de localisation de la source est proportionnelle aux dimensions de la région active. D'autre part les enceintes d'un trio ne doivent pas être placées sur le même plan. Une des enceintes doit être placée en élévation par rapport aux autres. Enfin, les triangles actifs de plusieurs bases de haut-parleurs ne peuvent pas se chevaucher.

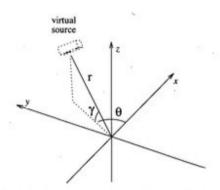


Figure D.10 : Placement de la source virtuelle dans un système de coordonnées cartésiennes.

Pour améliorer la précision en localisation dans de grands espaces, il est donc préconisé de réduire la taille des régions en augmentant le nombre de haut-parleurs dans cette zone. Cette réduction de la taille des triangles actifs permet aussi d'augmenter la distance entre l'auditeur et le haut-parleur. En effet, le signal est appliqué à seulement trois haut-parleurs en même temps, seule une différence de distance entre les haut-parleurs d'une même base et l'auditeur impliquerait une erreur de perception en localisation. En ce sens les distances entre haut-parleurs et auditeurs peuvent être plus grandes dans les extrémités d'un auditorium. Les différentes distances entre haut-parleurs peuvent être compensées par un retard (*Delay*) et par une adjonction de gain, ce qui permet un placement encore plus flexible des haut-parleurs.

Remarquons à notre lecteur que la technologie VBAP présente bien des limites, notamment dans une application de diffusion cinématographique. Il arrive effectivement, pour des raisons esthétiques, que le son soit diffusé uniformément par plusieurs enceintes, c'est le paramètre *Size* (développé en Annexe E.3.) ou encore le paramètre divergence dans les logiciels de panoramique classiques. Selon la direction du panoramique, le VBAP utilise un duo ou un trio d'enceintes qui rend impossible l'étalement de la source à plusieurs autres haut-parleurs. Dans son article de présentation universitaire ¹⁶⁵ présenté en 1999 à New York, Ville Pulkki

¹⁶⁵ Ville PULKKI, « Uniform spreading of amplitude panned virtual sources », Laboratory of Acoustics and

présente une méthode pour rendre l'étalement indépendant de la direction de la source et de son triangle actif approprié. La solution adoptée est de spatialiser le signal sonore dans de multiples directions proches les unes des autres afin de rendre l'étalement le plus homogène possible. Ce principe, connu sous le nom de MDAP (*Multiple-Direction Amplitude Panning*), contrôle l'étalement de la source sonore grâce à l'angle appliqué entre les multiples directions. Cet angle est appelé *Spread Angle* (angle d'étalement) et est applicable pour le VBAP 2D et 3D.

Enfin, Ville Pulkki décrit trois propriétés¹⁶⁶ émanant du fonctionnement des bases du VBAP :

- 1. Si la source virtuelle est située au milieu des trois haut-parleurs, les facteurs de gain des haut-parleurs sont égaux.
- 2. Si la source virtuelle est située sur une ligne entre deux haut-parleurs, le son ne sera diffusé que par ces deux haut-parleurs. Son fonctionnement sera équivalent à celui du VBAP bidimensionnel puisque le facteur de gain du troisième haut-parleur est zéro.
- 3. Si la source virtuelle est située suivant le même vecteur qu'un haut-parleur, celui-ci sera le seul à diffuser le signal. Ce qui peut poser des problèmes dans l'étalement spatial de la source et dans la coloration de celle-ci. Bien que cela dépasse notre domaine de recherche, notons qu'une étude¹⁶⁷ a été menée par l'auteur pour améliorer le timbre et la spatialisation des sources.

Pour résumer les différentes étapes du processus réalisé par le VBAP, le système appliquera les étapes suivantes dans une boucle infinie et ceci pour chaque source virtuelle déplacée :

Audio Signal Processing Helsinki University of Technology, proc. of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, 1999.

¹⁶⁶ Ville Pulkki, « Virtual Sound Source Positioning Using Vector Base Amplitude Panning », op. cit.

¹⁶⁷ Ville Pulkki, Matti Karjalainen et Vesa Välimäki, « Localization, Coloration, and Enhancement of Amplitude-Panned Virtual Sources », Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction, 1999.

- le vecteur correspondant à la source virtuelle est établi,
- les bases correspondantes sont sélectionnées,
- les facteurs de gains sont calculés
- et un fondu enchainé entre les facteurs de gain est réalisé au fil du temps.

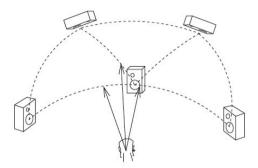


Figure D.11: Etalement d'une source sonore virtuelle par aiguillage de plusieurs directions activant un trio de haut-parleurs.

3. DBAP

Dans la plupart des techniques de diffusion multicanal, le positionnement de l'auditeur au *sweet spot* est requis. De nombreuses recherches ont été menées afin de créer un système permettant un placement plus strict des haut-parleurs et un positionnement plus libre de l'auditeur vis-à-vis du dispositif. Le DBAP (*Distance Based Amplitude Panning*), mis en œuvre par Trond Lossius, Pascal Baltazar et Théo de la Hogue¹⁶⁸, offre une alternative aux pratiques actuelles de spatialisation. Il s'agit d'un système basé sur une matrice (*Matrix-Based Spatialization Technique*) dans laquelle on renseigne la position des haut-parleurs dans la salle sans prendre le *sweet spot* comme point de référence. Les facteurs de gain pour chaque haut-parleur sont établis sur un modèle d'atténuation de distance entre la source virtuelle et l'emplacement réel des haut-parleurs.

¹⁶⁸ Trond Lossius, Pascal Baltazar et Théo de la Hogue, *DBAP-distance-based amplitude panning*, Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2009.

Le DBAP part du principe du panoramique d'intensité connu dans la stéréophonie. Néanmoins, il en étend sa définition en se libérant des contraintes de positionnement physique des haut-parleurs dans l'espace et relatif entre eux. Cette technique est aussi dérivée du VBAP où les facteurs de gain sont aussi calculés par rapport au positionnement des haut-parleurs. Pour faciliter la compréhension, nous expliquerons par la suite le fonctionnement de ce modèle en 2D, il reste bien entendu applicable et viable pour un univers tridimensionnel. Prenons une source virtuelle de coordonnées cartésiennes. Cette source virtuelle va être reproduite par N haut-parleurs, la position du dernier étant donnée par (x_i, y_i) . Le facteur de gain, calculé pour chaque haut-parleur, est déterminé par la distance entre la source et les haut-parleurs. La distance d_i , entre la source est le haut-parleur i, est donnée par la relation :

$$d_{i} = \sqrt{(x_{i} - x_{s})^{2} + (y_{i} - y_{s})^{2}}$$
(9)

Pour $1 \le i \le N$.

Comme dans le VBAP ou encore dans la stéréophonie classique, le gain de chaque haut-parleur est normalisé dans le but d'avoir un système équilibré à intensité constante. Si l'amplitude du $i^{ème}$ haut-parleur est g_i l'intensité I est donnée par :

$$I = \sum_{i=1}^{N} g_i^2 = 1 \tag{10}$$

Dans le DBAP, chaque haut-parleur est actif en continu ; le gain pour chaque hautparleur est inversement proportionnel à la distance entre le haut-parleur et la position de la source :

$$g_i = \frac{k}{d_i^a} \tag{11}$$

k est un coefficient qui dépend de la position de la source par rapport aux hautparleurs et a est un coefficient dépendant de la loi d'atténuation de la propagation du son par doublement de distance.

$$a = \frac{G_{att}}{20\log(2)} \tag{12}$$

 G_{att} correspond à l'atténuation en dB relative à la loi d'atténuation géométrique du son. Cette dernière évolue selon le modèle de propagation choisi, à savoir la propagation en champ libre ou en champ réverbéré. La combinaison de (11) et de (12) permet d'établir la formule permettant de calculer le gain nécessaire à chaque haut-parleur pour reproduire la position de la source virtuelle sonore :

$$g_i = \frac{1}{\sqrt{\sum_{i=1}^{N} \frac{1}{d_i^{a^2}}}}$$
 (13)

Cependant, si la position de l'auditeur est connue il est possible d'améliorer le rendu du procédé DBAP. En effet, jusqu'à maintenant aucun retard n'est appliqué aux haut-parleurs puisqu'aucun point d'écoute n'a été défini. Mais, si le point d'écoute est identifié, des retards sont ajoutés aux haut-parleurs pour que le son soit perçu en phase par l'auditeur. Notons aussi que le positionnement libre des haut-parleurs permet de s'adapter au positionnement conseillé par le VBAP permettant ainsi une utilisation combinée des deux techniques. En effet, le calcul des facteurs de gains dépend, tout comme en VBAP, de la position des haut-parleurs.

4. Dual Balance Panning

L e *Dual Balance Panning*¹⁶⁹ est défini comme étant une extension directe du panoramique stéréophonique classique. Son principe est visible dans toutes les installations cinématographiques classiques 5.1 ou 7.1. Il se base sur l'utilisation de deux paires de haut-parleurs (*Figure C.8b*). Le positionnement de la source sonore virtuelle dépendra du facteur de gain appliqué aux différentes enceintes. Ce facteur de gain est décomposé en deux valeurs s'échelonnant entre [0, 1] qui correspondent aux facteurs de gain pour un positionnement gauche-droite et pour un positionnement avant-arrière, respectivement.

5. Ambisonie

Nous procédons à un descriptif de ce procédé, car, bien que nous ne le supposions pas présent dans le principe de fonctionnement du Dolby Atmos, il présente des propriétés intéressantes permettant des passerelles avec les technologies décrites précédemment.

En 1970, Michael Gerzon fit une des premières approches d'un système d'encodage et de décodage d'un champ sonore en trois dimensions. L'objectif d'un tel procédé est de capter un champ sonore complexe et de le restituer très fidèlement. Sur ce plan, un tel système permet la restitution d'évènements sonores de la manière la plus homogène possible sur une sphère d'écoute. Actuellement, le HOA (*High Order Ambisonnic*) est à la base du paradigme *scene-oriented*.

Cette approche, plus connue sous sa forme restreinte au premier ordre, réalise un encodage directionnel minimal du champ sonore à travers une composante omnidirectionnelle W (pression) et trois composantes bidirectionnelles X, Y, Z (gradients de pression), la composante Z étant omise pour une restitution sur un

¹⁶⁹ Nicolas TSINGOS, Charles Q. ROBINSON et Poppy A.C CRUM, « Evaluation of panning algorithms for theatrical applications », *op. cit.*

dispositif horizontal. Le format B ainsi constitué transporte de façon explicite et concise les informations directionnelles de la scène sonore et peut donner lieu, au moyen d'un décodage, à une restitution sur des dispositifs multicanal à géométries variables (quatre haut-parleurs ou plus).

Bases théoriques :

L'ambisonie est un ensemble de techniques de synthèse, de capture et de reproduction du champ sonore basées sur une décomposition du champ acoustique en harmoniques sphériques. Cette représentation du champ acoustique, en une somme de fonctions spatiales est valable localement pour un espace défini autour d'un point donné assimilé à l'auditeur.

La représentation ambisonique traditionnelle est exprimée selon la pression p et la vitesse acoustique v au point d'écoute considéré. Exprimées sous forme scalaire, les quatre composantes WXYZ constituent le format B. Ces composantes sont respectivement équivalentes à une composante omnidirectionnelle W (pression) et trois composantes bidirectionnelles (X,Y,Z) (vélocité ou encore gradient de pression). Une spécificité majeure de l'approche de l'ambisonique repose sur le fait que les canaux transmis contiennent de manière explicite l'information directionnelle des images sonores dans la scène sonore à reproduire, indépendamment du dispositif de haut-parleurs employé. Mathématiquement, l'encodage directionnel d'une source sonore S dans la direction du vecteur unitaire — ou plutôt une onde plane d'incidence u portant un signal S — se traduit par l'équation suivante :

$$\begin{cases} W = S \\ X = \sqrt{2} \cdot \sin(\theta) \cdot \cos(\delta) \cdot S \\ Y = \sqrt{2} \cdot \sin(\theta) \cdot \cos(\delta) \cdot S \\ Z = \sqrt{2} \cdot \sin(\delta) \cdot S \end{cases}$$
(14)

Le vecteur d'incidence u est décrit par (θ, δ) en coordonnées sphériques. Le facteur de normalisation $\sqrt{2}$ a été introduit, à l'origine, pour assurer des canaux (W, X, Y) de puissances moyennes équivalentes dans le cas de sources sonores horizontales réparties uniformément dans toutes les directions, ou encore d'un champ diffus horizontal. Ces équations définissent les harmoniques sphériques de premier ordre.

La technologie de diffusion de l'ambisonique impose un nombre d'enceintes supérieur ou égal au nombre de microphones utilisés pour la captation. De plus, elle impose que ces enceintes soient uniformément réparties dans l'espace. Ceci constitue le principal défaut de l'ambisonique : la lourdeur de mise en place, et une forte contrainte de placement. En effet, la reproduction spatiale est cohérente dans une zone de *sweet spot* assez étroit, particulièrement au stade de l'ordre 1. La décomposition de l'espace dans une série d'harmoniques donnée par la série de Fourrier-Bessel impose que pour reproduire correctement tout l'espace, il y a nécessité de reproduire une infinité d'harmoniques sphériques ; ce qui est, en pratique, impossible et encore moins à l'ordre 1.

Pour dépasser ces limitations, une nouvelle extension de l'ambisonique dans des ordres plus élevés, appelée HOA, a vu le jour depuis quelques années. Là où l'ambisonique d'ordre 1 décompose l'espace acoustique en trois dimensions avec les composantes X, Y et Z, issues d'un matriçage tout à fait théorisé, l'HOA ajoute à ce matriçage des composantes associées aux harmoniques sphériques supérieures selon la décomposition de Fourrier-Bessel. Pour plus d'informations sur les enjeux de l'ambisonie d'ordre élevé nous renvoyons le lecteur au travail de thèse de Stéphanie Salaün Bertet¹⁷⁰. Plus il y a d'harmoniques sphériques discriminées, plus la scène sonore est représentée de façon précise (exemple à l'ordre 4).

¹⁷⁰ Stéphanie Salaün Bertet, Formats audio 3D hiérarchiques : caractérisation objective et perspective ces systèmes Ambisonics d'ordres supérieurs, Thèse de doctorat en Acoustique, Institut National des Sciences Appliquées, Villeurbanne, 2009.

Une prise de son ambisonique exige le recours à une installation matérielle de prise de son spécifique. Un tel système peut se révéler inadapté aux conditions de prise de son au cinéma. Néanmoins, le cinéma peut bénéficier des apports de l'ambisonie en effectuant des traitements appropriés (encodage et décodage) aux moyens d'outils logiciels sur des prises de son monophoniques.

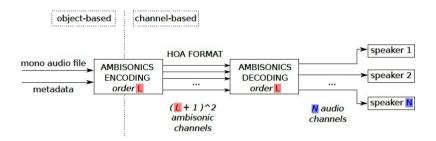


Figure D.12 : Schéma de principe d'une chaine d'encodage et décodage ambisonique.

6. VBAP, DBAP, HOA: DES PARALLÈLES INTÉRESSANTS

Durant l'étude de ces différentes techniques de spatialisation, nous avons pu entrevoir que des parallèles étaient possibles amenant l'éventualité d'une combinaison de ces trois techniques pour former un système hybride. Comme l'argumente Andrés Pérez-López dans son mémoire de Master¹⁷¹, l'approche hybride permet de cumuler les avantages des différents systèmes :

- HOA incorpore à la fois les modes orienté objet et orienté canal, tout comme le VBAP et le DBAP.
- HOA permet un mouvement des sources beaucoup plus homogène que ce proposent le VBAP et le DBAP.
- HOA incorpore les réverbérations et les sons ambiants de manière plus naturelle que ne le permettent le VBAP et le DBAP.

¹⁷¹ A. Pérez-López, *Real-Time 3D Audio Spatialization Tools for Interactive Performance*, Mémoire de Master, Université de Pompeu Fabra, Barcelone, 2014.

• Le VBAP permet une mise en espace ponctuelle et plus précise des sources sonores et ceci pour une zone d'écoute plus large que le HOA.

Cela permet aussi d'utiliser un système plutôt qu'un autre en fonction du contexte d'application et du type de contenu diffusé.

E OUTILS DE MISE EN ESPACE DÉVELOPPÉS PAR DOLBY

1. LE RMU:

Tout comme pour les systèmes de diffusion multicanal connus comme le 5.1 ou le 7.1, Dolby a dû développer des outils de spatialisation répondant aux critères du mixage orienté objet.

Au préalable, nous tenterons d'effectuer un résumé des différentes instances logicielles Dolby et de leurs fonctions principales avant d'aller plus en détails. Commençons par une présentation succincte du dispositif. Le flux audio, avant d'être diffusé en salle, se trouve sur un outil logiciel d'édition audio tel que ProTools ou encore Pyramix. Ce flux audio, afin d'être diffusé par le système Dolby Atmos, ne peut pas être directement assigné à un haut-parleur donné. Il doit tout d'abord être renseigné au moteur de spatialisation Dolby. Seul ce moteur de spatialisation peut adresser un flux audio au système de diffusion Dolby Atmos. C'est à ce niveau que l'utilisateur renseigne les données de position dans le plugin Dolby Atmos Panner. Ces données utilisateur sont alors traduites en métadonnées sous un langage propriétaire de Dolby avant d'être transmises au RMU et visualisé via l'application Dolby Atmos Monitor. Cette instance logicielle permet le contrôle visuel et audio des objets sonores. Il permet aussi de comprendre et de piloter le RMU qui traite conjointement les flux audio et leurs métadonnées. Enfin, le RMU adresse les différents flux audio qui lui parviennent aux différents haut-parleurs, via un traitement qui dépend à la fois de la configuration de la salle et des éléments utilisateurs renseignés.

Le RMU: (Rendering and Mastering Unit)

Le RMU est l'unité centrale du système de diffusion Dolby. Il s'agit d'un logiciel propriétaire de Dolby qui fonctionne sur un ordinateur marchant sous Windows. Il comporte deux cartes interfaces MADI qui lui permettent de gérer jusqu'à 128 canaux en entrée et 64 canaux en sortie. Le RMU est en quelque sorte un processeur de diffusion qui effectue du traitement du signal en temps réel. Premièrement, il gère la calibration de la salle qui lui est attribuée. Cette calibration est considérée par le RMU comme décrivant la configuration de la salle. En ce sens, il possède en mémoire toutes les informations relatives au nombre, aux modèles et aux caractéristiques des haut-parleurs, présents dans la salle. Il connaît également leurs positions, les distances qui les séparent les uns des autres, etc.

Cette machine dédiée effectue aussi des traitements d'ordre spatial; c'est à ce niveau que les métadonnées sont interprétées et le son rendu en fonction de ces informations spatiales préalablement recueillies. C'est pourquoi toute machine d'édition audio ou de mixage, traitant le mode objet, devra être reliée au RMU par liaison de type Ethernet. Ces informations de spatialisation sont traitées par le RMU en fonction de plusieurs paramètres. Le premier étant la configuration de la salle. En effet le RMU a une idée précise de la configuration de la salle qu'il est censé gérer. Au vu de ces informations et des métadonnées de positionnement utilisateur, le RMU viendra rendre (ou calculer) le ou les différents flux audio qu'il doit adresser au système de diffusion. Ce calcul est effectué grâce à un algorithme développé par Dolby. Il a pour but, entre autres, en fonction du nombre de hautparleurs et de leur positionnement, de recréer des points sources virtuels correspondant à l'information spatiale attendue par l'utilisateur. Le flux audio traité est alors acheminé aux systèmes électroacoustiques de la salle.

Mais le RMU ne se cantonne pas au traitement des flux et à leur diffusion. Il gère aussi tout ce qui relève de la distribution de la bande son dans la chaine de postproduction cinématographique. En effet, l'enregistrement final du mixage (avant l'envoi en DCP)en fin de session est ici réalisé par le RMU qui va créer un fichier (*Print Master*) comprenant le *bed* 9.1 et les objets ainsi que leurs métadonnées associées. Ce fichier comprend aussi les versions descendantes du mixage dans les formats Dolby Surround 7.1 et 5.1. Ces versions descendantes sont créées automatiquement par le RMU. Ce fichier est ensuite encapsulé dans le format MXF et encrypté par Dolby afin de préserver l'intégrité du contenu audio pour faciliter la livraison au DCP.

2. LE DOLBY LOCAL RENDERER: MODE D'EMPLOI ET UTILISATION

Principe général de fonctionnement :

Dans la chaine de postproduction audio, toutes les installations audio ne sont pas équipées de la même manière. Citons deux exemples : les salles de montage son et les auditoriums de mixage. Pour des raisons économiques et en rapport avec les besoins liés à l'exercice de ces différentes fonctions, une salle de montage son est de configuration technique moindre qu'un auditorium. Le système de diffusion y est généralement basique, dans le sens où il réduit au minimum le nombre nécessaire de haut-parleurs. Alors que dans un auditorium, les surfaces mises en jeux sont plus grandes, nécessitant des infrastructures visuelle et sonore plus conséquentes.

En ce sens, l'équipement des salles au format Dolby Atmos implique une installation technique lourde. Le but du logiciel *Local Renderer* est d'émuler et de remplacer le fonctionnement du RMU (*Figure E.13*).

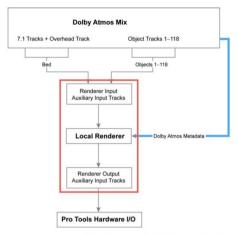


Figure E.13 : Schéma d'intégration du Local Renderer dans une configuration où ProTools contient les métadonnées objets.

Il s'agit d'une version logicielle du RMU qui fonctionne en tache de fond avec ProTools. Il s'adresse donc principalement aux salles de montage qui ne seraient pas équipées de la version matérielle. Bien que la version logicielle du RMU n'offre pas toutes les fonctionnalités de sa version matérielle, elle permet aux petites installations de bénéficier de l'environnement orienté objet proposé par le Dolby Atmos. Seules les fonctionnalités de spatialisation et de downmixage vers des versions orientées canal sont possibles. Cet outil sert donc au monteur son désireux de travailler sous un univers orienté objet (Dolby Atmos) sans avoir besoin des ressources du RMU.

Configuration Technique:

Le Local Renderer est constitué d'un ensemble de plugins d'extension .aax : Local Renderer Input e t Local Renderer Output installés sur le logiciel d'édition ProTools. Ces deux plugins communiquent avec le Local Renderer qui donne accès aux fonctionnalités du mode orienté objet. Ce logiciel permet, entre autres, de rendre effectifs les panoramiques effectués dans le Dolby Atmos Panner afin de préparer, si on le désire, le travail de spatialisation dès la salle de montage. Il

permet de rendre le mixage des différentes sources compatible avec le système de diffusion.

En outre, le *Local Renderer* requiert une configuration de haut-parleurs minimum pour permettre à l'utilisateur d'exploiter les ressources du panoramique Dolby Atmos. Ainsi, à l'ouverture du logiciel, nous avons le choix entre une configuration minimale 9.1¹⁷² et une configuration maximale correspondant à 15.1¹⁷³. Le *Local Renderer* peut alors supporter jusqu'à 16 haut-parleurs.

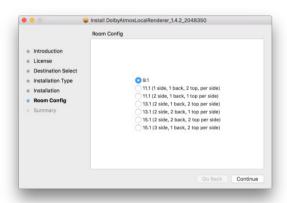


Figure E.14: Différentes configurations possibles du logiciel Local Renderer.

Une fois effectué le choix de la configuration de diffusion ainsi que l'installation du logiciel, l'utilisateur fait appel à seulement deux plugins : Local Renderer Input et Local Renderer Output. Le Local Renderer Input prélève un signal depuis ProTools pour l'envoyer vers le Local Renderer. Le Local Render Output prélève le signal traité par Local Renderer afin de le ré-envoyer dans Pro-Tools qui gère l'envoi des flux audio vers les sorties matérielles (haut-parleurs). En d'autres termes, les deux plugins font ce qu'ils disent ; le Local Renderer Input envoie les flux audio (bed et objet) vers le Local Renderer, le Local Renderer Output reçoit le mixage des flux

¹⁷² Il faut entendre le 9.1 comme une extension du format 7.1 correspondant à l'ajout de deux canaux zénithaux.

¹⁷³ Cette configuration correspond à trois enceintes frontales (LCR), trois enceintes par côté (soit un total de 6 pour la gauche et la droite), deux enceintes arrière (gauche et droite), deux enceintes par côté en élévation (soit un total de 4 haut-parleurs suspendus au plafond) et un caisson de renforcement des basses.

audio du *Local Renderer*. Basiquement les entrées audio du *Local Renderer* s'effectueront avec le plugin *input* et les sorties avec le plugin *output*.

Dans la suite de cette partie, nous expliquerons la configuration du logiciel d'édition nécessaire au fonctionnement du *Local Renderer*.

Commençons par créer des pistes auxiliaires sur ProTools :

- Un auxiliaire 7.1 correspondant à la partie horizontale. Cet auxiliaire constituera ce que Dolby appelle le *bed* et que nous nommerons 7.1 Bus.
- Un auxiliaire stéréophonique correspondant à la partie zénithale du *bed* 9.1 ; appelons cet auxiliaire OH Bus pour *OverHead Bus*.
- N nombre d'auxiliaires mono/stéréo correspondant au nombre d'objets mono/stéréo utilisés dans le projet. Ces auxiliaires serviront à envoyer le signal audio objet via le plugin *Local Renderer Input* au *Local Renderer* qui effectuera les traitements développés précédemment. Pour ceci les pistes auxiliaires devront comporter en insert le plugin *Local Renderer Input* en référençant la source type en tant qu'objet et en renseignant la source ID correspondant au numéro de l'objet. Ces auxiliaires porteront la dénomination *Objet n Rendering*.
- Un nombre d'auxiliaires dépendant du nombre de points sources utilisés dans le choix de configuration du format de diffusion. Dans le cas du 9.1, 10 auxiliaires seront donc utilisés. Ces auxiliaires servent à assigner le retour de traitement du *Local Renderer* au système de diffusion. En ce sens, chaque auxiliaire correspond à une enceinte physique. Par exemple il y aura une piste auxiliaire *Left* assignée à l'enceinte de gauche. Sur cette piste *Left* devra figurer en insert le plugin *Local Renderer Output*. Dans notre

exemple, la sortie numéro 1 du *Local Renderer* est renseignée dans l'onglet *Output Channel*.

Il faut comprendre que ces pistes auxiliaires serviront à l'envoi des flux audio vers le *Local Renderer* et au retour des flux du *Local Renderer* afin d'être assignées vers leur enceinte correspondante. L'avantage de cette méthode de *routing* est que les envois vers le *Local Renderer* s'effectuent via des bus, cette session peut être utilisée en phase de mixage et retourner en salle de montage sans changer le *routing* de sortie.

Le principe est simple, chaque piste audio doit être assignée à deux endroits différents. En sortie principale, il faut assigner la piste audio à sa sortie physique (en auditorium de mixage) correspondante. Par exemple, la piste Objet 1 aura pour sortie Objet 1 MADI 1. Ce système ne permet pas d'écouter cette piste via le *Local* Renderer. Cela nécessite d'envoyer en départ auxiliaire (send sur ProTools) le flux audio Objet 1 via un bus appelé Objet 1 bus vers la piste auxiliaire Objet 1 Rendering. L'auxiliaire Objet 1 Rendering devra, dans ses entrées y référencer le bus Objet 1 bus, rien ne sera référencé dans ses sorties car le flux audio passe par le Local Renderer Input afin d'être traité dans le Local Renderer et réinjecté dans ProTools via le Local Renderer Output. Le flux audio, une fois dans ProTools, sera routé grâce à des pistes auxiliaires vers les différentes sorties audio liées au système de diffusion. Ainsi, les sorties MADI n'étant pas disponibles dans la salle de montage, elles seront inactives (grisées) et l'auxiliaire envoie le flux audio vers le Local Renderer. En phase de mixage, les pistes auxiliaires d'envoi vers le Local Renderer seront inactives, par contre, les sorties MADI seront actives et permettront d'envoyer le flux audio à la console.



Figure E.15 : Configuration des pistes bed et objets ainsi que leurs bus associés.

3. LE DOLBY ATMOS PANNER

Présentation Générale :

Le Dolby Atmos est un plugiciel développé par Dolby pour fonctionner actuel-lement et exclusivement avec ProTools. Il s'agit d'un plugin natif d'extension .aax (64 bits et 32 bits) permettant d'effectuer des panoramiques dans le mode orienté objet du Dolby Atmos. C'est cette instance logicielle qui va créer les métadonnées de spatialisation et d'autres métadonnées propres à Dolby auxquelles nous n'avons pas pu avoir accès. En ce sens, ce plugin n'affecte pas directement le son sur la piste où il est actif, le traitement s'effectuera en tache de fond par une autre instance logiciel (RMU ou Local Renderer). Le son sera traité en fonction des données renseignées par l'utilisateur dans le plugin et le traitement s'effectuera avant l'envoi aux haut-parleurs.

Le plugin de Spatialisation :

Un outil de spatialisation a été développé pour permettre aux utilisateurs de déplacer un objet sonore dans un univers qualifié de 3D. L'interface est similaire à celles que mettent en pratique des systèmes de panoramique conventionnels (Stéréo, 5.1, 7.1). Cependant, à la différence des précédentes interfaces qui proposaient une spatialisation sur le plan horizontal, le modèle de mixage objet nécessite une composante Z (composante d'élévation). Le plugin *Dolby Atmos Panner* offre une spatialisation dans l'espace tridimensionnel défini par la configuration Atmos requise à la diffusion en salle. Ce plugin vient s'insérer sur toutes les pistes ou ensembles de pistes qualifiés d'objets. Ainsi la piste monophonique, stéréophonique ou bus prendra la dénomination d'objet.

Ce plugin est à activer en insert de la piste objet à spatialiser. Néanmoins, à la différence de l'outil de panoramique classique, le *Dolby Atmos Panner* n'agit pas directement sur le traitement du son. Les données intrinsèques à la spatialisation sont transmises numériquement au RMU (ou au *Local Renderer*). Celui-ci viendra ensuite appliquer le traitement du signal approprié à l'information de spatialisation de l'utilisateur.

Configuration et paramétrage du panner :

Comme nous l'avons vu, la spatialisation ne s'effectue pas selon une loi de panoramique classique s'appliquant sur le son de la piste concernée. Ici, il faut renseigner le numéro de l'objet concerné par le panoramique ; il ne suffit donc pas de l'insérer sur la piste que l'on veut spatialiser. Le bouton de configuration permet ainsi de renseigner le numéro de l'objet que l'on veut mettre en espace. C'est aussi dans cette rubrique que l'on peut assigner le plugin à un joystick ou à des boutons d'une console de commande.

Chemin du signal & différentes configurations possibles :

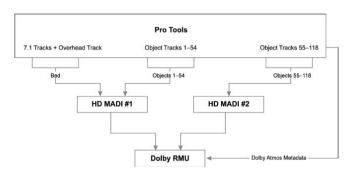


Figure E.16 : Exemple de chemin du signal audio et des métadonnées.

La *Figure E.16* s'applique essentiellement aux configurations n'impliquant pas de console de mixage, ce qui le réduit essentiellement à une utilisation en salle d'écoute ou encore en salle de montage (avec l'utilisation du *Local Renderer* à la place du RMU).

Du premier exemple le plus classique apparaissent alors deux autres manières de fonctionner. La première étant celle employée dans les débuts du Dolby Atmos, où, certaines consoles, notamment l'Euphonix System 5, ne permettaient pas une spatialisation en 3D, et n'étaient alors que des commandes du *panner*. Cela consiste à envoyer des données de commandes au RMU qui les retransmet à son tour au *Dolby Atmos Panner* (*Figure E.17*). Celui-ci retranscrit les informations de commandes en informations de spatialisation et les renvoie au RMU afin qu'il effectue le traitement adéquat. Dans cette configuration, l'ordinateur réalise la transcription des informations de commande en données de spatialisation et non la console de mixage. La session ProTools a aussi le rôle d'enregistrement des automations de spatialisation via le contrôle du plugin par la console (mode télécommande). Ce type de configuration permet l'utilisation de la console en tant qu'interface et libère le mixeur de l'utilisation de ProTools.

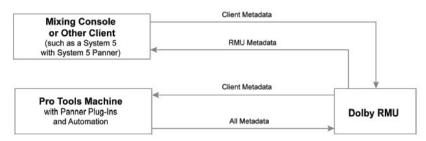


Figure E.17: Représentation du chemin du signal des métadonnées de spatialisation propriétaires aux consoles et au Dolby Atmos. Source:

Authoring for Dolby Atmos © Sound Manual

Dans notre explication ci-dessus, tout travail de spatialisation était effectué au moyen de logiciels d'édition. Or, en phase de mixage, les consoles doivent aussi pouvoir contrôler le mode objet. C'est pourquoi, les principaux constructeurs de consoles de mixage ont dû élaborer leurs propres logiciels de panoramique implémentant un système de panoramique à trois dimensions, tout en répondant aux attentes du mode orienté objet. Les consoles audionumériques, qui ne traitaient autrefois que des données audio, disposent maintenant de modules qui traitent les métadonnées directement issues du RMU. Les informations émises par les logiciels d'édition audio ne sont pas pour autant perdues. La centralisation des métadonnées par le RMU permet un échange bidirectionnel des informations de spatialisation. Ainsi, une automation, réalisée par ProTools, se retrouvera transcrite sur l'outil panoramique de la console et réciproquement.

Le constructeur anglo-saxon de consoles AMS-Neve fut l'un des premiers à implémenter un panner 3D dans sa console DFC Gemini. En étroite collaboration avec Dolby Labs, les fabricants de consoles tels que Harrison ou Euphonix, ont implémenté un outil de panoramique dédié à l'utilisation du Dolby Atmos. Cet outil a la particularité de fonctionner directement en lien avec le RMU sans passer par le Dolby Atmos Panner plugin. Les informations sont maintenant écrites dans les données d'automation de la console et de la session ProTools (Figure E.18).

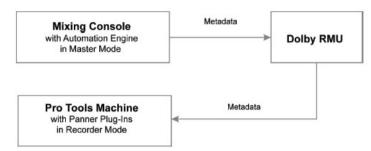


Figure E.18 : Configuration la plus simplifiée des échanges entre la console de mixage, le RMU et la station d'édition ProTools. Source : Authoring for Dolby Atmos © Sound Manual

L'enregistrement des données de spatialisation s'effectue non plus par le contrôle du plugin en télécommande mais via le mode *recorder* activé dans le panner de Dolby. Ce mode permet l'enregistrement des métadonnées Dolby Atmos de la console vers l'ordinateur en passant par le RMU. Cette configuration est assez intéressante notamment lorsque le travail de spatialisation a été commencé en phase de montage son. Les données de spatialisation sont alors lisibles en auditorium et modifiables par le mixeur. Ce mode de communication bidirectionnelle permet la transportabilité des panoramiques entre la salle de montage et l'auditorium.

Présentation de la fenêtre du plugin :

La fenêtre du plugin s'organise en plusieurs parties. L'interface utilisateur de spatialisation d'un objet sonore passe une représentation de l'objet par un point jaune ; la salle dans lequel il est spatialisé apparaît sous forme de carré. Ce carré, proposant une vision en hauteur de la salle, est orienté par la représentation d'un écran. Les carrés bleus représentent les groupes d'enceintes présents dans l'installation. Le point jaune représentant la position de la source sonore est situé par défaut au centre de la salle ce qui correspond aux coordonnées indiquées dans la *Figure E.19*. Le positionnement sur l'axe vertical est visualisé par le grossissement du point jaune. Cette orientation s'effectue de plusieurs manières et fait appel aux différents modes de panoramique en élévation proposés sur le côté droit de la

fenêtre. La première étant en mode *Elevation Mode Button* qui permet un contrôle libre du point jaune, l'élévation s'effectue avec le graduateur Z et ceci peu importe le positionnement du point jaune dans la salle. Ensuite dérivent plusieurs modes dont l'icône représente le mouvement de panoramique effectué :

- Le *Wedge Elevation* divise la salle en deux parties égales, l'avant et l'arrière salle. La montée en élévation s'effectue de manière triangulaire. Le déplacement du point jaune de l'avant vers l'arrière part du plus bas du système de diffusion, atteint le haut en milieu de salle et retrouve un positionnement bas à l'arrière de la salle.
- Le mode *Sphere Elevation* propose une trajectoire verticale sphérique selon le positionnement du pointeur jaune dans la salle.
- Le mode *Ceilling Elevation* propose une montée zénithale très rapide de l'objet sonore. Un point objet se situant au premier tiers de la salle sera perçu théoriquement comme provenant en hauteur.

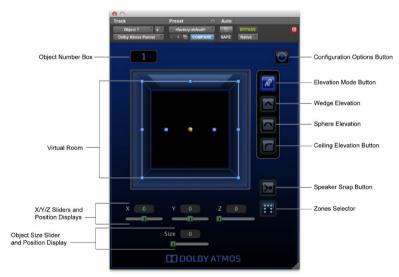


Figure E.19: Fenêtre du plugin Dolby Atmos Panner. Source: Authoring for Dolby Atmos © Sound Manual

Le paramètre *Size* permet de gérer la largeur de la source sonore. Il se traduit dans la salle par la répartition du son sur les enceintes ; à la base ponctuelle (un haut parleur ou un groupement de haut-parleurs utilisés pour effectuer le placement virtuel et ponctuel de la source), il peut être envoyé vers d'autres haut-parleurs adjacents comme le suggère la *Figure E.20*.



Figure E.20 : Fenêtre du plugin, Dolby Atmos Panner avec l'outil Size en action.

Rappelons à notre lecteur, l'existence de deux derniers modes qui s'éloignent des fonctions classiques d'un logiciel de panoramique : *Speaker snap* e t *snap zone selector*. Le mode *Speaker snap* permet d'assigner un son à une seule enceinte à la fois. Ce mode est dégagé de n'importe quelles lois de panoramique puisque le son n'est pas spatialisé selon les méthodes citées en Annexe D. Le mode *snap zone selector* permet de sélectionner plusieurs enceintes assurant la diffusion d'un même son. La source sonore sera diffusée au même niveau par toutes les enceintes de la zone sélectionnée. Ce mode s'apparente fortement à l'approche orientée canal puisque c'est l'utilisateur qui sélectionne les canaux de diffusion et non le processeur de diffusion.

Méthode de panoramique selon le type de pistes :

Le *Dolby Atmos Panne*r peut être utilisé sur des pistes objet monophoniques et stéréophoniques. La *Figure E.21* montre la fenêtre du plugin pour une piste objet stéréophonique. Ces modes sont paramétrables dans la fenêtre *linking* et proposent le déplacement libre des deux objets, leur déplacement miroir ou encore la reproduction du même déplacement dans l'espace. La version stéréophonique du *panner* reste très semblable à celle que l'on pourrait trouver dans le monde orienté canal. Remarquons qu'il n'existe pas de version multipistes du plugin, ce qui restreint les prises de sons multicanal au monde de l'orienté canal ou aux *bed*.



Figure E.21 : Fenêtre du plugin pour une piste objet stéréophonique.

F COMPATIBILITÉ DES MIXAGES, DOWNMIXAGE AUTOMATIQUE

La question de la rétrocompatibilité des mixages se pose conjointement à l'évolution des technologies de diffusion au cinéma. Une fois n'est pas coutume, la firme Dolby a proposé un système sous forme logicielle ou matérielle pour répondre à ces questions. Le but étant de lancer le nouveau système de diffusion sur le marché tout en restant compatible avec les salles qui n'en sont pas encore équipées. Pour amener les productions à effectuer leur film dans un nouveau format, il faut que le temps de post-production reste le même. En d'autres termes, les coûts relatifs à la conception de la bande son doivent être inchangés. Or, comment est-il possible d'effectuer un mixage dans le nouveau format et les mixages assurant la rétrocompatibilité dans les mêmes temps de postproduction? Une des solutions commerciales est le recours au downmixage automatique.

1. FONCTIONNEMENT GÉNÉRAL DES PROCESSEURS DE DIFFUSION ORIENTÉE OBJET

Dans cette partie, nous tenterons d'expliquer le plus succinctement possible le principe général de fonctionnement des downixeurs automatiques proposés par certains formats de mixage orienté objet. Notre démarche s'appuiera sur le communiqué de la nouvelle norme MPEG-H Audio 3D¹⁷⁴, de son procédé de spatialisation SAOC¹⁷⁵ ainsi que sur la documentation relative à l'évolution du standard de compression Dolby AC-4¹⁷⁶ développé conjointement avec la norme ETSI TS

¹⁷⁴ Jürgen HERRE, Johannes HILPERT, Achim KUNTZ et Jan PLOGSTIES, MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding, J. Audio Eng. Soc, 2015, vol. 62, nº 12, pp. 821-830.

¹⁷⁵ Jeroen Breebaart, Jonas Engdegard, Cornelia Falch, Oliver Hellmuth, Johannes Hilpert, Andreas Hoelzer, Jeroen Koppens, Werner Oomen, Barbara Resch, Erik Schullers et others, « Spatial audio object coding (SAOC)-The upcoming MPEG standard on parametric object based audio coding », Convention Paper 7377, Presented at the 124th Convention, Amsterdam, 2008.

¹⁷⁶ DOLBY LABORATORIES INC., *Dolby*© *AC-4*: *Audio Delivery for Next-Generation Entertainment Services*, Document de normalisation, 2015.

103¹⁷⁷. Nous conseillons vivement la lecture de ces documents pour celles et ceux qui souhaitent approfondir les fondements du downmixage automatique.

D'autre part, nous avons remarqué durant l'étude des bases du Dolby Atmos de nombreuses similitudes avec les systèmes classiques 5.1 et 7.1. En effet, premièrement, le VBAP développé en Annexe D est une évolution technologique des procédés traditionnels de spatialisation. Cette technique reste compatible avec ses prédécesseurs (panoramique d'intensité). Deuxièmement, comme nous le remarquons en Annexe B, le système Dolby Atmos, outre l'addition de plusieurs hautparleurs, conserve une base ressemblant fortement aux configurations 5.1 et 7.1. Enfin, en Annexe C, nous constatons que l'élargissement de la réponse en fréquences des enceintes arrière a déjà été engagé dans les systèmes 5.1 et 7.1. L'évolution que propose le Dolby Atmos se situe essentiellement dans le complément basses fréquences des enceintes arrière. Bien que de nombreux progrès aient été amenés avec ce principe de diffusion, de grandes analogies avec ses ancêtres facilitent selon nous la compatibilité descendante et permettent au downmixeur automatique un travail simplifié et plus efficace.

Depuis 2008, le SAOC est le format standard de spatialisation en mixages orienté objet. Il fut élaboré en collaboration avec l'institut de Fraunhofer, Dolby Labs et la branche recherche de la société Philips. Cette technologie propose des solutions d'encodage des flux audio orientés objet en un downmixage monophonique ou stéréophonique et un décodage permettant de reconstituer un mixage orienté objet cohérent. Cette norme, bien que s'éloignant de notre sujet, semble plus que nécessaire à l'heure où les mixages orientés objet nécessitent un nombre de pistes très important.

¹⁷⁷ ETSI, Digital Audio Compression (AC-4) Standard Part 2: Immersive and personalized audio, ETSI TS 103 190-2, 2015, vol. 1.1.1.

Le SAOC développe un algorithme de codage et de décodage sous forme matricielle, permettant à partir d'un flux orienté objet, de rendre un mixage sur une configuration variable de haut-parleurs. Cet algorithme est fixé dans une phase d'initialisation du programme et du système de reproduction. L'enjeu est d'orienter un flux audio comportant N objets vers une configuration électroacoustique comportant M canaux. Pour que l'acheminement du signal audio objet au(x) hautparleur(s) soit réalisé, une matrice de rendu est établie en lien avec le nombre de haut-parleurs, leur positionnement dans la salle, leur positionnement relatif et les métadonnées de positionnement des objets audio. Cette matrice¹⁷⁸, présentée dans le document de standardisation du MPEG-H Audio 3D¹⁷⁹ et détaillée dans le document de normalisation ETSI TS 103180, détermine un premier ensemble de lois définissant le rendu d'une entrée audio donnée vers une ou plusieurs sorties audio. Ces règles sont créées individuellement pour chaque entrée audio et peuvent être ajustées selon le système employé. Par exemple, le SAOC implémente des corrections visant à privilégier l'assignation d'un signal à une enceinte plutôt que la création de sources virtuelles via un réseau d'enceintes.

Certains processeurs de *downmixage* automatique proposent aussi un algorithme actif en addition au caractère passif décrit ci-dessus. Le but est de respecter au mieux les intentions artistiques mises en jeu dans le contenu audio du programme en fonction de conditions d'écoute données. Ceci est réalisé par une première phase d'analyse du contenu audio qui débouche sur son traitement dans une deuxième partie. Par exemple, l'algorithme développé dans la norme MPEG-H Audio 3D, permet d'éviter d'éventuels artefacts comme la suppression ou le filtrage d'un

¹⁷⁸ La matrice de rendu A oriente les N objets du mixage objet dans la configuration de haut-parleurs définie par Y. La matrice A est un ensemble de coefficients établis selon la configuration électroacoustique (nombres de haut parleurs, angulations entre eux) et les différentes entrées audio du programme. La matrice Y représente la matrice cible à alimenter selon la relation Y=A.S, où S représente la matrice de N objets.

¹⁷⁹ Jürgen HERRE, Johannes HILPERT, Achim KUNTZ et Jan PLOGSTIES, «MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding », op. cit., p. 12.

¹⁸⁰ ETSI, « Digital Audio Compression (AC-4) Standard Part 2 : Immersive and personalized audio », *op. cit.*, p. 87.

signal, conséquences d'un downmix passif où plusieurs signaux corrélés en phase, viennent s'additionner matriciellement au sein d'un ou plusieurs canaux. Ce qui, dans le cadre de l'évolution des systèmes audio 3D, devient de plus en plus fréquent. En effet, pour alimenter ces haut-parleurs additionnels présents dans les installations 3D, un même signal est copié, redistribué, filtré et traité par des délais ou des réverbérations. Le downmixeur actif adapte donc le signal afin de prévenir ce type de problèmes. Premièrement, une analyse de corrélation des signaux entrants est effectuée avant de l'intégrer dans le calcul de *downmix* en alignant la phase des différents signaux si nécessaire. Deuxièmement, afin de préserver l'énergie du signal entrant, alors pondéré par le calcul matriciel, une normalisation en puissance est effectuée.

2. FONCTIONNEMENT DU DOWMIXEUR DOLBY ATMOS:

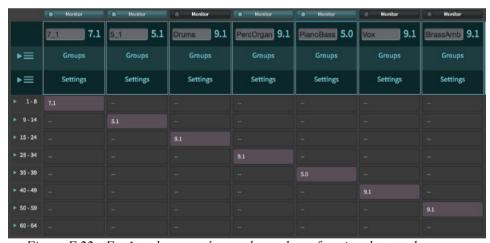


Figure F.22 : Fenêtre du paramétrage du rendu en fonction du type de source.

Dans le cadre de notre recherche et du système étudié, il est difficile de déterminer si le Dolby Atmos utilise un système de downmixeur actif. Malgré tout, il est sûr que le principe de base s'appuie sur un calcul matriciel. Dans la version actualisée des outils de mixage développés par Dolby ¹⁸¹, le RMU propose plusieurs manières

¹⁸¹ DOLBY LABORATORIES INC., What's new in Dolby Atmos mixing tools v1.6, Documentation de

de rendre son mixage orienté objet exportable vers des formats orientés canal 5.1 ou 7.1.

Dans cette version matérielle, il est alors possible d'effectuer des downmixages automatiques vers les formats suivants : 5.1, 7.1 et 9.1 (Cf. *Figure F.22*). L'intérêt se situe selon nous au niveau de la matrice de rendu qui propose ces différents formats en fonction de la source à downmixer (Dialogue, Musique, Ambiance, Effet) et du type de signal (*bed, Objet* ou *bed* et *Objet*). Ces différentes sources qui s'apparentent à la notion de *stem*¹⁸² peuvent subir des traitements additionnels en phase de downmixage :

- Possibilité d'appliquer un gain général sur l'ensemble des enceintes surround du format orienté canal choisi.
- Possibilité d'orienter le signal basses fréquences issu du bass management des enceintes arrière vers le canal LFE.
- Choix de la fréquence pour laquelle le contenu fréquentiel basses fréquences du contenu surround est envoyé au canal LFE ou tout simplement supprimé en fonction de la configuration utilisée.
- Un gain de +3 dB dans la chaine audio surround qui corrige la différence d'alignement sonore entre une configuration surround classique calibrée à 82 dB_{SPL} pour le 5.1 et 85 dB_{SPL} pour la configuration Dolby Atmos.
- Réglage de gain de la contribution du contenu audio zénithal. Ce paramètre permet de régler le niveau des informations zénithales lors d'un passage à des formats horizontaux tel que le 5.1 ou 7.1.

spécification technique, 2016.

¹⁸² La notion de *stem* dans le mixage cinéma classique représente un groupe de pistes configuré pour un format orienté canal. Ce groupe de pistes est alimenté par un seul type de sources. Par exemple, tous les dialogues d'un film sont orientés vers le *stem* dialogue. Le mixage final est issu de la sommation de plusieurs *stems* existant dans le projet mixé.

G <u>Déscription du schéma de principe du décodeur MPEG H Audio</u> 3D:

- La première étape soumet tous les flux audio au décodeur MPEG de compression de données. Les flux audio peuvent contenir plusieurs méthodes de spatialisation dont le scene-based sous le nom de HOA, objectbased pour les techniques VBAP et WFS, channel-based pour les mixages orientés canal.
- Le signal *channel-based* est orienté vers le système de reproduction utilisateur cible, en passant par un convertisseur selon la configuration.
- Le signal *object-based* est mis en espace grâce à ses métadonnées associées selon le système utilisateur cible.
- Les systèmes de spatialisation propriétaires SAOC (Spatial Audio Object Coding) calculent le rendu des objets et des canaux de leur flux audio, selon le système de restitution cible avec l'aide de leurs métadonnées propres.
- Un mixer virtuel propose de rendre toutes ces méthodes de spatialisation selon la configuration de l'utilisateur. Notons même que ce décodeur prévoit un mixage binaural pour les utilisateurs bénéficiant d'un casque audio.

H DESCRIPTIF DES FILMS UTILISÉS

1. Les Medouny, 2016, 16'

Au petit matin, Saïd se rend à la salle de boxe. Ses deux frères s'entraînent pour un combat à venir. De toute évidence les trois frères ne semblent pas entretenir une relation calme et paisible. La supériorité de l'ainé, Silvio, propulse ses frères cadets à suivre ses préceptes. Tandis que Malick suis l'entrainement poussé et exigent de Silvio afin de concourir à son prochain combat, Said le benjamin prend du recul par rapport à sa famille. Dans la voiture pour se rendre chez leur père Karim, ancien champion de boxe, personne ne parle, le malaise est palpable. Durant le déjeuner, la discussion, au début banale finit par s'envenimer quand le père doute des capacités



Figure H.23: Les Medouny, Simon Bonanni, 2016, Photogramme.

de Malick à remporter le combat. La tension monte, et leur père apprend qu'il ne lui reste que peu de temps à vivre. Le trio que forment les trois frères va devoir changer sa personnalité et son lien pour surmonter cette épreuve.

2. Bamako Saint Afrique, 2013, 14'

« Vous ne connaissez pas le Bamako-Saint-Afrique ? Le plus grand rallye du monde. Une course cent pour cent africaine. Une course qui traverse les terres pas toujours hospitalières de la France profonde. Un rêve pour Djibril et Ali les deux personnages principaux, l'équipage malien qui, pour la première fois, est en mesure de l'emporter. Mais du rêve au cauchemar, la frontière est souvent mince face au choc des civilisations »

Avec humour et légèreté, ce court métrage traite d'un sujet beaucoup plus grave qu'il n'en a l'air. Le réalisateur propose une version satirique de la vision européenne de la mort d'autochtones par accident durant les rallyes comme le Paris Dakar en fait chaque année le triste constat, par exemple. Afin de nous interpeller, et de dénoncer ce rapport inégal entre l'Afrique et l'Europe, les rôles s'inversent et les habitants de la région du limousin deviennent les spectateurs autochtones du rallye Bamako – Saint Afrique, non à l'abri d'accidents. D'autre part, cette comédie ouvre la réflexion sur une vision peu glorifiante des décisions politiques Françaises, notamment concernant l'immigration.



Figure H.24: Bamako Saint Afrique, Frédéric Jolfre, 2013, Photogramme.

¹⁸³ Synopsis original du film. Source Unifrance.

3. Les Saisons, 2016, 1 h 37'

L'hiver durait depuis 80 000 ans lorsque, en un temps très bref, une forêt immense recouvre tout le continent. Une nouvelle configuration planétaire et tout est bouleversé. Le cycle des saisons se met en place, le paysage se métamorphose, la faune et la flore évoluent. L'histoire commence...

Les hommes deviennent chasseurs-cueilleurs. Ils interviennent de plus en plus sur leur environnement et la déforestation commence. Au plus près d'espèces sauvages comme les ours, les chevaux, les lynx ou les loups, les réalisateurs reviennent sur l'histoire de ces territoires que l'homme partage avec les animaux, de la forêt profonde à la campagne riante. *Les Saisons*¹⁸⁴ est une épopée sensible qui relate la longue et tumultueuse histoire commune qui lie l'homme aux animaux.



Figure H.25: Les Saisons, Jacques Perrin et Jacques Cluzaud, 2016, Photogramme.

¹⁸⁴ Jacques Perrin et Jacques Cluzaud, Les Saisons, op. cit.