
COLIN Etienne

L'audio orienté objet : une solution face à la multiplicité des formats



Aix-Marseille Université

Département Sciences, Arts et Techniques de l'Image et du Son

Mémoire de Master « Cinéma et Audiovisuel »

2018-2019

29/03/2019

Etienne COLIN

**Département Sciences, Arts et Techniques de l'Image et
du Son (SATIS)**

Aix-Marseille Université

Mémoire de Master « Cinéma et Audiovisuel »

**L'audio orienté objet :
Une solution face à la multiplicité des
formats**

Travail réalisé sous la direction de : Frédéric BELIN

Mars 2019

Remerciements :

Merci à mon directeur de mémoire Frédéric Belin pour ses retours pertinents.

Merci à Jean-Michel Denizart pour son aide, sa bienveillance et pour l'intérêt qu'il a porté à ma recherche.

Merci à François-Xavier Ballon, Chris Baume, Thibaut Carpentier, Hervé Déjardin, Benjamin Duval, Nicolas Epain, Xavier Gibert, Jean-Christophe Messonnier, David Paillet, Matthieu Parmentier et Olivier Warusfel, pour le temps qu'ils m'ont accordé.

Merci aux membres du consortium ORPHEUS et aux équipes de l'IRT à Munich.

Résumé:

Aujourd'hui les producteurs de contenus audiovisuels sont en constante recherche de nouvelles expériences utilisateurs pour répondre à la demande. Dans l'optique de diffuser un son toujours plus impressionnant, les producteurs se sont tournés très tôt vers des standards dit plus immersifs. Il en résulte une multiplication considérable des livrables (Stéréo, 5.1, Binaural, AC3 ...) pour une même production. Or, tous ces standards d'écoutes sont devenus difficiles à gérer entre eux. L'absence de compatibilité force les chefs opérateurs du son à différencier les workflows et leurs mixages sont détériorés par l'absence d'un dispositif de prise de son adapté et par des downmixs qui altèrent l'immersion et la précision dans l'espace. Le développement du son « orienté objet » semble apporter des solutions concrètes concernant la compatibilité entre les différents formats et l'interopérabilité d'une scène sonore. En effet, le concept d'objet sonore, largement démocratisé ces dernières années incarne son innovation dans l'invariabilité de la scène sonore. Grâce aux métadonnées intégrées à l'objet sonore, les mécanismes de restitution ou moteurs de rendus peuvent reproduire, recréer le mixage en s'adaptant aux conditions de restitution. Son fonctionnement implique cependant un bouleversement dans la manière de produire révolutionnant ainsi la chaîne de travail, des dispositifs de prise de son aux systèmes de restitution. L'enjeu d'une architecture de référence pour la production orientée objet contribuerait à évangéliser les professionnels du son quant aux potentialités créatrices offertes par cette nouvelle approche mais aussi à les initier aux nouveaux outils ainsi qu'aux nouvelles manières de produire qui en découlent. Outre certains défis techniques et la mise en concurrence des sociétés privées pour un standard unique, le véritable enjeu de l'implémentation d'un modèle objet repose sur la mise en place d'une économie et d'une culture de production en mode objet.

MOTS-CLES:

Audio Orienté Objet, NGA, Multicanal, Multiplication, Standards d'écoute, Systèmes de restitution, Chaîne de production objet, Compatibilité, Standardisation

Table des Matières

Introduction	1
I/ Le pari Schumpétérien de l'approche du son « orientée objet ».....	5
1) L'audio orienté objet une réalité actuelle	5
a) Une définition du concept déjà partagée	6
b) La proposition de nouvelles expériences utilisateurs	13
2) Constance de la scène sonore indépendamment de son système de restitution	17
a) L'origami ou la conception d'une scène objet.....	17
b) Une nouvelle esthétique du son spatialisé	22
II/ Un concept disruptif : le renouvellement du Workflow	24
1) Une révolution pour les Broadcasteurs	24
a) Une tendance à la production IP	24
b) L'architecture de référence de l'ORPHEUS	25
2) De la captation à la restitution d'une scène 3D.....	31
a) Créer une scène sonore orientée objet	31
b) Des perspectives d'applications.....	34
III/ Les défis à surmonter pour affirmer le modèle orienté objet	37
1) L'œuf ou la poule, des équilibres à dynamiser.....	37
a) Des défis techniques à surmonter	37
b) Un standard pour gouverner tous les formats	41
2) L'évolution des métiers de l'audiovisuel pour assumer les nouvelles méthodes de production.....	43
a) La mise en place d'une économie et d'une culture de production en mode objet.....	44
b) De nouvelles méthodes de travail	47
Conclusion.....	50
Annexes	51
Bibliographie	55
Netographie	56
Filmographie	57

INTRODUCTION

Par sa malléabilité, son universalité et ses qualités esthétiques la scène audio « orientée objet » permettrait de satisfaire nos besoins en termes de diffusion dans les décennies à venir.

Aujourd'hui les producteurs de contenus audiovisuels sont en constante recherche de nouvelles expériences utilisateurs pour répondre à la demande. Ces dernières années, les recherches se sont focalisées sur les qualités de personnalisation, d'accessibilité, d'interactivité, et d'immersion d'un contenu audio. Dans l'optique de diffuser un son toujours plus impressionnant, les producteurs se sont tournés très tôt vers des standards dit plus immersifs comme le 5.1, puis plus récemment vers le binaural, disponible depuis un smartphone ou un site internet. Cependant dans l'industrie audiovisuelle actuelle, la production audio est pensée suivant une approche orientée canal selon laquelle un canal audio viendrait alimenter un ou plusieurs haut-parleurs du système de diffusion. C'est pourquoi, cette production sera conçue pour un système de diffusion particulier et une fois rendue, elle s'avèrera ainsi dépendante de ce système. Pour un rendu en stéréophonie, deux canaux (gauche-droite) seront nécessaires, six canaux pour un rendu 5.1, deux canaux filtrés pour un rendu binaural et ainsi de suite. Il résulte alors une multiplication considérable des livrables (Stéréo, 5.1, Binaural, ...) pour une même production grâce auxquels les producteurs parviennent à rendre leurs contenus à la fois plus immersifs et interactifs tout en assurant leur accessibilité via divers supports notamment les applications pour smartphones et les sites Internet qui représentaient plus de 12% du volume total d'écoute de la radio en Mars 2017. Or, tous ces standards d'écoutes sont devenus difficiles à gérer. L'absence de compatibilité force les chefs-opérateurs du son à différencier les workflows. Il convient d'adapter la prise de son de manière à ce qu'elle reste commune à tous ces standards de diffusion. De plus, dans la mesure où l'on se focalise sur un seul de ces standards, le mixage final se trouvera altéré par le downmix et perdra de sa cohérence lors d'une restitution via un autre standard.

Réussir à obtenir un mixage qui reste cohérent sur tous les standards représente donc un enjeu de taille. Pour répondre à la problématique de la multiplicité des livrables, l'approche de l'audio orienté objet semble être une solution efficace. En effet, contrairement à l'approche orientée canal précédemment évoquée, l'approche orientée objet libère de cette notion de canal et va reposer sur la gestion d'objets sonores. Un objet peut être considéré comme une entité sonore accompagnée d'informations. Ces métadonnées descriptives informent sur le type du son (voix, ambiance, musique...), mais aussi de sa position dans un espace sonore 3D et de la

manière dont ces objets doivent être rendus lors de l'écoute finale en fonction du standard de diffusion. Cette approche permet ainsi l'adaptation du contenu audio à différents dispositifs de restitution et lieux de diffusion mais peut également enrichir considérablement l'expérience de l'utilisateur en lui offrant notamment différents degrés d'interactivité.

Depuis quelques années de nouvelles exigences émergent chez les auditeurs et en particulier leur désir de contenus plus immersifs. Pour répondre à ces demandes, les producteurs de contenus ont cherché à s'adapter en fournissant sur des plateformes Web des contenus intégrant du son dit 3D. On peut citer en France les lives *Sequences* produit par Milgram productions et disponibles sur la plateforme Culturebox, l'*Hyperradio* de Radio France et leur label *nouv0son* ou encore les reportages en binaural du RFI Labo. De plus, le multicanal s'est démocratisé dans de nouveaux domaines d'activités notamment à la télévision où les premiers mixages en 5.1 ont fait leur apparition conjointement à la vidéo HD lors de la coupe du monde de Football en Allemagne en 2006. L'arrivée de l'UHD a aussitôt initié une réaction chez les diffuseurs qui ont cherché à répercuter l'immersion sur le son à l'aide du son 3D. Des retransmissions de grandes compétitions sportives telles que la ligue des champions, par exemple, sont désormais mixées en Dolby *Atmos* à partir des demi-finales.

Lorsqu'un auditeur souhaite écouter un programme sur un standard particulier, par exemple sur ses enceintes de télévision en stéréo il achète un certain contenu. Maintenant s'il dispose d'un système Home cinéma 5.1, il sera obligé de racheter ce même contenu dans un format différent. De même pour le producteur de ce programme qui devra prendre le temps de produire ce nouveau contenu et qui sera naturellement confronté aux impératifs économiques de la production audiovisuelle. Afin de rentabiliser les coûts relatifs à une production multicanale, ce type de production doit être rendu sous différents formats afin de les rendre accessibles via une grande variété de supports (TV, Home-Cinéma, Application Smartphone, etc.). Pour une captation en direct, cela implique une séparation des Workflows pour différents livrables.

Pour certains événements, il peut être demandé à un ingénieur du son de livrer plusieurs versions (stéréo, Dolby E, binaural, Dolby AC3 ...). Pour ce type de captations, les chefs-opérateurs du son sont obligés de se poser la question de la transportabilité de leur mix sur les différents standards. Il est en effet compliqué de gérer tous ces standards en même temps et notamment à la prise de son où un système de prise de son natif, propre à un standard de diffusion, aura toujours une meilleure réponse. D'autant plus qu'il est compliqué de combiner

les systèmes de prises de sons natifs sur une même captation. Imaginez une tête binaurale, un couple stéréo, et un arbre 4.0 à placer au sweet spot au sein d'une même configuration... Des solutions techniques basées sur le fameux précepte du « qui peut le plus, peut le moins » ont été trouvées et depuis plusieurs années déjà, les métadonnées Dolby, MPEG ou DVB servent à réaliser le downmix stéréo d'un flux 5.1 reçu sur une TV. Cependant, on peut légitimement se poser la question de la qualité de ces downmix et de l'influence de la conversion sur la précision de la spatialisation. On se dirige donc vraisemblablement vers un format universel qui serait compatible avec tous les standards de diffusion tout en conservant le même contenu.

La notion de « Mix Objet », peut-elle devenir la prochaine étape pour un mixage commun à tous les standards de diffusion et ainsi parvenir à faire cohabiter ces différents formats ?

Le développement du son « orienté objet » semble apporter des solutions concrètes concernant la compatibilité entre les différents formats et l'interopérabilité d'une scène sonore. En effet, le concept d'objet sonore, a été largement démocratisé ces dernières années par quelques grandes sociétés de l'audiovisuel qui ont fait le pari de l'innovation. Celle-ci s'incarne notamment dans l'invariabilité de la scène sonore objet indépendamment de son système de restitution. Grâce aux métadonnées intégrées à l'objet sonore, les mécanismes de restitution ou moteurs de rendu peuvent reproduire, le mixage en fonction des indications fournies par le mixeur et cela en s'adaptant aux conditions de restitution ainsi qu'au contexte d'écoute.

Son fonctionnement implique cependant un bouleversement dans la manière de produire révolutionnant ainsi la chaîne de travail : des dispositifs de prise de son aux systèmes de restitution. Des professionnels et chercheurs se rassemblent donc dans des consortiums afin de vérifier la viabilité des nouveaux standards objets au sein du contexte de production actuel. L'enjeu d'une architecture de référence pour la production orientée objet serait de considérer les possibilités offertes par la création d'une scène sonore en trois dimensions intégrant des objets sonores. Par ce biais, elle contribuerait à évangéliser les professionnels du son quant aux potentialités créatrices de l'approche orientée objet et de les initier aux nouveaux outils ainsi qu'aux nouvelles manières de produire qui en découlent.

Les défis à surmonter pour affirmer le modèle orienté objet relèvent de plusieurs enjeux : techniques, économiques, sociologiques. Pour passer du statut d'invention à celui d'innovation, l'audio orienté objet doit pouvoir s'affirmer sur le marché. A cet égard, des sociétés privées comme Dolby, DTS ou Fraunhofer IIS cherchent à se démarquer auprès des

instances de normalisation, notamment Européenne, en vantant leur solution dans l'espoir qu'elle triomphe de ses concurrentes et devienne ainsi le standard dominant. Outre certains défis techniques, le véritable enjeu repose sur la mise en place d'une économie et d'une culture de production en mode objet. D'une part, il est encore difficile de présager si le saut qualitatif de l'audio orienté objet sera suffisant pour convaincre les utilisateurs et de la sorte inciter les diffuseurs à investir dans le domaine. D'autre part, le statu quo auquel se confronte actuellement l'industrie audiovisuelle s'en remet à des facteurs socio-économiques liés au processus de normalisation et aux incidences inhérentes aux nouvelles méthodes de production.

I/ LE PARI SCHUMPETERIEN DE L'APPROCHE DU SON « ORIENTEE OBJET »

Selon Joseph Schumpeter, économiste et professeur de sciences politiques du début du XXème siècle, l'innovation se caractérise par un passage, une rupture. Pour qu'il y ait innovation il faut que le changement provoqué soit d'importance, qu'il modifie les conditions de la production et du marché. Dans son rapport de Novembre 2018, l'EBU (European Broadcasting Union) définit trois approches fondamentales pour la mise en place des systèmes NGA (Next Generation Audio) à l'origine de contenus audio innovants offrant un degré d'immersion et d'interaction à ses auditeurs.

- L'approche orientée canal (Channel-based) dans laquelle un canal vient alimenter un ou plusieurs haut-parleurs.
- L'approche orientée scène (Scene-based) qui consiste à capter et rendre un champ sonore 3D à l'aide d'un maillage régulier de l'espace au sein duquel les données spatiales du son sont encodées selon les axes x, y et z. Ce maillage appelé HOA pour High Order Ambisonic peut être plus ou moins précis et se définit par son ordre ambisonique, qui représente le nombre d'harmoniques sphériques utilisé pour la création du champ sonore. Plus cet ordre est élevé, plus la résolution spatiale, autrement dit la précision en terme de localisation angulaire des sources augmente.
- L'approche orientée objet (Object-based) dans laquelle la scène sonore est composée d'éléments individuels dont les caractéristiques (Position, directivité, ...) sont définies en temps réel par des métadonnées.

Ainsi, l'audio orienté objet est une réalité actuelle et représente une innovation qui s'incarne dans la constance de la scène sonore indépendamment de son système de restitution.

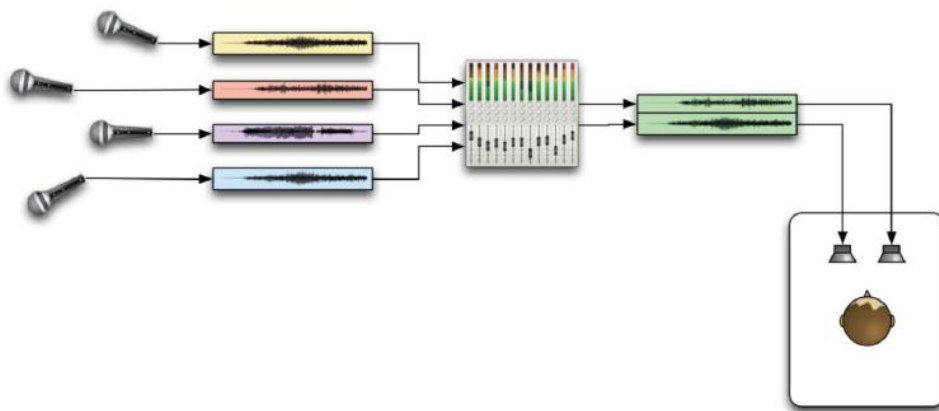
1) L'AUDIO ORIENTE OBJET UNE REALITE ACTUELLE

La conception de l'audio orienté objet existe et sa définition se trouve déjà partagée pour permettre de délivrer des expériences plus personnalisées à un auditeur.

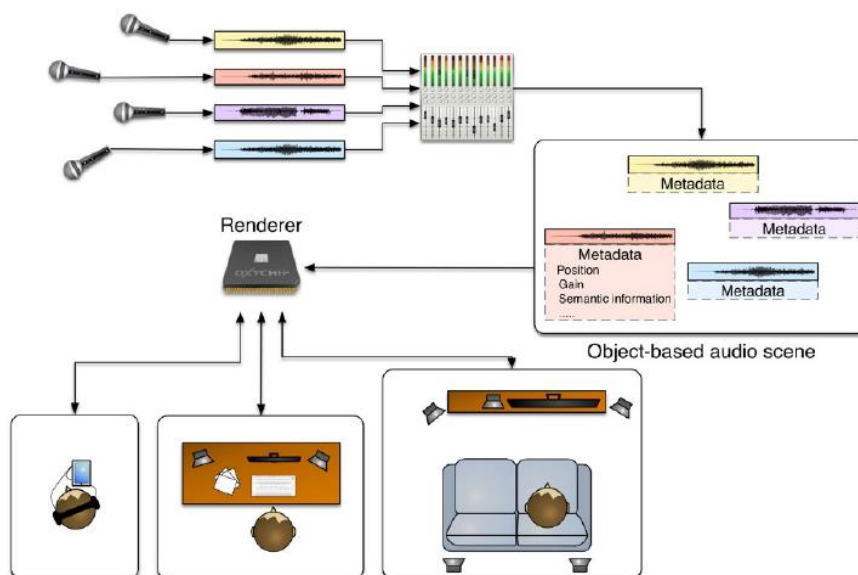
a) UNE DEFINITION DU CONCEPT DEJA PARTAGEE

Le concept d'objet sonore

Le concept d'objet sonore, déjà ancré dans le paysage audiovisuel international, repose sur des entités sonores accompagnées d'informations (métadonnées). Ces métadonnées informent un moteur de rendu, ou *Renderer*, de la manière dont ces objets doivent être restitués lors de l'écoute finale en fonction du lieu et du système de restitution. Par la séparation d'une source sonore de ses données de spatialisation l'approche orientée objet se démarque nettement de l'approche orientée canal et offre, outre la spatialité de nouvelles expériences à l'utilisateur notamment un certain degré d'interactivité et de personnalisation.



Production et Réception en approche orientée canal traditionnelle - Source : Results of ORPHEUS_Andreas SILZLE_19/06/2018



Production et Réception d'une scène objet - Source: Results of ORPHEUS_Andreas SILZLE_19/06/2018

L'approche objet se démarque dans sa capacité à capter des sons qui prennent leurs fonctions dans le mix à l'aune d'un contexte particulier. Prenons l'exemple d'une bande son de film, nous disposons des pistes sonores correspondants aux dialogues, des pistes correspondants aux musiques, des pistes correspondants aux effets ... Toutes ces pistes sont transmises séparément, en tant qu'objets, avec des métadonnées produites par le mixeur. Grâce à ces métadonnées, les mécanismes de restitution (moteurs de rendus) peuvent d'une part reproduire le mixage en fonction des indications fournies par le mixeur mais aussi permettre au spectateur d'interagir avec ces objets et sur leurs caractéristiques. Le moteur de rendu est l'outil qui va réaliser ces scènes pour un système de diffusion particulier. Son rôle est de faire le lien entre un mode (objets ou multicanal) indépendant du système de diffusion et le système de diffusion en lui-même.

La naissance de l'ADM

Les premières expérimentations sur de l'audio orienté objet ont été réalisées il y a plus de trente ans et les outils étaient très semblables à ceux utilisés aujourd'hui. Cependant l'engouement pour cette approche n'a été que très faible au départ et les industriels, préoccupés par d'autres enjeux économiques, n'étaient vraisemblablement pas prêts à investir dans le domaine. Pourtant, ces dernières années, les activités de Dolby autour de l'objet et le succès du Dolby *Atmos* au cinéma ont suscité l'intérêt des acteurs de l'audiovisuel, qui y ont perçu un intérêt applicable à leur champ d'activité. Par la suite, L'IRCAM (Institut de Recherche et Coordination Acoustique/Musique) s'inspirera de son plugin *Spatialisateur* ou *Spat*, créé au début des années 90 qui incluait déjà une technologie objet en interne pour spatialiser des sources sonores afin de développer un format Open Source nommé ADM pour Audio Definition Model. L'objectif était de créer un standard pour les métadonnées avec lesquelles il serait possible d'interagir à l'opposé du plugin *Spat*, qui recourt à ses propres métadonnées non standardisées. Dès 2015, des ingénieurs du son de Radio France ont commencé à étudier des scènes objet complexes avec l'automation des coordonnées de position d'objets sonores via le plug-in *ToscA* développé à l'IRCAM et basé sur de l'OSC (Open Sound Control). Ce plug-in est destiné à envoyer/recevoir des données d'automation. Deux ans plus tard, les premiers enregistrements en ADM étaient effectués.

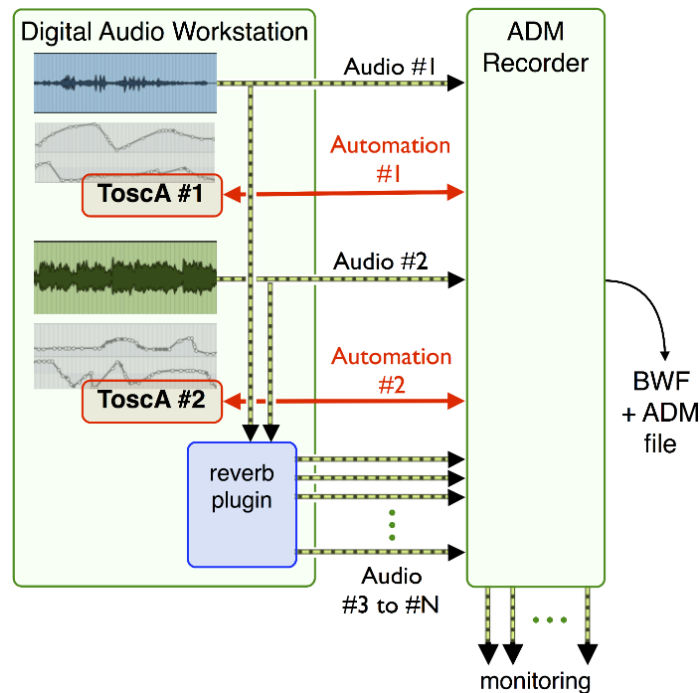
Audio Definition Model

L'ADM est un format basé sur de l'XML qui permet de générer et d'enregistrer des métadonnées descriptives des entités audio. Ces métadonnées se divisent en deux catégories

distinctes. La partie format qui définit les caractéristiques techniques de l’audio de manière à ce qu’il soit restitué correctement et la partie contenu qui ne peut être complétée qu’après la génération du signal audio et qui décrit le contenu d’un objet comme la langue d’une voix par exemple. Chacune de ces parties est composée de divers éléments que nous aborderons brièvement en ANNEXE 1 de ce mémoire en suivant l’ordre de la Recommandation ITU-R BS.2076-1 qui présente le premier modèle global de l’ADM en Juin 2017.

De nouveaux outils : les ADMix Tools

Pour tester et valider le format ADM, il a fallu développer des outils afin de mettre la théorie en pratique. Ainsi, Thibaut Carpentier, chercheur au CNRS et à l’IRCAM a développé les *ADMix Tools* pour expérimenter l’enregistrement, la lecture et la restitution de fichiers ADM.

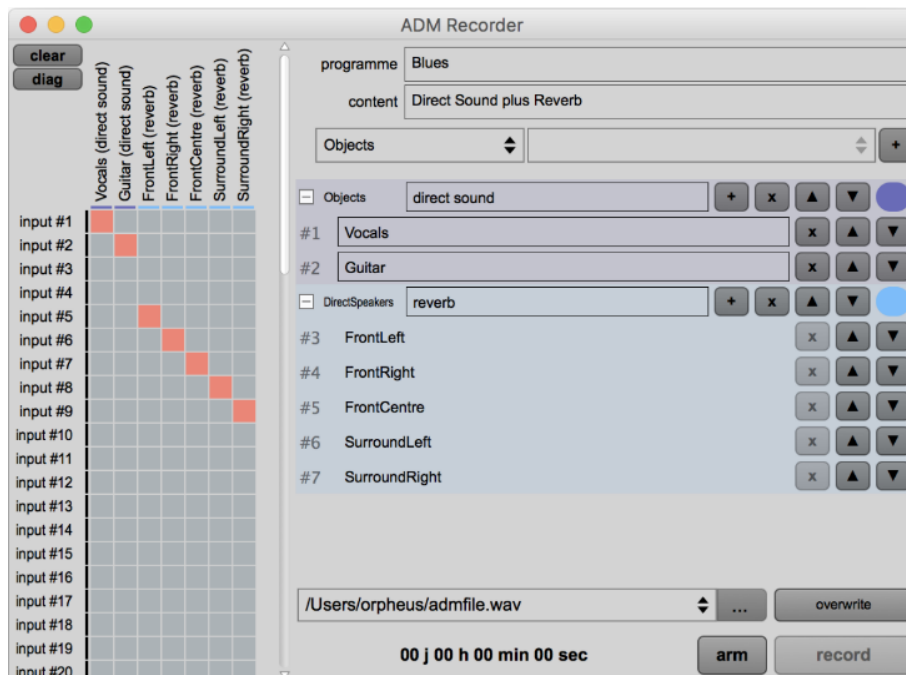


Synoptique d’enregistrement d’un fichier ADM – Source :

<https://forumnet.ircam.fr/product/spat-en/admix-en/>

L’ADM Recorder est le premier outil de la chaîne. Il permet de produire un fichier ADM compatible avec le standard à partir de flux audio et des métadonnées correspondantes. Dans cette configuration, les automatisations sont transmises à l’ADM Recorder à l’aide du plug-in *Tosca* en insert direct sur la piste de la station de travail (DAW). L’ADM Recorder interprète ces automatisations et enregistre les métadonnées dans un fichier ADM.

L'image ci-dessous présente l'éditeur de métadonnées et la matrice de Routing d'un programme 5.1 + 2 objets (*Vocals* et *Guitar*) de l'*ADM Recorder*.

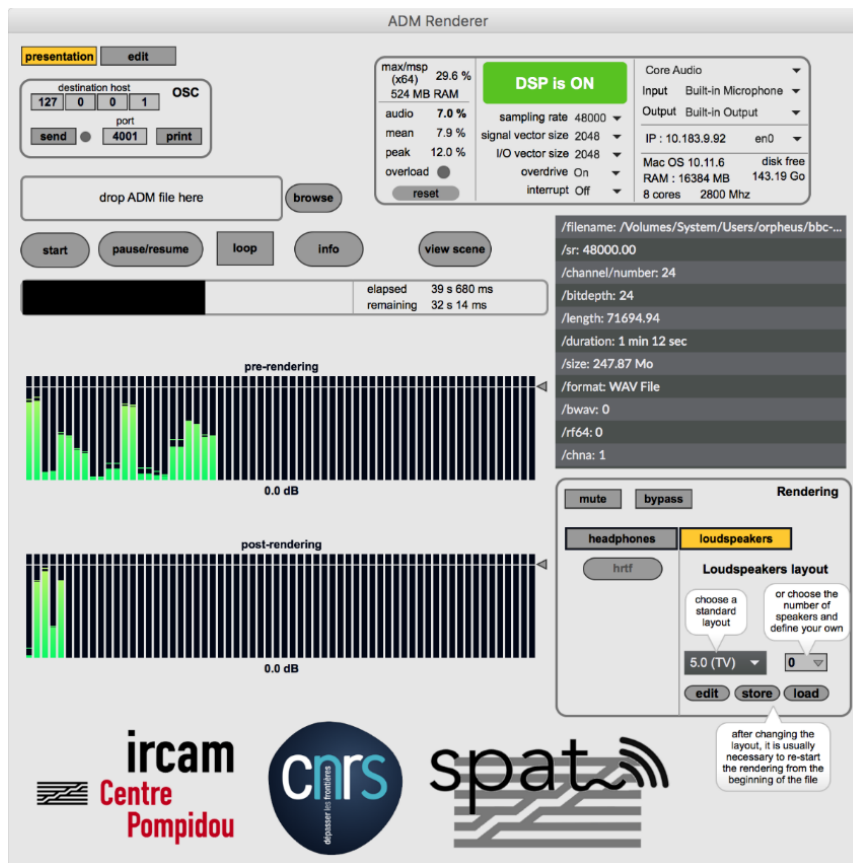


Editeur de métadonnées et Matrice de Routing de l'ADM Recorder - Source :

<https://forumnet.ircam.fr/product/spat-en/admix-en/>

L'*ADM Player* correspond à un simple lecteur du fichier ADM. Il lit simplement les flux audio et les métadonnées ADM. Il est utile pour utiliser un moteur de rendu quelconque qui interprètera les métadonnées et constituait une étape nécessaire en vue du développement du *Renderer*.

L'*ADM Renderer* permet de rendre le fichier ADM, c'est-à-dire lire les flux audio et interpréter les métadonnées en fonction du setup de restitution et permettre éventuellement une interactivité avec l'utilisateur.



Interface utilisateur de l'ADM Renderer – Source : <https://forumnet.ircam.fr/product/spat-en/admix-en/>

Un paradigme objet

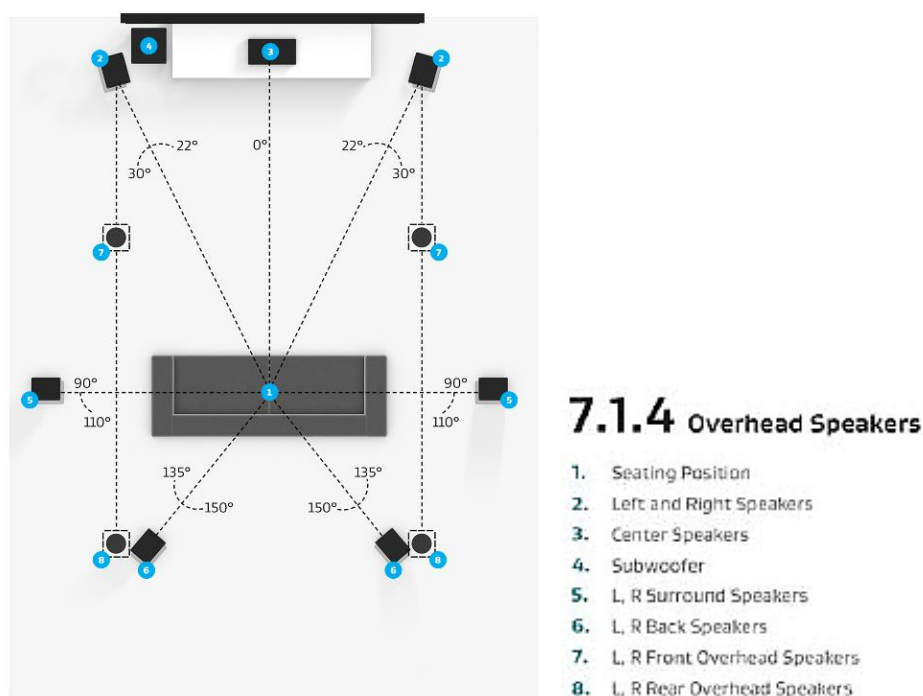
Les débuts de la production en mode objet se sont concentré sur un mode « pur objet » qui tend aujourd’hui à évoluer avec l’intégration des « beds » inspirée de la technologie du Dolby *Atmos* déjà largement généralisée dans la diffusion du son 3D au cinéma.



Principe de base du Dolby ATMOS – Source : [Dolby-Atmos-Next-Generation-Audio-for-Cinema-White-Paper](#)

« Il y a quelques années on travaillait en mode objet en concevant plutôt ça sur un mode « pur objet ». J'ai commencé à travailler avec des scènes où il y avait 24 objets dans les 24 directions j'allais utiliser » m'a révélé Jean-Christophe Messonnier : Ingénieur du son, Mixeur et Enseignant-Chercheur au Conservatoire National Supérieur de Musique et de Danse de Paris.

Aujourd'hui la production en NGA (Next Generation Audio) se base sur un mode objet seulement partiel dans le sens où il est couplé avec une grosse base multicanale ou Ambisonique. C'est d'ailleurs sur ce couplage que se sont créés les codecs Dolby AC-4 *Atmos* et MPEG-H développé par Fraunhofer IIS, Qualcomm et Technicolor. Afin de trouver le meilleur compromis entre le nombre de canaux et la définition de l'espace, une configuration optimale s'est finalement fondée sur un 7.1.4 + 4 objets pour un total de 16 canaux. Cette configuration s'inspire du 7.1 préexistant avec trois bases LCR à 60° dans le plan, un LFE, 2 bases *Sides* à 90°/110°, 2 bases *Rear* à 135°/150° et 4 bases en élévation à 45°.



Configuration de Haut-Parleurs en 7.1.4 – Source :

<https://westcoastavgallery.ca/dolby-atmos-home-theater-speaker-setup-guide/>

Ainsi, il est possible de rendre une ambiance immersive à partir d'un enregistrement multicanal ou HOA et de le coupler avec des objets, qui pour l'instant se cantonnent à des voix commentées. Pour ce qui concerne les 16 canaux du MPEG-H, il est possible d'intégrer 16 objets, 12 canaux et 4 objets ou un HOA d'ordre 3 et 4 objets.

Pour comprendre le rôle de l'ADM dans le paradigme objet actuel, on peut établir une comparaison culinaire. La recette d'un gâteau se compose d'une liste d'ingrédients et d'instructions sur la façon dont ils doivent être mélangés. L'ADM est similaire à cette liste d'instructions qui décrit précisément chaque ingrédient et sur la manière dont ces ingrédients doivent être combinés. Ainsi, le standard ADM définit le format, c'est-à-dire la manière dont les données sont stockées dans le fichier. En revanche il ne prescrit pas la manière dont les données doivent être interprétées et restituées. En effet, l'ADM est peu viable pour la diffusion car c'est un format non compressé. Un encodeur va alors se charger de convertir l'ADM dans un format compressé qui sera transmis puis restitué via le système de diffusion au moyen d'un moteur de rendu ou *Renderer* qui décodera l'information.

Le Renderer

A la suite des publications de l'EBU et de l'ITU qui en tant qu'instances de normalisation, reconnaissent l'ADM comme un standard viable pour la production en mode objet, ces instances ont volontairement laissé les constructeurs développer librement leurs *Renderers*. De ce fait, ils n'ont pas prescrit de façon de restituer l'ADM et ont incité à la mise en concurrence entre les constructeurs de moteur de rendus. Ainsi, même si les technologies utilisées (algorithmes de type VBAP) et les rendus finaux sont à peu près similaires entre les différents constructeurs, une part de leur succès sur le marché est attribuée aux capacités de ces sociétés à communiquer sur leurs produits. Il y a quelques mois l'EBU a finalement mis à disposition son propre *Renderer* l'EBU ADM *Renderer* (EAR)⁽¹⁾ qui sert de base et de modèle mais ne l'ont pas imposé et ont laissé les concurrents développer leurs propres moteurs de rendus. L'ambition était d'offrir un moteur de rendu Open Source qui compile les objets selon des algorithmes de référence. Au final, l'importance d'un format gratuit et agnostique s'avère cruciale pour pouvoir adresser tous les types d'encodeurs en gardant les mêmes métadonnées. Ces dernières années de nombreux constructeurs ont émergé sur le marché pour imposer leur Codec comme standard. Parmi eux, Dolby avec l'AC-4 *Atmos*, DTS avec le DTS-X ou le DTS MDA, et Fraunhofer

(1) EBU Tech3388_ADM *Renderer* for use in Next Generation Audio Broadcasting_BTF *Renderer* Group_03/2018

avec le MPEG-H sont les plus répandus. Chacun de ces codecs n'est donc pas standardisé par l'ITU, ce sont des formats propriétaires, brevetés et à utilisation commerciale.

Lorsqu'on évoque une conversion, il faut toujours considérer une perte d'information ou de précision. Afin d'optimiser cette conversion et par conséquent la restitution, les constructeurs ont défini des profils de compatibilité qui constituent en quelque sorte un dénominateur commun pour la conversion de l'ADM et le codec. L'objectif étant de définir certains presets d'usages des métadonnées soit des préconisations d'utilisations du fichier ADM pour que la conversion se fasse correctement. Par exemple, pour un rendu sur téléphone les capacités seront réduites et le producteur sera forcé de limiter le nombre d'objets dans une scène pour assurer un rendu le plus fidèle possible. Enfin, cette conversion n'est pas forcément faite en temps réel et c'est une des raisons pour laquelle l'*ADM Renderer* peut être utilisé par les mixeurs afin de monitorer directement au niveau de l'ADM non compressé.

Un essor international

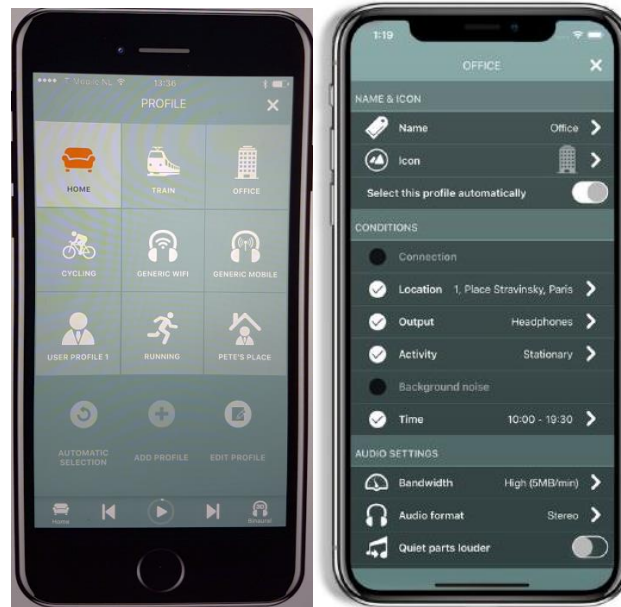
Ces dernières années, l'audio orienté objet a connu un essor fulgurant notamment grâce au cinéma avec le mixage en Dolby *Atmos* qui s'est ensuite étendu à d'autres domaines de l'audiovisuel et du multimédia. En effet, le codec Dolby AC-4 *Atmos* s'est imposé pour les nouveaux services de télévision numérique ATSC 3.0 aux Etats-Unis mais également dans les décodeurs Apple TV 4K ainsi que dans des Smartphones Haut de gammes comme le Samsung Galaxy S10. Fraunhofer IIS ont conquis l'Asie orientale avec le MPEG-H 3D et plus particulièrement la Corée du Sud avec le démarrage de leurs chaînes en Ultra HD. En effet, le MPEG -H, a été la première solution développée pour un téléviseur à l'occasion des Jeux Olympiques d'Hiver de PyeongChang en 2017. On peut également citer DTS qui a percé dans l'industrie du divertissement avec le DTS:X et le DTS MDA.

Parmi les innovations offertes certaines constituent des opportunités pour séduire les utilisateurs.

b) LA PROPOSITION DE NOUVELLES EXPERIENCES UTILISATEURS

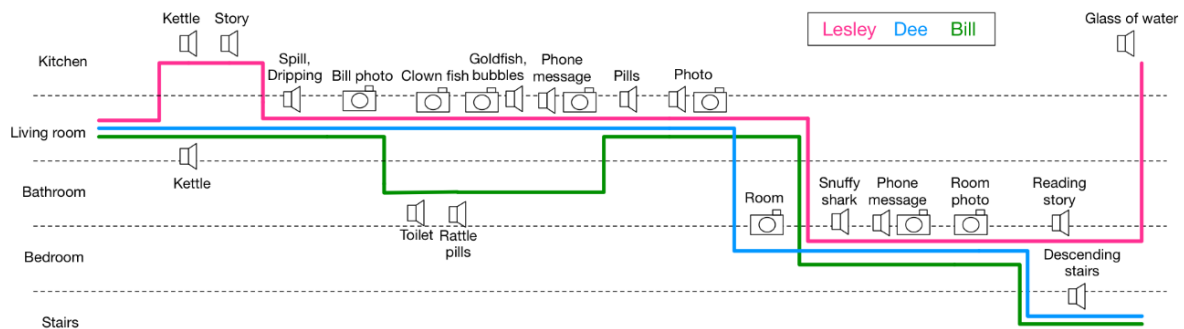
Parmi toutes ses qualités l'approche orientée objet séduit les diffuseurs par la possibilité qu'elle offre de fournir à l'auditeur de nouvelles expériences utilisateur fondées sur l'interaction avec le contenu audio. Tout d'abord, l'audio orienté objet permettrait de délivrer une expérience plus personnelle à un auditeur avec la capacité de choisir entre différentes langues ou sa capacité d'adaptation à différents lieux de restitutions. Ainsi, l'audio orienté objet propose d'adapter le rendu du mixage aux conditions d'écoute parfois délicates mais aussi de

l'optimiser en fonction de l'environnement qui devient de plus en plus diversifié et même parfois mobile (métro, avion, chambre à coucher...). Par exemple, il serait possible d'écouter un programme dans un environnement bruyant avec la possibilité de compresser le signal lors d'une écoute dans la rue et retrouver une pleine dynamique lors d'une écoute à la maison.



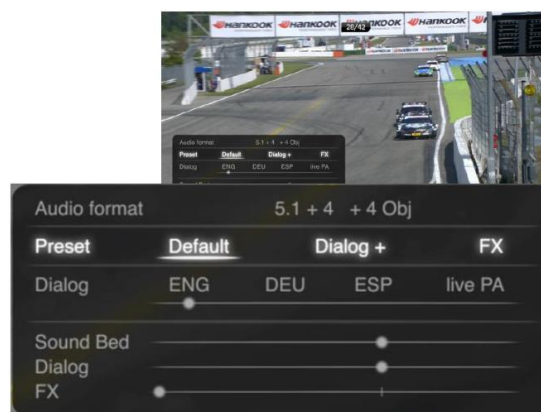
Exemples de presets d'écoutes sur l'application ORPHEUS – Source : Design and Implementation of a mobile object-based audio radio player_Niels Bogaards_elephantcandy_19/06/2018

De plus, l'audio orienté objet permettrait de fournir au public un certain niveau d'interaction à l'auditeur de par sa capacité à modifier à la volée un contenu audio. En fiction, cela engendrerait de nouveaux scénarii interactifs au sein de quels le spectateur pourrait choisir une scène plutôt qu'une autre et inciterait de cette manière les producteurs à être plus créatifs sur les expériences qu'ils peuvent rassembler et livrer au public.



Exemple de scénario Interactif: *The Mermaid Tears* fiction radiophonique de Jessica Mitic et Chris Baume - Source: *The Mermaid's Tears: Creating the world's first live interactive object-based radio drama*_Dr. Chris Baume_BBC_19/06/2018

On peut citer plusieurs autres exemples d'interactions. Premièrement, il serait possible d'ajuster la durée d'un programme en fonction du temps disponible. Prenons l'exemple d'un contexte radiophonique, une émission radio d'une durée d'une heure. Avec un contenu sous forme d'objets, le moteur de rendu peut fabriquer automatiquement un résumé de 15 min de l'émission à partir des métadonnées renseignées par le mixeur/producteur. Deuxièmement, il serait envisageable de fournir un contenu en adéquation avec les préférences d'un auditeur. L'auditeur aurait ainsi la possibilité de choisir la langue et de modifier la balance entre les commentateurs et l'ambiance du stade lors de la retransmission d'une captation sportive.



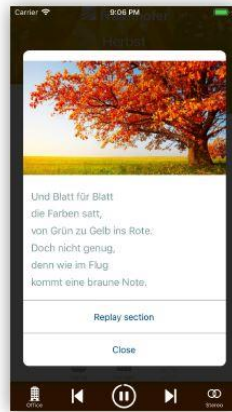
Menu des paramètres audio MPEG-H – Source: *MPEG-H Audio System for Broadcasting*_Jan Plogsties_Fraunhofer IIS_2015

Enfin, figure également le renforcement de l'intelligibilité, une fonctionnalité nécessaire pour les spectateurs dont l'ouïe commence à montrer certains signes de faiblesse. Il serait donc possible d'augmenter des rapports de niveaux entre dialogues et musiques ou ambiances au sein même d'un programme. Troisièmement, il permettrait de retrouver des émissions à partir de requêtes sur un moteur de recherche qui redirigent directement vers l'audio à l'aide de l'outil

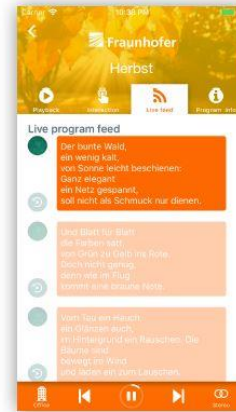
de transcription de texte. De manière réciproque, tous les auditeurs pourraient recevoir des métadonnées au fil d'une émission de radio avec la possibilité de suivre des liens relatifs au contenu de l'émission comme une navigation Wikipédia par exemple.



Time Scrubbing



Points Of Interest



Text Transcript



3D object
interaction



narrative
trajectory



variable-length
playback

Interfaces d'interactions proposées par l'application ORPHEUS – Source : Design and Implementation of a mobile object-based audio radio player_Niels Bogaards_elephantcandy_19/06/2018

Toutes ces nouvelles expériences utilisateurs pourraient bientôt être à disposition du grand public et accélérer la démocratisation de ces programmes en audio orienté objet sous réserve que le retour soit positif du côté des auditeurs. Cependant, la véritable innovation pourrait reposer sur l'universalité du mode objet. Une scène sonore en mode objet est en théorie, indépendante du système de restitution, c'est-à-dire qu'il serait possible de traiter la spatialité d'un objet sans se soucier de son format de restitution.

2) CONSTANCE DE LA SCENE SONORE INDEPENDAMMENT DE SON SYSTEME DE RESTITUTION

La nouveauté consiste à redéfinir le mode de conception d'une scène sonore. L'approche orientée objet nous permettrait donc de penser l'audio d'une manière malléable par rapport à la spatialité des sons utilisés. Celle-ci constituerait donc une véritable révolution au regard de l'approche canal traditionnelle et par conséquent une solution indéniable à la multiplicité des formats. Et cette évolution modifiera aussi les préceptes de l'esthétique du son.

a) L'ORIGAMI OU LA CONCEPTION D'UNE SCENE OBJET

L'universalité d'une scène objet peut être comparable à un origami. Une conception de papier en trois dimensions à partir d'une feuille en deux dimensions. Cette logique s'adapte à un contenu audio stéréo qui peut fournir une conception 3D et réciproquement en fonction du système de restitution sur lequel il sera diffusé. Par son universalité le mode objet semble être la solution la plus efficace et pérenne. Pour Jean-Christophe Messonnier, le fait de produire en mode objet, permettrait de satisfaire les besoins en termes de diffusion dans la décennie voire les décennies à venir. En fixant une définition maximum à une scène objet dite scène de base, il est ensuite possible de restituer des versions réduites qui seront perçues comme des downmix de cette version maximale. Cette scène de base contient toute l'information de l'espace sonore que l'on veut représenter. Elle est ainsi composée d'un certain nombre d'objets.

En théorie, l'audio orienté objet permettrait donc de créer des contenus immersifs destinés à de multiples formats et de ce fait à de multiples supports sans avoir à créer de versions individuelles pour chaque standard de diffusion et ainsi régler de manière élégante toutes les questions de downmix. Pour Jean-Christophe Messonnier, il demeure indispensable de conserver la possibilité de diffuser en stéréophonie qui reste le mode dominant car elle est à fois efficace et économe. L'objectif est alors de pouvoir faire un enregistrement qui puisse remplacer la stéréo de manière vraiment satisfaisante sur l'intégralité des formats multicanaux qu'on utilise. Cette adaptation s'applique particulièrement bien au cinéma où la stéréo n'est pas viable au car elle repose sur un « sweet spot ». Pour restituer un effet dans une salle de cinéma, généralement très absorbante, il faut donc une base centrale et des ambiances plutôt englobantes. Cependant, lorsqu'un nouveau système de diffusion est mis sur le marché, il se doit d'être compatible avec les salles qui ne sont pas encore équipées. De plus, pour amener les productions à effectuer leur film dans ce nouveau format, il faut généralement que le temps, et par conséquent, le coût de post-production reste le même. C'est la raison pour laquelle, le

« downmix automatique » à partir d'une scène objet constitue une solution adaptée à la rétrocompatibilité des mixages non seulement au cinéma mais également pour les systèmes de diffusion Home-cinéma. Cette conversion s'appliquerait ainsi à tous les standards de diffusion et assurerait une compatibilité par réduction d'informations. Parfois, on peut même considérer un changement de nature de l'information par rapport à la scène objet originale dans le cas où on lui applique des filtres binauraux qui vont permettre de l'optimiser pour le casque ou des filtres transauraux pour des enceintes.

Pour réaliser des mouvements d'objets les plus transparents possible dans l'espace sonore 3D et rendre une scène sonore cohérente, les moteurs de rendu se basent sur une technologie développée en collaboration par Dolby, Fraunhofer et Philips R&D et standardisée en 2008 : Le SAOC (Spatial Audio Object Coding)⁽¹⁾.

Le SAOC est un système de spatialisation propriétaire qui permet de rendre un mixage sur une configuration variable de haut-parleurs à partir d'un flux audio orienté objet. Pour y parvenir, il utilise des algorithmes d'encodage et de décodage des flux audio fondés sur une matrice orientant le nombre d'objets vers le nombre de canaux rendus disponibles par le système de restitution. Afin d'acheminer au mieux le signal audio au(x) haut-parleur(s), une matrice de rendu est établie à partir du nombre de haut-parleurs, de leur positionnement dans la salle, de leur positionnement relatif et des métadonnées de position des objets audio. Cette matrice détermine ainsi un ensemble de lois qui vont définir le rendu d'un signal audio vers une ou plusieurs sorties audio tout en s'adaptant au système de diffusion employé. Ces lois sont créées individuellement pour chaque entrée audio et sont ajustables en fonction du système de restitution de l'auditeur. Par exemple, le SAOC intègre des solutions visant à privilégier l'assignation d'un signal à une enceinte plutôt que la création de sources virtuelles via plusieurs enceintes de manière à privilégier le timbre du son diffusé. Outre ce caractère passif, certains processeurs comme celui du MPEG-H intègrent des algorithmes actifs basés sur une analyse du signal audio entrant afin d'éliminer d'éventuels problèmes liés à des signaux corrélés en phase au sein d'un même canal. Ce downmixer actif se révèle d'une importance capitale pour obtenir des versions réduites cohérentes. En effet, au sein d'un système de son 3D, un même signal est souvent copié, redistribué, filtré et traité par différents délais et réverbérations et ceci peut être à l'origine de nombreux artefacts.

(1) ENGDEGARD Jonas, RESCH Barbara, FALCH Cornelia, HELLMUTH Oliver, HILPERT Johannes, HOELZER Andreas, TARENTIEV Leonid, BREEBAART Jeroen, KOPPENS Jeroen, SCHUIJERS Erik, OOMEN Werner, *Spatial Audio Object Coding (SAOC) – Upcoming MPEG Standard on Parametric Object Based Audio Coding*

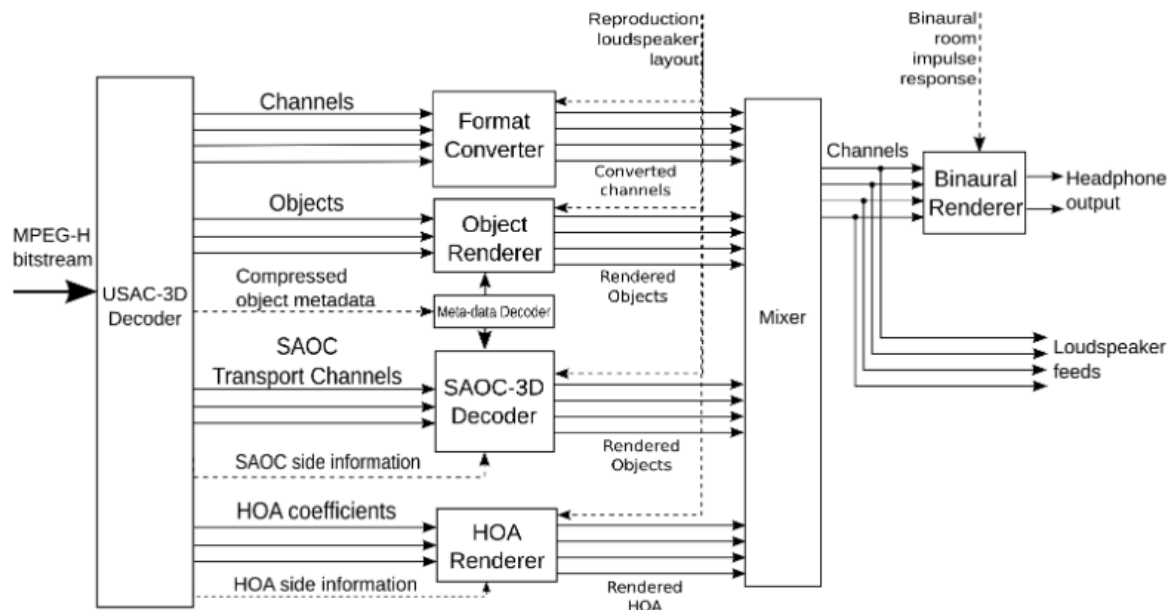


Schéma du décodeur du MPEG-H – Source : AES137_MPEG-H Audio – The New Standard for Universal Spatial / 3D Audio Coding_ Jürgen Herre, Johannes Hilpert, Achim Kuntz, Jan Plogsties_Fraunhofer IIS

Toutes ces technologies d’encodage et de décodage ont été exploitées dans le but d’améliorer la rétrocompatibilité des mixages et de la même manière l’esthétique des downmixs. L’objectif final serait d’acquérir un maximum de précision dans la provenance du son par rapport au point d’écoute. L’information décrivant l’espace sonore est contenue dans l’information de position des objets et dans l’information portée par les objets. Il devient alors possible de définir une source sonore quelconque en temps réel en tant qu’objet selon trois caractéristiques principales :

- La position de la source sonore : ses coordonnées spatiales (azimut, distance, élévation)
- Les caractéristiques de la source sonore : sa directivité, sa taille dans l’espace sonore, son étalement.
- Les paramètres de la salle de diffusion : la taille de la salle, sa réverbération éventuelle...

Ces caractéristiques sont définies par 15 sous-éléments dans les *audioBlockFormats* de l’ADM⁽¹⁾. Tout d’abord, les coordonnées de position dans l’espace 3D à savoir l’azimut en degrés (entre -180 et 180, l’élévation en degrés (entre -90 et 90) et la distance par rapport à l’origine en valeur absolue. Ensuite viennent les informations relatives aux dimensions de

(1) *Audio Definition Model* Recommendation ITU-R BS.2076-1 06/2017

l'objet soit l'ampleur horizontale et l'ampleur verticale en degrés ainsi que la profondeur en rapport (de 0 à 1) à la distance. Il est possible de définir ces valeurs dans un domaine cartésien en le spécifiant à l'aide du sous-élément *cartesian*. De plus, on peut appliquer une valeur de gain linéaire à l'objet et définir la balance entre son direct et son diffus (entre 0.0 et 1.0). Le sous-élément *channelLock* permet de cantonner l'objet à un seul canal ou haut-parleur de diffusion afin de privilégier le timbre de l'objet et d'éviter les sources virtuelles. Son attribut *maxDistance* définit une distance autour de l'objet à partir de laquelle un haut-parleur peut s'attribuer le signal de cet objet. Par défaut, cette distance est infinie. Ensuite vient le sous-élément *objectDivergence* à l'origine de la directivité de la source sonore. Sa valeur peut varier entre 0 (aucune divergence) et 1 (divergence maximale). Ses attributs *azimuthRange* et *positionRange* correspondent respectivement à la valeur (de 0 à 180°) et à la position de (0 à 1) de l'angle de divergence. Le sous-élément *jumpPosition* évoqué précédemment permet d'interpoler les valeurs de position entre les différents *audioBlockFormats*. Si *jumpPosition* est réglé sur 1 la position changera instantanément à partir du bloc précédent. S'il est réglé sur 0, alors l'interpolation de la position durera pendant toute la longueur du bloc. Son attribut *interpolationLength* permet de définir la durée (en secondes) de l'interpolation entre les deux positions instanciées par deux blocs successifs. La durée de l'interpolation doit toujours être inférieure ou égale à la durée minimale d'un bloc. Ensuite, s'applique des sous éléments relatifs à la zone de restitution d'un objet. En effet, à l'aide du sous-élément *zoneExclusion* il est possible de définir une zone de haut-parleurs dans la salle dans laquelle un objet ne doit pas être restitué. Cette zone est caractérisée par des attributs cartésiens (x,y,z) mais aussi sphériques (Élévation et Azimut). Enfin, viennent des informations concernant la relation éventuelle d'un objet avec une image avec l'élément *screenRef* et l'importance de cet objet dans la scène sonore établie sur une échelle de 0 à 10. Cette échelle d'importance appliquée aux objets est particulièrement pratique pour établir des règles de priorité entre les objets lors de la restitution d'une scène. Toutes ces métadonnées sont détaillées plus précisément dans la Recommandation *ITU-R BS.2076-1*⁽¹⁾. De la sorte, chacun de ces éléments et sous-éléments peut être modifié en temps réel et offre une infinité de possibilités dans la création d'une scène sonore orientée objet universelle sur tous les standards de diffusion.

Un saut qualitatif dans l'esthétique du son spatialisé ?

Cependant, on peut se demander si la réduction d'une scène objet pour des standards plus traditionnels comme le 5.1 ou la stéréophonie est réellement d'une qualité esthétique supérieure

(1) *Audio Definition Model_Recommandation ITU-R BS.2076-1, 06/2017*

à celle d'un downmix classique. Pour cela, plusieurs campagnes de tests ont été réalisées notamment une pour l'évaluation du moteur de rendu MPEG-H qui a rassemblé plus de 4000 réponses à partir d'une dizaine de tests. Ces tests prenaient en compte le rendu d'une scène objet réalisée en 22.2, en débit constant (512kbits/sec) sur différents systèmes de restitution via Haut-parleurs au « sweet spot » et hors « sweet spot » ainsi qu'en binaural au casque. Ces tests ont démontré que les écoutes d'une scène objet hors « sweet spot » n'ont révélé aucun problème lié à une dégradation de la qualité sonore ni même après la conversion de cette scène en binaural simplifié/optimisé.

Une étude a également été menée par Florent Denizot, ancien étudiant à l'ENS Louis-Lumière, dont il explique les tenants et les aboutissants dans son mémoire de fin d'études « *L'approche orientée objet au cinéma : entre surenchère technologique et outil de création sonore* » qu'il est intéressant de consulter. L'objectif de l'étude était de donner une notation de la perception des rendus de mixages objet issus de quatre séquences de fictions, d'évaluer les downmix à partir de formats supérieurs et de les comparer à des downmix traditionnels. Cette étude s'est centrée sur quatre différences significatives : l'immersion, la qualité du timbre, la lisibilité du son et l'adéquation à l'image.

Les résultats de cette étude ont révélé que l'immersion se trouvait décuplée lors de l'écoute d'un mixage objet et de son downmix 5.1 associé comparé à un mixage traditionnel 5.1. Les vingt-deux personnes qui ont participé à ce test ont décelé plus de précision dans la spatialisation des éléments et ont décrit le downmix objet comme étant plus fidèle au réel. En effet, les downmix traditionnels sont souvent à l'origine d'une altération du timbre lors de la diffusion qui se répercute sur la description de l'espace jugé plus confus et moins immersif. Il apparaît ainsi que la qualité du timbre soit améliorée en mixage objet ainsi que pour le downmix objet. En effet, le mixage objet et son downmix associé se sont révélés concluant dans ce cadre cinématographique du fait que le mixage objet permet un dégagement de la zone frontale de l'écran et diminue nettement l'inter-masquage des sources lors de la restitution. En revanche, aucune différence significative n'a été décelée quant à la lisibilité des sons mixés dans les différents cas étudiés. En ce qui concerne, l'adéquation à l'image Florent Denizot précise qu'il est indispensable d'appréhender la création d'une scène sonore dès le montage son pour éviter les phénomènes d'inéquation à l'image. Il apparaît effectivement à travers cette étude que l'exagération d'effets immersifs, particulièrement au cinéma, peuvent devenir nuisible à la mise en scène sonore. Il en conclut que malgré l'engouement que peut susciter les potentialités multiples de spatialisation offertes par l'approche orientée objet, il convient toujours de servir

la mise en scène. Cela soulève des interrogations sur la légitimité du son 3D dans le processus de création.

b) UNE NOUVELLE ESTHETIQUE DU SON SPATIALISE

Au cours de la présente recherche, il apparaît important d'évoquer l'apport du son 3D pour une production radiophonique ou audiovisuelle et de questionner sa légitimité dans la création. En effet, l'intérêt du son spatialisé va dépendre du projet, il peut apporter une plus-value dans l'immersion mais peut aussi se révéler inutile sur certains programmes.

Dans son ouvrage *MP3 : The meaning of a format*, Jonathan Sterne, professeur à l'Université McGill et fondateur des *Sound Studies* se pose la question suivante : Comment se représenter ce qu'est une meilleure définition ? Il explique que la recherche de la meilleure définition est souvent perçue comme une quête de l'immersion et de la richesse pour l'expérience utilisateur finale. Pour Jonathan Sterne Cette quête de la définition passe souvent par la quête du réalisme ou « verisimilitude »⁽¹⁾. En effet, la représentation d'un espace sonore le plus proche possible de la réalité correspond souvent à un gage de définition et le rêve d'un son immersif fidèle à la réalité est devenu un argument indispensable pour le marketing d'une nouvelle technologie de diffusion. L'approche objet collerait donc avec cette évolution des esthétiques où le côté produit (grande largeur de spectre, dynamique contrôlée ...) et spatialisé est maintenant considéré comme le son réaliste et immersif. Cette nouvelle esthétique du son est induite des nouvelles attentes des spectateurs qui tendent vers l'hyper-réalisme. Michel Chion affirmait dans *L'Audio-vision* « Personne ne se plaint d'une non-fidélité par excès de définition »⁽²⁾. Certaines musiques contemporaines comme la musique pop ou la musique électroacoustique s'affranchissent du réalisme ce qui donne des licences pour être créatif au niveau de l'espace. Cependant, il est primordial de se poser la question de cette légitimité du son 3D dans l'optique où des objets spatialisés viendraient compléter la création d'une scène sonore dans un contexte audiovisuel où l'image demeure frontale.

Une scène objet peut connaître une multitude de développements possible et l'on peut se perdre rapidement dans les potentialités narratives offerte par le son 3D. De surcroît, si le son est associé à une image, il est d'autant plus important de se focaliser sur un mode d'écoute pour développer une scène et de prendre en considération le cadre audiovisuel. Le son doit naturellement garder une certaine cohérence avec l'image et apporter un réel intérêt narratif

(1) STERNE Jonathan, *MP3 : The Meaning of a format* – 2012 - 359 pages - page 4

(2) CHION Michel – *L'Audio-Vision 4^{ème} ed.* - Armand Colin - 2017 - 240 pages

afin d'éviter l'aspect gadget souvent reproché aux nouvelles technologies de diffusion, lorsqu'elles tentent de se substituer à la mise scène. Pour François-Xavier Ballon, ingénieur du son et mixeur à France Télévisions, le choix du standard joue forcément sur le récit d'un quelconque programme. Il évoque le binaural natif qui pose la question des rapports de distances difficiles à percevoir en fonction des axes. Un « excès de définition » pourrait donc nuire à l'expérience utilisateur.

C'est pourquoi, il est fondamental de réfléchir à une scénographie sonore de manière à garder un espace sonore cohérent et ainsi appréhender le mode objet comme un outil esthétique plutôt que comme une simple technique. L'immersion ne vient pas d'un bombardement de sons provenant de toutes les directions mais plutôt de la précision dans le placement de sources sonores ponctuelles choisies. L'esthétique du son 3D doit donc reposer, selon Claude Bailblé, sur l'immersion du spectateur dans une géode sonore avec une zone frontale écoutée par le spectateur et une zone hors-champ ou zone d'ambiance, entendue mais inécoutée par le spectateur⁽¹⁾. L'approche orientée objet dégage ainsi deux nouvelles conceptions de la création sonore que sont la synthèse d'espaces sonores virtuels et la captation in-situ d'espaces sonores réels via des enregistrements HOA. Ces conceptions utilisées à bon escient, apporteraient une plus-value incontestable sur l'immersion du spectateur.

L'intégration de ces nouvelles conceptions pourrait faire leur apparition au sein de la chaîne de production avec l'arrivée de nouveaux outils destinés à la production en mode objet. En outre, il est difficile de penser un son orienté objet s'il est monté dans un environnement orienté canal. Une transition pourrait donc être amorcée vers un mixage orienté objet unique compatible sur tous les standards de diffusion. Cependant, cette transition vers l'approche objet serait synonyme d'un bouleversement considérable au sein de la chaîne de production c'est la raison pour laquelle elle se doit d'être anticipée.

(1) BAILBLE Claude – *L'image frontale, le son spatial*, in Cinéma et dernières technologies, Paris : De Boeck Université, INA, 1998, 226 pages

II/ UN CONCEPT DISRUPTIF : LE RENOUVELLEMENT DU WORKFLOW

Le milieu de production des programmes de stock s'est déjà familiarisé avec le mode objet et s'adaptera assez logiquement à mesure que les outils nécessaires feront leur apparition. De plus, l'investissement dans le domaine de la production et la durée du mixage ne seraient pas plus élevés en éliminant la phase des multiples rendus coûteux et chronophages. En effet, le mixeur va seulement livrer la sortie des sous-groupes plutôt que le mix final sous différents formats. Cependant l'implémentation de l'audio orienté objet au sein d'une chaîne de travail de type broadcast représente un véritable enjeu pour les broadcasteurs qui souhaiteraient amorcer la transition.

1) UNE REVOLUTION POUR LES BROADCASTEURS

La tendance des broadcasteurs à la production IP favoriserait l'implémentation d'une chaîne de travail intégralement orientée objet.

a) UNE TENDANCE A LA PRODUCTION IP

Depuis déjà quelques années, la tendance du broadcast s'est orientée vers le transport de signaux en réseau IP avec la standardisation de l'AES67 et des réseaux audio comme *Ravenna* ou encore *Dante* développé par Audinate et aujourd'hui omniprésent dans des prestations importantes de broadcast ou de sonorisation. Les captations actuelles, notamment sportives, présagent de la disparition du car-régie et d'une customisation des moyens en fonction du type de prestations. Après l'aménagement de nombreux stades en France, équipés en fibres optiques, les broadcasteurs TV ont franchi une étape capitale vers l'avenir de la captation sportive. Cette configuration leur permet de réaliser des économies considérables du fait qu'ils n'ont plus besoin de déplacer d'équipes ni de car-régie. Les signaux des caméras et micros sont directement numérisés et multiplexés dans la fibre afin d'être acheminé via IP dans une régie fixe où la réalisation se fera en direct grâce à un Mélangeur IP. Ce mode de production appelé « Remote production » a déjà été mis en œuvre par France Télévision pour les Jeux Olympiques de Rio. France TV a maintenu ses plateaux à Paris et n'a envoyé qu'une équipe très restreinte de techniciens sur place pour assurer la responsabilité des caméras et micros. Le patch relié à Internet a permis de transmettre les images et les sons récupérés à Paris. Auparavant, les consoles étaient conçues selon un standard (Analogique, Numérique, AES, MADI) autour duquel les chefs-opérateurs s'articulaient pour la prestation. Aujourd'hui, les standards

évoluent malheureusement trop vite pour qu'un car-régie puisse se construire sur un standard au risque d'être déjà dépassé lors de sa mise en service.

L'enjeu d'une chaîne de production intégralement objet permettrait de résoudre les problématiques liées aux standards c'est dans cette optique que s'est construit le projet ORPHEUS.

b) L'ARCHITECTURE DE REFERENCE DE L'ORPHEUS

ORPHEUS est un projet de recherche européen de 30 mois partiellement financé par le programme Horizon2020 de l'Union Européenne et dédié à l'amélioration de la gestion du contenu audio pour créer de nouvelles expériences utilisateur. Les dix partenaires du consortium souhaitent développer, implémenter et valider une nouvelle chaîne de production sonore intégralement orientée objets.



*Membres du consortium ORPHEUS – Source : Results of ORPHEUS_Andreas
SILZLE_19/06/2018*

L'objectif du consortium était de vérifier la viabilité de l'ADM dans un contexte de production. En effet, à son lancement il n'existait aucun outil et l'ADM n'avait jamais été mis en pratique. Les membres du consortium se sont donc attelés au développement de cette chaîne de production orientée objet mais aussi à la production de contenu en audio orienté objet et à la validation de ce contenu auprès de l'utilisateur. Les fruits de leur travail et les nouveaux outils développés ont été présentés lors d'un workshop le 15 Mai 2018 à l'IRT à Munich. En outre, leurs résultats sont détaillés dans deux rapports techniques de l'EBU le *Technical Report 042*⁽¹⁾ qui explicite l'architecture de la chaîne de production orientée objet de référence et l'*EBU TECH3388*⁽²⁾ qui donne les spécifications de l'EAR, le moteur de rendu de référence pour l'ADM.

(1) EBU Technical Report 042_ORPHEUS Project_05/2018

(2) EBU Tech3388_ADM Renderer for use in Next Generation Audio Broadcasting_BTFRenderer Group_03/2018

La chaîne de production de référence résultante s'articule autour de cinq étapes distinctes : l'enregistrement, la pré-production/mixage, la réalisation en studio, la distribution et enfin la réception. L'architecture de référence de l'ORPHEUS pour un Workflow intégralement orienté objet se trouve en ANNEXE 2.

L'approche du NGA (Next Generation Audio) est prévue pour pouvoir combiner des enregistrements issus de différentes approches (canaux-objets-HOA) au sein d'une même scène.

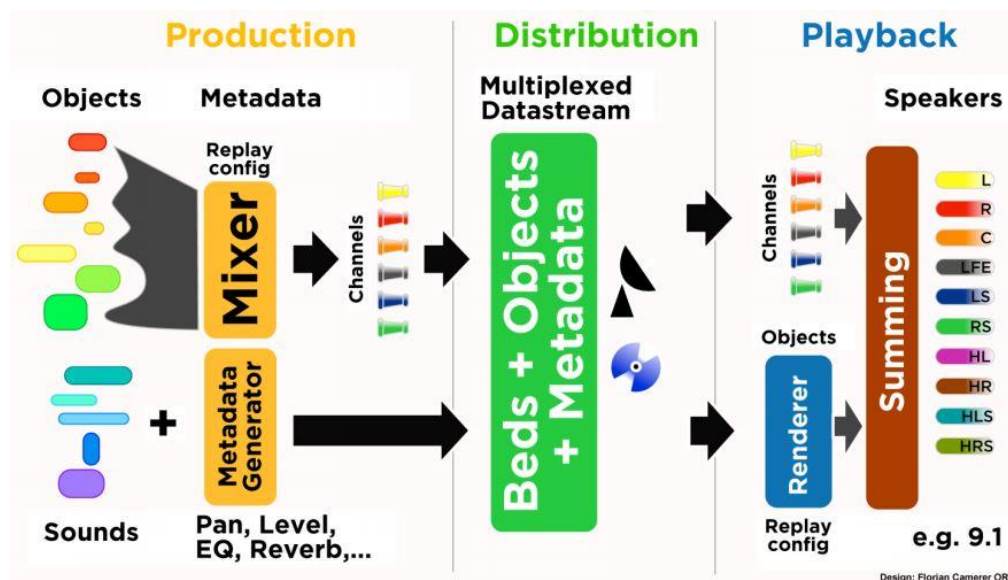


Schéma de production NGA combinant Objets et Beds multicanaux - Source : Florian Camerer ORf_EBU Technical Report 045_AS_01/2019

Il serait donc possible d'utiliser différents systèmes de prise de son et notamment des systèmes multicanaux ou 3D comme l'ORTF 3D, l'Eigenmike ou le Sennheiser AMBEO afin de réaliser des beds d'ambiance. Ensuite les enregistrements se feraient directement sur une station de travail (DAW) à l'aide de plug-ins VST Steinberg ou AAX Avid. Les équipes de b<>com et Fraunhofer ont donc développé des outils pour l'analyse, le processing la conversion en HOA d'enregistrements réalisés à partir de microphones 3D ainsi que des outils pour les restituer sur différentes configurations de haut-parleurs. L'enregistrement d'un contenu orienté objet peut se faire indépendamment du DAW mais dans ce cas le contenu enregistré est transféré sous la forme d'un fichier multicanal avec des métadonnées ADM au format BW64 standardisé *ITU-R BS.2088*⁽¹⁾.

(1) Recommendation ITU-R BS.2088-0_Long-form file format for the international exchange of audio programme materials with metadata_10/2015

Pré-production/Mixage

La phase de pré-production et de mixage propose un mode de montage son et de mixage incluant l'édition des métadonnées propres aux objets au sein même du DAW. Ces expérimentations ont été réalisées sur la station de travail *Sequoia* développée par MAGIX. Le rôle de l'étape de pré-production est de réaliser un montage et un mixage des objets sonores et des autres signaux au sein d'une station de travail orientée objet « object-based DAW ». Celle-ci serait complétée par différents outils dans le but d'importer, monter, monitorer et exporter du contenu orienté objet. Les composantes principales de cette station de travail orientée objet s'articulent autour de :

- La possibilité d'importer, décoder, encoder et exporter du contenu orienté objet ainsi que leurs métadonnées.
- La possibilité d'accéder à du contenu non-objet et à le convertir automatiquement en contenu objet en lui associant des métadonnées objet requises.
- La possibilité d'appliquer des effets et plug-ins, notamment pour le processing des HOA.
- La possibilité de monitorer le contenu orienté objet sur différentes configuration de haut-parleurs ou en binaural ce qui suppose l'intégration d'un Renderer.

Le studio de radio

L'étape du studio se caractérise par la composition de la scène objet dans le but de réaliser un seul programme. Contrairement à un studio traditionnel, les sources audio sont séparées et il convient de générer les métadonnées correspondantes à la façon dont elles doivent être restituées au sein du programme. Pour ce faire, l'ingénieur du son devra utiliser une interface intégrant différentes fonctionnalités :

- La possibilité de créer une nouvelle scène objet.
- La possibilité d'ajouter et de supprimer des objets.
- La possibilité de modifier le gain et la position des objets dans l'espace 3D.
- La possibilité de marquer chaque objet en fonction de son contexte d'utilisation.

Cette interface correspondrait idéalement à une table de mixage complétée d'une interface graphique et d'un contrôleur physique. L'interface de pré-production quant à elle serait destinée au producteur pour lui permettre de déterminer les données d'édition du

programme comme par exemple la création d'une séquence de chapitres ou de scènes dans le programme ou alors d'intégrer des liens URL en relation avec le contenu audio.

Dans le cas d'une production Live, les signaux ne sont pas enregistrés au préalable mais arrivent via des flux audio PCM et complétés par des flux de métadonnées ADM. C'est au cours de cette étape que les sources seraient mixées, spatialisées et monitorées.

L'étape du studio permettrait la transmission de l'audio et les métadonnées via des réseaux IP comme l'AES67 pour l'audio où l'UMCP (Universal Media Composition Protocol) pour le transport de métadonnées. Pour ces transmissions d'audio orienté objet via réseaux IP, l'ORPHEUS se base sur les spécifications NMOS (Networked Media Open Standards) pour définir les sources audio et le timing des événements. En outre, l'étape du studio doit permettre à l'ingénieur du son d'enregistrer et de relire des flux audio mais aussi des flux de métadonnées. Pour cela, il pourra enregistrer des séquences, « *sequence store* » selon l'ORPHEUS, et interagir avec elles à l'aide d'une interface API développée par la BBC appelée « *Media Access API* ». Enfin, elle se doit d'inclure une phase de vérification et de contrôle avec la possibilité pour l'ingénieur du son de monitorer son signal et de le soumettre à des tests de Loudness.

La distribution

Par la suite, la phase de distribution définit les modules et outils nécessaires à la distribution d'un contenu audio orienté objet, du diffuseur à l'utilisateur final. Comme évoqué précédemment, pour distribuer ce contenu audio orienté objet il est obligatoire de le convertir dans un format adapté au transport. L'étape de distribution récupère donc un flux audio PCM par AES67 et un flux de métadonnées basé sur le standard ADM à partir d'un réseau IP privé. Une grande partie de cette distribution se fait sur le Web mais il est indispensable de rester compatible avec les modes de diffusion traditionnels. Pour la diffusion Web, l'architecture de l'ORPHEUS se base sur le protocole de streaming DASH et sur un CDN (Content Diffusion Network). Deux options ont été privilégiées pour le moment, une distribution AAC pour les navigateurs HTML5 et une pour les players supportant le MPEG-H. En ce qui concerne les systèmes de distribution traditionnels (DAB+ ou DVB), une rétrocompatibilité des mixages objet avec des systèmes stéréo, 5.1 et binaural a été pensée et certaines métadonnées audio choisies seraient transportées dans des modèles préexistants pour ces systèmes. Ces métadonnées sont spécifiées dans la phase de « filtrage » où seules les métadonnées utiles à l'utilisateur sont isolées avant d'être distribuées. En fonction des capacités des formats de distribution, le contenu audio et les métadonnées devront parfois être traités ou convertis selon

le profil de compatibilité du codec. Ce traitement peut consister à la réduction du nombre d'objets ou le regroupement de plusieurs objets sur une piste. Ensuite les métadonnées de l'ADM seront converties selon le profil de compatibilité du codec utilisé (MPEG-H dans le cadre de l'ORPHEUS) qui spécifiera alors ses propres métadonnées.

L'interface utilisateur

Ultérieurement, la phase de conception de l'interface utilisateur est décisive pour permettre à l'auteur de créer de l'interaction avec les utilisateurs ou ne serait-ce que pour leur fournir un programme adapté à leurs préférences comme le choix de la langue, celui d'un ajustement de niveau ou celui d'une lecture non-linéaire. Le design de cette interface a une influence particulièrement importante sur l'expérience utilisateur finale. Elle doit donc être considérée en amont.

La réception

Enfin, la phase de réception détermine les solutions pour une réception optimale et personnalisée du contenu audio orienté objet pour l'utilisateur final. Ces solutions diffèrent dans leurs capacités de rendu en fonction de la bande passante réseau disponible, du nombre de canaux de sortie audio et de la puissance de traitement. Ils diffèrent également en termes d'interface utilisateur et de fonctionnalités proposées en matière de personnalisation/interactivité puisqu'elles ne s'adressent pas aux mêmes contextes d'écoute. Cependant, toutes ces solutions comprennent deux composants majeurs : un module de décodage et de rendu et un module de personnalisation. Le module de décodage/rendu décode les flux audio compressés et adapte la restitution au type de scène et au système de diffusion de l'utilisateur. Le module de personnalisation permet quant à lui, la modification des métadonnées afférentes aux fonctionnalités d'interactions avec l'utilisateur. Il commande également l'interface utilisateur qui interprète et affiche les métadonnées relatives au contenu tel que le titre, l'auteur, le choix de la langue et la transcription du texte. Ces solutions de restitutions ont toutes été testées par le consortium ORPHEUS avec notamment le développement par elephantcandy d'une application sur iOS intégrant les fonctionnalités de l'audio orienté objet. La restitution via périphérique audio hardware a été également testée à l'aide d'un périphérique construit par Trinnov et destiné à des systèmes home-cinéma hauts de gamme pour restituer une scène sur plusieurs configurations de haut-parleurs incluant l'élévation. Ces solutions ont été testées partir d'un profil de compatibilité du MPEG-H (Low Complexity Profile level 3) limitant la scène à 16 signaux simultanés sur un total de 32 canaux,

objets et/ou signaux HOA. Enfin, la réception via un navigateur Web HTML5 a été expérimentée car elle correspond au mode de diffusion actuel le plus accessible dans le monde. Cependant, aucun player Web n'intègre de Renderer pour le moment, notamment à cause de la concurrence entre les différents constructeurs que nous aborderons plus tard. Le consortium ORPHEUS s'est donc concentré sur l'implémentation d'un moteur de rendu basé sur JavaScript à partir d'une plateforme WAA (Web Audio API) dédiée au traitement de l'audio dans les applications Web. Les formats supportés par ce moteur de rendu sont le binaural au casque et des configurations de haut-parleurs 2D et 3D conventionnels avec une spatialisation basé sur la technologie de panning d'amplitude VBAP.

Pour conclure, le consortium ORPHEUS a réalisé des études plus approfondies sur des aspects pratiques caractéristiques au milieu de production broadcast :

- Un système d'archives pour les fichiers audio en BW64.
- Le déclenchement à un timing précis de la lecture de fichiers ADM ou d'ADM sérialisé pour les applications de streaming ou de diffusion live.
- La compatibilité de l'architecture de référence avec d'autres formats et protocoles (AC-4 *Atmos*, DTS:X).
- La compatibilité de l'architecture de référence avec les systèmes de distributions existants (DAB, DVB, ATSC) et avec les plateformes hors-IP.
- La gestion des débits binaires variable pour le transport de l'audio orienté objet encodé.
- La gestion du transport des métadonnées additionnelles aux métadonnées de bases.
- La gestion du Loudness des programmes audio orientés objet.

Les résultats du consortium ORPHEUS ont révélé les potentialités de l'audio orienté objet pour la création de nouveaux contenus interactifs et adaptables aux systèmes de diffusion et aux conditions d'écoute des utilisateurs. Avec des équipes de la BBC, un projet pilote d'une émission de radio orientée objet dont différents événements sont transcrits en objets en direct jusqu'à la conversion, se sont révélés concluants sur l'implémentation de cette chaîne de production.

Le consortium a donc prouvé la viabilité d'un processus intégralement orienté objet. Malgré tout, ils démontrent que l'implémentation de cette technologie provoquerait de profonds changements dans les architectures et infrastructures existantes des diffuseurs.

2) DE LA CAPTATION A LA RESTITUTION D'UNE SCENE 3D

L'implémentation de l'audio orienté objet impliquerait un bouleversement dans la manière d'appréhender une scène sonore.

a) CREER UNE SCENE SONORE ORIENTEE OBJET

Ces changements au sein du workflow interviendraient dès la prise de son. En effet, dans un workflow classique, le dispositif de prise de son est pensé de manière à ce qu'il soit compatible avec le downmix. Lors du passage de la stéréo au 5.1, des expérimentations ont été faites pour faire du 5.1 avec deux couples AB (avant, arrière) mais les downmix stéréo résultants étaient générateurs de problèmes de phases entre les différents composants. En effet, ces systèmes de prises de son étaient variables par rotation, il a donc fallu adapter le dispositif pour que le mixage 5.1 soit compatible avec la stéréo. Ainsi les couples AB ont été remplacés par des dispositifs tels que l'Arbre de Decca avec la possibilité de les compléter avec des systèmes surround (Omni8+, Croix IRT, Carré Hamasaki). Ces dispositifs ont apporté des nouvelles possibilités comme la disposition de « micros antennes » mixtes composés de deux microphones omnidirectionnels et un microphone plus directif pour capter le son direct.

Les problématiques liées à la transition de la stéréo au 5.1 sont identiques à celles que se posent désormais les ingénieurs du son pour assurer la transition vers la création de scènes audio 3D. L'enjeu principal de ces scènes 3D est de pouvoir assurer une prise de son favorisant la décorrélation des sources ou qui garde un bon rapport entre corrélation et décorrélation. Par bon rapport, nous entendons palier les problèmes liés à la présence de deux sources sonores séparées sur plusieurs canaux lors de l'enregistrement. L'idée phare est de capter la plupart des sources en contrôlant leur corrélation ou leur décorrélation pour que la scène finale soit invariante par rotation. A partir du moment où ces scènes sonores sont invariantes par rotation, il deviendra possible de les restituer sur n'importe quel système de diffusion. L'intérêt de ces scènes 3D est aussi de créer des espaces cohérents. C'est-à-dire que l'auditeur doit être à l'aise avec cette spatialisation et celle-ci doit rendre des caractéristiques de l'environnement acoustiques et valoriser l'aspect spatio-temporel du son.

Dans ce but, Jean-Christophe Messonnier et ses collaborateurs ont publié en 2016 un rapport à l'AES décrivant de nouveaux dispositifs de prise de son appliqués à l'approche orientée objet dans un contexte de captation de concerts⁽¹⁾. Ces nouveaux dispositifs sont basés

(1) MESSONNIER Jean-Christophe, LYZWA Jean-Marc, DEVALLEZ Delphine; DE BOISHERAUD Catherine, *Object-Based Audio Recording Methods*, AES E-Library, 2016

sur un système de rampes. Ces rampes équipées de microphones plus ou moins directifs ont la caractéristique de séparer les directions de captation du champ acoustique auquel on s'intéresse en plus de garder un bon rapport entre la corrélation et la décorrélation des sources. Pour composer une scène, il est alors idéal d'avoir des sons directs complètement décorrélés ainsi que des réverbérations, qui par principe, sont complètement décorrélées même si elles représentent le même espace. Une difficulté survient lorsqu'on veut représenter des scènes continues. Elle consiste à représenter des zones où il y aurait une continuité des sons afin de rester proche d'une logique orientée objet. Pour Jean-Christophe Messonnier, adopter une logique du mode objet consisterait à justement démultiplier les approches et à penser de manière différente les systèmes de captation. Cette différence d'orientation intellectuelle pousse à remettre en question l'apport de chaque système et de l'adapter en fonction de ce qui nous intéresse. L'approche objet serait donc à l'origine d'une grande variété de systèmes possibles et transposable en fonction du type de captation. Jean-Christophe Messonnier établit même un parallèle avec les enseignements tirés des problématiques de la captation multicanale qui se sont avérés bénéfiques pour la captation en stéréophonie pure. Les ingénieurs du son s'interdisaient des pratiques avec les couples AB qui ont été approuvées à la suite d'expérimentations en multicanal. Ainsi, par la décorrélation des sources, l'utilisation de rampes pour une production en mode objet serait également compatible avec des formats de base comme la stéréophonie, à la fois sur enceintes et au casque. Il serait même envisageable voire bénéfique d'utiliser ces rampes pour des captations en stéréo pure. Pour Jean-Christophe Messonnier, il est indispensable que ces nouveaux modes de production ne provoquent pas une perte de qualité par rapport à un travail en stéréophonie.

En ce qui concerne l'enregistrement, Jean-Christophe Messonnier avait déjà anticipé que la création d'une scène objet pouvait se diviser en plusieurs étapes et notamment en différenciant la phase d'enregistrement des objets sonores, la phase de mixage et celle de la composition de la scène. Il prend l'exemple de la captation d'un opéra en audio orienté objet. Chaque source est soumise à diverses couches de traitement (Egalisations, Dynamiques etc..) avant d'être enregistrée post-FX et post-Fader. Chaque son enregistré est alors considéré comme un objet non-standard ou objet préformé, appartenant à une grosse scène. Ces sons sont ensuite regroupés au sein d'objets standards qui seront spatialisés lors d'une deuxième phase destinée à la composition de la scène objet par réduction de la scène originelle. Cette phase de composition devient une sorte d'empilement de différents conditionnements du son avec l'ajout d'effets spatiaux comme des réverbérations artificielles. L'ajout d'effets spatiaux contribue à

changer la nature des objets de base et on peut le considérer comme générateur de nouveaux objets intégrés à l'espace. On peut également citer la possibilité d'anticiper une réverbération 3D dès la prise de son et de recréer cette réverbération à partir de trois appoints mono, quitte à la réduire ensuite en stéréo.

La réverbération pose des enjeux cruciaux pour la création d'une scène objet. Le consortium ORPHEUS avait également anticipé ces problématiques et avait sollicité les équipes de l'IRCAM pour étudier la représentation de la réverbération dans un contexte objet et ainsi entrevoir des pistes d'évolution pour l'ADM. Pour la production d'une scène sonore 3D, une réverbération enveloppante s'impose à la fois à la captation et à la restitution. En effet, la zone du son direct peut se limiter à un axe entre les enceintes gauche et droite auquel on ajoute des objets mais ce sont les ambiances et la réverbération qui sont ensuite à l'origine d'une spatialité plus ou moins marquée. La plupart du temps un mélange entre l'acoustique de la salle et des réverbérations 3D artificielles est réalisé. Cependant les questions de réverbération imposent des choix par rapport aux systèmes de prise de son. Soit les pistes sont enregistrées séparément (une pour chaque source) et dans ce cas-là la création d'un espace 3D sera nécessaire pour chacune de ces pistes ce qui implique un travail très lourd pour l'ingénieur du son. Soit, les pistes sont enregistrées avec une base acoustique commune déjà présente. Celle-ci va se prêter au mode objet car on aura une captation de la pièce en 3D qui donnera une empreinte particulièrement enveloppante en fonction des qualités acoustiques de la salle. Dans ce cas, la réverbération artificielle 3D devient négligeable et l'acoustique peut suffire aux enjeux de la captation. En revanche, les systèmes principaux de prise de son ne captent pas toute la réverbération. Il convient donc de se poser une question essentielle : quel système de prise de son sera affecté au son direct et quel système de prise de son sera affecté au son réverbéré ? On pourrait alors utiliser un système de prise de son pour l'avant avec un meilleur accès au son direct (exemple des microphones HF pour un Opéra) afin de garder une bonne présence qu'on associerait avec un système de captation de la réverbération qui serait diffusée partout. Par ce biais, la balance son direct/son réverbéré serait respectée et la spatialité de la scène sonore 3D résultante serait cohérente. Néanmoins, dans une scène 3D les flux de réverbération doivent être séparés pour retranscrire au mieux l'espace. C'est pourquoi, sur un système de prise de son 3D, le son réverbéré n'est pas seulement capté par l'avant. Dès la prise de son, il faut arriver à distinguer à l'aide de différents systèmes de captation ce qui vient de l'avant, ce qui vient d'en haut, ce qui vient de l'arrière...

En conclusion ces associations de systèmes de prise de son pourraient donner ce type de dispositif pour la captation d'une scène 3D :

- **Avant** : Rampes de microphones plus ou moins directifs pour le son direct
- **Arrière** : deux microphones cardioïdes orientés vers le fond ou des microphones omnidirectionnels très éloignés de la scène dans des grandes salles
- **Côtés** : Microphones omnidirectionnels pour garder une connexion avec ce qu'il se passe à l'avant.
- **Élévation** : Système de quatre microphones bidirectionnels suspendus - Carré Hamasaki - plus ou moins orienté vers les sources.

b) DES PERSPECTIVES D'APPLICATIONS

Suite à l'engouement des diffuseurs pour l'audio 3D, de nombreuses expérimentations ont été réalisées dans l'optique d'étudier la viabilité de l'audio orienté objet au sein d'un workflow de captations live.

En 2017, David Paillet, ingénieur du son pour AMP Visual TV a participé à des tests des nouveaux équipements Dolby *Atmos* pour d'un match de hockey sur glace avec des ingénieurs du laboratoire Dolby dans le cadre du Consortium *4Ever*. Ce fut la première captation encodée en Dolby AC4 avec transmission et réception par satellite. Bien que le Dolby AC4 n'affecte que 2 canaux dans les enceintes overhead contrairement aux 4 canaux préconisés par la norme *Atmos*, les diffuseurs ont fait un premier pas vers la distribution d'un contenu audio 3D en direct à l'aide des encodeurs DP590, DP591, DP 580. L'encodeur Dolby DP 591 encode à partir du format MADI en sortie de console qui amène des possibilités allant jusqu'à 64 canaux au lieu des 8 disponibles en Dolby E. L'intérêt de ces nouveaux codeurs pour un diffuseur était de pouvoir transporter une VI 7.1 standardisée Dolby *Atmos*, une VI stéréo, un mixage complet (Version Internationale + Commentaires Anglais mixés), un mixage complet Français ainsi que des mixages dans des langues différentes si nécessaires. La prise de son quant à elle reprenait les mêmes codes qu'en 5.1 avec des signaux d'ambiances placées à l'arrière, à noter l'ajout de plusieurs couples placés en élévation et dirigés vers le bas de la patinoire. Pour l'instant, les équipements Dolby pour les diffuseurs ne sont pas à vendre mais uniquement fournis en prestation de services par Dolby avec un technicien venant sur place pour valider la qualité de la cabine de mixage, les prises de son, etc. Cependant, cette situation est amenée à évoluer.

En effet, l'avenir du Dolby *Atmos* en production TV passerait comme lors des coupes du monde de football précédentes, par une cabine de mixage distante, agréée, recevant les canaux discrets via IP et pouvant encoder directement en AC4.

David Paillet a également été chargé de la captation et de la retransmission audio des sites de « Figure Skating » et de « Short Track » aux Jeux Olympiques de PyeongChang sur lesquels étaient disposés dix couples de microphones sur deux niveaux (haut et bas) avec un microphone *Surround 5.1* DPA s'ajoutant aux ambiances très hautes afin de réaliser un mixage en Dolby *Atmos*. De plus, il a participé à une captation en 22.2 du « Short Track » avec la NHK, service de radiotélédiffusion public Japonais, qui malgré la grande précision de sa spatialité et son immersion proche du réel n'est pas prête de s'exporter chez les téléspectateurs à cause de la trop grande complexité de sa configuration pour l'équipement des utilisateurs.

Une autre expérimentation de l'approche objet au sein d'une captation a été réalisée par France Télévisions lors des Internationaux de France de Tennis à Roland Garros en 2018. L'ingénieur du son du car principal récupérait un bed d'ambiance à partir d'un set de microphones ORTF 3D Schoeps. De plus, il avait défini en tant qu'objets les voix commentées Anglaise et Française. Après les traitements d'égalisation et de dynamique correspondants le niveau de la voix ne varie plus, le chef opérateur du son n'a plus à gérer les commentaires et peut se concentrer sur le jeu et ouvrir les micros du bon côté du terrain de manière à éviter les effets d'écho. En sortie du car, il envoie donc sept mixages à partir de sa console : une version 3D avec deux objets (Commentaire Anglais et Commentaire Français) encodée en MPEG-H, une VI 5.1, un 5.1 + Commentaire Anglais, un 5.1 + Commentaire Français, une VI Stéréo, un stéréo + Commentaire Anglais et un stéréo + Commentaire Français.

De même pour les championnats d'Europe d'athlétisme 2018 à Berlin où France Télévisions n'a envoyé que des ingénieurs du son de manière à tester trois chaînes d'encodage NGA toutes limitées à 16 canaux (15 pistes audio, 1 piste de métadonnées) :

- MPEG-H (Multicanal + Objets)
- MPEG-H (HOA + Objets)
- Dolby AC4 *Atmos*

La première chaîne d'encodage distribuait un bed 7.4 et 4 objets qui correspondaient aux voix commentées (1 : Commentaire Français, 2: Commentaire Anglais, 3: Audio Description Français, 4: Audio Description Anglais). Pour la chaîne d'encodage HOA, le bed multicanal 7.4 est remplacé par un HOA d'ordinaire d'ordre 2 qui correspond à 9 canaux. Cependant

Qualcom, qui a développé la partie ambisonique du MPEG-H, parvient à simuler un HOA d'ordre 3 en exploitant les 11 pistes disponibles en diminuant la dimension verticale dans la sphère du bas de l'ordre 3 (jugée peu intéressante) et en l'exploitant pour augmenter la dimension horizontale.

Ces expérimentations variées proposent toutes des solutions pour fournir un contenu audio plus immersif et personnalisable mais pour l'instant les diffuseurs Live sont forcés d'envoyer des mixages complets aux téléspectateurs européens car leurs téléviseurs ne sont pas encore équipés de moteurs de rendu. Le modèle objet semble donc apporter des solutions prometteuses pour les diffuseurs néanmoins des freins subsistent à l'intégration de l'audio orienté objet.

III/ LES DEFIS A SURMONTER POUR AFFIRMER LE MODELE ORIENTE OBJET

Les deux premières parties avaient pour objet d'exposer le fonctionnement, les potentialités, les intérêts et la manière d'implémenter l'audio orienté objet au sein d'une chaîne de production pour les diffuseurs de programmes Live ou de stock. Cependant, pour que l'audio orienté objet s'affirme en tant que modèle, il est indispensable de ne pas se limiter à une approche techno-centrée. Il convient d'étudier aussi le contexte socio-économico-culturel qui influe particulièrement sur l'approbation d'une technologie, de son déploiement sur le marché et qui détermine le passage du statut d'invention à celui d'innovation.

1) L'ŒUF OU LA POULE, DES EQUILIBRES A DYNAMISER

A l'heure actuelle, tous les diffuseurs s'observent et n'arrivent pas à s'entendre sur ce qui amorcera concrètement la production en audio orienté objet. D'une part, le manque d'équipements pour les auditeurs et le manque d'outils pour développer du contenu n'incitent pas les producteurs à investir pour fournir des expériences nouvelles aux auditeurs. D'autre part, le manque de contenus proposés par les producteurs (seulement quelques démos issues des filiales de recherches technologies comme celles de Radio France, d'ARTE ou France Télévisions) n'incite pas les constructeurs à développer des outils et le public à s'équiper. Toutefois, l'intégration d'outils ADM dans des stations de travail comme Pyramix ou bientôt Pro Tools aura forcément un impact en faveur de la démocratisation de la production en mode objet.

a) DES DEFIS TECHNIQUES A SURMONTER

De la même manière, le perfectionnement de certains aspects techniques de la chaîne de production suscitent quelques réticences du côté des diffuseurs qui préfèrent temporiser en attendant une version définitive du processus avant d'investir. Je précise qu'à la date où j'écris ce mémoire, toutes les problématiques citées ci-dessous sont actuellement à l'étude et avec la vitesse du développement de ces technologies ces problèmes d'ordres techniques sont amenés à être solutionnés au cours des mois à venir, si ce n'est plus tôt.

La finalisation du protocole ADM

Le protocole ADM est effectivement en constante évolution suite à un travail en collaboration avec l'EBU. C'est seulement une fois finalisé qu'il sera possible de définir des presets d'usage des métadonnées à destination des ingénieurs du son. Cela suscite donc une

impatience chez les producteurs de contenus qui aimeraient commencer à travailler en mode objet. Ceux-ci attendent la finalisation du protocole ADM pour réfléchir à la manière d'implémenter le Recorder ADM dans leurs chaînes de travail et réaliser des archives d'enregistrements ADM qui seront compatibles lorsque les premiers encodeurs MPEG-H ou Dolby *Atmos* sortiront sur le marché.

L'optimisation de la conversion de l'ADM

L'ADM n'est pas destiné à la restitution, il nécessite donc une conversion en un format (MPEG-H ou *Atmos*) pour être restitué et cette conversion implique forcément de la distorsion et de la perte d'informations. L'approximation dans le codage induit une perte de qualité ou de précision de la spatialisation dans le rendu. Les codecs de Dolby et Fraunhofer sont donc régulièrement mis à jour dans l'optique de convertir au mieux les métadonnées ADM et en particulier d'optimiser les profils de compatibilité indispensables pour conserver l'intégrité d'une scène et obtenir un rendu cohérent quelque-soit le standard de diffusion.

Le développement d'un player Web

Pour l'instant, aucun navigateur Web grand public n'a intégré de Renderer. En effet, pour la consommation de vidéos sur le Web ou sur les applications, le problème des licences intervient. Pour intégrer le player à une application il est nécessaire de payer la licence du propriétaire et aucun navigateur n'accepte de payer pour intégrer cette technologie. Prenons l'exemple suivant. Si la société Google souhaitait intégrer le player MPEG-H dans son navigateur Chrome, cela signifie qu'elle accepterait de reverser un euro à Fraunhofer pour chaque téléchargement de Chrome ce qui représenterait une somme colossale (environ 100 millions d'euros). C'est pourquoi, chacun essaie de faire des partenariats avec des constructeurs de smartphones et de téléviseurs qui paieraient la licence pour intégrer le décodeur dans leurs appareils. C'est ce qui s'est passé lors du partenariat entre Samsung et Dolby pour intégrer le décodeur *Atmos* dans leurs smartphones Galaxy S10.

Enfin, il existe la potentialité qu'une société influente comme Facebook, Google ou Amazon s'intéresse au format Open Source de l'EBU et à son Renderer EAR et l'utilise pour faire évoluer leur codec (gratuit) *Opus* pour lui permettre de manipuler de l'ADM. Cette hypothèse serait dramatique pour Dolby, DTS et Fraunhofer car cela signifierait que personne ne paierait pour leurs solutions. Dolby et Fraunhofer peuvent perdre beaucoup d'argent sur les navigateurs Web si *Opus* s'empare d'un format Open Source. En revanche, ils gagneront toujours 1€ par téléviseur vendu intégrant leur décodeur. En effet, les fabricants de téléviseurs exploitent uniquement des technologies qui fonctionnent contrairement aux fabricants de navigateurs internet. Ceux-ci sont plus ouverts aux expérimentations car ils détiennent un plan

de secours si cela ne fonctionne pas correctement. C'est pourquoi, les fabricants de téléviseurs ont tendance à choisir des solutions éventuellement payantes car ils peuvent établir un contrat de service qui leur assure la sécurité que le code ne « plantera » pas.

Sur ce marché, ce n'est plus la rémunération de la propriété intellectuelle qui entre en compte mais la rémunération du travail effectué. Aujourd'hui, le code du Dolby AC3+ est tombé dans le domaine public. Dolby a établi une nouvelle stratégie qui consiste à vendre des mises à jour du codec AC3+ qui sont achetées par les constructeurs de téléviseurs qui par sécurité préfèrent avoir la dernière version du code.

Quels paramètres seront à la portée de l'auditeur ?

La phase de réception inclut l'enjeu primordial du design de l'interface utilisateur. Pour Hervé Déjardin, responsable Recherche et Développement à Radio France, il convient de créer de l'interactivité ludique et d'éviter autant que possible l'interactivité technologique. L'objectif pour les distributeurs est d'amener à l'utilisateur la meilleure qualité possible, le plus simplement possible et le plus démocratiquement possible. Toutes les options d'intelligibilité, d'acoustique, d'amélioration de la perception du rendu doivent être automatisées dès que c'est faisable.

La recherche autour du design de cette interface vise un maximum d'accessibilité et la simplicité pour l'auditeur. Prenons par exemple la volonté pour un auditeur d'augmenter la prééminence des dialogues. Dans une logique de simplification, l'interface intégrerait seulement un bouton sur lequel appuyer plutôt que d'aller chercher dans des menus le niveau des dialogues. En effet, les décodeurs intégrés dans les téléviseurs ne permettent pas forcément d'accéder aux fonctionnalités avancées de manière efficace. Alors que les constructeurs de téléviseurs montrent une certaine réticence à ajouter des boutons sur leurs télécommandes, l'interface pour les utilisateurs demeure essentielle.

Malgré tout, elle est toujours sujette à des évolutions. Dolby offre gratuitement la licence AC4 aux propriétaires des licences AC3 et AC3+ mais une interface est nécessaire pour que l'utilisateur puisse avoir accès à ces nouveaux réglages et pour pouvoir interagir avec les objets. Pour l'instant aucun constructeur de téléviseur n'a fabriqué d'interface pour le gérer.

Le développement d'une interface pour les mixeurs

L'enjeu d'une interface Hardware entre l'ADM et les codecs à disposition des mixeurs se pose également. Afin de maîtriser la spatialisation des objets, le monitoring et la conversion de l'ADM, une interface physique intuitive doit être pensée. Cependant, des pourparlers subsistent

afin de statuer sur son apparence ainsi que sur les paramètres qui seront mis à disposition des mixeurs.

L'évolution des outils de mixage

Dès ses premières applications, les scènes sonores 3D ont posé de nombreuses problématiques quant aux outils de mixages nécessaires à leur conception. Dans son récent rapport, le consortium Edison 3D présente une chaîne de production NGA qui intègre de nouveaux logiciels Open Source pour la création et l'archivage de scènes 3D. L'objectif phare était d'uniformiser des outils de création en son 3D de manière à les intégrer dans des stations de travail. Au sein de sa chaîne de production, le projet Edison 3D a également expérimenté les aspects d'interaction avec l'utilisateur, l'enregistrement de position d'objets, le monitoring des scènes sonores sur différents systèmes de restitution (stéréo, 5.1, binaural, WFS, HOA) ainsi que l'export de fichiers basés sur le standard ADM.

A l'heure actuelle, les outils de mixage en mode objet sont pensés pour être incorporés dans des stations de travail (Pyramix, Nuendo, Pro Tools, Sequoia ...). L'intégration de ces outils dans les stations de travail (DAW) les rendraient compatibles avec une production orientée objet et contribuerait à démocratiser le modèle. Une tendance se dessine également vers l'aménagement des consoles de manière à les équiper d'outils intégrés capables de générer les métadonnées objet. Chaque constructeur de moteur de rendu essaie alors d'influencer les développeurs de stations de travail et les fabricants de consoles de façon à ce qu'ils intègrent ce moteur de rendu dans leurs produits. C'est aussi par ce biais que des constructeurs comme Dolby ou Fraunhofer pourraient imposer leurs solutions sur le marché de la production objet.

Les dispositifs de restitution : un moteur de rendu propre à chaque « device »

Avec l'audio orienté objet, un saut technologique majeur est annoncé pour la télévision avec l'arrivée d'un moteur de rendu audio placé dans les récepteurs. En revanche, un investissement majeur est nécessaire pour implémenter un décodeur dans tous les types de « devices ». Pour pouvoir décoder en fonction du système d'écoute, le moteur de rendu doit prendre conscience de la configuration dans laquelle il est intégré (TV stéréo, Home Cinéma, Binaural sur smartphone) mais aussi de la position des haut-parleurs et des caractéristiques de la salle d'écoute. Le monde de l'informatique et des télécoms possède déjà une longueur d'avance sur la télévision puisque tablettes, smartphones et applications web intègrent déjà des moteurs de rendu audio capables d'offrir de nouvelles expériences utilisateurs à partir de formats objet. En effet, les processeurs y sont plus faciles à programmer, et les éditeurs de contenus peuvent choisir d'y intégrer les fonctionnalités qu'ils désirent. Avec la sortie de l'EAR, le moteur de rendu de référence Open Source de l'EBU, il est désormais possible

d'offrir un moteur de rendu qui compile les objets selon les algorithmes de référence. Pour Matthieu Parmentier, responsable Recherche et Développement de France Télévisions, la vraie question n'est pas de savoir si les fabricants de téléviseurs décideront de l'incorporer dans leurs produits du fait que les moteurs de rendu vont commencer par coloniser les applications pour smartphones. La transition sera donc d'abord amorcée par les smartphones haut de gamme, puis s'imposera dans les moyennes gammes puis dans les téléviseurs au fur et à mesure que les contenus objet se démocratiseront. Pour lui, le saut qualitatif des nouvelles expériences utilisateur fournies par l'audio orienté objet poussera naturellement le marché à s'y adapter.

Au-delà des considérations techniques, la démocratisation du modèle objet se confronte aux controverses liées au processus de standardisation.

b) UN STANDARD POUR GOUVERNER TOUS LES FORMATS

Nous avons évoqué précédemment le changement de paradigme fondé sur le standard ADM puis encodé et restitué au moyen de codecs propriétaires ou Open Source. Contrairement, au cinéma où la bande passante est illimitée, l'avènement du codec s'impose dans des domaines comme le broadcast ou le multimédia. Tout le monde semble trouver un intérêt à ce qu'un format unique s'impose pour éviter les problèmes de compatibilité douteuse avec d'autres formats. De plus, la plupart des diffuseurs et producteurs de contenu européens attendent de savoir lequel sera le plus performant avant de s'équiper. Peut-on alors se diriger vers un seul codec standardisé pour l'audiovisuel et le multimédia ?

Dans son ouvrage *MP3, The meaning of a format*, Jonathan Sterne définit un standard comme un format opérable sur tous les systèmes et régissant tous les protocoles nécessaires dans ce but. Il évoque le fait qu'un standard émerge en réponse à divers problèmes précis dans un environnement commercial particulier mais qu'il résulte autant de la politique que de la technologie, de l'économie ou de la culture. Dans son ouvrage, il prend l'exemple du MPEG qui illustre bien la manière dont les différentes valeurs influencent le développement d'un standard et quels différents intérêts y sont représentés. Il évoque la question d'un compromis dans les décisions techniques en relation avec les politiques des pays, leurs cultures et les intérêts des grandes entreprises. Malgré tout, pour Jonathan Sterne, les entreprises qui participent volontairement à l'établissement d'un standard souhaitent imposer leur solution sur le marché et les gouvernements sont souvent impuissants face aux lois du marché.

« Lorsque des compagnies ou industries instaurent leurs standards, il n'y a pas la place pour un débat public. Les standards fixés par les sociétés tendent à une certaine forme de conservatisme étant donné que leur objectif est que les acteurs influents restent influents de manière à rendre les marchés émergents stables et prédictibles »⁽¹⁾.

La représentation de nombreuses entreprises au sein de groupes de standardisation traduit donc leur volonté de contrôler le marché suite à l'émergence d'un nouveau marché. Sterne établit un parallèle avec le meeting à Hanovre au sein du *Moving Picture Expert Group* en décembre 1988 pour établir un standard commun pour la compression des données audio numériques. De nombreuses sociétés étaient représentées de manière à maîtriser l'arrivée des équipements numériques et éviter la création de standards multiples concurrents qui auraient rendu le marché trop imprévisible. La création de standards propriétaires a été souvent décriée du fait qu'elle pousse les entreprises à établir des stratégies de communication de manière à garder le contrôle sur le marché. De plus, cela contraint des ingénieurs à agir comme consultants pour vanter les mérites de leur solution plutôt que de se concentrer sur la recherche. Certains considèrent donc ces stratégies de marketing comme un gaspillage de ressources voire une entrave au progrès⁽²⁾. Sterne se place donc en fervent défenseur de politiques internationales sur la standardisation et pour un regain d'influence des instances de normalisation comme l'ISO (International Organization for Standardization) ou l'ITU (International Telecommunications Union) tant que les grandes sociétés interviendront dans le processus de normalisation avec pour seul objectif d'asseoir leur domination sur le marché.

Aujourd'hui les standards propriétaires peuvent donc être amenés à dominer le marché aussi facilement que des standards élaborés par des organisations dédiées. L'implémentation d'un codec pour l'audio orienté objet est soumise aux mêmes problématiques. L'ADM a été standardisé pour la production mais DTS, Fraunhofer et Dolby se concurrencent pour s'imposer sur le marché de la restitution. Leurs standards respectifs (DTS:X, MPEG-H et Dolby AC4 *Atmos*) se développent en parallèle. Les codecs libres comme celui de l'EBU pourraient également surprendre par leurs compétences et semer le trouble au sein de la compétition.

MPEG-H s'impose en Corée du Sud avec les Jeux Olympiques d'hiver puis au Japon où le MPEG-H a été choisi pour la télévision numérique, ce qui a lancé la production en Audio orienté objet. Les constructeurs de téléviseurs japonais peuvent désormais intégrer le MPEG-H dans leurs produits. Dolby se développe en Amérique du Nord avec l'AC4 *Atmos* choisi pour

(1) STERNE Jonathan, MP3 : The Meaning of a format, 2012, 359 pages, page 130 – Traduit de l'Anglais

(2) STERNE Jonathan, MP3 : The Meaning of a format, 2012, 359 pages, page 133

l'ATSC 3.0, soit les nouveaux services de télévision numérique aux Etats-Unis. En revanche, en Europe, les instances de normalisation n'a pas encore tranché sur l'un des standards. Des discussions tendues sont en cours afin de choisir un standard pour la TNT. Chaque constructeur cherche alors des arguments pour faire pression auprès de la Commission Européenne qui choisira selon différents facteurs dont la qualité. Cette concurrence ralentit le déploiement de l'audio orienté objet mais elle pousse aussi au progrès car chaque constructeur cherche à offrir le codec le plus performant pour s'imposer sur le marché européen. Les instances de normalisation et de régulation attendent donc de voir quel format va s'imposer avant de faire un choix. Par conséquent, aucun diffuseur européen ne peut se décider à investir sans que les autorités de régulations aient choisi le format de diffusion, ce qui retarde la démocratisation de la production objet en Europe. Une grande majorité des professionnels se remémore la concurrence entre la VHS et le Betamax pour la cassette vidéo et espèrent que le codec le plus performant s'imposera et non pas celui dont le constructeur a le plus gros poids politique ou marketing.

Mais avant que ne s'impose le moteur de rendu dans les domaines de l'audiovisuel, il va bien sûr falloir évangéliser les professionnels du son et de la diffusion, mais aussi apprendre à maîtriser ce concept hérité du jeu vidéo.

2) L'EVOLUTION DES METIERS DE L'AUDIOVISUEL POUR ASSUMER LES NOUVELLES METHODES DE PRODUCTION

« A terme on pourrait imaginer que ce type de moteur [de rendu] soit ainsi utilisé, mais la question n'est pas que technique, elle est aussi socio-économique. Comment mettre en place une économie et une culture de production en mode objet, qui fasse qu'on puisse diffuser ce type d'outils ? Il faut que le public accroche à ce genre d'écoute, qu'on développe des standards correspondants et souhaiter que la concurrence entre fabricants ne bloque pas cette dynamique. » **Jean-Christophe Messonnier**

Par cette affirmation, Jean-Christophe Messonnier entend qu'il ne faut pas se limiter à une approche techno-centrée et que le déploiement de l'audio orienté objet est soumis au contexte socio-économico-culturel. En effet, l'implémentation du modèle objet dans les chaînes de production impliquerait un investissement technique avec le renouvellement du workflow, mais pour se faire il nécessite d'instaurer au préalable une économie et une culture du mode objet à la fois auprès des producteurs mais aussi auprès des utilisateurs. Les possibilités

créatrices du mode objet sont tout autant restreintes par des contraintes économiques que par les conventions et les différentes idéologies de pratiques des professionnels

a) LA MISE EN PLACE D'UNE ECONOMIE ET D'UNE CULTURE DE PRODUCTION EN
MODE OBJET

« *On est techniquement capable de produire en mode objet mais les moyens de travail et les conditions d'écoute du public ne sont pas au rendez-vous.* » **Xavier Gibert**

Xavier Gibert, Responsable du RFI Labo, confiait dans une interview que pour l'instant les contenus en son immersif correspondait plus à une vitrine technologique pour les diffuseurs qu'à une véritable volonté de leur part d'investir dans l'esthétique sonore. La plupart des diffuseurs ne possèderaient pas cette culture de la qualité et c'est plutôt l'aspect économique qui les pousseraient à amorcer un changement vers l'audio orienté objet. On peut expliquer que l'aspect qualitatif ne préoccupe que très peu les productions. En effet, les diffuseurs publics sont réticents à investir beaucoup de moyens pour produire du contenu spatialisé que les spectateurs écouteront dans de mauvaises conditions. De plus, dans la production audiovisuelle actuelle, l'audio est considéré comme secondaire par rapport à la vidéo du fait que les progrès réalisés sont plus visibles.

De manière à impliquer les diffuseurs et les créateurs de contenu dans l'évolution le mode de production dans le sens de l'audio orienté objet, il est de mise d'instaurer une culture du mode objet propre au contexte socio-économico-culturel actuel. Aujourd'hui, par la multiplicité des supports, des usages et des modes de consommations, il n'est plus possible de se contenter d'un mixage unique pour tout le monde. La multiplicité des formats pour une seule production constituait une solution à l'évolution des modes de consommation. Cependant chaque nouveau format impose aux chaînes de télévision, aux prestataires techniques, aux sociétés de post-production, d'archivages, l'investissement dans de nouveaux équipements, des installations avec remplacements d'ancien matériel, des formations du personnel. De plus, multiplier les mixages est un travail considérable pour les producteurs. L'audio orienté objet permettrait ainsi d'alléger le travail des intervenants en fournissant un mixage unique invariable quel que soit le format ou le système de restitution. Actuellement, pour les producteurs, les aspects d'accessibilité et d'interaction avec le contenu suscitent donc plus d'intérêt que la possibilité de fournir des contenus audio immersifs.

Pour initier cette transition, il convient donc de cibler les attentes des producteurs et les nouvelles expériences qu'ils souhaiteraient intégrer dans leurs programmes. Auparavant, aucun

mixage ne pouvait s'adapter de lui-même, l'ingénieur du son essaie d'optimiser dynamique, spectre et réverbération au mixage puis au mastering. Dorénavant, la possibilité de personnaliser le mixage en fonction de ses préférences (proéminence des dialogues, choix de la langue...) est un aspect séduisant de l'approche objet. De plus, les aspects de non-linéarité des programmes s'adaptent complètement aux nouveaux usages et à une manière d'écouter la radio qui repose désormais sur les podcasts et les applications smartphones.

Une fois cette transition amorcée, le mode de production n'aurait pas d'inflation sur le coût de production d'un contenu et sur la durée de son mixage éliminant ainsi la phase des multiples rendus coûteux et chronophages. En effet, pour la production fiction ou documentaire le mixeur va seulement livrer la sortie des sous-groupes plutôt que le mix final sous différents formats. Ceci offrirait en plus de nouvelles perspectives comme fournir la possibilité à un téléspectateur de gérer la balance de niveau entre les différents Stems (Stem voix et un Stem Ambiances + Musiques + Effets par exemple). Pour une diffusion live, la notion de débit dynamique suscite un intérêt particulier pour des diffuseurs comme France Télévisions. L'objectif pour les diffuseurs serait de ne plus payer plus pour un débit non utilisé. Prenons l'exemple d'une chaîne de France Télévisions qui, pour un programme distribue actuellement un mixage VF, un mixage VO (ou VF si VO inexistante) et un mixage avec Audio Description. Ces trois mixages nécessitent un débit 384 kbits/s. Pour Matthieu Parmentier, 128 kbits/s seraient nécessaires car 90% du temps les deux autres mixages distribués sont réalisés à partir d'une copie de la VF. L'audio orienté objet offrirait donc la possibilité de déclarer seulement trois composantes et ainsi d'économiser du débit de manière à transmettre de l'audio 3D par exemple. De plus, la version Audio Description ne serait plus mixée mais la voix commentaire serait simplement définie en tant qu'objet et ajouté à la VF avec des métadonnées de mixage. De cette manière, un spectateur malvoyant pourrait alors augmenter le volume du commentaire voire le déplacer dans l'espace sonore pour éviter d'éventuels conflits avec les voix dialoguées.

Pour Hervé Déjardin, l'audio orienté objet serait également un moyen formidable de stimuler les productions européennes. Le mode objet offrirait la possibilité de produire en plusieurs langues le même programme et de laisser l'auditeur manipuler les objets correspondants. Hervé Déjardin évoque le potentiel de banques de contenus en format objet (régies par des droits) qui permettrait de valoriser le contenu d'une production en ayant la possibilité d'extraire tout ou partie du contenu originel. Prenons l'exemple d'un programme de fiction créé en langue française, la BBC aurait la possibilité de récupérer la production à partir d'une banque d'archives de productions européennes et de ne payer uniquement que pour

enregistrer la voix anglaise. En effet, en développant des productions européennes basées sur un cercle vertueux de collaboration il deviendrait plus aisé de donner de multiples vies à une production collaborative.

Au-delà de l'aspect de production, cette culture du mode objet doit indéniablement passer par l'utilisateur. Le saut qualitatif sera-t-il suffisant et les nouvelles expériences seront-elles assez concluantes pour convaincre le public ? Nous nous méfions souvent du caractère gadget des nouveaux standards de diffusions, ce qui peut légitimement créer des réticences chez les diffuseurs. La production en VR a d'ailleurs été victime de cette réticence et s'est transformé en phénomène de « hype ». Il est donc nécessaire de tester ces nouveaux dispositifs pour étudier et établir des hypothèses sur la demande du public.

Matthieu Parmentier rassemble régulièrement des groupes appelés « Focus Group » dans le but de tester les nouveaux dispositifs. Une phase de test s'est principalement concentrée sur l'intelligibilité des dialogues dans une scène objet avec la possibilité de modifier le rapport entre les dialogues et les autres éléments sonores. Ce genre de démarches a également été conduit sur les productions de la BBC et notamment une évaluation par des auditeurs de la fiction radiophonique *The Mermaid's Tears*, la première fiction radiophonique interactive en direct au monde intégralement réalisée à partir d'audio orienté objet. Les résultats de cette évaluation ont révélé que 88% des sondés étaient favorables au développement de ce type de programmes. Pour Hervé Déjardin, le succès de l'approche objet auprès des utilisateurs sera issu de l'immersion par la perception et/ou par la narration. Pour cela, il y a des enseignements à tirer du jeu vidéo qui propose déjà ce type d'interactions ludiques. De la même manière, le développement à moindre coût de nouveaux dispositifs d'écoute en son 3D toujours plus performants, du casque aux barres de sons, contribuera à la démocratisation de l'audio orienté objet.

L'implémentation de l'audio orienté objet aura un impact fort sur les manières de produire et donc sur l'évolution des métiers au sein de la chaîne de production. La spatialisation des éléments sonores influence la manière d'appréhender la création sonore ce qui fait varier les frontières entre les corps de métiers. Ce fut déjà le cas à la fin des années 1980, à l'arrivée du mixage multicanal dans la chaîne de post-production au cinéma. Il en a résulté la création du métier de monteur son. Avec l'arrivée de l'audio orientée objet, les métiers du son sont amenés à évoluer notamment avec la gestion des métadonnées objet.

b) DE NOUVELLES METHODES DE TRAVAIL

Suite aux travaux du consortium ORPHEUS, le nouveau workflow de la production en mode objet semble se dessiner. L'ADM est amené à devenir le premier format de production, d'archivage et d'échange en amont de l'encodage Dolby ou MPEG-H intégré aux stations de travail audionumériques. Les futures scènes objet seraient basées sur un bed d'ambiance HOA ou multicanal auxquels seraient ajoutés des objets.

Comme évoqué précédemment, la création d'une scène objet implique la maîtrise de nouveaux dispositifs de captations ainsi que des choix inédits en terme de prise de son. François-Xavier Ballon, ingénieur du son et mixeur pour France TV affirmait que les premiers ingénieurs du son qui ont dû travailler en stéréo se posaient les mêmes questions avec l'utilisation des couples et que l'évolution des dispositifs de prise de son allait logiquement faire évoluer le métier de chef-opérateur du son. Pour Jean-Christophe Messonnier, l'important est de pouvoir faire un enregistrement qui puisse remplacer le format stéréo de manière vraiment satisfaisante tout en restant compatible avec les standards de diffusion actuels. Le dispositif de prise de son doit rester compatible de manière à ce que lorsqu'on crée une scène 3D, celle-ci reste satisfaisante sur un système 5.1 ou stéréo. Cela passerait donc par une évolution des systèmes de captation avec un ensemble de rampes de micros acoustiques couplés avec des réverbérations 3D.

L'évolution des méthodes de travail se répercute sur le mixage avec l'arrivée de nouveaux outils d'ici quelques mois. Pour la partie software, la transition se ferait simplement avec des mises à jour des stations de travail de manière à intégrer l'approche objet. Cette transition a déjà commencé avec l'avènement des stations de travail adaptées à la production en mode objet comme Pyramix 10 qui met à disposition des mixeurs des bus pour les beds en 7.4, des bus « objet » et un panoramique 3D pour faciliter leur spatialisation. Le bus « objet » est un bus mono indépendant du mix principal qui contrôle des flux audio et des métadonnées. Il est donc possible de monitorer le mix avec ou sans les objets. Le principe est d'aller chercher dans ce bus mono les métadonnées de panoramique et de volume qu'il applique au moment du rendu. De même, Pro Tools a signé un contrat avec Dolby qui permet d'intégrer l'*Atmos Production Pack* qui utilise les panoramiques 3D de Pro Tools. Dolby gère donc à la fois les bus (« objets ») et les beds via un logiciel externe qui dialogue avec Pro Tools. On peut également citer Nuendo qui a déjà intégré le moteur *Atmos* et devrait bientôt intégrer l'ADM. En revanche, la partie hardware s'annonce plus compliquée avec les enjeux relatifs à l'intégration et au design des interfaces physique pour les mixeurs.

Pour assurer une rétrocompatibilité de la scène 3D il sera assurément nécessaire de faire des compromis au sein du mixage notamment pour optimiser l'intelligibilité. En effet, il est difficile de croire qu'un downmix stéréo puisse sonner comme un mixage appréhendé en stéréo. L'audio orienté objet va donc contraindre les ingénieurs du son à optimiser et vérifier leurs mixages sur des formats descendants pour acquérir de l'expérience et du savoir-faire et apprendre à maîtriser le compromis. Pour anticiper le changement de paradigme, les ingénieurs du son seront forcés de sortir de leur zone de confort et de revoir leurs habitudes. Effectivement, une part de la compatibilité des formats est faite par les moteurs de rendu mais l'autre part sera portée par le changement de manière de travailler des ingénieurs du son que ce soit lors de la prise de son ou du mixage. Des répercussions seraient donc à prévoir sur le métier de chef-opérateur du son.

Comme évoqué auparavant, l'anticipation de la post-production d'une scène objet se réalise dès l'enregistrement. Dans ce but, le chef-opérateur du son serait amené à gérer des objets en direct et à réaliser un conditionnement des sons par différentes couches de traitements en prévision du mixage. On est amené à penser que l'approche orienté objet reportera beaucoup de travail en post-production cependant les objets sonores gérés en direct offrirait une grande souplesse à l'ingénieur du son dans la conception de sa scène objet. Ainsi, il peut être amené à envoyer des flux audio, à exclure du contenu, en rajouter, en substituer ou modifier les caractéristiques des objets en direct. Il est donc parfaitement envisageable d'enregistrer un multipiste tout en créant simultanément des mouvements d'objets. Il serait ensuite possible de le relire par la suite, de le « bouncer » dans plusieurs formats ou de le redistribuer via MPEG-H ou AC4 *Atmos*.

La post-production serait, elle aussi, destinée à subir des bouleversements majeurs et notamment concernant le monitoring de ces scènes 3D. Le mixage de ces scènes objet est pour l'instant destiné à une définition maximale de 15 canaux soit 7.1.4 + 4 objets cependant il serait possible de réaliser une pré-production dans un bed 7.1.4+4 virtuel.

Jean-Christophe Messonnier évoquait une pré-production en Dolby *Atmos* sur un système virtuel dans l'optique d'appréhender un mixage pour le cinéma. Pour lui, ces nouvelles méthodes offrent de nombreux intérêts en exploitant des artifices propres à chacun. Il serait donc envisageable de travailler en 7.1.4 avec un système 5.1 dans la mesure où une réduction du 7+4 sur un système 5.1 est possible. Ceci tout en prévoyant tout de même de sortir son mix en 7+4 quitte à avoir un moteur de rendu binaural pour monitorer les effets d'élévation et penser

les trajectoires en mode objet. Il existerait donc plusieurs types de monitoring qui permettraient de définir la scène qui sortira finalement indépendante du système de restitution. Pour Jean-Christophe Messonnier le mode de réalisation en binaural est un mode très économique qui permet de rendre des scènes complexes avec seulement un moteur de rendu, un headtracker et un casque. Ainsi, il serait possible de commencer à travailler des scènes objet en binaural avant de les monitorer sur des systèmes de restitution plus imposants. La post-production peut donc être anticipée dans des configurations plus économiques comme le binaural à l'heure où les auditoriums de mixage sont de plus en plus chers à louer.

Enfin, on peut se poser la question d'une éventuelle répercussion sur la stratégie de mise en scène sonore. L'avènement de l'audio orienté objet peut-il être à l'origine d'un basculement du mode stratégique de la mise en scène ? Cette stratégie se faisait auparavant par des acteurs et à des temps différents et elle tend à se concentrer sur un seul acteur en une synthèse ultérieure, lors de la post-production. Il s'avère pourtant indispensable d'anticiper la mise en scène sonore et la composition des scènes sonores objet dès la prise de son. En fait, cette recherche de scénographie sonore devrait idéalement se constituer avec tous les acteurs de la chaîne de production son de manière à penser en amont les possibilités techniques et esthétiques apportées par le mode objet.

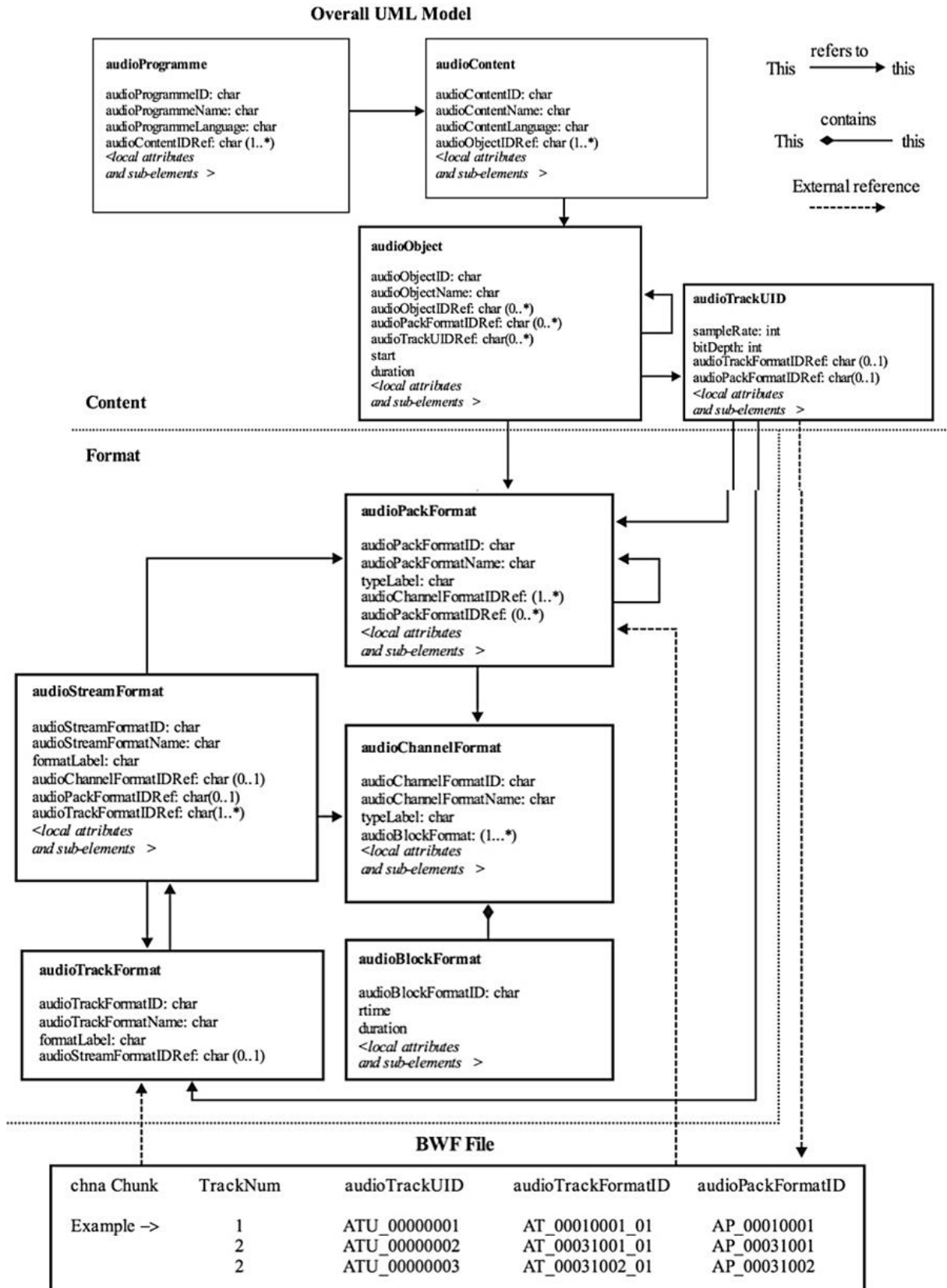
CONCLUSION

L'approche de l'audio orienté objet est déjà inventée. Elle s'inscrit dans un contexte de production où la multiplicité croissante des supports, des usages et des modes de consommations de contenus audio est propice à l'innovation. Par l'invariabilité des scènes sonores et les nouvelles expériences utilisateurs qu'elle propose, l'approche objet semble convenir pour résoudre les problèmes de compatibilité entre les formats tout en fournissant à l'auditeur un contenu plus immersif et personnalisable. De par son concept disruptif, les professionnels du son ont été amenés à penser une nouvelle chaîne de production intégrant à la fois les problématiques inhérentes à la production de contenu, à sa distribution et à sa réception par l'auditeur. Au-delà de l'aspect technique, le modèle objet devra se confronter à des obstacles d'ordres économiques et culturels pour s'implémenter dans les infrastructures de production et de diffusion.

D'une manière générale, on peut constater que l'innovation sonore se complexifie par l'interrelation nouvelle entre les technologies développées (dispositifs de prise de son, outils de mixages, distribution web, ...), par la diversité des acteurs qui interviennent dans le processus de développement, par l'articulation encore plus forte avec les enjeux économiques et par le rôle éventuellement secondaire de la dimension artistique. L'intérêt esthétique n'est plus un aspect essentiel dans le processus d'innovation et les standards d'aujourd'hui sont issus de compromis, de négociations entre les préoccupations techniques, industrielles et esthétiques. L'anticipation stratégique devient donc plus aléatoire, plus éthérée, plus difficile à envisager pour les chercheurs et pour les professionnels de l'audiovisuel. C'est l'heure où la recherche et la production audio sont à un point de bascule imminent de leur histoire, et aucun acteur n'a réellement le pouvoir, seul, de porter les décisions qui bouleverseront l'économie et la culture de l'industrie audiovisuelle. Le futur se recompose dès à présent sous nos yeux et il va falloir savoir distinguer les événements décisifs qui vont structurer l'avenir.

ANNEXES

ANNEXE 1 : Schéma de référence de l'ADM – Source : Audio Definition Model_Recommandation ITU-R BS.2076-1_06/2017



La partie format des métadonnées se compose des éléments suivants : l'élément *audioTrackFormat* est utilisé pour décrire le format de l'audio. C'est un élément indispensable pour décoder correctement de l'audio codé au préalable (Non PCM). Il se réfère directement à l'élément *audioStreamFormat* qui va décoder un flux de données directement à partir d'une combinaison de pistes (ou une piste) afin de restituer un canal, un objet ou un HOA. L'élément *audioChannelFormat* représente une séquence unique d'échantillons audio sur lesquels certaines actions peuvent être effectuées, telles que le mouvement d'un objet, qui sera rendu dans une scène. L'*audioChannelFormat* est subdivisé dans le domaine temporel en un ou plusieurs *audioBlockFormats* qui représentent une séquence d'échantillons d'une durée fixe et dont les paramètres instanciés, y compris la position, sont fixes. C'est à partir du *Type definition* issu de l'*audioChannelFormat* qu'il est possible de décrire l'approche dans laquelle on souhaite travailler (DirectSpeaker, Matrix, HOA, objet ou binaural) et c'est ce qui déterminera les paramètres qui seront accessibles dans les *audioBlockFormats*. Chaque nouvel *audioBlockFormat* se succède dans et transporte alors des paramètres susceptibles d'être modifiées en temps réel (gain, phase, delay, etc ...). Parmi elles demeurent les coordonnées spatiales (cartésiennes ou polaires) d'une piste dans l'espace sonore 3D ce qui permet de définir en temps réel la position de l'audio associé.

```

7518 </audioBlockFormat>
7519 <audioBlockFormat audioBlockFormatID="AB_00031006_0000014b" rtime="00:00:44.20266" duration="00:00:00.04266">
7520 <gain>1</gain>
7521 <color>0.741 0.443 0.412 1.000</color>
7522 <position coordinate="X">-0.0479936</position>
7523 <position coordinate="Y">-0.0715853</position>
7524 <position coordinate="Z">0</position>
7525 <jumpPosition interpolationLength="0.03000">1</jumpPosition>
7526 <cartesian>1</cartesian>
7527 <audioObjectInteraction onOffInteract="1" gainInteract="1">1</audioObjectInteraction>
7528 </audioBlockFormat>
7529 <audioBlockFormat audioBlockFormatID="AB_00031006_0000014c" rtime="00:00:44.24533" duration="00:00:00.04266">
7530 <gain>1</gain>
7531 <color>0.741 0.443 0.412 1.000</color>
7532 <position coordinate="X">-0.0469358</position>
7533 <position coordinate="Y">-0.07231</position>
7534 <position coordinate="Z">0</position>
7535 <jumpPosition interpolationLength="0.03000">1</jumpPosition>
7536 <cartesian>1</cartesian>
7537 <audioObjectInteraction onOffInteract="1" gainInteract="1">1</audioObjectInteraction>

```

Exemple d'une succession d'audioBlockFormats issu d'une scène objet

On observe ici, qu'entre ces deux *audioBlockFormats* la position de l'objet concerné varie sur les axes X et Y soit en azimuth et distance dans un système de coordonnées polaire. Dans le cas où l'objet considéré est en mouvement, chaque *audioBlockFormat* du fichier ADM possède une valeur différente des coordonnées de sa position qui dépendra de l'amplitude et de la rapidité du mouvement. Ces valeurs sont ensuite interpolées selon un facteur défini au préalable par le sous-élément *jumpPosition* et son attribut *interpolationLength* réglé sur 0.03 dans notre cas. Enfin, l'élément *audioPackFormat* permet de regrouper un ou plusieurs *audioChannelFormats* qui vont ensemble dans une configuration stéréo ou 7.1 par exemple.

Toutes les métadonnées et sous-métadonnées propres à la caractérisation d'un objet sont détaillées dans la documentation *ITU-R BS.2076-1*.

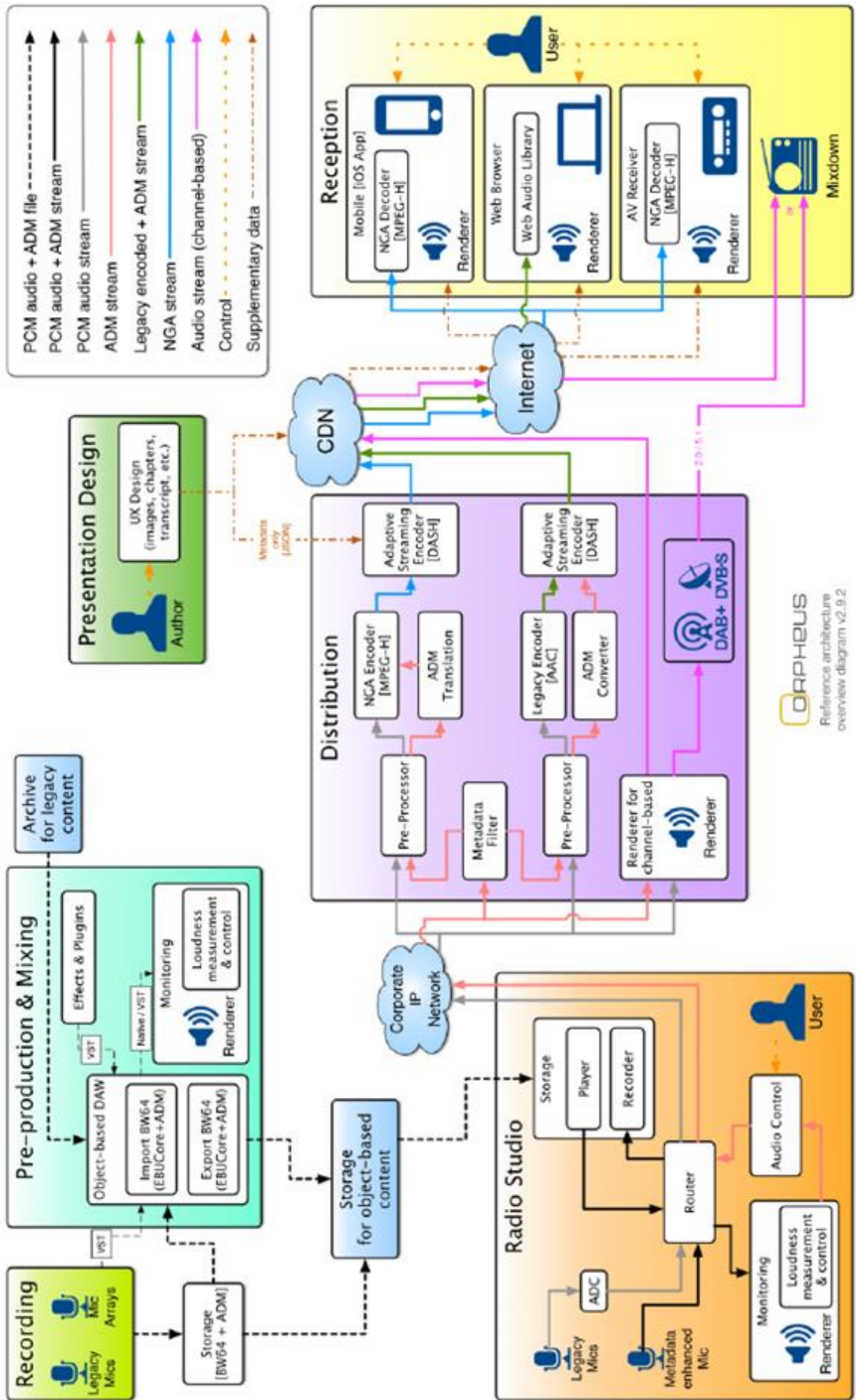
En ce qui concerne la partie contenu, l'élément *audioObject* établit la relation entre le contenu, le format via des packs audio et les ressources à l'aide des UID des pistes que nous aborderons par la suite. Les *audioObjects* peuvent être imbriqués et ainsi ils peuvent se référer à d'autres *audioObjects*. Les attributs et les sous-éléments inclus dans l'*audioObject* sont à l'origine des interactions avec l'auditeur. De cette manière, la position et le gain d'une source sonore résulte de la combinaison des attributs des sous-éléments de position et de gain de l'*audioBlockFormat* et de tous les changements causés par l'interaction induit par l'auditeur dans la hiérarchie des *objetsAudioObjects* qui se réfèrent à l'*audioBlockFormat*.

L'élément *audioContent* décrit le contenu d'une composante d'un programme (par exemple, dialogue, ambiance ou musique de fond). Il se réfère aux *audioObjects* pour lier le contenu à son format. Cet élément comprend un attribut qui intègre les métadonnées relatives au Loudness. L'élément *audioProgramme* se réfère à un ensemble d'un ou plusieurs *audioContents* qui sont combinés pour créer un programme audio complet. Il contient les timecodes de début et de fin du programme, qui peuvent être utilisés par exemple pour l'alignement avec le timecode d'une vidéo. Des métadonnées de Loudness sont également incluses pour permettre l'enregistrement de l'intensité sonore du programme.

Enfin l'*audioTrackUID* définit simplement une piste ou une ressource dans un fichier ou un enregistrement d'une scène audio. Cet élément contient les informations sur le format et la fréquence d'échantillonnage de la piste. Il contient également des sous-éléments qui permettent à des fichiers non BW64 d'accéder aux métadonnées de l'ADM via un nouveau RIFF chunk appelé <chna> et qui garantit l'universalité du modèle.

ANNEXE 2 Architecture de référence de l'ORPHEUS pour un Workflow intégralement orienté objet

Source : EBU Technical Report 042_ORPHEUS Project_05/2018



BIBLIOGRAPHIE

- a) Recommendation ITU-R BS.2076-1, *Audio Definition Model*, 06/2017, 106 pages.
- b) Ircam, CNRS, STMS, *ADMix Quick Start Manual*, 2017, 18 pages.
- c) ENGDEGARD Jonas, RESCH Barbara, FALCH Cornelia, HELLMUTH Oliver, HILPERT Johannes, HOELZER Andreas, TERENCEV Leonid, BREEBAART Jeroen, KOPPENS Jeroen, SCHUIJERS Erik, OOMEN Werner, *Spatial Audio Object Coding (SAOC) – Upcoming MPEG Standard on Parametric Object Based Audio Coding*, AES, 15 pages.
- d) HERRE Jürgen, HILPERT Johannes, KUNTZ Achim, PLOGSTIES Jan, *AES137 MPEG-H Audio – The New Standard for Universal Spatial / 3D Audio Coding*, Fraunhofer IIS, 12 pages.
- e) *Dolby-Atmos-Next-Generation-Audio-for-Cinema-White-Paper*, Dolby, 18 pages.
- f) EBU Tech3388, *ADM Renderer for use in Next Generation Audio Broadcasting*, BTF Renderer Group, 03/2018, 57 pages.
- g) EBU Technical Report 042, *Example of an end-to-end OBA broadcast architecture and workflow*, ORPHEUS Project, 05/2018, 35 pages.
- h) EBU Technical Report 045, *Why Broadcasters need an open, codec-independent workflow for NGA production deployment*, AS, 01/2019, 13 pages.
- i) BLEIDT Robert, *Introducing the MPEG-H Audio Alliance’s New Interactive and Immersive TV Audio System*, Presentation at the HPA Tech Retreat, Palm Springs, Fraunhofer IIS, 02/2015, 29 pages.
- j) PLOGSTIES Jan, *MPEG-H Audio System for Broadcasting*, ITU-R Workshop “Topics on the Future of Audio in Broadcasting”, Fraunhofer IIS, 2015, 17 pages.
- k) HASSLER Markus, *The future of illustrated sound in programme making*, Westdeutscher Rundfunk Koeln, 15/07/2015, 34 pages.
- l) EBU Tech3392, *ADM Broadcast production profile*, version 0.9, BTF Renderer Group, 10/2018, 26 pages.
- m) SILZLE Andreas, *The Future of Audio is Object-based, Results of ORPHEUS*, Final ORPHEUS Workshop, IRT/BR, Munich, 15/05/2018, 28 pages.
- n) WEITNAUER Michael, *The ORPHEUS Reference Architecture*, Final ORPHEUS Workshop, 15/05/2018, 18 pages.
- o) FUCHS Harald, *MPEG-H Audio The new TV system for Personalized and Immersive Audio*, Final ORPHEUS Workshop, Fraunhofer IIS, 15/05/2018, 31 pages.

- p) HÔ Lidwine Michael, *ORPHEUS Workshop Presentation*, Final ORPHEUS Workshop, France Télévisions, 15/05/2018, 19 pages.
- q) VOPEL Marius, *Object-based and variable length production in Sequoia*, Final ORPHEUS Workshop, MAGIX, 15/05/2018, 17 pages.
- r) BAUME Chris, *The Mermaid's Tears: Creating the world's first live interactive object-based radio drama*, Final ORPHEUS Workshop, BBC, 15/05/2018, 19 pages.
- s) BLEISTEINER Werner, *Approaches to make programs variable in length and depth*, Final ORPHEUS Workshop, BR, 15/05/2018, 22 pages.
- t) BOGAARDS Niels, *Design and Implementation of a mobile object-based audio radio player*, Final ORPHEUS Workshop, elephantcandy, 15/05/2018, 16 pages.
- u) COX Trevor, *Media Device Orchestration (MDO)*, Final ORPHEUS Workshop, S³A Future Spatial Audio, University of Salford, 15/05/2018, 27 pages.
- v) PERIAUX Bergame, OHL Jean-Luc, THEVENOT Patrick, *Le Son Multicanal*, Dunod, 2015, 448 pages.
- w) STERNE Jonathan, *MP3 : The Meaning of a format*, 2012, 359 pages.
- x) DENIZOT Florent, *L'approche orientée objet au cinéma : entre surenchère technologique et outil de création sonore*, 2016, 214 pages.
- y) MESSONNIER Jean-Christophe, *L'audio orienté objet et la scène sonore*, associationdesmixeurs.fr, 2017, 4 pages.
- z) MESSONNIER Jean-Christophe, LYZWA Jean-Marc, DEVALLEZ Delphine; DE BOISHERAUD Catherine, *Object-Based Audio Recording Methods*, AES E-Library, 2016
- aa) STEFANI Benoît, *L'audio face à la 4K*, Mediakwest.com, 2015

NETOGRAPHIE

- (1) tvtechnology.com (15/12 /2017) : TVTechnology rassemble toutes les innovations dans le monde du broadcast, de la télévision numérique et des nouveaux médias.
- (2) Ircam.fr (10/01/2019) : L'IRCAM est un institut de recherche, de production mais aussi lieu de formation, de consultation et d'expérimentation, avec la volonté d'une interaction entre musique et scientifiques.
- (3) Orpheus-audio.eu (22/06/2018) : ORPHEUS est un projet de recherche européen dédié à l'amélioration de la gestion du contenu audio. Ils sont à l'origine de l'approche des médias orientés objet.

- (4) Mpeghaa.com (22/06/2018) : Le système audio MPEG-H offre un son immersif non compressé en 7+4. Il est conçu pour fonctionner dans les systèmes de diffusion existants et futurs, de la contribution à la transmission.
- (5) Itu.int (10/01/2019) : L'ITU est l'institution spécialisée des Nations Unies pour les technologies de l'information et de la communication (TIC).
- (6) Ibc.org (19/02/2019) : IBC 365 est une plate-forme communautaire en ligne fournissant des connaissances commerciales pour l'industrie mondiale des médias, du divertissement et de la technologie.

FILMOGRAPHIE

- (1) *The Mermaids Tears*, Jessica MITIC et Chris BAUME, 2017 (Fiction Radiophonique)

CONTACTS

François-Xavier Ballon – Ingénieur du Son, Mixeur à France Télévisions

Thibaut Carpentier - ingénieur CNRS-STMS

Hervé Déjardin – Responsable R&D Radio France

Benjamin Duval – Ingénieur R&D Trinnov

Nicolas Epain - Ingénieur de recherche à l'IRT b<>com

Xavier Gibert - Responsable Technique du RFI Labo

Jean-Christophe Messonnier – Ingénieur du son, Mixeur, Enseignant-Chercheur au Conservatoire National Supérieur de Musique et de Danse de Paris.

David Paillet - Chef Opérateur Son AMP Visual TV

Matthieu Parmentier – Responsable Recherche et Développement de France Télévisions

Olivier Warusfel – Chercheur IRCAM-STMS

RESUME :

L'approche de l'audio orienté objet semble être une solution efficace aux problématiques liées à la multiplicité des livrables. Celle-ci libère de la notion de canal et va reposer sur la gestion d'objets sonores. Elle permet l'adaptation du contenu audio à différents dispositifs de restitution et lieux de diffusion mais son implémentation dans la chaîne de production implique une révolution des outils et de méthodes de travail.

MOTS CLES :

Audio Orienté Objet, NGA, Multicanal, Multiplication, Standards d'écoute, Systèmes de restitution, Chaîne de production objet, Compatibilité, Standardisation

ABSTRACT :

Object based audio seems to be an efficient solution to solve multiplicity of the formats issues. This one releases from the channel notion and relies on sound objects handling. It allows the versatility of an audio content with different restitution setups and listening contexts but its implementation in production workflow involves a turmoil in tools and working methods.

KEY WORDS :

Object Based Audio, NGA, Multichannel, Multiplicity, Broadcasting standards, Restitution setups, Object based workflow, Versatility, Standardisation