

Méthodes de prise de son en multicanal / Objets

Jean-Christophe Messonnier, Jean-Marc Lyzwa et Alexis Ling,
Conservatoire National Supérieur de Musique et de Danse de Paris

Résumé

Depuis les années 1990, le service audiovisuel du Conservatoire de Paris travaille sur les systèmes de captation et de restitution de l'espace sonore dédiés à l'enregistrement de la musique. Ces recherches ont abouti à des méthodes adaptées à l'audio 3D, comme l'intégration du mode multicanal/objet, le contrôle du mixage via un rendu binaural avec headtracker, ou des techniques transaurales. En parallèle, nous avons pratiqué sur de nombreuses productions des systèmes de captation permettant de post-produire sous ces formats. L'évolution récente des moyens de diffusion audio et audiovisuelle comme le développement du format Dolby Atmos pour la musique donnant une actualité à ces méthodes, il a semblé opportun de faire un point sur celles-ci, en décrivant les différentes raisons de cette démarche.

Abstract

Since the 1990s, the audiovisual department of the Conservatoire de Paris has been working on music specific systems for recording and diffusing the sound space. This research has resulted in methods particular to 3D audio, such as the integration of multichannel/object mode, mixing control via binaural rendering with headtracker, or transaural techniques. At the same time, we have put to test, on many productions, capture systems allowing post-production in these formats. The recent evolution of the means of audio and audiovisual diffusion, such as the development of the Dolby Atmos format for music, gives a topicality to these methods, and it seemed appropriate to take stock of them, by describing the different reasons for this approach.

Depuis les années 1990, le service audiovisuel du Conservatoire de Paris travaille sur les systèmes de captation et de restitution de l'espace sonore. Ses recherches ont conduit à la mise en œuvre de techniques de prise de son et de post-production permettant d'améliorer, de préciser et de stabiliser la sensation d'enveloppement et d'immersion tout en renforçant la lisibilité et la compréhension de l'œuvre restituée. Travailler la prise de son et la post-production musicale, c'est attacher une importance toute particulière à la restitution spectrale, à la dynamique, à l'espace et à son ressenti, à la définition des sources pour une bonne intelligibilité et compréhensibilité des différents instruments ou groupes instrumentaux constituant l'œuvre. L'industrie cinématographique a très largement contribué aux développements et à la popularisation des techniques de restitutions spatiales et multicanales sans toutefois définir de canon esthétique. Après avoir expérimenté en multicanal depuis l'avènement du 5.1 et intégré le groupe Bili (<http://www.bili-project.org/le-projet/>) consacré

au binaural, nous pratiquons depuis de nombreuses années des prises de son musicales, diffusées sur le site du Conservatoire de Paris, intégrant ces techniques.

Cet article propose une synthèse de notre démarche. Nous avons développé plusieurs outils (moteurs de rendu transauraux et binauraux avec headtracker, base de données de HRTF) nous permettant de progresser dans la qualité de nos productions et nous décrivons leur utilisation, ainsi que la manière dont nous intégrons les techniques classiques d'enregistrement.

Stéréophonie à deux canaux et multicanal

La stéréophonie à deux canaux permet, à partir de deux enceintes disposées à $\pm 30^\circ$ et d'une position précise de l'auditeur, de donner à percevoir des sources fantômes en continu dans l'espace, d'une enceinte à l'autre, en fonction de la différence d'informations entre les deux oreilles perçue par l'auditeur (différence d'intensité et différence de temps). Pour plus d'informations sur ce phénomène, on peut se référer à Blauert (1983) et Plenge & Theile (1987). Elle permet également de percevoir les sources en profondeur, mais cette propriété est conservée en multicanal et elle ne constitue donc pas une différence entre les deux formats. On peut optimiser le signal à deux canaux, à la prise et en post-production. Cette attitude est largement dominante en musique, à la fois avec l'emploi d'un système principal microphonique à deux canaux par beaucoup d'ingénieurs du son en musique classique, et par la pratique du *mastering* sur le signal à deux canaux. Plus généralement, le *monitoring* de référence est dans la plupart des cas un système à deux canaux, sur enceintes ou au casque.

Cependant, dans la production du son pour le cinéma, un autre modèle a émergé et est maintenant largement dominant. Le signal *master* est ainsi constitué d'un ensemble de signaux décrivant la scène sonore, s'adaptant à toute une gamme de systèmes, allant de la très grande salle de cinéma équipée de dizaines d'enceintes, à l'écoute domestique pour laquelle un *downmix* deux canaux est réalisé. Nous reviendrons sur les caractéristiques de ces signaux, mais nous devons auparavant poser le débat.

Premièrement, le modèle de la production cinéma est-il exportable en partie pour la musique ? Y-a-t-il un intérêt à pratiquer le son multicanal pour la musique ? La question mérite actuellement d'être posée, car l'industrie musicale a commencé récemment à produire au format Dolby Atmos, et que la société Apple transforme ce format en binaural avec headtracker vers ses écouteurs nomades, afin que l'auditeur perçoive la scène sonore comme étant fixe, lorsqu'il bouge la tête. Nul ne connaît l'avenir socio-économique de ces systèmes lorsqu'ils sont appliqués à la musique, mais ils impliqueraient alors une production en multicanal qui changerait radicalement les pratiques si elle s'imposait.

L'expérience que nous avons acquise depuis trois décennies dans le domaine de la captation et de la restitution en multicanal, mais également dans le développement d'outils spécifiques aux techniques en audio 3D, nous permet de donner un avis esthétique (pourquoi le faisons-nous ?) et technique (comment le faisons-nous ?). L'apport du multicanal est simple : il permet une immersion sonore beaucoup plus perceptible que la stéréophonie à deux canaux, l'immersion étant définie comme la sensation d'être à l'intérieur de l'espace restitué plutôt que devant cet espace. Cette différence d'immersion n'est pas forcément sensible dans des lieux d'écoute de petits volumes, où la réverbération de la salle crée l'essentiel de cette

immersion et où, par conséquent, l'apport du multicanal n'est pas décisif. En effet, le signal émis par les enceintes est diffusé à travers l'acoustique du lieu d'écoute et l'auditeur est alors en présence du son direct des enceintes (entre $\pm 30^\circ$), mais également du son réverbéré de la salle (en plus de celui diffusé par les enceintes sur 360°). La perception des plans sonores qui en résulte dépend donc de ces deux facteurs (Messonnier & Moraud, 2011). Mais, dès que le volume de la salle augmente assez pour que le niveau de la réverbération soit suffisamment bas à la position d'écoute, l'immersion est créée par le signal direct du système d'écoute et le multicanal est dans ce cas un atout décisif. Cette immersion va de pair avec une possibilité accrue d'analyse du champ acoustique due au démasquage binaural (Moore, 1989) et les systèmes multicanaux sont donc plus riches en quantité d'information perçue. L'immersion est liée à une utilisation de l'espace sonore qui peut être réalisée, soit par les sons directs, soit par l'ambiance, soit par la réverbération, et une infinité d'esthétiques peuvent être développées à partir de cette possibilité d'utiliser toutes les directions de l'espace sonore.

Le multicanal peut donc être, à notre sens, un atout. La question est donc la suivante : peut-on pratiquer la stéréophonie en multicanal pour la musique en ayant exactement la même qualité que celle de la stéréophonie à deux canaux ? Les systèmes de diffusion du son ont évolué. Indépendamment des dispositifs multicanaux, deux systèmes ont vu leur popularité exploser : l'enceinte Bluetooth avec un retour à la monophonie, et le casque audio pour la lecture de médias en mobilité. Pour les enceintes Bluetooth, la compatibilité mono est donc un enjeu. L'usage du casque ouvre aux techniques binaurales, avec ou sans headtracker, et cela nécessite un contenu multicanal. En parallèle avec la question précédente : continuer à produire comme avant n'est-il pas un problème ? Pour permettre cette présentation, nous devons tout d'abord rappeler quelques définitions.

Les modes de spatialité : multicanal ; mode objet ; ambisonics (Messonnier et al., 2016)

Comme décrit dans la norme EBU (2014) spécifiant les modes de spatialité pouvant être inscrits dans les signaux audio (Audio Definition Model, ADM), on dénombre au moins trois manières de spatialiser les sons. Ces données sont stockées dans la partie réservée aux metadata des fichiers Broadcast Wave File (BWF). Le même type de données est utilisé dans les systèmes propriétaires Dolby Atmos DTSX et MPEGH, permettant de plus de diffuser le signal à un débit raisonnable par des systèmes de compression de données. Ces trois modes de spatialité sont les suivants : le multicanal, le mode objet et l'ambisonie.

Le multicanal (Channel Based Audio), dont la stéréophonie à deux canaux fait partie, définit des canaux normalisés (L, R, C, etc.) correspondant à des directions de l'espace EBU, 2014). Ces canaux reçoivent des signaux répartis entre eux, soit par les systèmes de prise de son (un couple AB réparti par exemple le son entre L et R dans la stéréophonie à deux canaux), soit par les systèmes de post-production (potentiomètre panoramique). Si un signal n'est envoyé que sur un seul canal, ce signal se comporte comme un objet fixe. Un objet (Object Based Audio) est un signal défini par sa position spatiale, par exemple -15° azimut, 0° d'élévation, à un temps (time code) donné. Les positions des objets peuvent donc évoluer avec le time code et donc avec l'automatisation.

L'ambisonie (Scene Based Audio) est basée sur une analyse régulière de l'espace sonore avec une précision donnée par un paramètre appelé ordre (voir aussi ce numéro, Jullien, chapitre 5,

pp. XX-XX). Chaque ordre renvoie donc à un ensemble de signaux complémentaires entre eux et représentant l'espace sonore, indépendamment du système de monitoring (en son 3D cela implique un minimum de canaux au 1^{er} ordre : 4 canaux, 2^{ème} ordre : 9 canaux, 3^{ème} ordre : 16 canaux, 4^{ème} ordre : 25 canaux, 5^{ème} ordre : 36 canaux, 6^{ème} ordre : 49 canaux...)

Ces différents modes de spatialité sont combinables entre eux dans l'ADM (Audio Definition Model, cf. ci-dessus), mais les formats propriétaires n'admettent qu'une base multicanale (appelée *Bed*) et un certain nombre d'*objets* (le nombre dépendant des systèmes). Lorsque nous parlions du multicanal dans la partie précédente, il s'agissait d'une combinaison de multicanal et de mode objet.

Les moteurs de rendu

En ambisonie, on a donc l'idée d'une scène sonore idéale, que l'on va devoir adapter au système de monitoring que l'on utilise. Lorsque l'on réalise une scène complexe mélangeant multicanal et mode objet, on est confronté à une situation similaire : la réalisation de la spatialité de la scène dépend du système de monitoring que l'on utilise et on doit avoir un dispositif permettant d'adapter l'information spatiale contenue dans le mixage (dans les metadata) à ce système de monitoring : c'est la fonction du moteur de rendu. Par extension, nous appellerons aussi "une scène" toute combinaison de multicanal, mode objet et ambisonie, nécessitant un moteur de rendu pour être réalisée.

Downmix et Upmix

Le moteur de rendu permet d'adapter un contenu multicanal/objet d'un système à un autre. Si le canal prévu dans le mixage existe, le signal sera diffusé sur l'enceinte correspondant à ce canal. Si le canal prévu n'existe pas, le signal prévu sera diffusé comme une source fantôme entre les deux enceintes correspondant aux canaux existants les plus proches, ou sera projeté dans le plan horizontal s'il n'y a pas d'enceintes en élévation dans la zone d'azimut concernée. S'il y a plusieurs plans d'élévation, le moteur réalisera une interpolation entre ces différents plans.

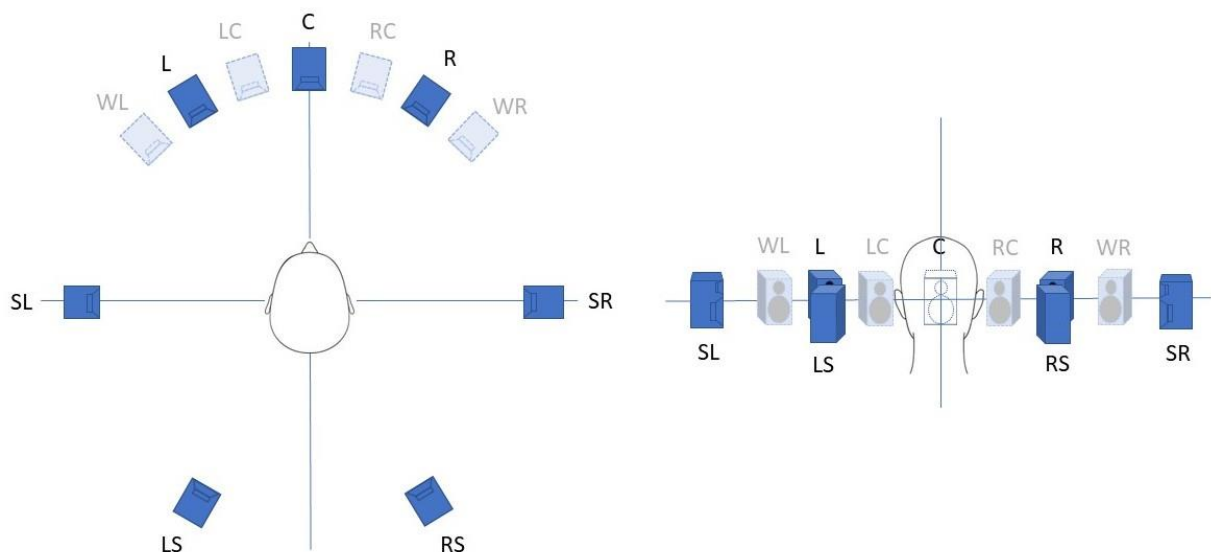


Figure 1 : Format 7.1 avec sources fantômes (en pointillés). Exemple : Un canal Lc ou l'objet correspondant (Left Center, -15° azimuth ; 0° élévation) sera diffusé sur l'enceinte Lc si elle existe, et entre les enceintes L (-30° ; 0°) et C (0° ; 0°) si elle n'existe pas.

L'optimisation du signal deux canaux

La méthode la plus simple du *downmix* deux canaux consiste à utiliser des panoramiques d'intensité classiques pour tous les angles compris entre -30° et $+30^\circ$ et pour la zone symétrique à l'arrière (entre -150° et $+150^\circ$), qui est repliée sur l'avant. Les objets et les canaux dont les angles sont compris entre -30° et -150° sont fixés sur le canal de gauche, les objets et les canaux dont les angles sont compris entre $+30^\circ$ et $+150^\circ$ sont fixés sur le canal de droite. On obtient ainsi un signal stéréophonique pouvant être lu classiquement au casque comme sur enceinte, mais la scène est tronquée au-delà de $\pm 30^\circ$.

On peut optimiser la scène sonore au casque ou sur enceinte en utilisant les techniques binaurales/transaurales. C'est l'objet des outils Transpan et MyBino, des moteurs de rendu transauraux et binauraux, dont le développement est issu d'une collaboration entre le service audiovisuel du Conservatoire et différents partenaires. Les techniques binaurales, pour une écoute au casque, utilisent des filtres reproduisant l'effet de la tête et des pavillons des oreilles sur le champ acoustique pour chacune des directions de provenance du son, nommés HRTF (Head Related Transfer Function). Les techniques transaurales, pour une écoute sur haut-parleurs, ajoutent à ces HRTF des filtres annulant les trajets croisés des enceintes (signal sortant du haut-parleur gauche allant sur l'oreille droite de l'auditeur, et signal sortant du haut-parleur droit allant sur l'oreille gauche), afin de retrouver ces HRTF aux oreilles de l'auditeur. Il y a un enjeu spectral dans le choix de ces HRTF. Pour Transpan comme pour MyBino, un travail important a été réalisé pour que la sonorité des filtres se rapproche le plus possible de la sonorité du signal, l'idée étant de garder l'équilibre spectral du signal stéréophonique tout en optimisant la spatialité de la scène, pour le casque ou pour le système d'enceintes.

Le moteur de rendu prend en entrée les différents signaux et affecte à chacun une voie de traitement qui va réaliser le filtre (HRTF + annulation des trajets croisés si transaural) pour une direction donnée. Il faut donc faire correspondre les réglages des directions à la configuration que l'on utilise. Pour prendre un exemple, on peut alimenter le moteur de rendu par un bus multicanal 11.1.4 et le régler sur les angles correspondants :

L (-30° ; 0°) R ($+30^\circ$; 0°) C (0° ; 0°) LFE (0° ; 0°) WL (-45° ; 0°) WR ($+45^\circ$; 0°) SL (-90° ; 0°) SR ($+90^\circ$; 0°) LS (-150° ; 0°) RS ($+150^\circ$; 0°) TL (-45° ; 45°) TR ($+45^\circ$; 45°) LC (-15° ; 0°) RC ($+15^\circ$; 0°) TRL (-135° ; 45°) TRR ($+135^\circ$; 45°)

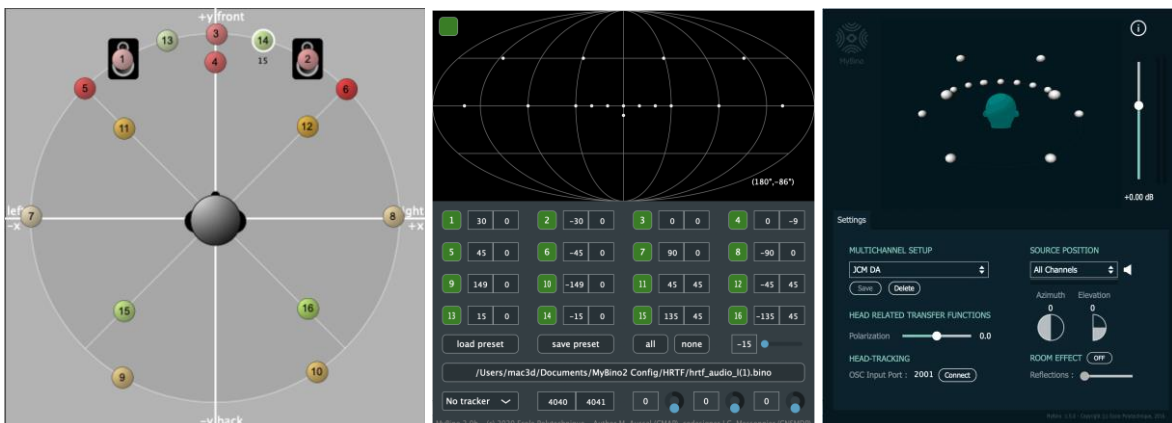


Figure 2 : Représentation des directions des canaux sur trois moteurs de rendu différents : Transpan, MyBino2, MyBino1 (Note : l'ordre du bus correspond à la modification d'un bus 9.1.6 dans Pyramix, qui existe en natif et permet, de façon extrêmement pratique, une affectation des sons aux différents canaux, ce qui explique ce choix).

Idéalement, chacun des sons de la scène doit être traité par une seule HRTF pour que l'on ait une direction précise. Si un son est situé entre deux canaux, par exemple entre LC et C, il en résulte un flou de localisation, puisque la somme des HRTF (0° ; 0°) et (-15° ; 0°) ne correspond pas exactement à la HRTF ($-7,5^\circ$; 0°). On a par conséquent le choix : utiliser plus d'objets, pour chacun des sons à spatialiser, ou accepter ce flou sur certains sons, pour toujours garder le même schéma de travail pour tous ses mixages.

Le signal ainsi obtenu peut être utilisé comme un signal stéréophonique traditionnel, même si la spatialité est optimisée soit pour le casque, soit pour les enceintes. Le moteur de rendu MyBino2 est conçu pour générer un signal binaural optimisé pour une écoute au casque, qui puisse être entendu comme un signal stéréophonique traditionnel sur enceintes, l'image sonore étant alors anamorphosée.

Il est également possible d'utiliser un headtracker avec le moteur binaural (Hedrot pour MyBino1 (<https://alexisbaskind.net/fr/hedrot-head-tracker/>), Feichter pour MyBino2). Ce dispositif envoie les coordonnées d'orientation de la tête de l'auditeur au moteur de rendu, afin que celui-ci change les HRTF produites pour que les sources paraissent à l'auditeur comme étant fixes dans l'espace lorsqu'il bouge la tête, comme dans un espace sonore réel. Le contrôle de la stabilité de la scène passe par la maîtrise de l'information fournie par le headtracker et par l'individualisation des HRTF, tous les individus ne percevant pas la même localisation pour une HRTF donnée. Ce point constitue une direction de recherche encore en cours, mais MyBino apporte des débuts de réponse. Ce dispositif permet d'améliorer le réalisme des sensations sonores (Hendrickx *et al.*, 2017), notamment pour les sources situées dans la zone frontale. On peut formuler l'hypothèse qu'il existe une ambiguïté pour ces sources, le cerveau ayant des difficultés à déterminer leur position, avant ou arrière, puisque les différences de phase et d'intensité perçues par nos deux oreilles sont identiques dans ces deux cas. Cette ambiguïté est levée par le dispositif car, lorsqu'on tourne la tête, les sources se déplacent en sens inverse si elles sont à l'avant ou à l'arrière.

Nous utilisons ces différentes réalisations du mixage pour le contrôler. Chacune des réalisations est vérifiée sur un système particulier (mono sur une seule enceinte, deux canaux sur enceintes, deux canaux au casque, multicanal dans différents formats) et renseigne sur un aspect particulier du mixage, les localisations et le déploiement des réverbérations dans l'espace conditionnant les relations de masquage entre les différents sons qui constituent la scène. On peut passer d'un mode de monitoring à un autre en fonction de son environnement de travail. Ces modes de monitoring sont complémentaires.

La spatialité des scènes

Les considérations précédentes ne font pas d'hypothèses sur les signaux correspondants aux différentes directions de l'espace. Dans le contexte de la production de musique en général, et de notre approche en particulier, une partie de ces signaux correspond aux sons directs des différentes sources constituant la scène sonore, et une autre partie de ces signaux correspond à la réverbération de la scène. Cette réverbération n'est pas forcément homogène : elle peut être constituée de plusieurs réverbérations ayant chacune des caractéristiques spécifiques, comme

par exemple différents temps de réverbération, pour différents groupes de sources, ayant chacune une répartition spatiale différente. Cette spatialité de la réverbération est une des possibilités nouvelles offertes par ces nouveaux formats et elle ouvre de nouveaux champs esthétiques.

On peut considérer le format 11.1.4 comme un format 7.1.4 dont on a augmenté la définition dans la partie avant par l'ajout de sources supplémentaires WL, WR, LC et RC. Le fait de matérialiser autant que possible la direction des sources par leur présence sur un canal (ou un objet) unique, et non pas comme une source fantôme entre deux canaux, permet d'une part, d'avoir une scène plus stable en multicanal (si l'auditeur se déplace, les sources fantômes se déplacent mais les sources réelles restent à leur place) et d'autre part, d'avoir des scènes binaurales plus définies en évitant de mélanger plusieurs HRTF pour une direction de source, ce qui est inévitable si un signal est présent sur deux canaux en même temps. Il est donc nécessaire de disposer d'un certain nombre de canaux, ou d'objets, afin d'obtenir une bonne définition pour les sons directs, mais il n'est pas indispensable de tous les utiliser pour diffuser la réverbération de la scène.

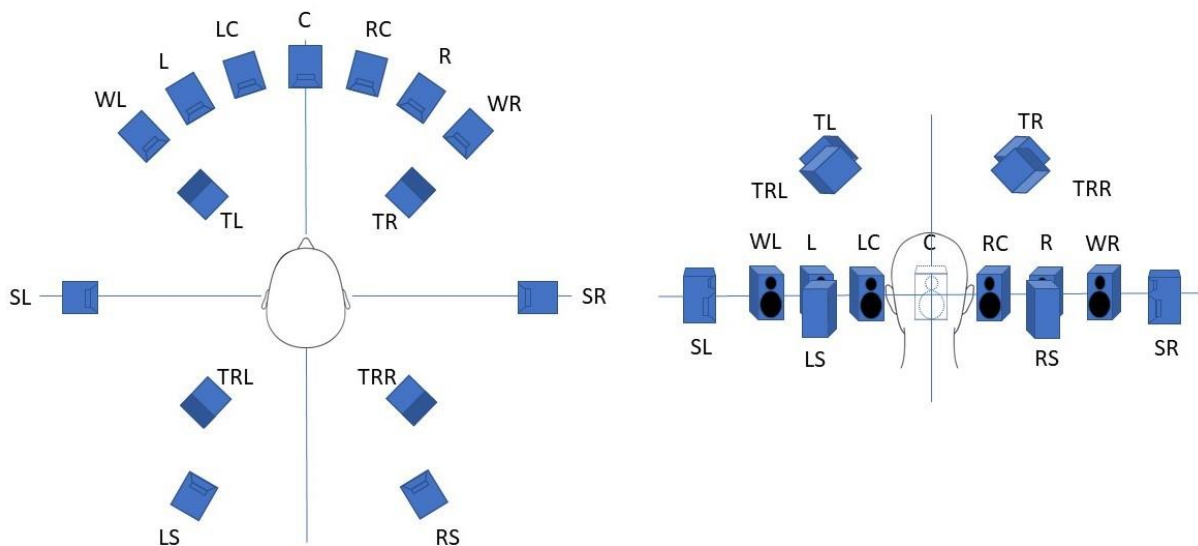


Figure 3 : Format 11.1.4

Un des avantages des formats immersifs est de permettre une différenciation plus forte des directions spatiales de la réverbération. Dans une acoustique naturelle d'une salle de grand volume, la réverbération qui vient de l'avant n'est pas la même que celle qui vient des côtés, de l'arrière ou du haut, car le jeu des réflexions sur les parois de la salle est différent suivant la position des sources, leur directivité et l'acoustique même de la salle. On peut utiliser les canaux L, R, SL, SR, LS, RS, TL, TR, TRL et TRR pour répartir la réverbération sur des directions occupant tout l'espace d'écoute de manière homogène, car on a ainsi 6 bases équiréparties de 60° dans le plan et 4 bases de 90° en élévation (les angles des canaux étant réglés comme sur la Figure 1).

Cette richesse de l'acoustique est très bien perçue en acoustique naturelle mais est limitée et altérée en stéréophonie à deux canaux. Il ne s'agit pas de prétendre que l'on peut rendre cette complexité telle qu'elle est à l'origine, car le multicanal implique un couplage avec la salle de diffusion qui va forcément modifier le rendu de la scène, mais il sera possible d'enrichir l'écoute en bénéficiant d'un plus grand nombre de ressources pour restituer cette complexité.

Chacun est libre de jouer ou non sur le naturel de la représentation, et les options de répartition de la réverbération dépendent bien sûr du type de musique que l'on produit, mais,

encore une fois, la nouveauté réside dans la possibilité de différencier la provenance spatiale des sons.

Les systèmes de captation et de post-production

Un des aspects fondamentaux de notre démarche est de rechercher une continuité entre les méthodes que l'on utilise pour la stéréophonie à deux canaux et celles que l'on peut utiliser en multicanal ou en mode objet. Il nous semble important par exemple de continuer à exploiter les mêmes microphones que nous utilisons pour la stéréophonie à deux canaux car ils constituent pour nous un des éléments centraux de l'esthétique que nous cherchons à développer, quel que soit le format pour lequel on travaille.

Les méthodes sont un sujet sensible car elles représentent l'approche personnelle de chacun et chacune. Le fait de partager notre expérience permet de respecter ces approches et peut donner à réfléchir sur l'évolution de ses méthodes personnelles pour aborder ces formats ou offrir des éléments de comparaison si on les pratique déjà. De grandes lignes peuvent être partagées, qui dépendent notamment des styles musicaux enregistrés. Concrètement, chacun des objets de la scène est capté par un ou plusieurs microphones (cas du piano par exemple) et la méthode de base de captation des objets est donc identique à celle des microphones d'appoint en stéréophonie. Il est même usuel, pour nous, de construire une scène de jazz uniquement à partir de microphones d'appoint et de réverbérations artificielles.

On trouve des réverbérations artificielles multicanales, comme le SPAT, que nous utilisons beaucoup. On peut construire l'espace sonore à partir d'un ensemble de réverbérations 4 canaux ou 2 canaux réparties dans l'espace. La société AudioEase propose avec l'Altiverb des réverbérations 4 canaux à convolution ; TC Electronic, avec son System 6000, propose des moteurs de réverbérations multicanales ; et on peut également citer la DRE 777 de Sony, dont la production a été abandonnée, mais dont les réponses impulsives sont excellentes et toujours utilisées. Pour les réverbérations à deux canaux, tous les systèmes disponibles sur le marché sont utilisables. On peut construire une réverbération multicanale avec plusieurs réverbérations à deux canaux possédant des critères communs (même temps de réverbération, par exemple). Toutes les techniques des musiques actuelles amplifiées sont utilisables, il suffit donc juste d'ajouter à ces traitements des choix plus larges de répartition spatiale qu'en stéréophonie à deux canaux.

Ces éléments ont en commun d'être tous suffisamment décorrélés et donc de pouvoir être additionnés sans problèmes de phase. La constitution de la scène va dépendre du positionnement spatial de chacun de ces éléments et la réduction stéréophonique peut très bien correspondre à ce que l'on aurait choisi en stéréophonie à deux canaux. Les réverbérations seules peuvent être immersives en multicanal et redevenir des réverbérations stéréophoniques à deux canaux lors de la réduction. L'emploi des techniques binaurales et transaurales nous permet de définir des scènes sur plus ou moins 45°, plutôt que plus ou moins 30° pour les sons directs, afin de les rendre plus immersives. En transaural, la scène sonore sera plus large que le cadre des enceintes, et le format binaural, lorsqu'il est écouté au casque, permettra une meilleure externalisation des sources.

On peut adjoindre à ces éléments des ambiances (captation de la réverbération de la salle) ou des ailes (captations latérales de la scène sonore et captation de la réverbération de la salle) qui elles aussi seront décorrélées. Il existe de nombreux choix et de nombreuses nuances pour

capter les ambiances, en fonction de la directivité et du placement des microphones. On peut, par exemple, utiliser des ambiances omnidirectionnelles situées au-delà de la distance critique, placées, lors du mixage, à plus ou moins 90° dans la scène, et des ambiances directives (cardioïdes ou autres) orientées vers le fond de la salle, placées à plus ou moins 150°. Pour les canaux d'élévation, on peut utiliser des ambiances diffuses si l'acoustique de la salle est intéressante, ou utiliser des réverbérations artificielles si l'on souhaite modifier l'impression acoustique de l'auditeur. On peut bien évidemment mélanger ambiance naturelle et réverbération artificielle sur un même canal.

Si l'on place ces réverbérations, naturelles ou artificielles, avec des gains similaires, dans des directions régulièrement réparties de l'espace (dans notre cas $\pm 30^\circ$, $\pm 90^\circ$, $\pm 150^\circ$ dans le plan horizontal, et $\pm 45^\circ$, $\pm 135^\circ$ à 45° d'élévation), on obtient une réverbération immersive. La partie la plus délicate va concerner les secteurs où il y a aussi du son direct (traditionnellement l'avant-scène).

Une première solution consiste à associer des appoints et des ambiances avant (par exemple des microphones cardioïdes orientés vers la scène, mais distants entre eux). La localisation sera alors donnée par la direction de l'appoint. On peut aussi souhaiter capter une source, ou un ensemble de sources, d'un point plus éloigné, d'une part, parce que l'on ne souhaite pas placer d'appoints sur chacune des sources (cas d'un orchestre symphonique par exemple) et d'autre part, lorsqu'un éloignement du microphone est préférable pour une question de qualité de timbre. Les sources musicales n'ont pas la même qualité de son direct à différentes distances car les points de rayonnement s'y additionnent différemment. La qualité de leurs timbres (indépendamment de la réverbération qui y participe) est optimisée pour une certaine distance de perception. C'est donc à chacun de décider, en fonction de l'ensemble de ses contraintes, de ses conceptions et de ses goûts propres, s'il préfère, selon les cas, utiliser un appoint ou enregistrer la source par un système placé plus loin et captant une partie de la scène sonore, voire toute la scène sonore.

La première différence d'un système frontal, utilisé en multicanal, comparé à un système principal, utilisé en stéréophonie à deux canaux, est qu'il ne capte qu'une partie de la réverbération, les autres directions de celle-ci devant être réparties sur d'autres directions de la scène afin d'avoir un résultat immersif. Il devra donc être placé plus près ou être constitué de microphones plus directifs. La deuxième différence est qu'il devra résister aux différentes réductions (downmixs) et que la sommation des éléments du système ne devra pas poser de problème de phase. La troisième est que, si l'on veut utiliser une version binaurale du mixage, on a tout intérêt à matérialiser chacun des signaux que l'on veut ensuite traiter par des HRTF pour avoir une meilleure définition de la scène sonore. L'interpolation entre les HRTF $\pm 30^\circ$ de deux signaux issus d'un couple stéréophonique ne conduira pas au même résultat que la perception de sources fantômes à partir de deux enceintes situées à $\pm 30^\circ$ diffusant ces signaux. Dans le deuxième cas, le cerveau dispose des informations issues des mouvements de la tête et travaille avec ses propres informations correspondant aux HRTF, ce qui n'est généralement pas réalisé dans le premier cas.

Les systèmes que nous avons adoptés et qui répondent aux critères précédents sont des antennes acoustiques, dont la théorie et l'application aux systèmes multicanaux a été décrite de manière exhaustive par Mike Williams (2000). Les antennes acoustiques destinées à l'enregistrement multicanal sont initialement prévues pour capter tout l'espace et nous avons, en première approche, utilisé des antennes à 5 canaux omnidirectionnelles pour les systèmes 5.1.

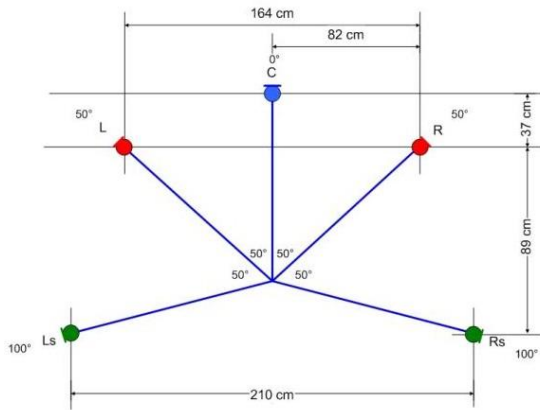


Figure 4 : Exemple Antenne pour un 5.1 – 5 microphones DPA 4041



Figure 5 : Installation à Notre-Dame de Paris

Dans la mesure où une antenne destinée à un grand nombre de canaux est très encombrante, on peut aussi remplacer les parties de l'antenne correspondant aux directions de l'espace où il n'y a pas de sources sonores directes par des ambiances.

Une autre de nos influences, importante, vient des systèmes que Kimio Hamazaki a développé pour les systèmes de restitution multicanaux en 22.2 (Hamazaki, 2006). Ces systèmes sont constitués d'un mélange d'antennes et de systèmes d'ambiance et nous avons travaillé en ce sens en essayant de rendre ces systèmes plus discrets visuellement. Les ambiances peuvent être placées par exemple sur les côtés afin de ne pas être trop visibles par le public ou par les caméras.

On peut choisir une antenne constituée de microphones de directivités différentes afin que la réverbération captée par l'antenne soit moins présente sur les canaux centraux que sur les côtés, ce qui permet de contrôler sa répartition spatiale.

Ces possibilités permettent de construire dans chacune des situations des systèmes différents, en fonction du but sonore recherché et des contraintes de production (présence de la vidéo ou du public et donc contraintes visuelles, contraintes dues au temps d'installation, à des limitations techniques, etc.). L'important est de définir un signal pour chacune des directions de la scène. Mais ce signal peut être défini par un appoint, une ambiance, un élément d'antenne ou de la réverbération artificielle, ce qui offre un grand nombre de possibilités.

Ces systèmes ont des fondements très classiques, dans la possibilité de les réaliser. Nous gardons d'une part, la maîtrise de la microphonie. D'autre part, nous pouvons capter le champ acoustique de près ou de loin en utilisant un système principal ou des appoints et des ambiances, comme nous le faisons avant. Le changement consiste uniquement à réaliser la direction du son différemment, en définissant quel signal correspond à toutes les directions d'une scène, plutôt qu'en définissant quelles relations (différences d'intensité et/ou différences de temps) il y a entre les sons présents sur deux canaux à la fois. Ceci ne définit pas une esthétique.

L'esthétique commence dans la mise en œuvre de ces moyens pour occuper l'espace. En partant du modèle cinématographique, on peut observer que cette esthétique peut être très différente suivant les films. Dans *Birdman* (2014) de A.G. Iñárritu, Michael Keaton fait le tour de la salle via les enceintes surround, ce qui est loin d'être habituel. Dans *The French Dispatch* (2020) de Wes Anderson, la musique est en mono à droite et les dialogues au centre, ce qui crée un parti-pris très fort. Chaque film a en fait ses spécificités et la manière de

répartir le son dans l'espace fait partie de son style, comme les choix de lumière, le format d'image, etc. Le champ d'expérimentation est à notre avis tout aussi ouvert en musique, avec des possibilités de spatialisation beaucoup plus importantes qu'auparavant.

Bibliographie

Blauert, J. (1997), *Spatial hearing*, London, The MIT Press.

HAMASAKI, K., HIYAMA, K. (2006), *Development of a 22.2 Multichannel Sound System*, Broadcast Technology n°25, <https://www.nhk.or.jp/strl/publica/bt/en/fe0025-2.pdf>

Hendrickx, E., Stitt, P., Messonnier, J.-C. , Lyzwa, J.-M. , Katz, B. F. G., de Boishéraud, C. (2017), *Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis*, J.Acoust. Soc. Am., vol. 141, pp. 3678–3688.

Messonnier, J.-C., Lyzwa, J.-M., Devallez, D., de Boishéraud, C. (2016), *Object-based audio recording methods*, AES 140th convention.

Messonnier, J.-C. & Moraud, A. (2011), *Auditory distance perception: criteria and listening room*, AES 130th convention, London.

Moore, B. C. J. (1989), *An introduction to the psychology of hearing*, Academic press, EBU TECH 3364 *Audio Definition Model. Metadata specification*, Genève, édit. 2014.

Plenge, G., & Theile, G. (1986), *Überlegungen zur Leistungsfähigkeit verschiedener stereofoner Verfahren*. Institut für Rundfunktechnik München.

Williams, M. (2004), *Microphone Arrays for stereo and Multichannel sound recording*, Vol II, Editrice II Rostro.

Biographies auteurs

Jean-Christophe Messonnier. Après une formation d'ingénieur du son à l'ENS Louis-Lumière et parallèlement à une formation d'ingénieur en acoustique au CNAM, terminée en 1994, Jean-Christophe Messonnier est entré au service audiovisuel du Conservatoire en 1988. Dans cette longue pratique au service du Conservatoire, il a cherché à comprendre comment on pouvait formaliser la transmission du son et comment l'acoustique pouvait aider à comprendre certains des problèmes reliés à cette transmission. C'est ce qu'il enseigne à la Formation Supérieure aux Métiers du Son du Conservatoire depuis 1992.

Jean-Marc Lyzwa exerce la fonction d'ingénieur du son depuis 1990 au service audiovisuel du Conservatoire de Paris et enseigne à la Formation Supérieure aux Métiers du Son les techniques de prise de son et de post-production. Investi depuis 1995 dans la restitution et la création de l'espace sonore en développant des techniques de captation et de post-production 3D en stéréophonie à deux canaux et en multicanal, il a collaboré avec l'équipe espaces acoustiques et cognitifs de l'IRCAM à la réalisation d'outils de traitement binaural / transaural, spécifiquement adaptés à la production musicale, notamment avec Transpan. Il participe au projet de recherche BiLi (Binaural Listening) de 2013 à 2017 et au projet de recherche PHEND (The Past Has Ears at Notre-Dame) depuis 2021.

Alexis Ling est responsable du service audiovisuel du Conservatoire depuis 2020, mettant à profit les connaissances qu'il a acquises lors de son parcours : licence EEA à la faculté d'Orsay, Maîtrise de Sciences et Techniques image et son à l'université de Brest, Formation Supérieure aux Métiers du Son au Conservatoire de Paris, puis activité d'ingénieur du son en *freelance* dans le secteur de la musique classique, puis en *broadcast* pour le groupe Canal+. Au sein du Conservatoire, il enseigne et mène une réflexion sur la façon dont les outils et les activités audiovisuelles, en production et en recherche, peuvent accompagner le travail pédagogique et valoriser le travail des étudiants et des étudiantes.