

SON *SURROUND* SUR SUPPORTS STANDARD : UNE ENQUETE TECHNOLOGIQUE POUR LE MUSICIEN

Benoît Courribet
CICM
Université Paris 8
MSH Paris Nord
benoit.courribet@gmail.com

Julien Bréval
CICM
Université Paris 8
MSH Paris Nord
julien@breval.com

Timothée Baschet
CICM
Université Paris 8
MSH Paris Nord
timbaschet@gmail.com

RÉSUMÉ

Dans cet article, nous présentons les technologies *surround*, dans le cadre des supports standard. Après avoir précisé le cadre de notre enquête et effectué un bref historique du son *surround*, nous abordons les formats 5.1 sur support DVD audio et vidéo, les conversions de format, la rétrocompatibilité, l'*upmix* et le *downmix*, ainsi que les techniques de prise de son *surround* coïncidentes s'appuyant sur une modélisation simple du champ acoustique. La conclusion porte sur les perspectives de formats ouverts.

1. INTRODUCTION

Cet article s'adresse au compositeur et au créateur sonore désireux de s'approprier les technologies et les formats audio multicanal apparus ces dernières années et rassemblés sous l'appellation *surround*. En effet, l'avènement de telles technologies est apparu comme une opportunité dans le cadre d'un travail sur des espaces de diffusion multicanal dépassant la stéréophonie classique. La possibilité de fixer sur un support standardisé et accessible une création musicale auparavant réservée au cadre du concert ou de l'installation spécifique a motivé de nombreux musiciens à s'intéresser aux technologies *surround*. Ainsi dès 1998, Bertrand Merlier avait pointé le potentiel et les limitations de ce genre de techniques pour le musicien [20]. Au-delà de ces limitations, il nous semble de nos jours encore utile d'aborder ces questions, pour ceux qui pourraient être amenés à travailler avec ce type de support et à utiliser des dispositifs de diffusion contraints par les standards des équipements destinés au grand public.

Cet article est un approfondissement de certains points, à l'occasion d'une étude en cours sur les formats audio et sur la cohérence de l'image et du son dans les pratiques audiovisuelles, effectuée dans le cadre du projet HD3D / Initiative pour une Industrie Ouverte du pôle de compétitivité francilien Cap Digital, accueillie au sein de l'axe de recherche « Industries culturelles et art » de la MSH Paris Nord.

Nous n'avons donc pas pour objectif d'étudier ici des questions en relation avec le champ très vaste de "l'espace sonore", tel qu'il est exploré par ailleurs dans notre équipe [3][23][24], et de façon générale en musicologie et en sciences de l'art. Nous faisons ici le choix de nous limiter à une enquête technologique sur le potentiel du son *surround* dès lors que le musicien se

trouve pragmatiquement confronté aux contraintes imposées par le milieu industriel.

Notre enquête s'est appuyée d'une part sur la littérature scientifique et technologique accessible et d'autre part sur une série de tests, que nous avons effectuée en studio et sur le terrain, de prises de son multicanal (*Schoeps Double MS*, *Soundfield ST350*, *Holophone H2 Pro*, *Holophone H4 Super Mini*, système FLRB maison), de tests et développements logiciels de décodage de *format B*, tests et de développements logiciels d'*upmix* et de *downmix*. Nous avons également examiné les usages et les pratiques du milieu professionnel des métiers du son (utilisateurs, techniciens, directeurs techniques, constructeurs).

Dans les limites de ce travail, nous n'aborderons pas l'appropriation du format 5.1 par les compositeurs de musique électroacoustique dans des environnements de diffusion en concert de type acousmatique. Nous ne ferons pas de commentaires non plus sur les enregistrements de musique électroacoustique en 5.1 déjà accessibles sur le marché. Nous écarterons également de notre enquête l'ensemble des techniques de spatialisation (édition de trajectoires) déjà bien connues et en cours d'appropriation par les musiciens utilisant des moyens informatiques.

Dans ce qui suit, après un rapide historique, nous aborderons les formats 5.1 sur support DVD audio et vidéo, les conversions de format, la rétrocompatibilité, l'*upmix* et le *downmix*. Nous présenterons également des techniques de prise de son *surround* coïncidentes s'appuyant sur une modélisation simple du champ acoustique. Nous évoquerons dans la conclusion les perspectives de formats ouverts.

2. HISTORIQUE DU SON *SURROUND*

Dans les salles de cinéma et les environnements de diffusion domestiques, le son enregistré est à l'origine diffusé sur un seul haut-parleur. Dès les années 30, la stéréophonie fait ses débuts avec les travaux d'Alan Fletcher (Bell Labs) et d'Alan Blumlein (EMI), deux approches très différentes. La stéréophonie s'imposera ensuite jusqu'à nos jours.

En 1940, le film *Fantasia* (Walt Disney / Bell Labs) est la première oeuvre *surround* montrée au grand public. Il s'agit d'un cas unique de film itinérant (la troupe se déplaçait avec tout le matériel nécessaire), dont le son est joué en direct par un orchestre spatialisé sur huit canaux, avec des effets, ce qui nécessita d'inventer des technologies comme le panoramique d'amplitude et les magnétophones multipistes.

Dans les années 70 apparaît la quadraphonie (quatre enceintes) pour le grand public, qui est alors un échec à la fois technique et commercial majeur. En effet, d'une part, la position des enceintes est mal choisie en termes d'image sonore et de compatibilité avec la stéréophonie classique ; d'autre part, un très grand nombre de formats propriétaires et non compatibles avec toutes les machines de lecture apparaissent alors, certains d'entre eux utilisant un matriçage affectant beaucoup trop la qualité sonore.

À cette même époque, la théorie *Ambisonics* fait son apparition mais, trop complexe pour être utilisée en pratique, elle ne resurgira que beaucoup plus tard, entre autres grâce à la simplification obtenue par les développements des technologies numériques. Pour un état de l'art complet à ce sujet, voir l'excellente thèse de Jérôme Daniel [7].

Ainsi, par la suite, c'est le son *Dolby* qui s'impose, pour la première fois dans un film en 1976, et jusqu'au *Dolby Digital* en 1992 ; le premier mixage pour une installation domestique utilisant ce dernier format sort en 1995. Entre temps, de nombreux autres formats *Dolby* ont existé : *Dolby Surround Pro Logic*, *Pro Logic II* et *Pro Logic IIx* notamment. On trouve de nombreuses variantes du *Dolby Digital* aujourd'hui, ainsi que des formats *Dolby* adaptés à la Haute Définition.

Le son *DTS* apparaît en 1993 au cinéma et en 1997 pour les installations domestiques. À l'image de *Dolby*, il existe un très grand nombre de variantes du *DTS*, y compris pour la Haute Définition.

Ces technologies propriétaires sont devenues un standard pour la diffusion de programmes audiovisuels en salle et en environnement *home cinema*. À l'heure des technologies ouvertes, il n'est pas inutile de faire le point pour le musicien.

3. LES FORMATS 5.1 SUR SUPPORTS DVD

Si l'utilisation d'un ordinateur personnel et d'une interface audio multicanal permettent au musicien de s'affranchir de l'utilisation d'un format normalisé spécifique pour diffuser du son *surround* (notamment dans le cadre d'un concert), la volonté de faciliter la diffusion des oeuvres conduit à considérer les formats de diffusion *surround* dominants. Aussi, on s'intéressera dans cet article aux formats audio 5.1 disponibles dans le cadre de la production et de l'utilisation de DVD-Audio et DVD-Vidéo.

Notons que dans le cas de la diffusion HDTV, les formats audio s'uniformisent autour d'un flux unique en 5.1 encodé avec une technologie *Dolby*.

3.1. Dispositif de diffusion et format 5.1

On désigne par le terme "système 5.1" un système de diffusion audio composé de 5 haut-parleurs principaux et d'un haut-parleur dédié aux fréquences graves (caisson de basses). Si la position du caisson de basses n'est pas spécifiée en raison du caractère non directif des basses fréquences, les 5 haut-parleurs principaux doivent être disposés suivant la configuration présentée sur la Figure 1. et répondant aux exigences de la norme ITU-R BS.775 [16], qui décrit les configurations multi haut-parleurs, qu'il s'agisse d'installations avec ou sans écran.

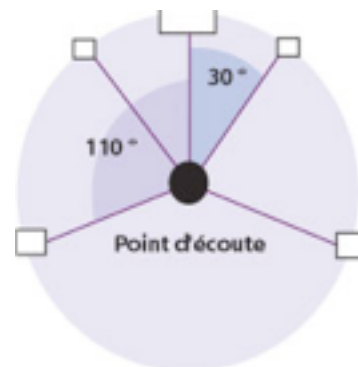


Figure 1 Disposition des haut-parleurs recommandée par la norme ITU-R BS.775.

Majoritairement utilisé pour la projection cinématographique et vidéo, ce système a été conçu de manière à assurer une compatibilité avec la diffusion de programmes stéréophoniques, ce qui se traduit dans la disposition des deux haut-parleurs frontaux, seulement espacés d'un angle de 60 degrés. Le haut-parleur central sert en général à la diffusion des dialogues. Les haut-parleurs *surround* servent à diffuser les ambiances sonores du film et sont plutôt considérés, de par leur disposition, comme des haut-parleurs latéraux que comme des haut-parleurs arrière. Enfin, le caisson de basse sert à restituer des effets sonores nécessitant beaucoup de puissance dans les graves (explosions, etc).

Les 6 haut-parleurs d'un système 5.1 diffusent le contenu de 6 canaux que nous nommerons de la façon suivante :

- L pour le canal gauche
- R pour le canal droit
- C pour le canal central
- LFE (*Low Frequency Effects*) pour le canal de basses
- Ls pour le canal *surround* gauche
- Rs pour le canal *surround* droit.

3.2. Les formats audio sur DVD-Audio et DVD-Vidéo

3.2.1. DVD-Audio

Le format obligatoire pour le DVD-Audio est le *LPCM* (*Linear PCM*) ; il permet d'obtenir des contenus audio non compressés à des fréquences d'échantillonnage de

44.1, 48, 88.2, 96, 176.4, 192 kHz et des résolutions de 16, 20 ou 24 bits, et ce jusqu'à 6 canaux¹.

Afin de maximiser la quantité de contenu audio stockable sur un disque, un algorithme de compression sans perte (permettant une compression des données jusqu'à 50%) peut être utilisé : il s'agit du *MLP* (*Meridian Lossless Packing*), une technologie sous licence *Dolby*.

D'autres formats optionnels peuvent être présents sur des disques DVD-Audio : *Dolby Digital*, *DTS* et *MPEG audio*. Cependant le format *Dolby Digital* est nécessaire pour les contenus audio associés à de la vidéo (!).

3.2.2. DVD-Vidéo

Dans le cas d'un support DVD-Vidéo, même si les spécifications prévoient la possibilité d'utiliser des formats audio non compressés², la tendance générale est tout autre. En effet, le débit global de données étant limité, le choix d'un contenu audio multicanal non compressé se ferait au détriment de la qualité de la compression vidéo, d'une part, et de la possibilité de stocker plusieurs pistes audio, d'autre part (versions étrangères, commentaires audio). Les seuls formats multicanal 5.1 lisibles par les lecteurs DVD sont, outre le *LPCM*, le *Dolby Digital* et le *MPEG-2*. Le *DTS* et le *SDDS* sont des formats supportés de manière optionnelle.

Ces derniers correspondent à des méthodes de compression destructive, basées sur un modèle psychoacoustique de l'oreille humaine. On citera les taux de compression pour les réglages standard des 2 algorithmes les plus utilisés : 10:1 pour le *Dolby Digital* (soit une réduction de 90%) et 6:1 pour le *DTS*.

On en conclut qu'à moins de sacrifier la qualité de la vidéo (choix qui à notre connaissance n'a jamais été fait), il n'est pas possible de travailler avec un format de haute qualité (soit sans compression, soit avec des caractéristiques supérieures à celles du CD Audio) en multicanal sur un support DVD-Vidéo.

3.3. Perspectives des nouveaux formats

L'avènement des nouveaux standards HD-DVD et Blu-Ray donne, à travers un débit plus important de données, la possibilité d'avoir accès à une piste multicanal 5.1 (ou plus) non compressée. Cependant, on assiste à une persistance des formats propriétaires notamment pour assurer une rétrocompatibilité. Ainsi, le format *Dolby True HD* (*LPCM* avec compression de données *MLP*) est le format choisi comme format obligatoire sur ces deux nouveaux supports.

¹ Le débit maximal pour un flux *LPCM* étant de 9.8 Mb/s, des restrictions s'appliquent quant aux combinaisons de ces différentes caractéristiques : ainsi, les fréquences d'échantillonnage de 176.4 et 192 KHz ne sont disponibles que sur 2 canaux.

² Le débit maximal pour l'audio sur le support DVD-Vidéo étant de 6144 Mb/s, il est possible d'utiliser de l'audio au format non compressé *LPCM* jusqu'à 8 canaux en 16 bit/48 kHz ou un nombre moindre de canaux à une résolution plus élevée.

Il convient ici de noter que des formats audio ouverts existent et que les résultats qu'ils permettent d'obtenir au niveau de la qualité sonore sont tout à fait équivalents à ceux des formats propriétaires que nous avons détaillé. On citera par exemple le *Vorbis* multicanal.

Il est également intéressant de présenter la position du DCI (*Digital Cinema Initiative*) à ce sujet [8]. Le DCI stipule que le son ne doit subir aucune compression entre la création du master numérique et la diffusion en salle. On aboutit donc à un format multicanal non compressé. On peut penser et espérer que cette tendance va se propager jusqu'à la diffusion pour les particuliers et ainsi devenir une pratique dominante.

4. CONVERSIONS DE FORMATS, RETROCOMPATIBILITE

Il est parfois nécessaire de diffuser un contenu multicanal sur un dispositif stéréophonique classique. Par exemple, les lecteurs DVD peuvent décoder une piste audio multicanal, mais il est possible que ceux-ci soient reliés à un système audio stéréophonique (une simple télévision par exemple). Détaillons les différents cas de figure pouvant se présenter.

4.1. Remixage

Les formats DVD (Audio, Vidéo, HD-DVD, Blu-Ray) permettent de stocker plusieurs pistes son sur un support, et donc de sélectionner une de ces pistes son à la lecture. Ainsi, il est possible de stocker une piste 5.1 et une piste stéréophonique et, à la lecture, de choisir la piste la mieux adaptée au dispositif de restitution. La meilleure solution dans ce cas de figure consiste à effectuer un remixage de la pièce afin d'en proposer une version stéréophonique. Libre ainsi au compositeur de décider de la manière dont son travail doit être projeté dans l'espace stéréophonique. Le cas de figure inverse est possible (passer de la stéréo au 5.1) et on recommandera également d'effectuer un remixage pour dispositif *surround*.

4.2. Downmix

Une des particularités des différents systèmes de décodage audio est de pouvoir effectuer un *downmix* automatique en temps réel. Le *downmix* est une opération permettant de passer d'un flux ou un fichier audionumérique à n canaux à un flux ou un fichier audionumérique à m canaux, m étant strictement inférieur à n . Plutôt que de ne préserver que les canaux L et R du mixage *surround*, ce qui poserait un problème majeur dans le cas d'une bande-son pour le cinéma (l'ensemble des dialogues étant diffusés par le canal C, ceux-ci disparaîtraient lors d'une diffusion stéréophonique), un mixage des différents canaux *surround* est réalisé. Ainsi, le canal C est envoyé, avec éventuellement une atténuation, vers les canaux L et R, ainsi que vers les canaux Ls et Rs.

Les éléments auxquels accorder une attention particulière sont donc les suivants :

- Les trajectoires des sons spatialisés. Dans le cas de techniques de panoramique d'intensité, toute information quant à la répartition avant/arrière sera perdue ; la perte du canal C peut également conduire à une distorsion de la perception de l'espace frontal. Dans le cas d'autres techniques de spatialisation, en fonction de l'utilisation d'effet Doppler, de décorrélations, etc., on peut aboutir à des modifications timbrales non souhaitées.
- Les problèmes d'opposition de phase entre deux répliques d'un même signal diffusées dans deux canaux différents. En effet, prenons pour exemple le cas d'un son diffusé en phase dans le canal L et avec une inversion de phase dans le canal Ls. Lors du *downmix*, les deux signaux s'annuleront.

4.3. *Upmix*

L'opération inverse est également possible et permet la diffusion d'une pièce stéréophonique à travers les 6 canaux d'un système 5.1. Cependant, il est difficile de trouver des informations sur les techniques utilisées par les différents systèmes de décodage de formats propriétaires. Dans le cas de l'*upmix* réalisé par le décodeur *Dolby Prologic II*, par exemple, des déphasages des signaux L et R sont introduits pour alimenter les canaux Ls et Rs. Divers outils logiciels existent pour réaliser soi-même un *upmix* sans avoir recours à un remixage des sources. Quelques tests à l'aide de certains de ces outils nous ont permis de déceler des modifications timbrales importantes. En effet, certaines techniques utilisées par ces outils introduisent souvent des artefacts affectant le timbre original, notamment les techniques de décorrélation, d'analyse / synthèse spectrale, de réverbération et de convolution. Ainsi, ces outils nous semblent plutôt adaptés à une accentuation de la sensation d'espace (*spaciousness*) qu'à une adaptation de pièce musicale. Les résultats obtenus, notamment sur des sons percussifs, laissent entrevoir des problèmes de timbre et de localisation spatiale. Nous travaillons actuellement à la réalisation d'un prototype de processeur logiciel d'*upmix* pour le musicien.

5. TECHNIQUES DE PRISE DE SON SURROUND COÏNCIDENTES

Dans le cadre de notre étude sur le son *surround*, nous sommes également intéressés à la prise de son. Pour un support multicanal, il existe trois façons d'aborder l'enregistrement. Premièrement, on peut partir d'un enregistrement monophonique ou stéréophonique déjà existant et en faire un *upmix*. Deuxièmement, on peut enregistrer en mono ou en stéréo et réaliser une spatialisation dite artificielle (généralement à l'aide de techniques de panoramiques d'intensité). Troisièmement, on peut effectuer une prise de son multicanal. Cette

dernière approche est la moins connue des musiciens en raison de sa relative nouveauté. Ainsi, dans cette partie, nous présentons quelques techniques d'enregistrement et de restitution de champs acoustiques, en insistant sur les plus simples à mettre en œuvre.

On peut enregistrer un espace acoustique en multicanal par diverses méthodes. Beaucoup de systèmes de prise de son multicanal font partie de la catégorie des *arbres* de microphones : on utilise plusieurs microphones disposés spécialement [10][22][25]. D'une part, ces méthodes impliquent l'utilisation d'un dispositif encombrant et d'autre part, de nombreuses études ont décrit ou comparé ces systèmes (voir par exemple [15]) ; ainsi, nous n'allons pas les détailler ici, malgré leur intérêt considérable.

5.1. Un modèle simple du champ acoustique

Un autre moyen d'enregistrer un espace acoustique est de passer par un modèle encodé du champ acoustique. Tout d'abord, il faut obtenir un enregistrement audio PCM multicanal de ce champ en forme encodée, grâce à l'une des techniques décrites ci-dessous. Ensuite, en utilisant un décodage spécial, on peut reproduire l'enregistrement sur divers systèmes de diffusion.

Dans le modèle *Ambisonics* (voir par exemple [14]), le champ acoustique est divisé en harmoniques sphériques. À l'ordre 1, le modèle comprend quatre composantes (une composante omnidirectionnelle dite W et trois composantes orthogonales en figure de "8", dites X, Y et Z). Cet encodage est appelé *format B* et peut correspondre à quatre pistes PCM enregistrées.

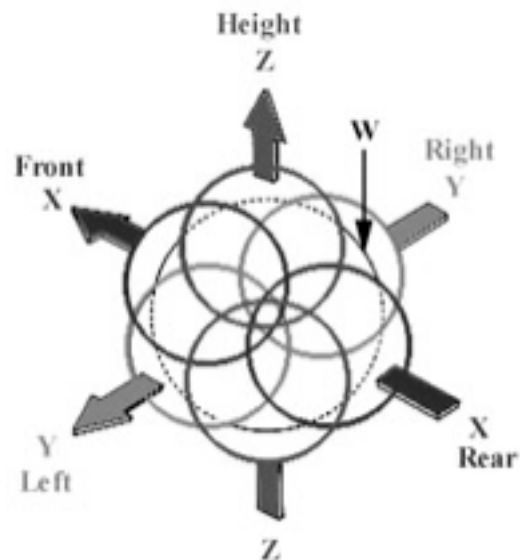


Figure 2 Représentation en trois dimensions de l'encodage du champ acoustique dans le modèle *Ambisonics* à l'ordre 1.

Dans la plupart des systèmes de diffusion actuels, la verticalité n'est pas reproduite, ainsi on peut omettre la composante Z : on parle alors de *format B horizontal* [2]. Par conséquent, on peut y associer un enregistrement de trois pistes seulement.

Notons que certains systèmes de prise de son utilisent le modèle *Ambisonics* à des ordres supérieurs, ce qui permet d'améliorer la résolution de l'encodage du champ acoustique. En particulier, on peut citer à ce sujet les travaux sur la Haute Résolution Spatiale de l'équipe d'Arnaud Laborie [17][18][19]. Nous ne détaillerons pas, ici, les techniques de prise de son basées sur les modèles d'ordre supérieur.

5.2. Techniques d'enregistrement

5.2.1. Enregistrement direct en forme encodée

On peut enregistrer directement en *format B*, au moyen d'un ensemble de microphones coïncidents comprenant un microphone omnidirectionnel et trois microphones en figure de "8". Or, d'une part il est difficile de positionner quatre microphones coïncidents de ce type, d'autre part la verticalité n'est souvent pas nécessaire. Ainsi, on peut se limiter à deux microphones en figure de "8", correspondant à X et Y. Cette technique s'appelle *format B natif horizontal* ou *Nimbus-Halliday arrangement* (développée par Jonathan Halliday pour la compagnie de disques Nimbus Records).

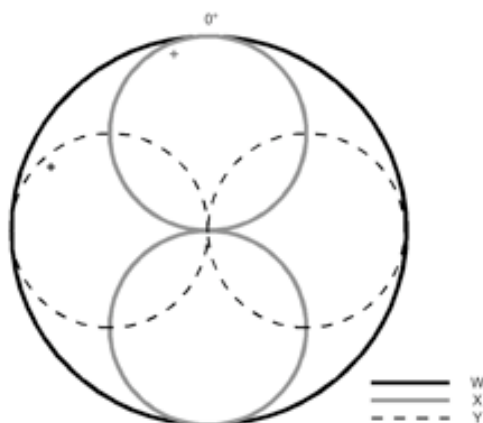


Figure 3 Diagramme polaire d'un enregistrement en *format B natif horizontal* avec trois microphones (vue du plan (XY), de dessus).

Pour réaliser ce type d'enregistrement, on doit respecter les conditions suivantes :

- les deux microphones en figure de "8" doivent être identiques ;
- le microphone omnidirectionnel, en dehors de sa très bonne réponse en fréquence pour les graves et de sa directivité, doit avoir des caractéristiques les plus proches possibles de celles des microphones en figure de "8" ;
- en particulier, il faut au minimum connaître les sensibilités des microphones afin de pouvoir les compenser si c'est nécessaire (les trois microphones doivent avoir la même sensibilité).

Notons qu'il existe aussi aujourd'hui un microphone spécial comprenant les trois capsules rassemblées dans un même corps : le *Josephson C700S*.

Les avantages de ce type de prise de son sont les suivants :

- c'est la seule technique qui permet d'obtenir le *format B horizontal* sans aucune conversion nécessaire ;
- on peut choisir les microphones qu'on veut tant qu'on respecte les conditions énoncées ci-dessus ;
- il est généralement facile de positionner les trois microphones de telle sorte qu'ils soient coïncidents et qu'aucune capsule ne soit masquée par le corps des autres microphones dans le plan horizontal ;
- le microphone omnidirectionnel permet une bonne reproduction des basses fréquences.

On note cependant les inconvénients suivants :

- au-delà d'une certaine fréquence, un microphone omnidirectionnel a tendance à devenir un peu cardioïde, ainsi un son ne sera pas enregistré de la même manière selon qu'il vienne de l'avant ou de l'arrière ; le fait d'utiliser une très petite capsule permet de corriger ce problème, mais les microphones à très petite capsule ont souvent un niveau de bruit propre trop élevé ;
- il est souvent difficile d'ajouter un quatrième microphone dans le cas où on veut également enregistrer la dimension verticale du son.

5.2.2. Enregistrement FLRB (Front Left Right Back)

Une variante consiste à utiliser quatre microphones cardioïdes coïncidents identiques, notés Cf, Cb, Cl et Cr, pointant vers l'avant (Cf), l'arrière (Cb), la gauche (Cl) et la droite (Cr).

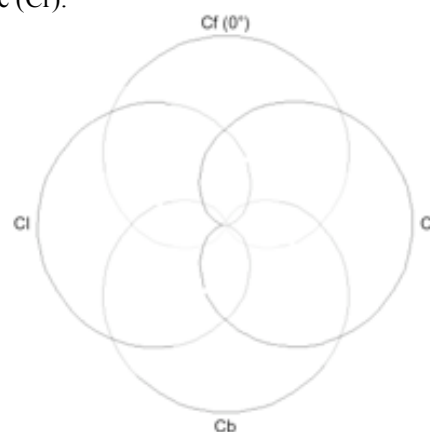


Figure 4 Diagramme polaire d'un enregistrement avec quatre microphones cardioïdes, destiné à une conversion en *format B horizontal* (vue du plan (XY), de dessus).

Notons qu'on peut également utiliser deux microphones stéréo XY cardioïdes si pour chacun d'entre eux les deux capsules sont placées à 90 degrés (l'un de

ces microphones correspondra à Cf et Cl, l'autre à Cr et Cb).

Il existe aussi un microphone rassemblant les quatre capsules cardioïdes nécessaires dans un même corps : le *Line QM12i*.

Pour obtenir le *format B horizontal*, on doit passer par la conversion suivante :

$$\begin{cases} W = Cf + Cb \text{ ou } W = Cl + Cr \\ X = Cf - Cb \\ Y = Cl - Cr \end{cases} \quad (1)$$

Dans l'équation (1), on choisit $W = Cf + Cb$ ou $W = Cl + Cr$ en fonction de la position des sources les plus importantes de l'espace acoustique.

Les avantages de cette technique sont les suivants :

- un son est enregistré de la même manière selon qu'il vient de l'avant, de l'arrière, de la droite ou de la gauche ;
- on peut choisir les microphones qu'on veut tant qu'ils sont identiques et cardioïdes ;
- il existe un très grand nombre de microphones cardioïdes.

Mais il faut aussi noter les inconvénients suivants :

- parmi les nombreux microphones cardioïdes existants, peu d'entre eux sont suffisamment petits pour permettre à la fois une bonne coïncidence et une absence de masquage des capsules par les corps des microphones ;
- cette prise de son ne reproduit pas très bien les basses fréquences puisqu'il n'y a pas de microphone omnidirectionnel ;
- on a besoin d'utiliser quatre microphones pour le plan horizontal, alors que les autres techniques n'en utilisent que trois ;
- il est pratiquement impossible d'ajouter deux autres microphones cardioïdes ni même un microphone en figure de "8" dans le cas où on voudrait aussi enregistrer la verticalité du son.

5.2.3. Enregistrement en format A

On peut également utiliser un microphone spécial appelé *microphone de format A* [5], qui comprend quatre capsules cardioïdes coïncidentes disposées sur les faces d'un tétraèdre. Les quatre signaux issus de ce microphone, appelés *format A*, doivent être encodés en *format B*. Cette conversion peut être effectuée soit en appliquant les filtres théoriques décrits dans [5], soit en utilisant des filtres issus de mesures pratiques [13]. Les microphones de *format A* sont fournis avec un encodeur logiciel ou matériel permettant de réaliser cette opération.

À côté des modèles haut de gamme (*Soundfield MKV* par exemple), on trouve actuellement un seul modèle accessible (le *Core Sound TetraMic*), mais deux autres modèles bon marché seront bientôt disponibles

(*Soundfield SPS200* ; *Oktava 4D preamp* à utiliser avec des capsules *Oktava MK-012* cardioïdes).

Les avantages de ce type de captation sont les suivants :

- un microphone de *format A* enregistre toujours la hauteur (composante Z) avec la même qualité que pour les autres directions ;
- les capsules ont un espacement égal en trois dimensions ;
- la calibration nécessaire des capsules [13] permet d'obtenir un appariage parfait, mais aussi une correction de la réponse en fréquence (par exemple, on peut compenser le manque de basses fréquences si les quatre capsules cardioïdes sont calibrées par rapport à un microphone omnidirectionnel de mesure).

On note cependant les inconvénients suivants :

- la conversion du *format A* en *format B* est une opération complexe, qui nécessite le plus souvent l'utilisation de convolutions, implémentées dans le domaine spectral, ce qui se traduit par une latence importante (et parfois une consommation de processeur importante) si on veut une qualité correcte ;
- la qualité de la conversion dépend de la méthode utilisée ;
- en ce qui concerne le plan horizontal, les signaux obtenus pour le *format B* sont moins bons que ceux enregistrés en *format B natif horizontal* [2].

5.2.4. Enregistrement double M/S

La technique dite *double M/S* permet aussi d'obtenir un enregistrement en *format B horizontal* [26]. Elle comprend deux microphones cardioïdes (notés Mf et Mr) et un microphone en figure de "8" (noté S). Les conditions sur les choix des microphones sont similaires à celles données précédemment pour le *format B natif*.

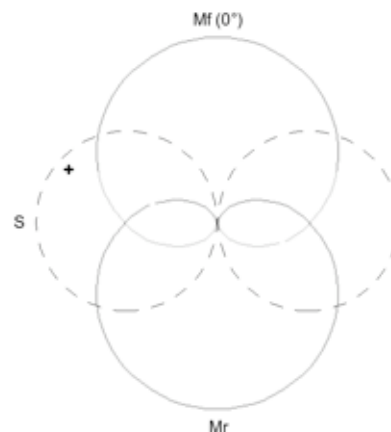


Figure 5 Diagramme polaire d'un enregistrement double M/S, destiné à une conversion en *format B horizontal* (vue du plan (XY), de dessus).

Pour passer du *double M/S* au *format B horizontal*, on utilise la conversion suivante :

$$\begin{cases} W = Mf + Mr \\ X = Mf - Mr \\ Y = S \end{cases} \quad (2)$$

La prise de son *double M/S* offre les avantages suivants :

- un son venant de l'avant ou de l'arrière est enregistré de manière identique (ce qui, rappelons-le, n'est pas le cas pour le *format B natif*) ;
- on peut choisir les microphones qu'on veut.

Il faut cependant compter avec les désavantages suivants :

- la reproduction des basses fréquences n'est pas parfaite puisqu'il n'y a pas de microphone omnidirectionnel ;
- il existe relativement peu de microphones permettant à la fois une bonne coïncidence et aucun masquage des capsules par les corps des microphones ;
- il est souvent difficile d'ajouter un troisième microphone en figure de "8" dans le cas où on veut également enregistrer la verticalité du son.

5.3. Décodage

Quelle que soit la manière par laquelle il a été enregistré, le *format B* obtenu doit être décodé avant d'être diffusé sur des enceintes. Il existe plusieurs approches très diverses du décodage (voir par exemple [9], [12] et [26]). Le décodage est d'une importance capitale dans un enregistrement de ce type, et le choix d'une méthode ou d'une autre conditionne le rendu sonore de manière considérable, en particulier pour des questions de restitution de l'espace sonore. La technique de décodage décrite ici suit les recommandations de [11] et de [26]. Notons qu'elle peut être utilisée pendant l'enregistrement ou en post-production.

Nous avons implémenté un prototype de décodeur générique de *format B* utilisant cette technique, adapté aux systèmes de diffusion horizontaux allant du 1/0/0 au 3/2/1. Ce décodeur a été programmé dans l'environnement *Max/MSP* [6] et pourra être bientôt téléchargé sur le site du CICM [3]. Notons qu'il existe d'autres décodeurs similaires gratuits [4] [21].

Notre décodeur est basé sur la synthèse de microphones coïncidents dits virtuels. Cette synthèse est effectuée à partir de mixages des canaux W, X, Y et Z, sachant qu'on peut ignorer Z si le système de diffusion est horizontal. Chaque microphone virtuel est défini par deux caractéristiques : sa directivité et son angle de positionnement. On peut dire, par exemple, qu'un microphone virtuel supercardioïde positionné à 90 degrés permet d'obtenir un signal très similaire à celui enregistré par un véritable microphone supercardioïde placé à cette position.

La synthèse des microphones virtuels repose sur deux principes :

- le mixage d'un microphone omnidirectionnel et d'un microphone en figure de "8" coïncidents permet d'obtenir un microphone virtuel ayant l'une des directivités suivantes (en fonction des coefficients de mixage) : omnidirectionnel, cardioïde large, cardioïde, supercardioïde, hypercardioïde, en figure de "8", ainsi que toutes les directivités intermédiaires ;
- le mixage de deux microphones en figure de "8" disposés à 90 degrés, par exemple l'un pointant vers l'avant et l'autre vers la gauche, permet d'obtenir un microphone virtuel en figure de "8" pointant dans n'importe quelle direction entre l'avant et la gauche (notons que pour obtenir toutes les directions en trois dimensions il suffit d'ajouter un troisième microphone coïncident en figure de "8", orthogonal aux deux autres).

En utilisant ces deux principes, on peut créer un microphone virtuel ayant une directivité et un angle arbitraires. Si on note M le signal associé à ce microphone, on peut écrire (si on se limite au plan horizontal) :

$$M = aW + b(cX + dY) \quad (3)$$

a, b, c et d étant quatre coefficients de mixage.

En théorie, on peut synthétiser en parallèle une infinité de microphones virtuels de ce type mais il faut garder à l'esprit qu'ils sont tous coïncidents.

L'intérêt de cette synthèse est de pouvoir s'adapter au système de diffusion. Dans notre décodeur, nous avons choisi de diffuser dans chacune des enceintes le signal correspondant à un unique microphone virtuel, ceci afin d'éviter les problèmes de *crossstalk* (diaphonie) entre les canaux [26].

Ainsi, pour une diffusion stéréophonique 2/0 (deux enceintes frontales), on utilise deux microphones virtuels (un par enceinte). Ceci correspond à la diffusion directe dans deux enceintes d'un enregistrement en couple XY avec deux microphones physiques coïncidents (un microphone par enceinte). Notons cependant qu'un avantage du décodage est de pouvoir varier l'angle entre les deux microphones ainsi que leurs directivités après l'enregistrement. Avec des microphones physiques, il faudrait au contraire refaire l'enregistrement (éventuellement avec des microphones différents).

Ce principe de décodage peut facilement être étendu aux diffusions multicanal, en choisissant des angles et des directivités appropriés pour chacun des microphones virtuels. Ainsi, notre décodeur fonctionne pour les systèmes de diffusion suivants : 1/0 (monophonie), 2/0 (stéréophonie à deux enceintes), 3/0, 2/2 (quadriphonie), 3/2 (*surround*), avec ou sans le caisson de basses.

Notons que pour le canal LFE éventuel, on utilise toujours le signal W seul, sur lequel il faut appliquer un filtre passe-bas si le caisson de basses n'en est pas équipé.

Un second avantage du décodage est donc de pouvoir diffuser un même enregistrement en *format B* (qui ne comprend que trois ou quatre pistes) sur n'importe quel système de diffusion. Cependant, pour les systèmes comprenant approximativement 8 enceintes ou plus (ceci en trois dimensions), il est souhaitable de passer à un encodage du champ acoustique plus précis (ce qui demande un système de prise de son et un décodage différents).

Toujours en agissant sur la synthèse de microphones virtuels, il est possible d'implémenter des effets de spatialisation : panoramique, rotation ou modification de l'élévation de la scène sonore par exemple (dans ce dernier cas, il faut obligatoirement disposer du canal Z). On peut aussi déplacer chaque ensemble coïncident de microphones virtuels plus ou moins loin de la source sonore enregistrée.

5.4. Caractéristiques des enregistrements *surround* coïncidents

Les caractéristiques d'une prise de son coïncidente en *format B* dépendent en grande partie des réglages du décodeur [25][26]. Dans le cas d'un décodage pour un système à deux enceintes, on se trouve dans la même situation que pour une prise de son en stéréophonie XY (coïncidente), avec des caractéristiques similaires.

Pour un décodage multicanal, on donne les caractéristiques suivantes (qui doivent être lues comparativement aux autres techniques de prise de son multicanal) :

- la **spatialité** (capacité à restituer l'espace dans lequel l'enregistrement a été effectué) est souvent limitée, mais peut être améliorée si on ajoute un couple de microphones omnidirectionnels espacés dans la prise de son ;
- la **localisation** (capacité à donner à l'auditeur la possibilité de déterminer la position des sources sonores dans l'espace où l'enregistrement a été effectué) est bonne ;
- l'**immersion** (capacité à donner à l'auditeur l'impression de se trouver, au moment où il écoute, dans l'espace où l'enregistrement a été effectué) est limitée, mais peut être améliorée si on ajoute un couple de microphones omnidirectionnels espacés dans la prise de son ;
- la **zone d'écoute satisfaisante** est petite, mais peut être agrandie si on ajoute un retard (de l'ordre de 10 millisecondes) entre les canaux avant et arrière, ce qui est possible dans la plupart des décodeurs ;
- la **compatibilité au *downmix*** (possibilité de diffuser un décodage prévu pour un système multicanal sur un système à deux enceintes ou une enceinte) est très bonne, même en mono ; cependant, on peut obtenir un "downmix" encore meilleur en faisant un décodage différent adapté à un moins grand nombre d'enceintes (par exemple, on ne garde souvent que W pour une diffusion en mono) ;

- si la prise de son ne comprend pas de microphone omnidirectionnel, il arrive que les basses fréquences soient mal reproduites ; le fait d'ajouter un microphone omnidirectionnel permet une reproduction parfaite des basses, et si on désire aussi reproduire leur espace sonore on peut recommander l'ajout d'un couple de microphones omnidirectionnels espacés ; rappelons enfin que les microphones de *format A* utilisent souvent une égalisation spectrale, réalisée en amont de l'encodage en *format B*, qui permet de corriger le manque de basses fréquences dû à la directivité cardioïde des capsules.

Il va de soi que ces caractéristiques sont très générales et varient largement en fonction des contextes d'enregistrement, de production en studio et de diffusion. Cependant, on donne souvent des caractéristiques similaires pour les prises de son coïncidentes utilisant deux microphones (stéréophonies XY et M/S).

6. CONCLUSION

Cet article, outre le fait d'analyser les différentes techniques permettant de créer un flux multicanal, s'est proposé d'étudier les moyens de stockage existant afin, d'une part, de garantir la fidélité de l'œuvre originale (en préservant la meilleure qualité possible) et, d'autre part, de rendre accessible ce contenu à un large public, donc de le diffuser à travers des supports physiques standard.

Il existe aujourd'hui des supports de données numériques standard permettant de stocker un très grand volume de données et de les lire avec un débit suffisant. Ainsi, pour la première fois, il n'y a plus aucune limite technologique imposant de compresser — avec une perte de qualité — un signal sonore multicanal afin qu'il tienne sur des supports standard et qu'il puisse être lu par des dispositifs standard.

D'ailleurs, dans le contexte du cinéma numérique, le DCI recommande que le son ne soit pas compressé de la création du master numérique jusqu'à sa diffusion en salle, pour des raisons de qualité mais aussi de simplicité, puisque certains processus d'encodage et de décodage ne sont ainsi plus nécessaires.

Le fait qu'un format soit propriétaire ou ouvert n'a rien à voir avec la qualité du son. Cependant, un format ouvert, c'est-à-dire un format ne dépendant pas d'un logiciel spécifique et dont les spécifications techniques sont publiques, sans restriction d'accès ni de mise en œuvre, permet par sa transparence une réutilisation dans la recherche et le développement audionumérique par un grand nombre d'utilisateurs. Au contraire, un format propriétaire n'est connu que par ses créateurs, qui peuvent ainsi l'exploiter commercialement sans être concurrencés (à condition que ce format devienne un format standard utilisé par la communauté professionnelle). Ainsi, le label *Dolby* est souvent un argument commercial évoquant auprès du grand public une pseudo-garantie de qualité, parfois proche du fétichisme [1].

Cependant, pour qu'un format ouvert s'impose, il faut qu'il soit simple et utile. En effet, de nombreux nouveaux formats ouverts n'apportent aucune valeur ajoutée par rapport aux formats déjà existants (ouverts ou propriétaires); de plus un trop grand nombre de formats complique les échanges de fichiers. Enfin, un format doit être reconnu par la communauté professionnelle comme un standard.

Dans ce contexte, la question des formats ouverts ou propriétaires reste en suspens. Plutôt que de préconiser l'un ou l'autre de ces choix, spécifier la qualité requise pour les essences sonores est essentiel pour le musicien, et surtout possible, tant les capacités de stockage et de débits de lecture des supports numériques actuels sont élevés.

7. RÉFÉRENCES

- [1] Adorno T.W., *Le caractère fétiche dans la musique*, Paris, 2001.
- [2] Benjamin E., Chen T., « The native B-format Microphone, Part I », *119th AES Convention*, New York, 2005, Preprint No.6621, 2005.
- [3] CICM, Centre de recherches informatiques et création musicale, Université de Paris VIII, MSH Paris Nord, <http://cicm.mshparisnord.org/>
- [4] Courville D., *plug-ins B2X*, <http://www.radio.uqam.ca/ambisonic/b2x.html>
- [5] Craven P. G. , Gerzon M. A., *Coincident microphone simulation covering three dimensionnal space and yielding various directionnal outputs*, United States Patent 4042779, 1977.
- [6] <http://www.cycling74.com/products/maxmsp>
- [7] Daniel J., *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia*, Université de Paris VI, 2000.
- [8] *Digital Cinema Initiative*, www.dcinovies.com
- [9] Di Liscia O. P., *Sound spatialisation using Ambisonic*, 2001.
- [10] DPA microphones A/S, *Microphone university*, <http://www.dpamicrophones.com>
- [11] Farrar K., schémas reproduits dans [22], pp. 206-208
- [12] Farina A., Ugolotti E., « Software Implementation of B-Format Encoding and Decoding », *109th AES Convention*, 1998.
- [13] Farina A., *A-format to B-format conversion*, 2001. <http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm>
- [14] Gerzon M. A., « Periphony: with-height sound reproduction », *JAES*, 21, pp. 2-10, 1973.
- [15] Hauptmikrofon web site, « Comparison of 7 surround main microphones », *ORF seminar 2001* <http://www.hauptmikrofon.de/orf.htm>
- [16] *ITU-R BS.775, spécification du format 5.1*, www.itu.int/md/R03-SG06-C-0260/en
- [17] Laborie A., Bruno R., Montoya S., « A New Comprehensive Approach of Surround Sound Recording », *114th AES Convention*, 2003.
- [18] [18] Laborie A., Bruno R., Montoya S., « High Spatial Resolution Multichannel Recording », *116th AES Convention*, 2004.
- [19] Laborie A., Bruno R., Montoya S., « Designing High Spatial Resolution Microphones », *117th AES Convention*, 2004.
- [20] Merlier B., « À la conquête de l'espace », *JIM 98*, CNRS-LMA Marseille, 1998.
- [21] McGriffy D., logiciel et plug-in *Visual Virtual Microphone*, <http://mcgriffy.com/audio/ambisonic/vvmic/> et <http://mcgriffy.com/audio/ambisonic/vvmicvst/>
- [22] Rumsey F., *Spatial Audio*, ed. Focal Press, 2004.
- [23] Sedes A. (sous la direction de), *Espaces sonores, actes de recherche*, EMT, MSH Paris Nord, Paris, 2003.
- [24] Sedes A., *De la notion d'espace sonore à la visualisation du son, un parcours musicologique, technologique et artistique*, thèse d'Habilitation à Diriger des Recherches, université de Paris VIII, Saint-Denis, 2007.
- [25] Schoeps GmbH, *Overview of surround recording methods*, 2006. http://www.schoeps.de/PDFs/SCHOEPS_surround-brochure.pdf
- [26] Wittek H., *Double M/S - a Surround recording technique to put to test*, 2006, http://www.schoeps.de/E-2004/PDFs/Schoeps_DoubleMS_Paper.pdf

L'ensemble des sites Internet cités a été consulté pour la dernière fois en janvier 2008.