

Utilisation de techniques binaurales et transaurales en prises de son et en post-productions multicanales 5.1.

Conférence AES - Sao Paulo – Brésil – mai 2009

Jean-Marc LYZWA : Ingénieur du son

Conservatoire national supérieur de musique et de danse de Paris (Cnsmdp)

Alexis BASKIND : Réalisateur en informatique musicale et ingénieur du son

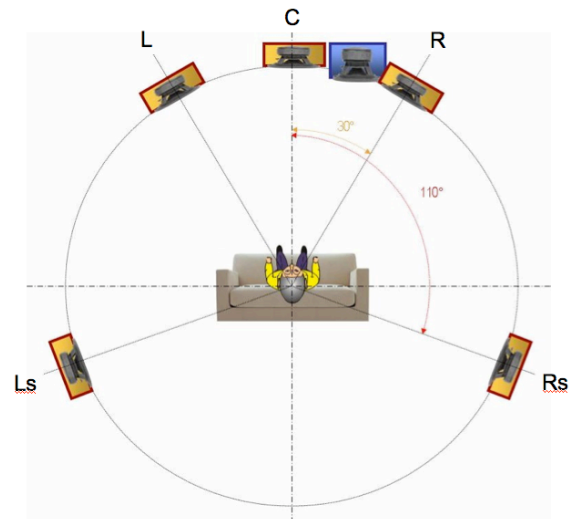
<http://alexisbaskind.net>

Introduction

Depuis déjà plusieurs années, une réflexion est menée au Cnsmdp sur les techniques de prise de son et de post-production à mettre en oeuvre afin de créer une immersion sonore plus réaliste pour l'auditeur en ce qui concerne la musique enregistrée. L'objectif est d'améliorer, de préciser et de stabiliser la sensation d'enveloppement tout en renforçant la lisibilité et la compréhension de l'œuvre restituée.

Les différentes évolutions du son cinématographique et des techniques audionumériques ont conduit à la normalisation d'un format : le système multicanal 5.1. Il constitue aujourd'hui la référence commerciale des systèmes de restitution en multicanal. Il est largement distribué et les ventes de systèmes dédiés augmentent sans cesse avec le passage progressif aux téléviseurs à écrans plats. Les grandes enseignes de matériel HiFi proposent sur le marché une palette de systèmes 5.1, relativement diversifiée en termes de qualité et de prix.

Initialement associée exclusivement à l'image, cette norme 5.1 obéit à une logique frontale de la scène sonore, restituée sur le secteur LCR.



Le système 5.1

Elle est extrêmement plus fragile si cette scène musicale est ouverte, voire éclatée dans l'espace et répartie autour de l'auditeur.

C'est précisément ces cas particuliers qui suscitent notre intérêt.

Obtenir un enveloppement stable du champ sonore s'avère indispensable pour la HDTV et

plus particulièrement pour la 3D Full HD, mais également pour la restitution d'œuvres musicales spatialisées. La WFS et le 22.2 difficilement exploitable dans un salon, le 5.1 reste par conséquent le système de restitution multicanal le plus approprié et le mieux adapté pour le grand public.

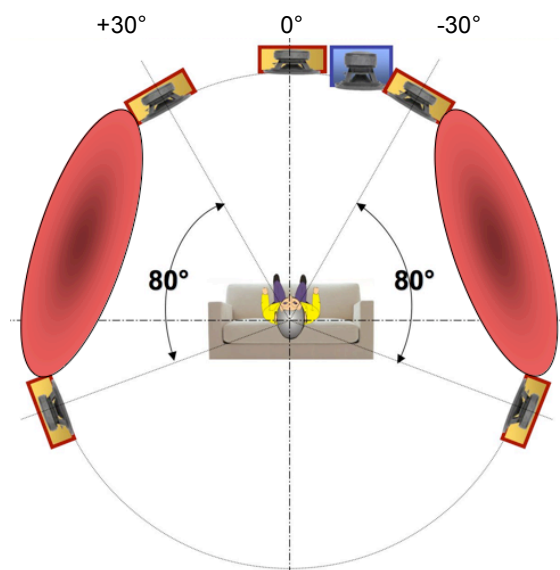
En comparaison avec la stéréophonie à deux canaux, le système multicanal 5.1 se caractérise par deux améliorations essentielles : l'une favorisant la restitution d'un enveloppement réaliste, l'autre permettant une localisation plus ou moins précise des sources sonores à l'intérieur de l'espace restitué.

Le système 5.1 permet donc de restituer de manière optimum une scène frontale, en plaçant l'auditeur dans la perspective d'une écoute réaliste en salle. Dans ce cas, l'auditeur est enveloppé par la réverbération du lieu d'enregistrement ou par une réverbération artificielle. La scène sonore est alors limitée sur le secteur frontal.

Le système 5.1 est également censé permettre la restitution d'une œuvre éclatée dans l'espace et par conséquent d'une image sonore enveloppante sur 360°.

Cependant, le système 5.1 reste par bien des points un système de restitution imparfait par son manque d'homogénéité.

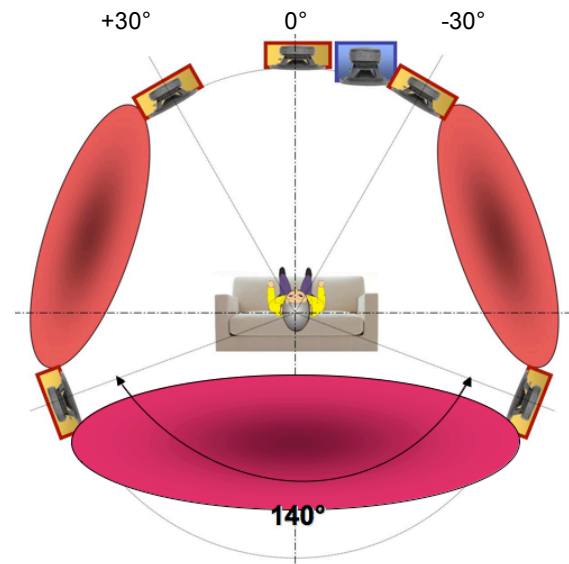
Une instabilité de localisation des sources concerne tout d'abord les secteurs latéraux du système. La distance des deux groupes de haut-parleurs entre le secteur frontal et le secteur arrière est importante, son ouverture est de 80°.



Limitations de la norme ITU 5.1 - Secteurs latéraux

Les sources sonores restituées dans ces secteurs sont des sources fantômes produites par un point d'émission avant ($\pm 30^\circ$) et un point d'émission arrière ($\pm 110^\circ$). Excentré par rapport à notre axe d'écoute (0°), notre système auditif rencontre des difficultés pour la perception cohérente de ce type de signal d'informations.

Avec une ouverture très importante de 140° , le secteur arrière est quant à lui extrêmement instable.



Limitations de la norme ITU 5.1 - Secteur arrière

Si l'on s'écarte latéralement de l'axe d'écoute, la source sonore restituée à l'arrière bascule très rapidement sur le haut-parleur arrière le plus proche. Le 6.1, avec un canal central arrière peut alors constituer une solution plus stable.

Notre processus de perception spatial est plus précis à l'avant et sur le plan horizontal que sur les cotés et à l'arrière. Le système de restitution 5.1 repose donc sur une certaine incohérence en l'absence de l'image.

Pour le secteur frontal, le secteur où l'oreille en localisation est la plus précise, on se retrouve avec trois points de diffusion, concentrés sur 60° (L (30°) / C (0°) / R (-30°)). On obtient par conséquent une très bonne stabilité pour ce secteur.

Seuls deux points de diffusions positionnés à $\pm 110^\circ$ à l'arrière (Ls (110°) / Rs (-110°)), assurent un relais avec le secteur avant pour couvrir les secteurs où l'oreille est la moins précise : les cotés et l'arrière. L'ensemble de cette zone fragile d'un point de vue perceptif couvre tout de même un secteur de 300° .

Conscient des défauts que compte le système multicanal 5.1 spécifiquement développé à l'origine pour l'audiovisuel et fixé par la norme ITU¹, la prise de son et la post-production à mettre en œuvre pour ce système imposent donc de nouvelles méthodes de travail, mais peuvent également conduire à une nouvelle orientation esthétique de l'image sonore.

L'orientation du projet s'articule sur l'association de différentes techniques propres à la prise de son et innovantes en post-production afin de proposer à l'auditeur une restitution de l'œuvre enregistrée enveloppante, stable, et homogène.

Notre culture de la scène sonore en musique, imposée par l'architecture de la plupart des salles, fixe de manière générale une représentation essentiellement frontale. La stéréophonie à 2 canaux installe l'auditeur efficacement dans cette représentation.

Les productions musicales en multicanal 5.1 de manière générale ont très souvent tendance à restituer de manière optimum une scène frontale, en plaçant l'auditeur dans la perspective d'une écoute réaliste en salle. L'auditeur est dans ce cas, enveloppé par la réverbération du lieu d'enregistrement ou par une réverbération artificielle. La sensation de profondeur de la scène sonore est généralement accentuée. Et le canal central vient renforcer la stabilité de la scène sonore.

Or, le multicanal 5.1 offre la possibilité d'orienter l'auditeur vers une nouvelle lecture de l'enregistrement et de l'œuvre musicale. D'une restitution réductrice de l'espace à 60° en stéréophonie à deux canaux, le multicanal 5.1 permet un déploiement de l'espace sur 220° voire sur 360° avec plus ou moins de stabilité.

Les sources sonores reproduites ne sont plus concentrées et délimitées sur le secteur frontal, mais peuvent s'étaler dans l'espace. On limite de ce fait les effets de masquage. La restitution en multicanal 5.1 permet d'adapter une image sonore en extra largeur, elle propose donc une séparation et une meilleure localisation spatiale des différents instruments qui composent la formation musicale enregistrée.

D'un point de vue esthétique, l'approche recherchée consiste à placer l'auditeur au plus près de ce que peut percevoir le chef d'orchestre lorsqu'il dirige et ceci dans une acoustique réaliste, restituée ou artificielle.

La direction de travail formulée repose sur l'hypothèse que le compositeur dans l'action de

l'écriture développe une image spatiale mentale très proche de ce que peut percevoir et ressentir le chef d'orchestre lorsqu'il dirige.

À la suite de nombreux entretiens avec des compositeurs, eux-mêmes chef d'orchestre, il en résulte que c'est très fréquemment le cas. Aujourd'hui, il reste à déterminer si cette hypothèse peut être généralisée.

À travers cette nouvelle restitution sonore c'est donc une nouvelle perception de l'œuvre enregistrée qui est proposée à l'auditeur.

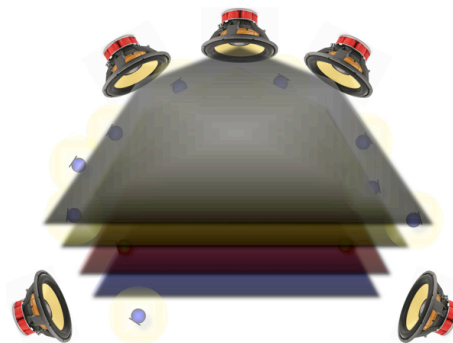
En élargissant l'espace stéréophonique, on améliore considérablement la clarté et la précision de l'œuvre musicale restituée. L'analyse de l'enregistrement dévoile plus précisément l'intention compositionnelle et la conception du compositeur.

Le multicanal 5.1 offre le choix de nouvelles directives pédagogiques et propose une perception novatrice de la scène sonore restituée, une analyse plus détaillée de l'œuvre enregistrée, une approche et un ressenti perceptif différent de l'œuvre musicale.

Il peut permettre à l'auditeur de décrypter plus facilement toutes les subtilités de l'écriture musicale.

L'approche vise à stabiliser et préciser les sources présentes dans ces zones sectorielles qualifiées de fragiles en créant un enveloppement immersif réaliste et stable dans le plus grand respect de la restitution des timbres.

Pour cela, il est par conséquent fondamental de concevoir de nouveaux outils de spatialisation pour la prise de son et la post-production, visant à pallier les défauts de la norme 5.1 en termes de stabilité et de précision de localisation latérale et arrière, avec la mise en œuvre d'une méthode construite sur la superposition de différentes couches, une combinaison de techniques communément utilisées en prise de son. C'est ce que nous nommerons par techniques mixtes ou combinatoires.

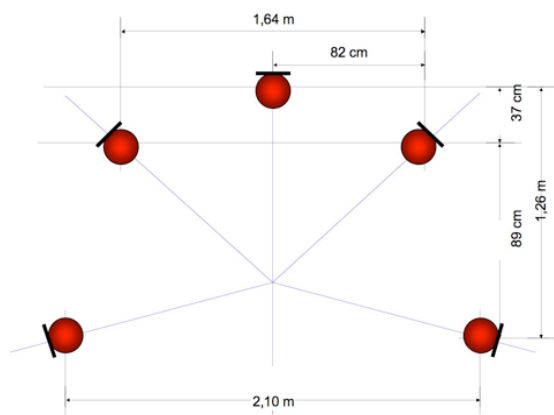


Superposition de différentes techniques de spatialisation

¹ Recommandation ITU-R BS.775

1^{ère} couche : C'est tout d'abord la configuration d'un quintuplé microphonique. Ce système reproduit l'empreinte acoustique de la scène sonore à restituer. Il structure l'espace. Il est le système principal dans le processus de prise de son multicanal.

Ce quintuplé microphonique est constitué de 5 omnidirectionnels (4041 ou 4006 DPA). Il est spécifiquement assemblé pour le système 5.1 normalisé. L'assemblage de ce quintuplé repose sur la théorie des angles de prise de son. Le point de départ est de considérer que la prise de son multicanal 5.1 est une extension de la prise de son stéréophonique à 2 canaux. On se retrouve avec un système d'assemblage de type MMAD (Multichannel Microphone Array Design), exclusivement réalisé avec des omnidirectionnels.



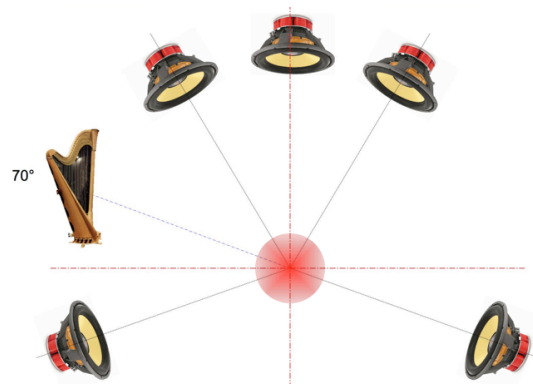
Quintuplé microphonique

Le principe est d'éviter tout chevauchement des angles de prise de son. Ce qui permet de raccorder parfaitement les couvertures frontale, latérales et arrière.

Si le système recrée des chevauchements, il apparaît une distorsion angulaire, un tassement du centre du segment. Ce trou central entraînera une condensation des informations sur les haut-parleurs.

Le choix de microphones omnidirectionnels quant à lui, se justifie pour le rendu fréquentiel très naturel et régulier dans l'axe du microphone mais également hors de son axe. Les problèmes liés à la diaphonie sont plus ou moins lissés. C'est bien entendu un choix esthétique personnel.

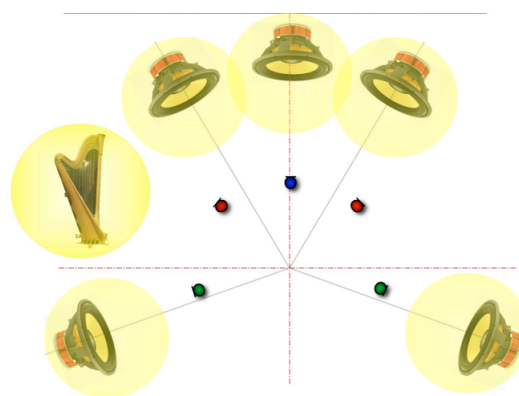
Prenons par exemple une source sonore positionnée physiquement à 72° par rapport à l'auditeur.



Exemple : restitution d'une harpe positionnée à 70° de l'auditeur

L'image d'une source sonore, captée par le quintuplé microphonique est restituée par l'ensemble des 5 haut-parleurs.

Cette source se retrouve localisée dans le système de diffusion très grossièrement à la place qu'elle occupe par rapport au système de captation. Mais ce système à lui seul reproduit une image relativement instable suivant la distance de la source sonore à capter.



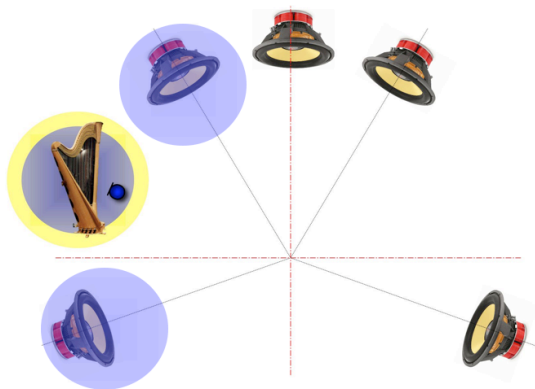
1^{ère} couche - le quintuplé microphonique

Le quintuplé microphonique omnidirectionnel peut évidemment être remplacé par tout autre système de captation multicanal. Le choix de ces systèmes est actuellement extrêmement diversifié. Nous pouvons citer entre autres : les systèmes basés sur le principe ambisonique comme les microphones Soundfield ou DPA, le système Trinov Audio, le système HOA, les systèmes basés sur le principe binaural, avec une tête artificielle, les systèmes basés sur le principe du double MS, avec le système schoeps, le MMAD, la croix IRT, l'OCT surround de Schoeps, l'ASM 5 de Brauner, le Holophone H3D, ...

Dans la pratique, il est peu fréquent d'envisager une prise de son utilisant un système de

captation unique. L'utilisation d'appoints microphoniques s'impose dans la plupart des projets de prise de son. Mais alors, comment gérer ces appoints dans une production en 5.1 ?

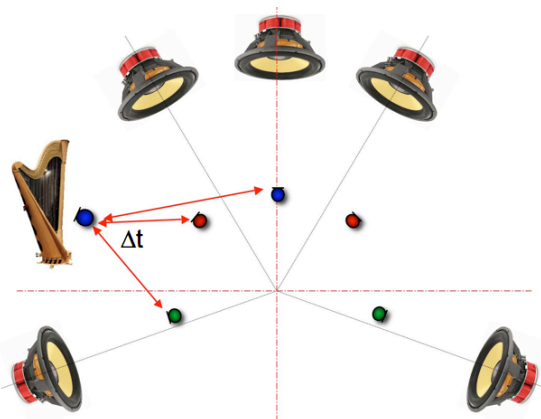
2^{ème} couche : Les microphones d'appoints constituent la seconde couche de post-production. Ils vont être mixés sur l'image du quintuplé. Ils sont positionnés dans l'espace restitué 5.1 au panoramique d'intensité. Ils précisent la localisation et définissent les différentes sources sonores pointées. L'appoint de la source sonore à souligner sur notre exemple est mixé au panoramique sur le secteur latéral gauche.



2^{ème} couche - les appoints

Cependant, la harpe captée par l'appoint arrive en avance de son image sonore restituée par le quintuplé microphonique, de par la distance qui les sépare.

Chaque appoint est par conséquent, recalé en temps et donc décalé par rapport au système dit principal. L'image sonore globale restituée du mixage entre le quintuplé et les différents appoints devient alors fusionnelle et cohérente.



L'appoint est remis en temps par rapport au quintuplé

Son image est calée sur l'image de cette source restituée par le quintuplé microphonique.

Cet appoint va préciser et consolider la position de la source. On peut pivoter la tête pour regarder cette image sonore : la source sonore soulignée est alors toujours perçue à son azimut d'origine, si l'on est situé proche du sweet spot. Cette construction reste fragile suivant la position d'écoute adoptée.

Le principe est simple, il consiste à ajuster l'appoint microphonique par rapport au quintuplé microphonique et de le positionner dans l'espace au panoramique d'intensité pour que les 2 images spatiales coïncident, ce qui permet d'apporter une précision spatiale et une précision spectrale en limitant toute forme d'incohérence.

3^{ème} couche : Elle est réalisée en post-production à l'aide d'un outil de traitement du signal. C'est l'élément nouveau introduit. Elle combine 2 traitements transauraux qui visent à stabiliser les zones dites fragiles du système de restitution 5.1, les latéraux et éventuellement l'arrière. Cet outil de traitement transaural est spécifiquement adapté au système de restitution multicanal normalisé.

Ce sont les signaux des appoints à spatialiser sur les secteurs latéraux qui sont traités.

Le traitement transaural de l'appoint se combine à l'image sonore restituée en panning d'intensité de ce même appoint.

Ces deux couches transaurales viennent se superposer et se combiner sur l'image sonore de l'instrument souligné par l'appoint et restituée par le quintuplé microphonique. Cette méthodologie a pour rôle de fixer, figer et structurer l'image sonore à l'intérieur de l'espace de restitution 5.1. C'est le principe de ce que nous nommerons : technique combinatoire.

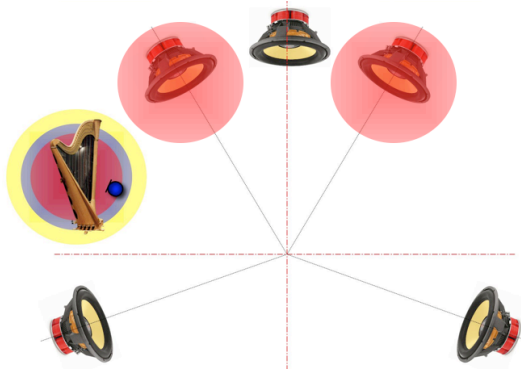
L'utilisation du traitement transaural permet de dépasser les limitations du 5.1, notamment l'instabilité perçue lors des déplacements avant/arrière et le manque de précision latérale et arrière.

L'utilisation du panoramique d'intensité permet de dépasser les limitations du transaural, en particulier l'instabilité engendrée par rotation de la tête, mais également le repliement partiel de la scène vers l'avant ainsi que la coloration spectrale pouvant être générée avec ce type de traitement.

Afin de limiter l'instabilité spatiale que peut engendrer le traitement transaural, notamment lors des mouvements de la tête, un double traitement transaural est appliqué.

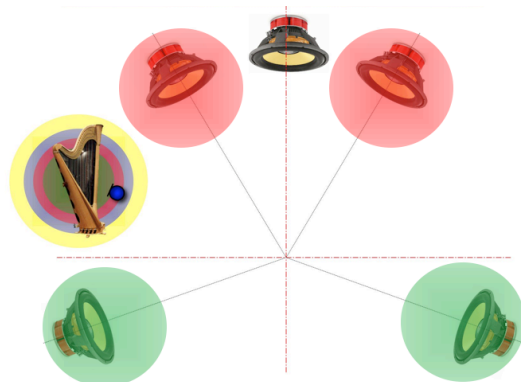
- Un premier traitement est appliqué sur le secteur avant.

L'azimut du traitement est calé sur la position de l'appoint donné par le panoramique d'intensité. La superposition des 2 couches, l'une issue du traitement transaural et l'autre du panoramique d'intensité permet de stabiliser l'image sonore de la source soulignée.



Double traitement transaural - frontal

- Un second traitement transaural est appliqué sur le secteur arrière. Il vient renforcer la stabilité de l'image de la source sonore soulignée. Cette technique fixe l'instrument souligné et contribue à la perception d'un enveloppement précis et réaliste.



Double traitement transaural : frontal & arrière

L'auditeur peut alors regarder et suivre de manière analytique une source isolée de la scène sonore sans déformation de l'espace restitué.

Une fois mixées, ces différentes techniques de spatialisation favorisent la construction d'une image sonore globale plus stable. La zone d'écoute s'en retrouve élargie.

4^{ème} couche : Les réverbérations artificielles en multicanal constituent la dernière couche. Elles permettent d'intégrer au mieux les différents éléments qui viennent se superposer. Elles

permettent de renforcer la notion d'espace et d'enveloppement.

la superposition des différentes couches de mixage augmente naturellement le nombre de paramètres de réglages et requiert un ajustement précis de ces différentes couches à la fois temporellement et spatialement.

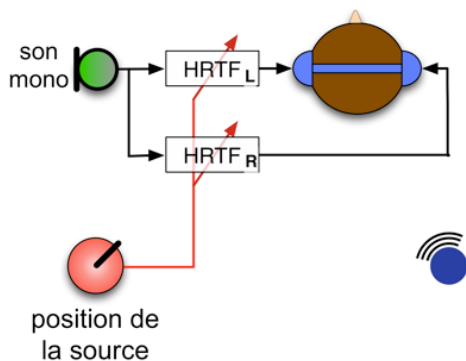
En particulier, en vertu de la loi du premier front d'onde qui détermine la localisation perçue des sources sonores il convient de contrôler avec précision l'ordonnancement temporel des informations délivrées respectivement par le système de prise de son principal (quintuplé, réseau microphonique, système ambisonic soundfield ou HOA,...), les traitements en panoramique d'intensité des différents appoints et les traitements transauraux effectués sur les microphones d'appoint de sorte à stabiliser l'image sonore à l'intérieur du système de restitution. De cet ordonnancement découleront simultanément la qualité de l'image spatiale (localisation et stabilité en azimut, proximité, élévation et stabilité) et la qualité spectrale.

Le Traitement transaural

Le traitement binaural / transaural permet de dépasser les limitations du panoramique 5.1. Il stabilise la source sonore sur les latéraux par rapport aux déplacements avant/arrière de l'auditeur.

Cette technique, dont le principe est connu depuis les années 60 (première implémentation pratique brevetée par Atal et Schroeder en 1966), est constituée de deux éléments en cascade :

- la synthèse binaurale vise à simuler le trajet acoustique entre une source sonore placée réellement autour de l'auditeur et chacune de ses deux oreilles, afin de restituer au mieux les caractéristiques (différences interaurales de spectre et de temps) requises par le système auditif pour évaluer la position de la source. Cette simulation est effectuée par filtrage de la source monophonique à traiter, les filtres étant mesurés en chambre sourde avec un sujet réel ou une tête artificielle, pour le plus grand nombre de positions possible. Le résultat de la synthèse binaurale est donc destiné à une écoute au casque, car l'écoute via des haut-parleurs introduit elle-même des trajets acoustiques parasitant les indices spatiaux



Principe de la synthèse binaurale

- le traitement transaural vise à restituer un signal binaural sur des enceintes acoustiques, par la correction des trajets acoustiques mentionnés ci-dessus. En particulier, il s'agit de chercher au mieux à *annuler les trajets croisés* (« cross-talk cancellation ») qui sont les principaux responsables de la perturbation des indices spatiaux.

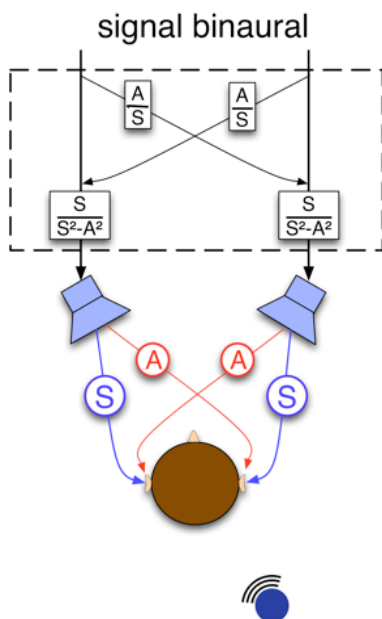


Schéma de principe du traitement transaural, implémentation symétrique

La conjonction de ces deux techniques permet donc en théorie une restitution des sources au-delà du traditionnel espace stéréophonique. Ainsi, il est possible de simuler la présence d'une source latérale avec des haut-parleurs frontaux ou arrières. En pratique, les points faibles de son application sont :

- sa sensibilité à la géométrie et aux caractéristiques acoustiques de la tête et du

torse de l'auditeur : pour une restitution optimale, il faudrait que chaque mixage soit adapté individuellement à chaque auditeur. En pratique c'est bien sûr impossible, donc on procède en utilisant une base de données de filtres convenant au plus grand nombre de gens possible. L'effet principal d'une inadéquation des filtres à l'auditeur est une perte d'efficacité de l'externalisation.

- la coloration entraînée par la multiplicité des filtres et les traitements par retards. Celle-ci est également partiellement compensée par l'utilisation en parallèle du panoramique d'intensité.
- sa sensibilité aux mouvements de l'auditeur : si celui-ci tourne la tête, ou se déplace de gauche à droite, l'effet est rapidement perdu. Ce défaut est compensé par l'utilisation conjointe avec un panoramique d'intensité, qui permet à la scène sonore de conserver sa cohérence.

En revanche la restitution transaurale est beaucoup plus robuste aux mouvements avant/arrière que ne l'est le panoramique d'intensité dans un environnement 5.1. L'utilisation conjointe des deux techniques vise donc à pouvoir compenser les défauts de l'une avec les atouts de l'autre.

Conclusion

D'innombrables débats se posent et s'opposent depuis plusieurs années déjà sur les différentes techniques de prise de son conciliables avec le système de restitution irrégulier que constitue le 5.1. De nombreuses théories ou réflexions autour de la captation sont développées et chacune se défend de présenter le meilleur système, la solution la plus adaptée pour retranscrire l'espace sonore de manière la plus réaliste et naturelle qu'il soit. Peu d'entre elles cependant, explorent l'ensemble des différentes techniques utilisées réellement par l'ingénieur du son dans la réalisation d'une production, notamment la mise en oeuvre et la gestion des microphones d'appoint, mais également les différentes méthodes de mixage et de la post-production dans leur ensemble.

En tenant compte des nombreux défauts du système de restitution décrits précédemment,

l'intention de ce projet est d'adapter et de créer des techniques de prise de son et de post-production aptes à améliorer la perception d'enveloppement, la précision et la stabilité de localisation sur les secteurs latéraux et sur l'arrière.

Aucun système n'étant réellement optimal pour la totalité des différents paramètres entrant en ligne de compte pour une production particulière dans une acoustique précise, les faiblesses d'une approche doivent être compensées par les qualités que peuvent apporter les autres techniques utilisées.

Les inconvénients d'une technique, vont sans cesse être compensés par les qualités combinées d'une autre, c'est le principe même de cette approche dite de techniques mixtes ou combinatoires.

La mise en œuvre de ces différents procédés vise à tirer le meilleur parti possible des contraintes imposées par la norme 5.1.

Les principales améliorations constatées mettent en évidence une très nette amélioration de la stabilité latérale, en particulier lors des mouvements avant et arrière de l'auditeur. La sensation d'enveloppement devient beaucoup plus homogène. On constate un élargissement de la zone d'écoute ainsi qu'un meilleur détachement des différents plans sonores.

Cependant, il n'est pour l'instant pas encore possible d'améliorer franchement la stabilité et la précision sur le secteur arrière.

Notre système d'analyse auditif permet de focaliser et de diriger notre écoute à l'intérieur ou sur l'ensemble de la scène sonore réelle ou reproduite.

Notre système de perception auditif est capable de naviguer à l'intérieur d'une image sonore restituée. Par cela rien ne doit faire obstacle entre le passage d'une écoute globale d'une scène sonore et l'écoute suivie d'un instrument ou d'un groupe instrumental à l'intérieur même de l'orchestre. L'ingénieur du son doit éviter d'imposer une image figée de la scène sonore en soulignant tel ou tel instrument par rapport à un autre.

Le mixage se résume en un subtil équilibre dans l'assemblage de différentes techniques de prise de son, dosées de telle manière qu'elles donneront à l'œuvre musicale restituée une lisibilité optimale et à l'auditeur la plus grande liberté d'écoute.

Chaque œuvre musicale est écrite avec des considérations spatiales particulières. Le principe de la restitution en multicanal 5.1 est

intimement lié à la sensation d'immersion et de reproduction d'une acoustique réaliste du lieu d'enregistrement ou artificielle. De ce point de vue, l'ouverture spatiale promise par la norme 5.1 ITU se doit d'apporter une évolution considérable par rapport à la stéréophonie à deux canaux.

jmlyzwa@cnsmdp.fr

baskind@no-log.org