

# Utilisation de techniques binaurales et transaurales en production multicanal 5.1

Alexis Baskind (1), Jean-Marc Lyzwa (2), Xavier Meunier (3)

- (1) IRCAM, 1, place Igor-Stravinsky, 75004 PARIS
- (2) CNSMDP, 209 avenue Jean Jaurès 75019 Paris
- (3) ENSLL, 7, allée du Promontoire, 93161 Noisy-Le-Grand

## 1/ Introduction

Nombre de standards multicanal proposés depuis 1980 sont issus du monde du cinéma, et à ce titre sont destinés à ajouter à la scène frontale une ambiance enveloppante diffuse. La norme 5.1 offre des possibilités supplémentaires, la présence de deux haut-parleurs arrières indépendants permettant de placer des sources virtuelles localisables tout autour de l'auditeur, et donc d'élargir l'espace stéréophonique par rapport à un système usuel à deux canaux. Ceci permet d'envisager des esthétiques de production radicalement différentes. Par exemple, le parti pris adopté lors du mixage de Répons (présenté également lors de ce forum) qui motive le travail présenté ici, consiste à développer une scène sonore cohérente avec ce qu'offrirait la place du chef d'orchestre, et à ce titre nécessite de pouvoir placer les solistes, les groupes instrumentaux et les sources électroacoustiques autour de l'auditeur.

Ces atouts sont néanmoins fréquemment nuancés : en effet, la disposition géométrique des haut-parleurs autour de la zone d'écoute est inhomogène, et privilégie malgré tout toujours les positions frontales par rapport aux positions latérales et arrières, car l'écart angulaire entre les haut-parleurs est très important (respectivement 80 et 140 degrés). Ceci est d'autant plus dommageable que l'effet de localisation par sommation [1], qui à la base de la stéréophonie, nécessite au contraire une résolution angulaire plus élevée sur les côtés que dans la zone frontale de l'auditeur. En pratique, cela se traduit fréquemment par une instabilité des sources qui ne sont pas positionnées au voisinage des haut-parleurs, dès lors que l'on a recours à des techniques basées uniquement sur la stéréophonie de temps ou d'intensité (réseau de microphones ou positionnement de sources monophoniques par panoramique d'intensité).

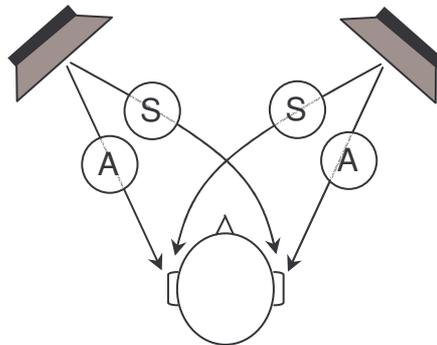
Les nombreuses recherches portant sur d'autres procédés de restitution spatiale dans le cadre du standard 5.1 prennent donc tout leur sens. Dans le cadre de ce projet, l'accent s'est porté sur le couple synthèse binaurale/traitement transaural, et sur les possibilités spécifiques offertes par un système de diffusion multicanal. A première vue, on peut se demander quels peuvent être les apports de cette technique connue de longue date : en effet, le traitement transaural, qui est utilisé avec succès dans des configurations à deux canaux (stéréophonie standard, stéréodipôles) pour élargir l'espace stéréophonique au-delà de la zone délimitée par les haut-parleurs, souffre néanmoins elle aussi de défauts liés à l'instabilité spatiale des sources : un déplacement latéral ou une rotation trop importante de l'auditeur par rapport à la position d'écoute optimale (*sweet spot*) entraîne inévitablement des biais sur les positions des sources, un repli de la zone arrière de la scène, très fragile, sur la zone frontale, et des problèmes de coloration. En revanche la technique est relativement stable à des déplacements vers l'avant ou l'arrière.

Ceci dit, le fait de disposer de plus de deux canaux, et notamment de sources physiques à l'arrière de l'auditeur, permet de reconsidérer dans une large mesure ces limitations. Nous nous sommes ainsi efforcés dans ce projet à développer des méthodes originales de positionnement de sources virtuelles en restitution 5.1 à partir de techniques transaurales et binaurales. Après un rappel des principes du traitement transaural, sont brièvement présentées quelques axes permettant de tirer parti des spécificités d'une configuration 5.1.

## 2/ Principe du traitement transaural

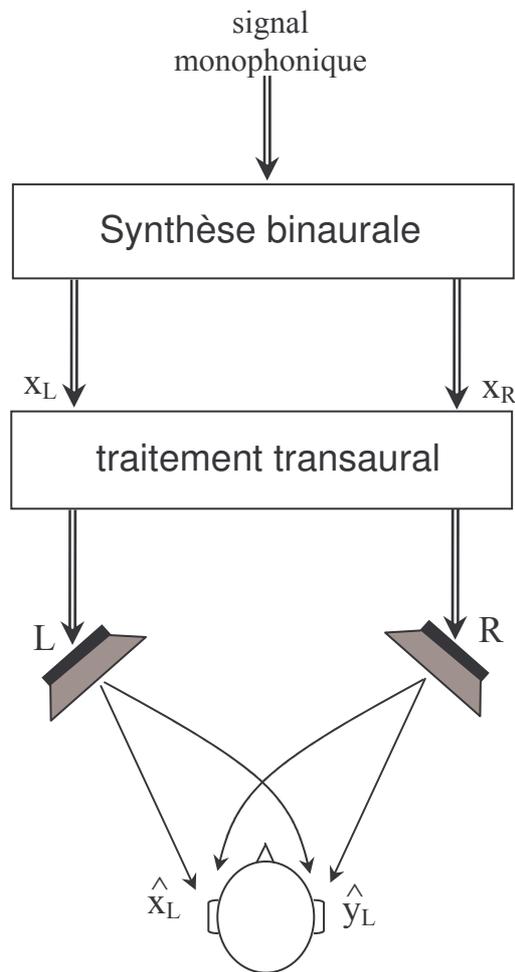
Dans le cadre d'une restitution à deux canaux, est apparue dès les années 1950 l'idée selon laquelle la méthode de restitution stéréophonique optimale consiste à présenter à l'auditeur des signaux de pression similaires à ceux qu'auraient captés chacun de ses tympans si celui-ci avait été présent lors de l'enregistrement. Cette intuition est à la base des techniques dites binaurales, se déclinant en deux catégories : les *enregistrements binauraux*, utilisant des têtes artificielles plus ou moins réalistes, et, lorsque l'évolution technique le permet, les méthodes de *synthèse binaurale*, visant à simuler cette même situation sur des signaux monophoniques au moyen de mesures effectuées en laboratoire, sur têtes artificielles ou sujets humains.

Le frein principal à l'essor de ces techniques est le fait que celles-ci sont naturellement vouées à une écoute au casque, qui permet de contrôler efficacement les signaux envoyés à chacun des tympans de l'auditeur. En situation d'écoute par haut-parleurs, l'information sonore destinée à chaque oreille est perturbée par la présence des ondes de contournement ou *trajets croisés* (figure 1), qui bouleverse l'organisation spatiale et entraîne un repli de l'ensemble de la scène dans la zone délimitée par les deux haut-parleurs.



**Figure 1 : trajets directs (A) et croisés (S) des haut-parleurs aux oreilles en restitution stéréophonique standard**

Pour pallier ce défaut, plusieurs méthodes ont été proposées très tôt [2,3,4,5], visant à corriger au mieux cette diaphonie acoustique. Le principe général de ces solutions, rassemblées sous le terme de *traitements transauraux*, consiste à insérer entre l'enregistrement binaural et les haut-parleurs un étage compensant au mieux les trajets croisés (figure 2). Dans l'idéal, il s'agit donc d'*inverser* le filtrage acoustique entre les haut-parleurs et les oreilles de l'auditeur.



**Figure 2 : principe de la correction transaurale : L et R sont les résultats du traitement des signaux  $x_L$  et  $x_R$ , issus généralement de synthèse binaurale ou d'un enregistrement binaural, de sorte que les signaux parvenant aux oreilles soient les plus proches possibles des originaux.**

Comme la plupart des problèmes d'inversion en traitement de signal, plusieurs obstacles se présentent :

D'une part, il est nécessaire de connaître avec précision la nature du trajet acoustique direct pour espérer le compenser : c'est pourquoi, en écoute transaurale plus qu'ailleurs, le respect des consignes de disposition des haut-parleurs et du point d'écoute (position, orientation) sont capitales. Dans ce cadre, et en supposant négligeable la réverbération dans le lieu d'écoute, il est possible d'appliquer une technique indépendante de ce dernier, et donc d'opérer un traitement non pas lors de la restitution (ce qui était techniquement impossible jusque récemment), mais directement lors de la production. Ceci nécessite donc au préalable la connaissance du trajet anéchoïque du son provenant d'une direction donnée à chacune des deux oreilles. Ce trajet est caractérisé par les fonctions de transfert de tête (*Head-Related Transfer Functions*, ou *HRTF*). Or, celles-ci sont étroitement liées à la forme du buste, et en particulier aux dimensions du crâne, ainsi qu'à la forme des oreilles et des épaules. Dans une situation de production où il n'est pas possible de proposer un traitement individualisé, il est donc de rigueur d'avoir recours à des HRTF relatives à un sujet (ou une tête artificielle) de dimensions moyennes.

D'autre part, la nature même de la propagation entre les haut-parleurs et les tympans de l'auditeur rend l'inversion difficile voire impossible au sens strict : en effet, la présence de zéros dans le filtre à inverser entraîne inévitablement des problèmes de coloration, en particulier dans les registres haut-médiums et aigus.

Un troisième obstacle d'importance réside dans la complexité de mise en œuvre de la méthode : même en supposant les filtres inverses calculés et affranchis des problèmes de coloration, il reste peu envisageable de l'implémenter de manière exacte, en particulier sous forme analogique. Ceci dit, les progrès techniques récents permettent d'appliquer un filtrage numérique par convolution en temps réel et à faible latence avec un coût de puissance relatif de plus en plus faible. De fait, on trouve sur plusieurs cartes son destinées au jeu vidéo des DSP munis entre autres de tels algorithmes.

Le problème se simplifie largement lorsque l'on suppose l'auditeur placé de manière symétrique par rapport aux haut-parleurs, ce qui est une règle inhérente à toutes les normes de diffusion stéréophonique (de la traditionnelle stéréophonie à deux canaux au multicanal). Cette hypothèse permet en effet de substituer à la structure de six filtres en treillis (*lattice*), proposée initialement par Atal et Schroeder [2], une structure dite par mélange (*shuffler*), dans laquelle les signaux sont tout d'abord soumis à un matriçage très simple (additions et soustractions uniquement), puis soumis à deux filtres uniquement, avant d'être à nouveau présentés à la même matrice. C'est cette structure qui est de rigueur actuellement dans la plupart des processeurs transauraux.<sup>1</sup>

### 3/ Application de techniques transaurales à un système 5.1

Le couple synthèse binaurale/traitement transaural, bien qu'offrant des perspectives impressionnantes, souffre donc néanmoins de plusieurs défauts. Les principaux obstacles techniques initiaux ont été peu à peu levés au cours des ans, mais subsistent les problèmes de coloration, et la forte instabilité par rapport au positionnement de l'auditeur, qui rendent actuellement assez difficile son utilisation en production en stéréophonie à deux canaux. Néanmoins, ces limitations sont à reconsidérer dans une large mesure en restitution multicanal. En effet, la présence de haut-parleurs supplémentaires répartis autour de l'auditeur offre une latitude supplémentaire, que nous avons cherché à exploiter. Trois exemples sont présentés ici :

#### 1/ Traitement mixte transaural/panoramique d'intensité par bande de fréquences

Il s'agit ici d'utiliser au mieux les avantages des deux techniques de positionnement de sources virtuelles mentionnées ici. L'utilisation du panoramique d'intensité souffre comme cela a été mentionné d'un manque de stabilité spatiale sur les côtés et à l'arrière lorsque l'auditeur se rapproche de l'un des deux haut-parleurs concernés. La restitution transaurale est elle (relativement) stable face à des déplacements de l'auditeur vers l'avant ou l'arrière. Le principe est donc d'appliquer l'une ou l'autre des deux techniques en fonction de la bande de fréquences concernées. Le traitement transaural est ainsi appliqué dans la zone 500 Hz/1.5 kHz, qui contient les indices les plus significatifs vis-à-vis de la localisation, alors que le panoramique d'intensité est appliqué dans les registres graves et aigus, afin de minimiser les problèmes de coloration. Ce principe n'a d'intérêt que pour des sources latérales, puisqu'il permet de minimiser l'influence de déplacements de l'auditeur vers l'avant ou l'arrière. La difficulté principale consiste à choisir les filtres pour réduire au maximum les éventuels problèmes de détimbrage au niveau des fréquences de recouvrement.

---

<sup>1</sup> Pour plus de précisions sur les questions d'implémentation, se référer notamment à l'article de Cooper et Bauck [5]

## 2/ Traitement mixte transaural/panoramique d'intensité en pleine bande avec un retard

L'idée de départ est identique (combinaison de deux techniques de positionnement), mais on privilégie ici une combinaison dans le domaine temporel : les deux traitements sont appliqués en pleine bande, le panoramique d'intensité étant en retard de quelques millisecondes par rapport au couple binaural/transaural, ce dernier fournissant de fait les principaux indices de localisation. En effet, le système auditif, en vertu de la « loi du premier front d'onde » [1], se base avant tout sur les informations les plus précoces pour estimer la position de la source, les informations ultérieures (à l'instar des premières réflexions dans un espace clos) jouant surtout sur le timbre.

## 3/ Deux couples transauraux

Il a déjà été mentionné que l'une des principales limitations des techniques transaurales en stéréophonie à deux canaux est l'instabilité spatiale des sources latérales et arrières. Puisque l'on dispose de sources physiques à l'arrière de l'auditeur, il est possible de minimiser ce défaut en introduisant à une seconde base transaurale, qui vient suppléer la base avant. La balance entre les deux couples est à régler en fonction de la position désirée de la source virtuelle.

## 4/ En pratique

Les principes évoqués ci-dessus ont été testés en situation de post-production, dans le cadre du mixage d'un enregistrement de « Répons » de Pierre Boulez effectué lors d'un concert à la cité de la musique en 2003. Les techniques utilisées sont issues du « Spatialisateur » développé au sein de l'équipe Acoustique des Salles de l'Ircam. Les traitements ont été implémentés sous forme de plug-ins expérimentaux utilisés sur une station de travail Pro Tools.

## Bibliographie

[1] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1983

[2] Schroeder, M.R. et Atal, B.S., *Computer simulation of sound transmission in rooms*, IEEE Conv. Record, 7:150-155, 1963

[3] Damaske, P., *Head-Related Two-Channel Stereophony with Loudspeaker Reproduction.*, J. Acoust. Soc. Am., 50(4):1109-1115, 1971

[4] Rossi, M., *Electroacoustique*, Presses polytechniques romandes, 1986

[5] Cooper, D. H. et Bauck, J. L., *Prospects for Transaural Recording*, JAES Vol. 37, No. 1/2, 1989