

survenus durant la propagation du son. L'opération est répétée pour un grand nombre de directions, de façon à baliser toute la sphère auditive et constituer une base de données de filtres directionnels.

Par la suite, lors de la synthèse spatiale, il suffit d'appliquer à tout signal monophonique, par exemple une voix captée par un microphone, le couple de filtres correspondant à la direction de la source virtuelle souhaitée. Les deux signaux obtenus, diffusés dans un casque, restituent de façon exhaustive les indices acoustiques responsables de la localisation auditive spatiale, et ce pour n'importe quelle direction présente dans la base de données.

Cette technique est réservée à la restitution sur un casque, faute de quoi, lors de la diffusion, le filtrage propre au trajet entre les enceintes et les oreilles de l'auditeur se superpose au filtre choisi et perturbe l'illusion sonore. En adaptant «en temps réel» le couple de filtres appliqué en fonction de la direction de la source ou des mouvements de l'auditeur, on obtient sans peine une source fictive en mouvement.

La synthèse binaurale partage les mêmes limites que l'échantillonnage: ne reposant sur aucune modélisation, elle se prête mal aux manipulations. Elle nécessite par exemple la constitution d'une lourde banque de filtres. En effet, bien qu'il soit possible de reconstruire une direction manquante par interpolation entre deux directions voisines, celles-ci ne doivent pas être éloignées de plus de 15 degrés sous peine de voir apparaître des effets indésirables. La reproduction de sources en mouvement nécessite même des mesures tous les cinq degrés environ, ce qui conduit à des bases de données de l'ordre de 200 à 600 couples de filtres pour couvrir l'ensemble des directions.

En outre, les filtres HRTF dépendent par nature de la morphologie de l'auditeur. Ainsi, l'utilisation de filtres «génériques», mesurés sur une tête artificielle, entraîne en général des erreurs de localisation, en particulier des confusions avant-arrière.

Cependant, le couplage d'un dispositif de synthèse binaurale avec un capteur d'orientation de la tête permet d'asservir de façon continue la scène sonore diffusée aux mouvements de l'auditeur. Cette rétroaction est déterminante pour l'auditeur, puisqu'elle lui permet de rétablir la corrélation entre ses propres mouvements et les variations des indices acoustiques transmis et réduit de façon considérable les confusions avant-arrière.

La technologie binaurale reste à ce jour le mode de restitution spatiale le plus rigoureux, et son utilisation se justifie non seulement dans le cadre d'études sur la perception et la cognition

auditive spatiale, mais également sur le terrain du jeu ou des installations sonores interactives.

Enfin, des approches plus récentes de la reproduction sonore s'attachent à modéliser et à reconstruire de façon précise les propriétés physiques du champ acoustique dans une zone plus ou moins large autour des auditeurs, en s'appuyant sur la résolution de l'équation d'onde. Parmi ces méthodes, on distingue les techniques holophoniques comme la *wave field synthesis* (WFS) et les approches par «décomposition harmonique», comme les formats Ambisonics. Développée dans les années 1980 à l'université de Delft, aux Pays-Bas, la WFS suscite aujourd'hui un large intérêt, car le grand nombre de canaux (jusqu'à plusieurs centaines) que nécessite cette technique de synthèse spatiale n'est plus un obstacle technique.

JOUER AVEC LES FRONTS D'ONDE

Le signal sonore d'une source primaire que l'on désire spatialiser est diffusé sur un réseau de haut-parleurs très proches les uns des autres.

Sous certaines conditions, la juxtaposition des fronts d'onde émis par chacune de ces sources secondaires est alors équivalente au front d'onde d'une unique onde acoustique, provenant d'une source primaire virtuelle. Un contrôle adéquat du retard temporel et de l'intensité du signal envoyé sur chaque haut-parleur permet d'obtenir la création d'un front d'onde commun qui a pour origine n'importe quelle position souhaitée (voir l'encadré page ci-contre). Ce traitement doit être répété de façon indépendante pour chaque source

composant la scène sonore.

En toute rigueur, la reproduction exacte du champ acoustique requiert une infinité de sources secondaires. En pratique, on se limite à créer des sources virtuelles dans le plan horizontal des auditeurs. Les sources secondaires sont alors elles-mêmes placées dans ce plan, et moyennant quelques termes correctifs, il suffit d'installer des haut-parleurs le long d'une couronne autour des auditeurs, ou d'une rangée devant eux.

De même, la distribution en théorie continue de sources secondaires est remplacée par un nombre fini de haut-parleurs régulièrement espacés. Cette discrétisation se traduit par une dégradation de la qualité de restitution spatiale dans les hautes fréquences. En pratique, pour assurer que l'indice interaural de retard soit reproduit de façon fiable dans la gamme de fréquences où il est primordial et sur l'ensemble de la région d'écoute, la fréquence limite est fixée au-dessus d'un kilohertz, ce qui conduit à un écartement maximal des haut-parleurs de l'ordre de 15 centimètres.

Si la stéréophonie s'appuie sur un modèle centré sur l'auditeur, la WFS adopte au contraire un point de vue exocentrique: le champ sonore

**L'holophonie
ou synthèse du champ
acoustique s'appuie sur
le formalisme ondulatoire
de Huygens**

est décrit et reproduit quel que soit le placement des auditeurs. La technique est donc particulièrement adaptée aux situations où les auditeurs sont multiples et en mouvement, comme les concerts, les installations sonores, les expositions... Cette technique a d'ailleurs été utilisée pour les expositions Dada, en janvier 2006, et Beckett, en 2007, au Centre Pompidou, à Paris.

LA DÉCOMPOSITION DU SON

L'autre approche pour reconstruire le champ acoustique est la décomposition en harmoniques (la famille des formats Ambisonics). Introduite dans les années 1970 par Michael Gerzon, alors à l'université d'Oxford, elle fait depuis une vingtaine d'années l'objet de nombreuses recherches. Elle repose également sur un formalisme physique, mais adopte un point de vue centré sur l'auditeur.

Par analogie avec l'analyse de Fourier, qui consiste à représenter un signal sous la forme d'une superposition de fonctions sinusoïdales, le champ acoustique entourant l'auditeur est représenté par une superposition de fonctions périodiques de l'espace (des harmoniques sphériques ou cylindriques), chacune associée à un canal audio (voir l'encadré page 75). Pour représenter un signal ayant une forte variabilité temporelle, on utilise des composantes sinusoïdales de fréquences élevées; de même, plus on désire décrire de façon précise l'organisation spatiale d'une scène sonore, plus il faut tenir compte des harmoniques spatiales d'ordre élevé et donc utiliser un nombre important de canaux. Pour encoder l'intégralité de l'information spatiale, il faudrait en théorie une infinité de canaux.

Le principal intérêt du format Ambisonics est que l'encodage est indépendant du dispositif de restitution et hiérarchique. Le signal envoyé à chaque haut-parleur est une combinaison linéaire des canaux correspondant à chaque harmonique, dont les coefficients ne dépendent que du nombre et de la position des enceintes autour des auditeurs. Pour reproduire une scène sonore avec une précision spatiale – un ordre d'harmonique – donnée, il suffit de disposer d'un nombre de haut-parleurs supérieur ou égal aux nombres de canaux. Si l'on ne dispose pas d'assez de haut-parleurs, la scène peut être néanmoins décodée sans déformation, mais avec une précision spatiale inférieure.

Cette technique est bien adaptée aux applications de réalité virtuelle. On peut en effet enregistrer une scène sonore complexe comportant de nombreuses sources en encodant son organisation tridimensionnelle, puis l'importer dans un environnement virtuel. La facilité de manipulation permet de réorienter la scène en 3D pour compenser le mouvement de la tête de l'utilisateur.

Les systèmes de spatialisation ne peuvent se contenter de simuler la direction des sources

composant une scène virtuelle. Ils doivent aussi rendre compte de l'interaction des ondes avec l'environnement, dont la réflexion et la diffraction sur les obstacles et les parois. Reproduire ces effets est primordial pour que l'auditeur puisse estimer correctement la distance de la source et construire une représentation auditive cohérente du lieu virtuel où il évolue.

Parmi les différentes approches, on retrouve la distinction entre la synthèse par échantillonnage, par modèle physique ou par analyse et synthèse des propriétés acoustiques de la salle virtuelle. Toutes aboutissent cependant à un filtre, par lequel le son brut émis par la source doit être transformé pour reproduire les effets de l'environnement. Une première possibilité est d'acquérir des réponses acoustiques d'environnements réels. C'est l'extension de la technique binaurale, à ceci près que l'enregistrement est réalisé non plus dans une chambre anéchoïque, mais dans la salle que l'on veut simuler. Le filtre obtenu consigne l'ensemble

Grâce au théâtrophone, Marcel Proust écoutait en stéréo des pièces de théâtre sans sortir de chez lui.

