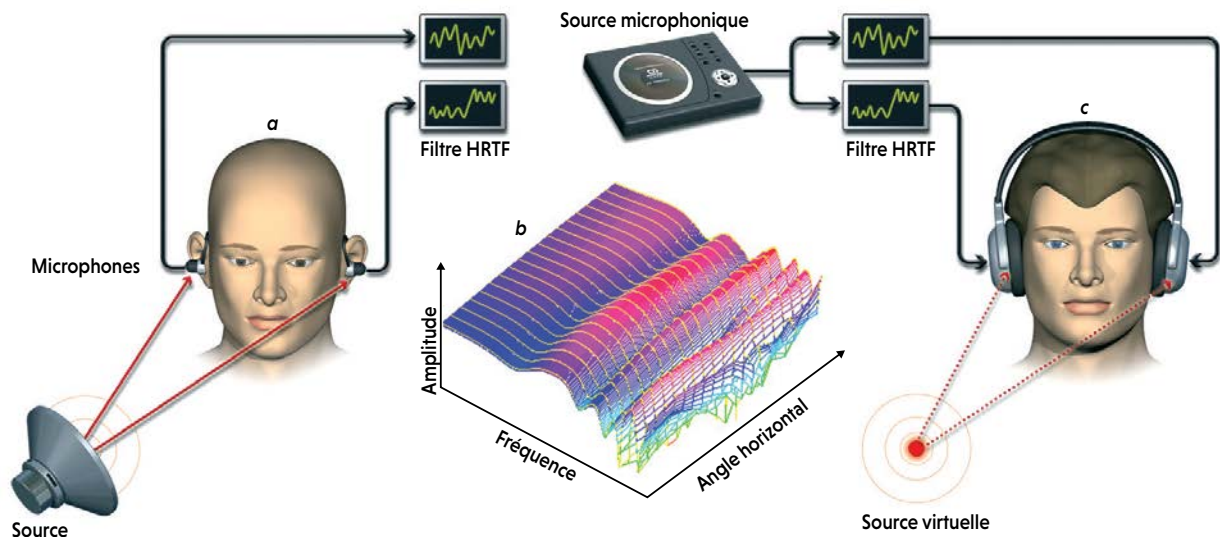




➤ considérablement enrichi le répertoire des timbres, mais aussi la mise en scène des sons dans l'espace. La spatialisation du son est devenue un aspect incontournable de la création musicale. Grâce à l'outil informatique, il est désormais possible de placer et d'animer des sources sonores de façon arbitraire, dans n'importe quelle direction, et ainsi de prolonger la scène traditionnelle où se trouvent les instrumentistes.

En parallèle, la spatialisation du son a connu une percée fulgurante chez le grand public par le biais des dispositifs de son 3D pour le home cinéma ou les jeux vidéo. Dans ces derniers, une composante sonore spatialisée est essentielle pour accroître le réalisme et l'immersion du joueur dans l'environnement de synthèse.

Quel que soit leur domaine d'application, les procédés de spatialisation font appel à plusieurs disciplines : informatique, acoustique, traitement du signal, psychoacoustique ou encore cognition. La synthèse de la direction et de la distance des sources sonores, de la propagation du son, ainsi que sa reproduction s'appuient à la fois sur la modélisation physique des phénomènes en jeu et sur des modèles de la perception auditive spatiale.

D'OÙ VIENT CE SON ?

Intéressons-nous d'abord à la perception spatiale du son, à la base de diverses techniques de spatialisation du son. Contrairement à la vue, qui chez l'être humain, souffre d'un champ limité, l'audition couvre l'ensemble de l'espace. Notre capacité à localiser des sources sonores repose sur la comparaison entre les signaux parvenant à nos deux oreilles, qui dépendent bien sûr de la position de celles-ci, mais aussi de la diffraction des ondes acoustiques sur notre tête. Ces phénomènes dépendent de la direction d'incidence des ondes et permettent de localiser la source.

Les performances de localisation du système auditif humain en azimut (dans le plan

La restitution binaurale. Pour une source donnée, on mesure d'abord les différences de temps d'arrivée et d'amplitudes aux deux oreilles d'un auditeur (a). Ces différences, qui permettent de localiser la source sonore, sont caractérisées par une paire de filtres dits HRTF (b). On construit de même des filtres pour toutes les directions. Pour spatialiser un son monophonique, il suffit ensuite d'appliquer au signal émis la paire de filtres correspondant à la direction souhaitée : tout se passe comme si le son reconstruit venait d'une source virtuelle située au même endroit que la source réelle d'origine (c).

horizontal) sont de l'ordre de trois degrés pour des sources voisines du plan de symétrie de la tête. Elles tombent à une dizaine de degrés pour des sources situées sur le côté.

Comment fonctionne la localisation en azimut ? Les deux indices fondamentaux sont la différence entre les temps d'arrivée aux oreilles (ITD, pour *interaural time difference*), traduisant leur écartement, ainsi que la différence d'intensité (ILD, pour *interaural level difference*), liée à l'obstacle de la tête. Ces indices dits binauraux décrivent les relations entre les signaux captés par les deux oreilles. Toutefois, des considérations géométriques simples révèlent qu'à un couple de valeurs données des indices interauraux de retard et d'intensité correspondent plusieurs directions possibles. Ainsi, pour chaque direction dans le plan médian, les différences interaurales sont nulles. Des mécanismes plus fins sont nécessaires pour différencier les directions dans cette zone de confusion et plus généralement permettre la localisation en élévation.

En raison de l'interaction complexe des ondes sonores avec notre corps (diffraction par le crâne, le pavillon de l'oreille, les épaules...), les variations de différences d'intensité dépendent de la fréquence du son. Dans les basses fréquences, les longueurs d'onde – plusieurs mètres – sont grandes par rapport aux dimensions de la tête. Celle-ci n'est donc pas un obstacle majeur et la différence de niveau sonore entre l'oreille du côté de la source (ipsilatérale) et l'oreille opposée (controlatérale) est faible.

Plus la fréquence du son est élevée – plus sa longueur d'onde diminue –, plus les caractéristiques morphologiques contribuent à différencier les signaux parvenant aux deux oreilles. Ainsi, chaque direction d'incidence s'accompagne d'une « coloration spectrale » particulière du son, que notre cerveau apprend à identifier. En d'autres termes, la perception auditive spatiale consiste à décoder l'effet de notre propre perturbation sur

le champ d'ondes acoustiques. Les indices spectraux sont notamment sollicités pour discriminer les directions avant, arrière et, de façon plus générale, dans le plan vertical. La nature plus subtile de ces indices explique les performances assez médiocres de la localisation en élévation (angle vertical), de l'ordre de 10 degrés de face, jusqu'à 30 degrés vers le haut ou l'arrière.

La perception de la distance dépend de plusieurs facteurs. L'éloignement de la source entraîne une baisse de l'intensité sonore et une atténuation des hautes fréquences liée à l'absorption par l'atmosphère. Ces indices ne sont cependant pas absolus; ils supposent eux aussi un traitement cognitif fondé sur une connaissance préalable de la puissance et du spectre émis par la source. Lorsque nous identifions une voix, un véhicule..., nous en comparons le niveau et le timbre par rapport à d'autres événements du même type, dans diverses situations spatiales (proche, derrière...), gardés dans notre mémoire.

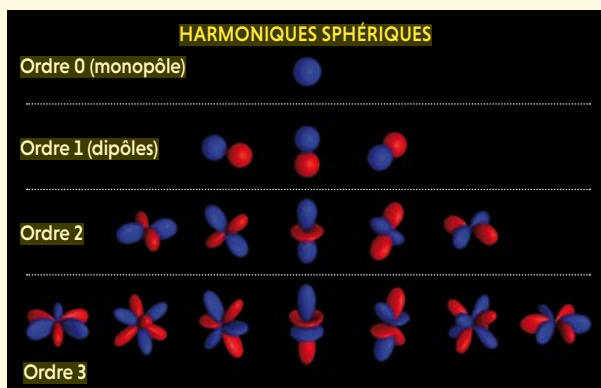
D'autres facteurs s'ajoutent dans le cas d'environnements réverbérants comme une salle, pour lesquels notre perception s'appuie sur le rapport entre l'intensité directe de la source, qui varie avec la distance, et l'intensité réverbérée, qui est uniformément répartie dans la salle.

Les systèmes de restitution sonore spatiale visent, de façon générale, à reproduire aux oreilles d'un auditeur les valeurs appropriées de ces indices de localisation selon la position souhaitée d'une source sonore virtuelle. Pour ces systèmes, on distingue trois familles: la stéréophonie et ses extensions, les techniques binaurales, réservées à la restitution sur casque, et les techniques de synthèse du champ acoustique. Au-delà de la fidélité avec laquelle elles reproduisent les indices de localisation, les mérites respectifs de ces différentes techniques se mesurent en termes de nombres de canaux à gérer et de contraintes sur la position et la mobilité de l'auditeur par rapport au dispositif de restitution.

LA DÉCOMPOSITION HARMONIQUE

La spatialisation par décomposition harmonique consiste à représenter l'organisation spatiale du champ acoustique d'un point de vue angulaire (sans notion de distance) sous forme d'une superposition de fonctions périodiques de l'espace, les harmoniques sphériques, chacune associée à un canal audio. L'harmonique d'ordre 0 (monopôle) correspond à une variation de pression homogène dans toutes les directions. Les harmoniques

d'ordre 1 (dipôles) caractérisent chacune un gradient de pression selon une des trois directions de l'espace. Les cinq harmoniques d'ordre 2, puis celles des ordres supérieurs, représentent des modes de vibration plus complexes; elles permettent de préciser de façon progressive la dépendance angulaire du champ acoustique. La description exacte de celle-ci requiert en théorie un nombre infini d'harmoniques, et donc de canaux.



EN HIFI ET EN STÉRÉO

Le modèle de restitution spatiale du son le plus utilisé encore aujourd'hui reste la stéréophonie. Deux haut-parleurs sont placés devant l'auditeur. En jouant sur le retard ou le niveau relatif des signaux envoyés à chaque haut-parleur, on contrôle de façon simplifiée les indices de localisation interauraux (ITD et ILD), ce qui donne naissance à une source sonore fictive. La direction de cette «source fantôme» est limitée à la zone encadrée par les deux haut-parleurs. L'utilisation d'un simple retard entre les deux canaux aboutit à une zone de restitution très limitée. La plage de variation du retard pour laquelle les signaux fusionnent pour créer une image fantôme est en effet très restreinte. Dès que le retard dépasse une milliseconde (soit un décalage de 30 centimètres environ), le signal en retard est inhibé: il n'intervient plus dans le traitement perceptif de la localisation (effet Hass). En conséquence, un auditeur décalé, même légèrement, de l'axe médian des deux haut-parleurs aura l'illusion que l'ensemble de la scène stéréophonique se déporte de façon abrupte vers le haut-parleur le plus proche. L'utilisation conjointe d'une différence de gain contribue à limiter cet effet, sans toutefois le compenser totalement.

Par extension, il est possible de reproduire un son spatialisé sur des ensembles de haut-parleurs, plans ou tridimensionnels, disposés autour de l'auditeur. L'exemple le plus courant est la configuration «5.1» (cinq haut-parleurs répartis dans le plan horizontal) des formats sonores cinématographiques, également utilisés dans les jeux vidéo. Les installations sonores plus ambitieuses nécessitent la mise en place de dômes constitués de quelques dizaines de haut-parleurs. Pour synthétiser la direction d'une source sonore, on sélectionne la paire ou le triplet de haut-parleurs entourant la direction désirée et l'on règle les