

Stéréophonie et perception spatiale

Roald Baudoux, version du 17/05/2006

1. Définition

De prime abord, la stéréophonie est définie comme l'ensemble des techniques électroacoustiques de captation, reproduction et diffusion visant à restituer une information auditive de **positionnement spatial et de relief sonore**. Cependant, l'usage fait que lorsqu'on parle de stéréophonie aujourd'hui, on sous-entend l'idée de restituer ce relief et ce positionnement grâce à une diffusion sur **deux** canaux seulement. Il a pourtant existé dans le passé des systèmes visant à cette restitution de positionnement et de relief avec une quantité de haut-parleurs plus importante. D'ailleurs, la racine grecque de ce mot signifiant « relief », un système à deux canaux devrait s'appeler « stéréophonie binaurale ». Aujourd'hui avec la diffusion très large des systèmes de type multicanal (5.1 et autres), la distinction est plus claire et les systèmes à plus de deux canaux sont regroupés dans la catégorie « multicanal », la stéréo étant donc associée par l'usage aux systèmes à deux canaux.

Il faut savoir que la stéréophonie n'est arrivée qu'assez tardivement dans l'histoire de l'électroacoustique (début dans les années 1960 et diffusion à grande échelle surtout dans les années 1970). Avant cela, la musique était composée ou enregistrée en monophonie. Les systèmes d'écoute n'avaient qu'un seul haut-parleur.

Il faut d'ailleurs songer qu'avoir deux canaux d'enregistrement ou de reproduction ne signifie pas nécessairement faire de la stéréophonie. En effet, si on envoie exactement le même son sur deux haut-parleurs d'une paire stéréophonique, un auditeur placé correctement entendra le son comme venant du point situé à mi-chemin entre les deux haut-parleurs. Cette situation correspond plutôt à de la MONOPHONIE même si elle est écoutée sur deux haut-parleurs.

Une autre situation qui n'est pas de la stéréophonie est celle où l'on écoute un son sur un haut-parleur et un tout autre son sur l'autre haut-parleur. Les deux haut-parleurs sont alors perçus comme deux sources ponctuelles différentes qui n'ont rien à voir l'une avec l'autre. Ce cas où il n'y a rien de commun entre les sons des deux canaux s'appelle de la BIPHONIE. La plupart des compositeurs utilisent un jour ou l'autre ce système, souvent en parallèle avec de la stéréophonie véritable.

Pour pouvoir parler réellement de stéréophonie, il faut que les deux canaux concourent à créer une IMAGE qui occupe l'espace entre les deux haut-parleurs.

Cette image peut être créée de plusieurs manières, soit à l'aide de manipulations en studio, soit dès la prise de son. Dans la plupart des cas, c'est dès la prise de son que l'on aura un son stéréophonique. Cependant, il arrive assez fréquemment qu'un compositeur cherche à créer une image stéréophonique à partir d'un son monophonique et il existe des opérations de studio le permettant.

Mais donc, l'essentiel du travail sur la stéréophonie se situe à la prise de son. C'est en effet le choix des microphones, la configuration du système de prise de son et la relation entre ce système et la « scène » à capter qui vont déterminer l'essentiel du résultat sonore final en ce qui concerne l'aspect spatial. Il est possible de créer de la largeur et du relief spatial artificiellement, mais il est en général plus intéressant (et plus facile) de capter l'espace dès la prise de son.

2. La perception spatiale sans stéréophonie

Il n'est pas nécessaire d'avoir deux canaux pour créer un jeu sonore sur l'espace. En effet, certains éléments de notre perception de l'espace fonctionnent avec une seule oreille (et donc un seul haut-parleur suffit).

Ainsi, comme les différentes fréquences audibles ne sont pas absorbées de la même manière par l'air et que l'oreille n'a pas la même sensibilité pour les différentes fréquences audibles, l'oreille peut estimer la distance d'une source suivant les différences de couleur du son, pour autant que le type de source soit connu par l'auditeur. Par exemple, cela fonctionne très bien avec de la voix parlée, car tout le monde a déjà entendu de la voix, mais cela fonctionne moins bien avec des sons qui ne correspondent pas à des sources naturelles connues (la plupart des sons synthétiques par exemple). Si le son ne correspond à rien de connu, cela peut fonctionner tout de même si on entend plusieurs fois de suite le son, avec des distances de captation différentes entraînant des colorations différentes qui seront perçues.

De plus, on associe à certaines images des distances différentes. Une voix chuchotée paraîtra automatiquement plus proche qu'une voix criée parce qu'en général on chuchote de près et on crie de loin.

De même, on a l'habitude d'associer une intensité caractéristique à la plupart des sons au travers desquels on reconnaît une source connue (voiture, conversation, craquement d'allumette, etc). Dès lors, si l'image de ce son est écoutée à une intensité différente, on peut avoir la sensation que la source s'est rapprochée ou éloignée.

Un autre phénomène perceptible avec une seule oreille est celui de la réverbération. La réverbération est cette prolongation du son qui est caractéristique des lieux comportant des surfaces peu ou prou réfléchissantes (soit la majorité des intérieurs ou des lieux urbains). Suivant qu'un son est proche ou lointain, la proportion entre le son direct et le son réfléchi sera différente. Il s'agit donc d'un indice lié à notre expérience vécue des différentes acoustiques des lieux que nous avons traversés.

En outre, la perception de la distance repose, dans un environnement réverbérant, sur le temps qui s'écoule entre l'arrivée au système auditif du premier front d'onde (son direct) et l'arrivée des premières réflexions.

On voit qu'il est donc possible de jouer sur certaines variables, sur certains indices pour utiliser l'espace en électroacoustique sans jouer nécessairement sur la stéréophonie.

Lorsqu'on parle d'écoute monaurale, on évoque les éléments qu'une oreille seule peut percevoir. L'écoute binaurale est celle qui fait appel aux deux oreilles. En effet, certains éléments ne sont perceptibles qu'au travers d'une comparaison des informations arrivant à l'une et l'autre oreille.

3. L'écoute binaurale naturelle

Cela peut sembler évident mais à bien y réfléchir, comment se fait-il qu'à partir de deux canaux diffusés sur deux haut-parleurs, on perçoit une image qui occupe tout l'espace entre les deux haut-parleurs ?

Pour le savoir, il faut se pencher sur la perception de la position spatiale en écoute BINAURALE NATURELLE. Il s'agit de la situation d'écoute que nous vivons lorsque nous écoutons avec nos deux oreilles et sans recourir à un système qui nous isole de l'environnement (comme un casque par exemple).

Comment fait notre perception auditive pour savoir qu'un son provient de tel côté plutôt que d'un autre ? La réponse réside dans les DIFFERENCES DE PERCEPTION entre les deux oreilles.

Si un son provient d'une source située sur la gauche d'un auditeur, l'onde sonore atteindra d'abord son oreille gauche, puis son oreille droite. Il y aura donc une DIFFERENCE DE TEMPS D'ARRIVEE entre les deux oreilles.

D'autre part, l'énergie qui parviendra à l'oreille droite sera moins importante que celle qui parviendra à l'oreille gauche, essentiellement parce que la tête elle-même absorbera une partie de l'énergie de l'onde sonore. Il y aura donc une différence de NIVEAU entre les deux oreilles.

Il faut savoir que les deux phénomènes ne jouent pas exactement pour toutes les fréquences audibles. En ce qui concerne la différence de temps entre les deux oreilles, elle vaut surtout pour les fréquences inférieures à 500 Hz. En effet, pour ces fréquences, la différence de temps est assimilée par notre système auditif à des différences de PHASE car les différence ne dépasse pas la valeur d'une période complète. Pour les fréquences plus élevées, la différence de temps peut correspondre à plus d'une période, voire à plusieurs périodes. Il devient alors impossible pour notre perception de considérer la différence de temps comme une différence de phase. Pour les cas où elle fonctionne, la différence de temps d'arrivée est très petite. En effet, un décalage d'une milliseconde entre les deux canaux suffit à créer la perception d'un positionnement sur une extrémité de la rampe stéréophonique.

En ce qui concerne la différence de niveau, elle joue plutôt sur les fréquences aigues. En effet, la tête absorbe surtout les hautes fréquences et n'affecte que peu les basses fréquences. La différence de niveau ne doit pas être très importante elle non plus : un décalage de niveau de 15 dB entre les deux canaux suffit à donner l'impression que la source est complètement localisée du côté du canal dominant.

Imaginons un exemple : une source est située à la droite d'un auditeur avec un angle de 55° par rapport à l'axe frontal et à une distance de 4 m. On constate que la différence de temps d'arrivée est de 4 m. Or, cette différence est interprétée comme **une différence de phase**.

Cette différence de phase est différente pour chaque fréquence, puisqu'elle vient de la différence de longueur d'onde. En effet, une distance identique correspond à une portion de longueur d'onde différente selon la fréquence. Voir le tableau ci-dessous.

Fréquence du signal	Longueur d'onde	Déphasage correspondant à la différence de temps d'arrivée de 0,5 ms.
125 Hz	272 cm	22,5 °
250 Hz	136 cm	45 °
500 Hz	68 cm	90°
1000 Hz	34 cm	180°
2000 Hz	17 cm	360°

Comment se fait-il que l'on interprète la différence de temps d'arrivée comme un déphasage ou une différence de longueur d'onde?

On sait que chaque fréquence possède une longueur d'onde, liée à la vitesse de déplacement du son dans l'air. Cette vitesse dénommée c (pour « célérité ») est de 344 mètres par seconde (mais elle varie avec la température et le degré d'hygrométrie). Attention, dans des milieux plus denses ce chiffre est beaucoup plus élevé.

La longueur d'onde (λ , lettre grecque « lambda ») est liée à la fréquence (f) et à la vitesse de déplacement (c) par l'équation suivante :

$$\lambda = c / f$$

Or, 0,5 millisecondes correspondent à une distance parcourue de 172 mm ($344 \text{ m} \times 0,0005$). Et ces 172 mm correspondent eux-mêmes à une portion de longueur d'onde des différentes fréquences. Par exemple, à 125 Hz, la longueur d'onde est de 2720 mm. $172/2720 \times 360$ nous donne la différence de phase en degrés.

$$172/2720 \times 360 = 22,7^\circ$$

A la lecture du tableau, on constate que la différence de phase est de plus en plus importante au fur et à mesure que la fréquence considérée est plus grande. Au-delà de 2000 Hz, cette différence dépasse 360° , c'est-à-dire une phase complète. Pour les fréquences plus élevées, il n'est donc plus possible pour l'oreille de se baser sur la différence de temps pour déterminer l'azimut.

Par ailleurs, dans le système auditif, le pavillon joue un rôle non négligeable, essentiellement dans la distinction entre sons venant de l'avant et sons venant de l'arrière. Cela repose sur des modifications spectrales caractéristiques des sons venant de l'arrière, le pavillon jouant un rôle de filtre.

4. L'apport de la stéréophonie

La stéréophonie va surtout permettre d'avoir une LOCALISATION EN AZIMUT des sons sur une RAMPE allant du haut-parleur gauche au haut-parleur droite ou d'occuper une certaine LARGEUR sur cette rampe. En utilisant des techniques spéciales, il est possible de placer des sons au-delà des enceintes (« extra-largeur ») mais ces techniques fonctionnent sur base de relations de phase particulières qui entraînent une plus grande fragilité de l'information sonore et elles sont à utiliser avec précaution, notamment en concert ou en diffusion radiophonique.

Dans une moindre mesure, la stéréophonie va aussi permettre d'avoir une perception de PROFONDEUR.

5. Système de captation pour écoute binaurale artificielle (c'est-à-dire pour une écoute au casque)

Tête artificielle, micros OKM.

La particularité de l'écoute au casque est que l'information sonore envoyée à une oreille ne parvient pas à l'autre oreille. Cela permet de créer des relations de phase entre les deux canaux qui sont impossibles ou problématiques avec des haut-parleurs. Par exemple, une opposition de phase entre deux canaux donne au casque une localisation très latéralisée, alors que sur des haut-parleurs, elle entraîne en sus la disparition d'une bonne partie du signal par annulation de phase.

6. Système de captation pour écoute stéréophonique (par haut-parleurs)

Note : le type de système influence le rapport d'intensité entre le son direct et le son réfléchi, donc la présence du son.

6.1. Systèmes à différence de temps/de phase

- a) AB : 2 micros omni distants de quelques dizaines de centimètres
- b) Système Decca : 2 micros omni/cardios distants de plusieurs mètres

6.2. Systèmes à différence d'intensité (couples dits *coïncidents*)

- a) XY
- b) Stereosonic
- c) MS (nécessite système électronique ou informatique de dématricage)

6.3. Systèmes mixtes temps + intensité

- d) ORTF (17 cm/110°), NOS (30 cm, 90°) (systèmes divergents)
- e) Couples à diffraction (sphère G. Theile - disque Charlin -17 cm, disque => avantages du PZM) similaire tête mais pas de profil de la tête ni de pavillon s.

7. Traitements de « stéréophonisation » de programmes monophoniques.

- a) Potentiomètre panoramique des consoles de mélange ou "pan" dans les logiciels.
- b) Ligne à retards (delay line) appliqué sur un seul des deux canaux
- c) Filtres passe-bande avec distribution alternée des bandes à gauche et à droite (à réaliser un avec un filtre graphique, par exemple le plug-in Q10 de Waves ou par FFT dans Max/MSP).
- d) Traitements restituant les modifications spectrales liées aux divers phénomènes acoustiques (absorption, diffraction, résonance) causés par les différentes parties de la tête (essentiellement le pavillon) : procédés Roland Sound Space (tridimensionnel) ou QDesign QSound, plug-in VST Steinberg Free-D (également tridimensionnel).

8. Systèmes d'écoute stéréophoniques avec haut-parleurs.

La qualité d'un système d'écoute stéréophonique repose sur cinq facteurs principaux :

- a) Ecartement entre les deux moniteurs. S'il est trop petit, il n'y a pas de place pour que l'image stéréophonique puisse se déployer et le moindre déplacement de la tête modifie profondément la perception de localisation. S'il est trop large, on peut avoir un « trou » au centre de la rampe stéréophonique, c'est-à-dire une zone dans laquelle il y a une atténuation du niveau anormale.
- b) Angle de positionnement des moniteurs. Les caractéristiques données par un fabricant pour une enceinte (et avant tout sa courbe de réponse) ne sont valables que dans un certain angle. De plus, l'angle de positionnement a une influence sur la diaphonie entre les deux canaux, qui doit être la plus petite possible (même s'il est impossible de l'éliminer totalement sur un système à haut-parleurs).
- c) Position de l'auditeur : outre la distance à l'axe imaginaire reliant les deux haut-parleurs, l'écart plus ou moins important de l'auditeur par rapport à l'axe de symétrie entre les deux enceintes

introduit une distorsion dans la perception de l'image. En effet, au plus on est proche d'un côté, au plus le son émis par celui-ci devient prépondérant dans la perception. C'est un peu comme si on regardait une photo non pas de face mais de trois-quart. Ce problème n'existe pas en monophonie ou en multiphonie (si elle est considérée comme une somme de monophonies).

- d) Acoustique du lieu d'écoute : au plus le lieu est réverbérant et surtout au plus long est le temps de réverbération, au plus la précision de l'image stéréophonique risque d'être compromise.
- e) Appariement des deux moniteurs : au plus les deux enceintes sont techniquement semblables, meilleure est l'image. Un point souvent ignoré mais néanmoins important est que les deux moniteurs devraient posséder une réponse en phase la plus identique possible.

9. Autres systèmes de captation/écoute

Au-delà des simples systèmes stéréophoniques, il existe aujourd'hui divers systèmes d'écoute (et de captation) reposant sur un plus grand nombre d'enceintes. Pour l'écoute domestique, les systèmes à 4, 5, 6 ou 7 canaux connaissent actuellement un très grand essor après avoir connu un grand succès au cinéma. Ces systèmes, appelés *surround* en anglais reposent sur un placement normalisé des enceintes autour des auditeurs. Notons que la traduction adoptée pour le mot *surround* en français est *multicanal*, ce qui crée un peu la confusion avec ce qu'on appelle la multiphonie dans le petit monde de la composition électroacoustique. En effet, s'il s'agit bien d'avoir un support pour l'enregistrement qui comporte plus de deux pistes (souvent huit), la disposition des enceintes dépend souvent du compositeur, de l'interprète ou de la configuration du lieu de projection. Par ailleurs, en composition, les huit pistes sont souvent utilisées soit comme autant de voix monophoniques, soit comme un assemblage de quatre stéréophonies soit selon d'autres configurations encore (cercle favorisant les trajectoires circulaires par exemple). En multicanal, les divers moniteurs sont essentiellement utilisés pour créer un seul champ acoustique qui rapproche de ce qu'on appelle l'ambiophonie, mais on pourrait utiliser ce système comme on le fait en électroacoustique (voir à ce sujet les expériences menées par des compositeurs tels que François Bayle, Annette Vande Gorne et d'autres pour diffuser des versions spatialisées de classiques du répertoire qui étaient en stéréo à l'origine).

Par ailleurs, l'*Ambisonics* est un système visant à créer un champ acoustique multidimensionnel. Il est possible de créer une composition en *Ambisonics* à l'aide de prises de son à canaux multiples ou par traitement de prises de son monophoniques. Bien que le format de diffusion (Format B) soit assez clairement défini, l'*Ambisonics* ne connaît pas encore une diffusion très large, peut-être parce qu'il s'agit d'un format qui n'a pas été créé par une firme industrielle.