Chapitre 2

Reproduction 3D d'un champ sonore

Assèi-te sus una formiguieira, Me dirâs qu'une es aquela que t'a picat.
(Assieds-toi sur une fourmillière, et dis-moi celle qui t'a piqué.)

Un loup est un loup - M. Folco - Ed. du Seuil, 1995

Sommaire

Dans tout ce qui suit, on entend par *reproduction 3D* ou *restitution spatialisée* d'un champ sonore, une restitution d'un champ acoustique qui *préserve les informations de localisation auditive*, de telle sorte qu'un auditeur soit capable d'identifier les positions des sources sonores et de suivre leurs déplacements dans les trois dimensions de l'espace[1]. En d'autres termes, dans la cadre de ce document, la *spatialisation sonore* ne concerne que le *positionnement des sources sonores* et la reproduction de l'effet de salle n'est pas considérée ici. Contrairement aux salles de concert où l'effet de salle est directement impliqué dans le message musical délivré aux oreilles de l'auditeur, l'effet de salle est en effet jugé nuisible dans le contexte des services de télécommunications, dans la mesure où il vient altérer l'intelligibilité de la parole et introduit des colorations spectrales sur les voix des locuteurs[2]. Cette différence provient principalement du fait que dans une application de télécommunication, la voix n'est jamais perçue directement par l'auditeur, mais toujours à travers une prise de son par un microphone. Dans une salle de concert, en revanche, l'auditeur perçoit le champ sonore avec ses propres oreilles et peut tirer parti de la sélectivité spatiale du système auditif pour extraire l'information utile du flux complexe d'évènements sonores auxquels il est soumis. Un microphone est incapable d'effectuer ce travail d'analyse et il est ensuite impossible de démêler a posteriori les différentes informations dans le signal unique transmis à l'issue de la prise de son. Par suite, dans toutes les applications de télécommunication, un effet de salle le plus mat possible est recherché, de façon à se rapprocher des conditions idéales d'une prise de son anéchoïque. Par exemple, le cahier des charges du système de visioconférence Varèse développé au C.N.E.T. spécifie un Temps de Réverbération (TR) inférieur à 300 ms de 125 Hz à 4 kHz.

2.1 Une expérience historique: le *Théâtrophone* de C. Ader (1881)

En guise de préambule à ce chapitre général sur les méthodes de spatialisation sonore, un constat frappant: l'idée d'une reproduction sonore 3D suit de quelques années seulement les inventions du téléphone (1876) et du phonographe (1877), puisque la première mise en œuvre d'un système de spatialisation sonore remonte à 1881 avec l'expérience du Théâtrophone de C. Ader [Hertz, 1981]. Il s'agissait de retransmettre un spectacle donné à l'Opéra de Paris dans des salles situées au Palais de l'Industrie. Dix microphones étaient placés sur le devant de la scène de l'Opéra (cf. Fig. 2.1a). Ils étaient connectés à des récepteurs téléphoniques situés au Palais de l'Industrie (cf. Fig. 2.1b). Un auditeur pouvait ainsi suivre le spectacle à distance au moyen d'une paire de récepteurs qu'il appliquait sur ses oreilles. L'effet de spatialisation était obtenu en reliant les deux récepteurs à des microphones distants de plusieurs mètres (cf. Fig. 2.2).

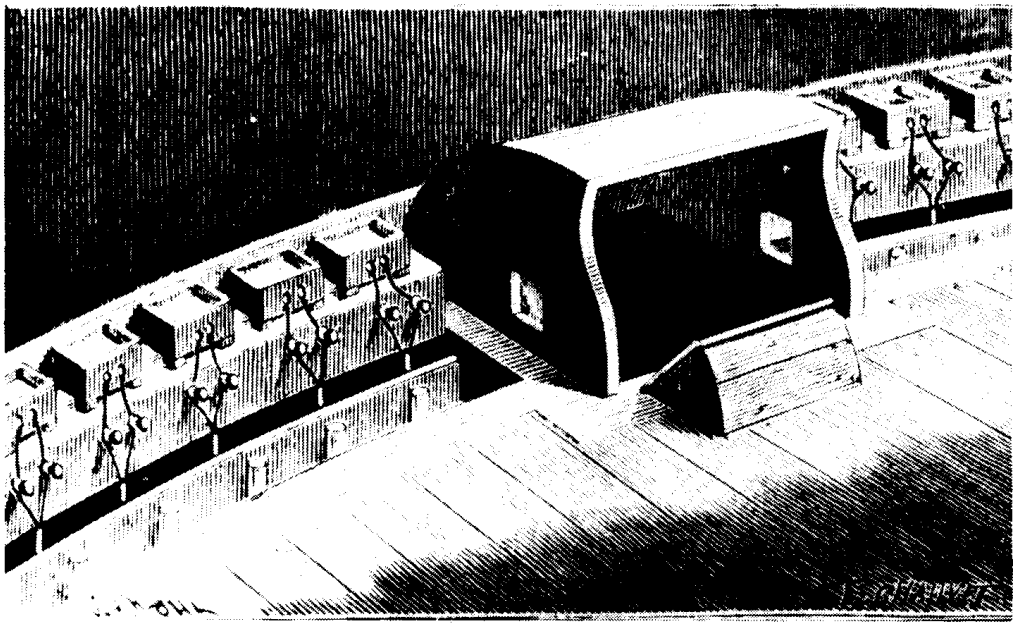
Il semble que les visiteurs du Palais de l'Industrie aient été fortement impressionnés par le rendu de la spatialisation sonore:

> "Every who has been fortunate enough to hear the telephones at the Palais de l'Industrie has remarked that, in listening with both ears at the two telephones, the sound takes a *special character of relief and localization* which a single receiver cannot produce. (...) As soon as the experiment commences the singers place themselves, in the mind in the listener, at a fixed distance, some to the right and others to the left. It is easy to follow theirs movements, and to indicate exactly, each time that they change their position, the imaginary distance at which they appear to be." [Hertz, 1981]

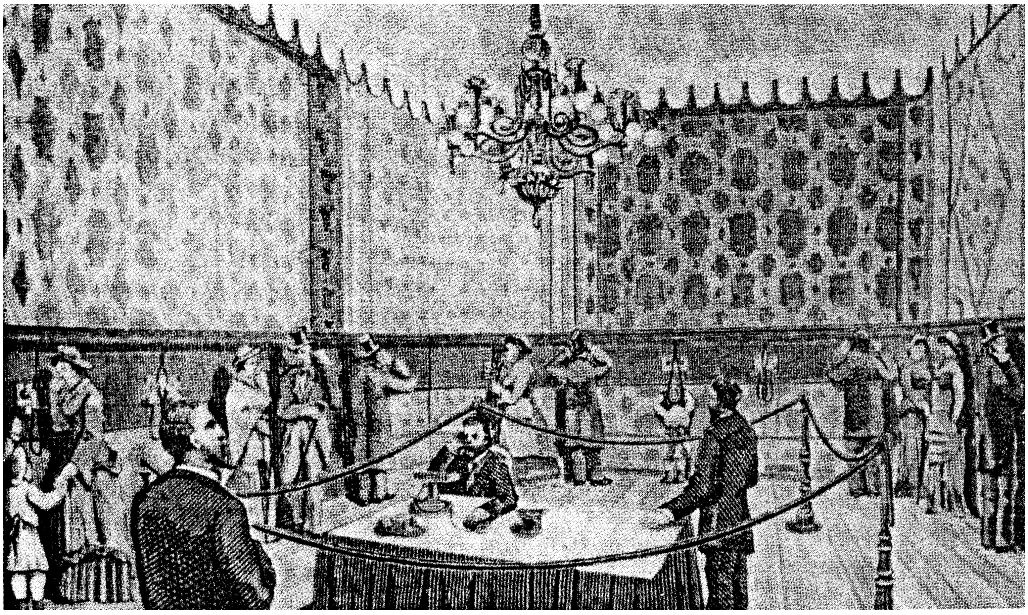
Dès l'apparition des techniques de prise et restitution sonore, l'idée de spatialisation est donc présente. Cette idée ne sera cependant pas développée avant le début des années 30 (près d'un demi-siècle plus tard!), où, sous l'impulsion du cinéma parlant, les équipes de H. Fletcher (Bell Telephon) ou de A. Blumlein (EMI) mèneront les premières recherches sur les systèmes stéréophoniques [Hugonnet & Walder, 1994]. Notre objectif n'est pas de retracer ici l'historique des techniques de spatialisation sonore: dans ce qui suit, nous nous proposons plutôt de faire un tour d'horizon des différents systèmes actuels qui permettent de reproduire

1. Dans son acception complète, une reproduction sonore *3D* implique une restitution dans tout l'espace autour de l'auditeur, c'est-à-dire dans ses trois dimensions. Cependant, il est possible de réduire le domaine où évoluent les sources sonores, par exemple à une demi-espace, voire un plan horizontal. On pourrait alors parler de "reproduction $3D^{\frac{1}{2}}$ ou 2D", en se référant au nombre de dimensions spatialisées de façon effective. Néanmoins, on préférera en général conserver le terme de reproduction 3D dans un souci de clarté.

2. L'effet de salle est en outre susceptible de renforcer les phénomènes de couplage acoustique entre les microphones et les haut-parleurs utilisés pour la prise et la restitution.



(a) Distribution des microphones sur le devant de la scène de l'Opéra (d'après [Hertz, 1981])



(b) Retransmission du spectacle au Palais de l'Industrie (d'après [Torick, 1998])

Fig. 2.1 - Expérience du Théâtrophone

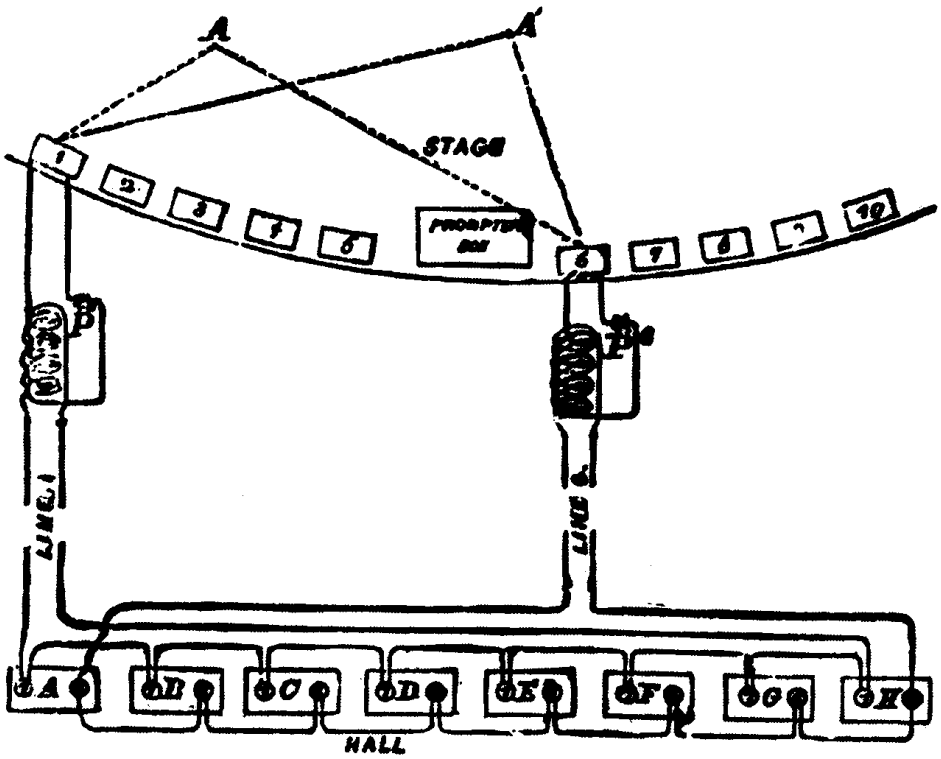


Fig. 2.2 - Expérience du Théâtrophone: Schéma de connection des microphones (Opéra) aux récepteurs téléphoniques (Palais de l'Industrie) [Hertz, 1981]

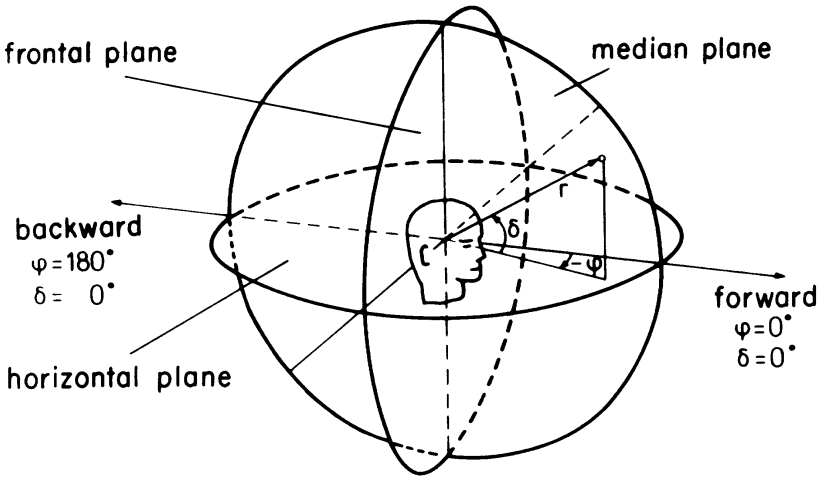


FIG. 2.3 - Localisation d'une source sonore dans l'espace à trois dimensions (la position de la source est repérée par ses coordonnées sphériques: rayon r, angle d'azimut φ et angle d'élévation δ): Localisation dans le plan horizontal (*horizontal plane*) et dans le plan médian (*median plane*) (d'après [Blauert, 1983]).

un champ sonore 3D et qui peuvent être éventuellement appliqués au concept de mur de téléprésence. Au préalable, il convient toutefois d'analyser comment l'effet de spatialisation sonore est perçu et interprété par le système auditif humain.

2.2 Perception 3D d'un champ sonore

Dans notre expérience quotidienne, nous sommes capables, les yeux fermés, de localiser les sources sonores qui nous entourent. Par quels mécanimes le cerveau identifie-t-il leurs positions à partir des seules informations contenues dans les signaux captés par les deux oreilles? On distingue communément trois types de mécanismes [Blauert, 1983]:

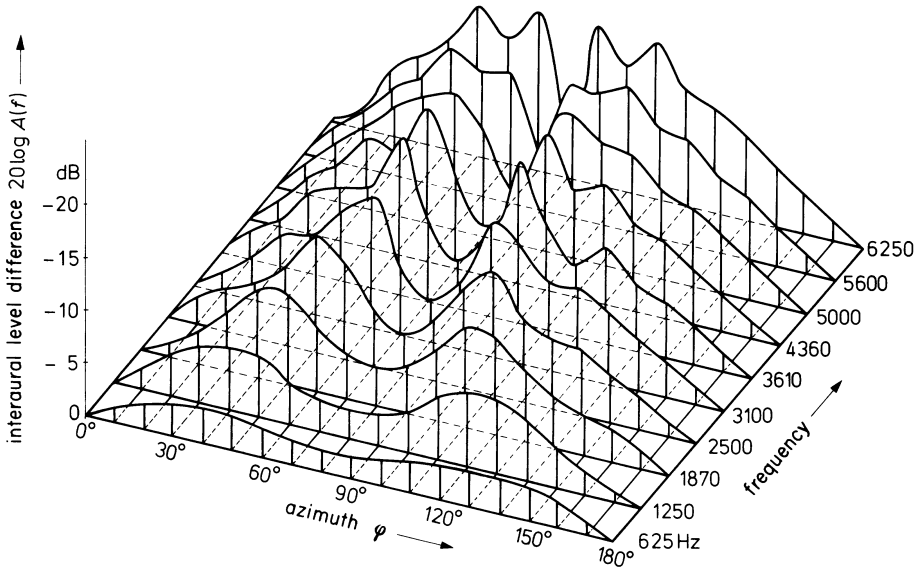
- la localisation dans le plan *horizontal* (cf. Fig. 2.3) qui permet d'identifier la position de la source sonore en azimut,
- la localisation dans le plan *médian*[3] (cf. Fig. 2.3) qui est utilisée pour repérer sa position en hauteur, c'est-à-dire en élévation,
- la localisation en *distance* pour évaluer la distance séparant la source sonore de l'auditeur.

2.2.1 Localisation dans le plan horizontal

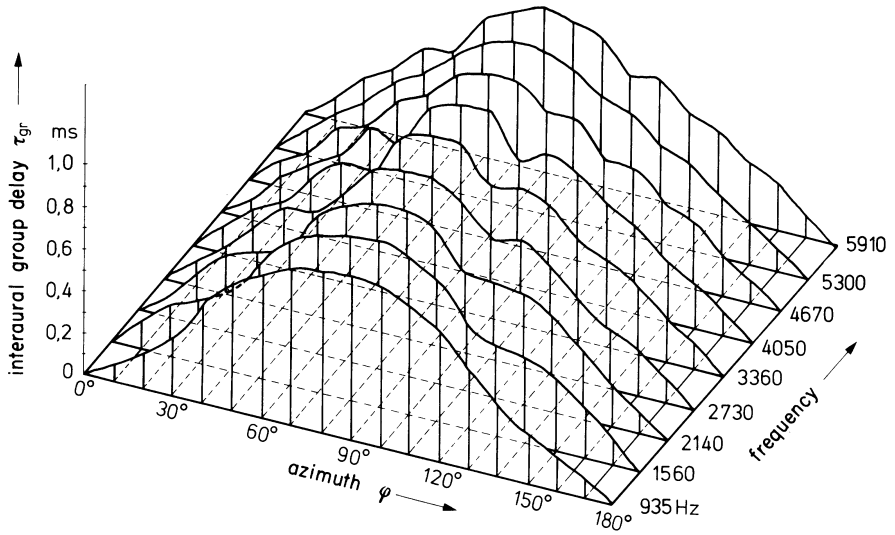
Pour localiser les sources sonores dans le plan horizontal, le système auditif utilise principalement les différences qu'il perçoit entre les signaux captés par les deux oreilles, c'est-à-dire les *différences interaurales*. Ces différences sont de deux sortes:

- les différences de *temps*[4] qui sont engendrées par les différences de trajet des deux ondes qui vont exciter les tympans (cf. Fig. 2.4a),

3. Le plan médian désigne le plan vertical qui sépare le corps humain en deux moitiés.
4. Différences de temps ou différences de phase, puisque les différences de temps se traduisent par des différences de phase pour un signal sinusoïdal.



(a) Différences Interaurales d'intensité



(b) Différences Interaurales de temps (retard de groupe)

FIG. 2.4 - Différences interaurales d'intensité et de temps calculées en modélisant la tête par une sphère rigide de diamètre 17 cm, les deux oreilles étant figurées par 2 points de sa surface situés en $(\varphi,\delta) = (100°,0°)$ et $(260°,0°)$, (d'après [Blauert, 1983]).

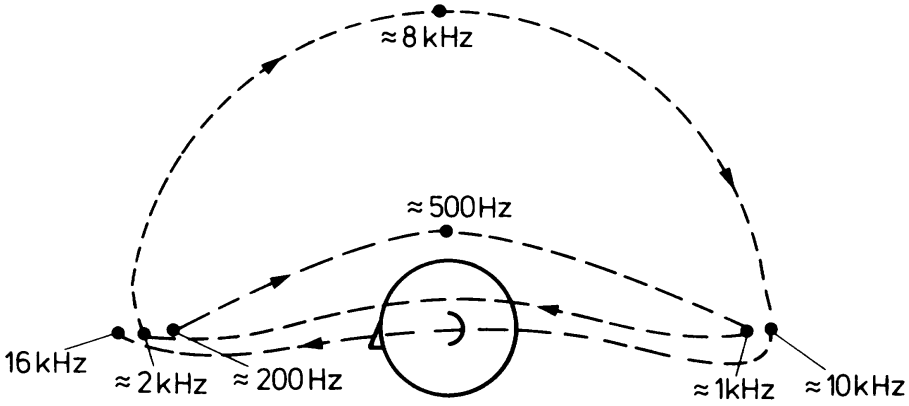


FIG. 2.5 - Expérience des "bandes directives" de J. Blauert: Lorsque l'on fait écouter des signaux à bande étroite de fréquence donnée et émis par une source fixe, l'évènement sonore est localisé indépendamment de la position de la source réelle et uniquement en fonction de la fréquence du son (d'après [Blauert, 1983]).

- les différences d'*intensité* qui proviennent essentiellement du phénomène de diffraction provoqué par la présence de la tête et qui ne sont donc perceptibles qu'aux hautes fréquences, à partir de 2 kHz environ (cf. Fig. 2.4b).

Il en résulte que les indices de localisation dans le plan horizontal dépendent de la fréquence: aux basses fréquences, l'azimut de la source est identifié sur la base des différences interaurales de temps, tandis qu'aux hautes fréquences interviennent les différences interaurales d'intensité.

2.2.2 Localisation dans le plan médian

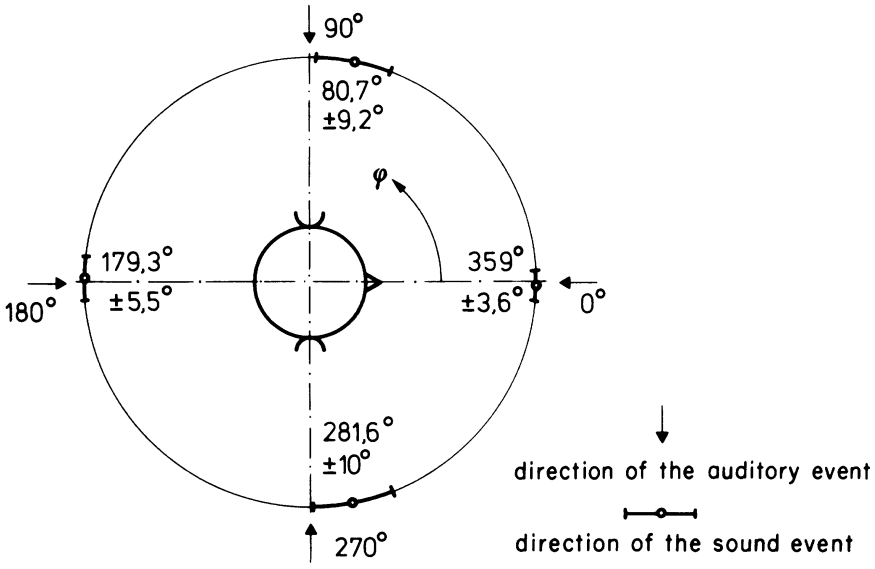
Dans le plan médian, il n'existe plus de différences interaurales. Aussi la localisation auditive est-elle basée sur des critères *monauraux*. Dans son trajet pour atteindre le tympan, l'onde acoustique subit des réflexions sur le pavillon de l'oreille, ainsi que sur les épaules et le torse de l'auditeur. Or, par un effet de filtrage en peigne, ces réflexions modifient le *timbre* du son perçu. De plus, comme la nature de ces réflexions dépend directement de l'incidence de l'onde, les modifications de timbre engendrées peuvent être interprétées pour identifier la hauteur de la source sonore. Ce phénomène a été mis en évidence par l'expérience des "bandes directives" de J. Blauert (1968) [Blauert, 1983]: en faisant écouter des signaux à bande étroite émis par une source fixe (un haut-parleur), on montre que la localisation de l'évènement sonore dans le plan médian n'est absolument pas reliée à la position de la source réelle, mais est uniquement déterminée par la fréquence du son (cf. Fig. 2.5). Ainsi un son de 1 kHz est systématiquement localisé derrière l'auditeur, tandis qu'un son de 8 kHz est perçu au dessus de sa tête, ceci quelle que soit la position du haut-parleur dans le plan vertical médian.

Contrairement aux différences interaurales de temps ou d'intensité qui relèvent de mécanismes innés, la localisation dans le plan médian est un mécanisme acquis, c'est-à-dire que l'individu apprend au cours de son existence à associer une coloration spectrale particulière à un hauteur donnée.

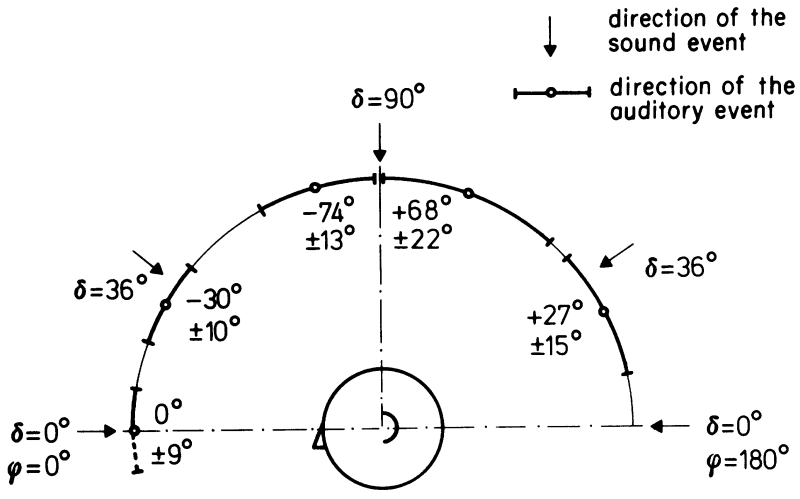
Par ailleurs, il faut aussi noter que la localisation dans le plan horizontal est nettement plus performante que la localisation dans le plan médian: une source sonore est localisée dans le plan horizontal avec une précision de l'ordre de quelques degrés seulement (cf. Fig. 2.6a), alors que l'erreur de localisation dans le plan médian est de l'ordre de une à plusieurs dizaines de degrés (cf. Fig. 2.6b).

2.2.3 Localisation en distance

Il est difficile pour le système auditif d'identifier la distance d'une source sonore dans *l'absolu*. Les indices visuels, ainsi que la connaissance du signal émis et de la source, jouent un grand rôle. L'évaluation des



(a) Localisation dans le plan horizontal



(b) Localisation dans le plan médian

FIG. 2.6 - Précision de la localisation dans le plan horizontal et dans le plan médian (d'après [Blauert, 1983]).

distances en *relatif* est mieux maîtrisée. Trois indices interviennent principalement [Blauert, 1983]:

- le *niveau sonore*: plus il est fort, plus la source est perçue proche,
- le *rapport entre l'énergie du champ direct et l'énérgie du champ réverbéré*: un rapport faible donne une sensation d'éloignement,
- le *contenu spectral*: les hautes fréquences étant plus fortement atténuées par la propagation dans l'air, un son riche en hautes fréquences renforce la sensation de présence de la source sonore[5]. Par ailleurs, pour des sources très proches, des distorsions spectrales liées au phénomène de champ proche (courbure du front d'onde) interviennent et peuvent influencer le jugement de la distance. En particulier, un renforcement des basses fréquence est observé.

On retiendra que, compte tenu de sa difficulté d'évaluation en situation réelle, la distance d'une source sonore est un paramètre délicat à simuler et à contrôler dans les champs sonores virtuels. En outre il est fortement influencé par des indices non auditifs.

2.2.4 Théorie des H.R.T.F

Il existe une théorie de la localisation auditive à la fois plus générale et plus complète que les indices interauraux et monauraux. Elle est basée sur des *fonctions de transfert* exprimant la propagation acoustique entre la source sonore située en un point donné $\vec{r}_s[r, \varphi, \delta]$ et les deux oreilles de l'auditeur. Il s'agit des *H.R.T.F* (*Head Related Transfer Function*) [Møller, 1992]. Ces fonctions de transfert modélisent l'ensemble des phénomènes qui vont affecter l'onde acoustique captée par le tympan et en particulier, elles rendent compte des phénomènes de diffraction par la tête, de réflexions sur le pavillon, le torse et les épaules de l'auditeur, ainsi que des temps d'arrivée de l'onde au niveau de chaque oreille. Les H.R.T.F contiennent donc les informations des différences interaurales de temps et d'intensité et les indices spectraux monauraux, mais, au delà de ces informations, elles traduisent de manière *exhaustive* le *codage acoustique* de la position de la source sonore.

2.3 Reproduction 3D d'un champ sonore: approche physique et approche psychoacoustique

L'analyse des mécanismes de localisation auditive suggère différentes pistes pour reproduire un champ sonore spatialisé. On distingue principalement l'approche *physique* et l'approche *psychoacoustique*.

2.3.1 Approche physique

L'approche physique ne prend pas en compte les mécanismes perceptifs de localisation auditive et se borne à reproduire le champ sonore à l'identique du champ acoustique original au sein d'une zone de dimensions finies. L'auditeur est ainsi plongé dans un champ en tout point identique à celui qu'il aurait perçu en présence des sources réelles et il est donc capable de localiser les sources sonores comme dans une situation réelle. L'holophonie [Jessel, 1973] définit par excellence la méthode la plus générale de reconstruction physique d'un champ sonore: équivalent acoustique de l'holographie, elle consiste à reproduire un champ acoustique à partir d'un enregistrement sur une surface. Le système ambisonique [Gerzon, 1992b] est un autre exemple de méthode basée sur une reconstruction physique du champ acoustique[6], cependant il repose sur une approche holophonique simplifiée en exploitant les propriétés des ondes planes et, de plus, la reconstruction du champ sonore n'est valable en pratique que sur une zone limitée correspondant au centre de la tête de l'auditeur étendu à son voisinage immédiat[7].

5. Ce phénomène est à l'origine du *filtre de présence* sur les consoles de mixage.

6. A proprement parlé, le système ambisonique n'utilise véritablenemt une approche de reconstruction physique que dans les basses fréquences. Dans les hautes fréquences, il a recours à des critères psychoacoustiques, comme nous allons l'expliquer par la suite (cf. Section 2.6). Néanmoins, il convient de noter que, dans l'ensemble du document, le terme "ambisonique" se réfère implicitement au système défini pour les basses fréquences.

7. Un des apports du présent travail de thèse a consisté à démontrer que le système ambisonique est un *cas particulier de l'holophonie*. On va en effet montrer que les équations de reconstruction du champ sonore qui définissent l'approche ambisonique

Dans son principe, la reconstruction physique d'un champ acoustique définit l'approche de reproduction sonore 3D la plus fiable en assurant une restitution parfaite des effets de spatialisation, puisqu'ils sont reproduits en "grandeur nature", mais elle suppose une reconstruction exacte du champ acoustique, ce qui requiert des moyens coûteux. Cependant cette approche présente plusieurs avantages: la *taille de la zone d'écoute* est arbitraire, elle peut être aussi étendue que l'on veut, à condition de disposer de moyens suffisants. En d'autres termes, la taille de la zone d'écoute n'est absolument pas imposée dans le principe intrinsèque de la méthode de spatialisation sonore. Par suite, il est possible de définir une aire d'écoute suffisamment grande pour accueillir plusieurs auditeurs qui seront libres de s'y déplacer, ce qui constitue un des points importants du cahier des charges dicté par le contexte de visioconférence (cf. Chapitre 1).

2.3.2 Approche psychoacoustique

Au contraire de l'approche physique, l'approche psychoacoustique cherche à utiliser les mécanismes de perception sonore 3D afin de simplifier le processus de reproduction. Par exemple, au lieu de reproduire le champ sonore sur toute une zone, on peut se contenter de le reproduire uniquement au niveau des deux oreilles de l'auditeur[8]. Les techniques *binaurales* se fondent sur cette idée. On note que, du moins en théorie, les effets de spatialisation sont restitués dans leur intégralité, dès lors que le champ acoustique est reconstruit exactement au niveau des oreilles de l'auditeur, de telle sorte que ses tympans perçoivent un champ identique à celui qu'auraient induits les sources réelles. En termes de moyens, le gain est très appréciable: il suffit de disposer par auditeur de deux microphones pour la prise de son et de deux sources (de préférence un casque) pour la restitution.

Cependant, cette réduction des coûts se paye par de sévères limitations. Un enregistrement binaural n'est en effet valable que pour une seule position d'écoute, ce qui exclut la possibilité de se déplacer. De plus, cet enregistrement porte la carte d'identité acoustique de l'individu sur lequel il a été réalisé. Or, si, entre différents individus, les H.R.T.F. présentent des points communs, elles possèdent dans leurs détails des différences interindividuelles qui sont essentiellement liées à la morphologie de chaque individu, notamment du point de vue des phénomènes de diffraction et de réflexion sur le corps, et qui constituent son identité acoustique. Lorsqu'on écoute un enregistrement binaural qui n'a pas été effectué dans ses propres oreilles, l'effet de spatialisation sonore est donc imparfaitement restitué.

Les technologies binaurales représentent l'étape ultime pour réduire les moyens de reproduction sans détériorer l'information spatiale. Toute simplification ultérieure se traduit nécessairement par une perte d'information. Cependant, même si une fraction de l'information *objective* n'est plus présente, sa disparition peut ne pas être perceptible d'un point de vue *subjectif, compte tenu des limitations des performances du système auditif*. L'information disponible dans les signaux captés au niveau des deux tympans est en effet destinée à être traitée par le système auditif pour en extraire des indices pertinents, à partir desquels s'opère le processus de localisation. Les différences interaurales de temps et d'intensité constituent deux exemples de ces critères perceptifs. Dès lors, on peut envisager de tenir compte des mécanismes de perception du système auditif afin d'identifier la quantité minimale d'information à reproduire pour obtenir un champ *psychoacoustiquement* identique au champ original, c'est-à-dire tel que l'oreille, en raison de la limitation de ses performances, soit incapable de les distinguer l'un de l'autre.

La *stéréophonie*[9] [Hugonnet & Walder, 1994] [Condamines, 1978] exploite cette idée, encore que de façon relativement approximative, dans la mesure où les effets de spatialisation sonore sont basés exclusivement sur des différences interaurales de temps ou d'intensité [Blauert, 1983]. Comme ces critères ne contrôlent la localisation que dans le plan horizontal, la stéréophonie n'est pas une technique de reproduction 3D à part entière: elle n'offre effectivement qu'une reproduction 2D. En outre, comme pour les techniques binaurales, la restitution n'est valable que pour une position d'écoute. En revanche, les différences interaurales de temps ou d'intensité présentent une assez faible variance interindividuelle et, par suite, l'effet de spatialisation est

peuvent être dérivées des équations holophoniques. Ce résultat sera détaillé au chapitre 7, dans lequel l'holophonie et le système ambisonique seront réunis sous une formulation générale de reconstruction physique de champ sonore.

8. De façon similaire, on peut exiger une reproduction fidèle du champ sonore sur une fraction du spectre seulement, afin de relâcher la contrainte sur le reste du spectre.

9. Bien qu'à l'origine, le terme de stéréophonie ait été utilisé pour désigner l'ensemble des techniques de spatialisation sonore 3D [Hugonnet & Walder, 1994], nous l'utiliserons dans tout le document dans son acception actuelle où il désigne une méthode de reproduction sonore spatialisée sur plusieurs haut-parleurs et dans laquelle l'effet de spatialisation est obtenu en contrôlant des différences de temps ou d'intensité entre les différents canaux (cf. Section 2.4).

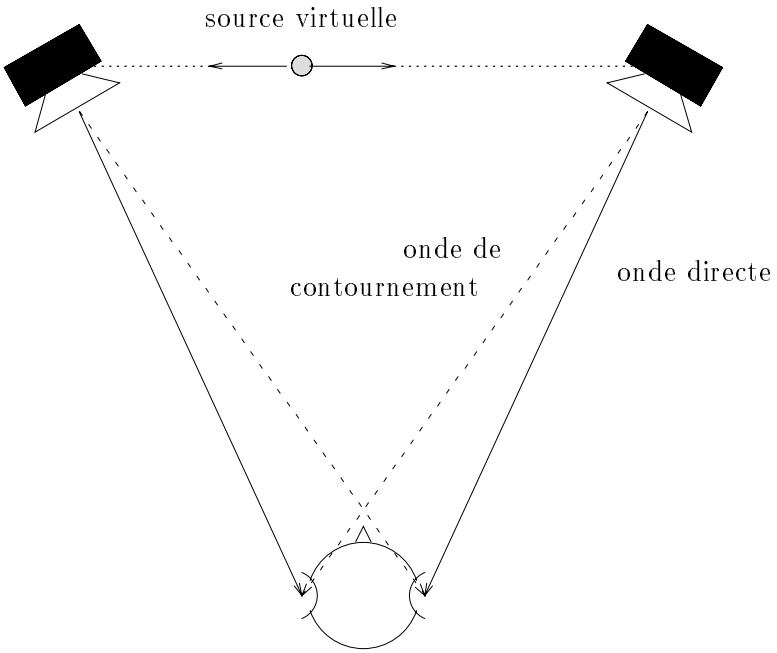


FIG. 2.7 - Système stéréophonique conventionnel

convenablement reproduit quel que soit l'auditeur.

En résumé, l'holophonie, le système ambisonique, les technologies binaurales et la stéréophonie définissent quatre approches de la spatialisation sonore. Dans ce qui précède, nous nous sommes contentés d'évoquer brièvement leurs principes, l'objectif étant de bien mettre en évidence en quoi elles dérivent d'une approche commune et comment elles s'en distinguent par des simplifications plus ou moins grossières. On retiendra notamment que les différentes méthodes psychoacoustiques ne sont que des cas particuliers de la méthode physique. A présent, ces différentes méthodes vont être détaillées, à la fois dans leurs principes fondamentaux et dans leurs performances en termes de rendu de spatialisation sonore, en adoptant l'ordre inverse de l'exposé précédent, c'est-à-dire en partant du particulier, la stéréophonie, pour aller vers le plus général, la reconstruction physique de champ sonore. Il importe de remarquer que, dans chacun des cas, les systèmes de prise et de restitution du son forment un tout indissociable, c'est-à-dire que l'on ne peut concevoir les moyens de reproduction indépendamment de la prise de son.

2.4 Stéréophonie

2.4.1 Principe

Un système stéréophonique conventionnel est constitué de deux hauts-parleurs disposés de telle sorte que la tête de l'auditeur et les deux sources forment les trois sommets d'un triangle équilatéral, les enceintes électroacoustiques étant orientées en direction de l'auditeur (cf. Fig. 2.7). Si les hauts-parleurs sont alimentés par des signaux identiques — en amplitude et en phase — , l'auditeur perçoit une source unique localisée au milieu des deux hauts-parleurs. Comme cette source n'a pas de support tangible, on parle de *source virtuelle*. Lorsque l'on introduit une *différence d'intensité* (ΔI) ou une *différence de temps* (ΔT) entre les signaux alimentant entre les deux haut-parleurs, la source virtuelle se déplace entre les deux-parleurs. Evidemment, le phénomène ne se produit que pour des faibles valeurs de ΔI et ΔT. En effet, pour des valeurs importantes de ΔI ou ΔT, la source virtuelle ne se déplace plus entre les deux haut-parleurs, mais reste localisée sur le

haut-parleur émettant le signal le plus fort ou le plus précoce. La valeur maximale des ΔI et ΔT dépend de la nature du signal. Pour un signal de parole, elle est respectivement estimée entre 15 et 17 dB pour le ΔI_{max} et entre 0.9 et 1.1 ms pour le ΔT_{max} [Hugonnet & Walder, 1994]. Pour des valeurs très élevées de ΔI et ΔT, l'effet de fusion disparaît et l'auditeur ne perçoit plus une seule source, mais deux sources dissociées localisées sur les deux haut-parleurs.

Dans son principe, le système stéréophonique s'inspire des mécanismes de localisation auditive dans le plan horizontal, dans la mesure où il est basé sur la gestion de différences d'intensité et de temps entre les signaux alimentant les deux hauts-parleurs. Cependant, cette gestion reste assez grossière. On note en outre que chaque oreille perçoit à la fois le signal émis par le haut-parleur gauche et celui émis par le haut-parleur droit. Cette diaphonie perturbe la perception des différences interaurales de temps ou d'intensité chez l'auditeur, mais, contrairement aux techniques binaurales où on cherche à s'en affranchir (cf. Section 2.5), elle est tolérée en stéréophonie. Il faut bien comprendre que l'écoute stéréophonique n'a aucun équivalent en termes de situation d'écoute réelle, puisqu'en présence de deux sources réelles, l'auditeur perçoit une source unique localisée en un point qui ne correspond à aucune des positions des sources réelles [Snow, 1953]! En fait, il semble que l'appareil auditif opère un travail de *fusion*[10] des sons en provenance des deux sources. Cette situation ne se rencontre jamais dans la nature. Certains auteurs suggèrent d'ailleurs que l'écoute stéréophonique ne serait qu'un artéfact de la perception [Polack, 1995]...

Il faut noter enfin que la restitution stéréophonique n'est pas à proprement parler un système de reproduction sonore en 3 dimensions, c'est-à-dire permettant de localiser une source en azimut, en élévation et en distance. Le système stéréophonique ne restitue en effet que deux dimensions de l'espace sonore, à savoir la localisation dans le plan horizontal. Encore faut-il remarquer que cette restitution n'est qu'incomplète étant donné que la source virtuelle ne peut se mouvoir qu'au sein de la portion d'espace compris entre les deux hauts-parleurs. En outre, même si quiconque ayant écouté une prise de son stéréophonique peut attester qu'une sensation de profondeur est présente et qu'il est possible d'identifier plusieurs plans sonores en fonction de la distance, il faut reconnaître que ce paramètre est mal contrôlé dans les systèmes stéréophoniques. En particulier, il est restitué de façon très inégale en fonction du type de prise de son adopté et du sytème électroacoustique de reproduction.

2.4.2 Prise de son

Les différences d'intensité ou de temps sont obtenues à la prise de son en enregistrant le champ sonore avec un couple stéréophonique constitué de deux microphones montés sur une structure mécanique. Les différences d'intensité sont introduites en utilisant des microphones directifs coïncidents, tandis que les différences de temps sont obtenues en espaçant les deux microphones d'une distqnce variant en tre une et plusieurs dizaines de centimètres.

On distingue trois catégories de systèmes stéréophoniques [Hugonnet & Walder, 1994]:

- la stéréophonie d'intensité (couple de microphones directifs coïncidents, cf. Fig. 2.8),
- la stéréophonie de temps (couple de microphones omnidirectifs et espacés de plusieurs dizaines de centimètres, cf. Fig. 2.9),
- la stéréophonie mixte (combinaison de ΔI et ΔT en utilisant un couple de microphones directifs non coïncidents, cf. Fig. 2.10).

10. Lorsque la fusion s'effectue sur des différences de temps, on parle d'*effet de sommation*.

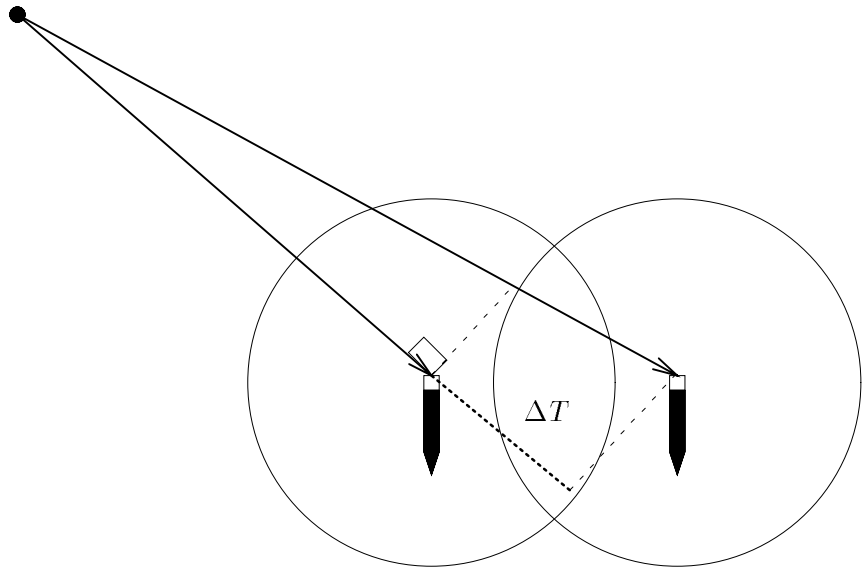
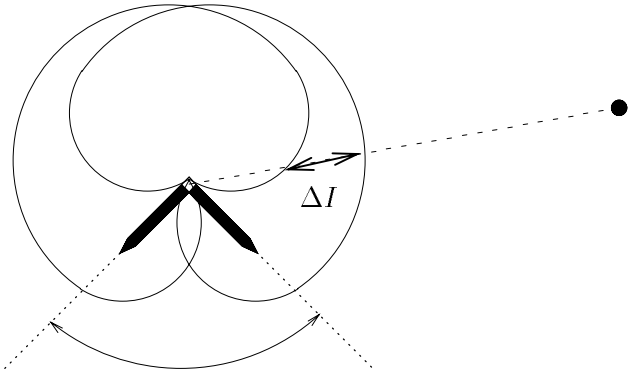
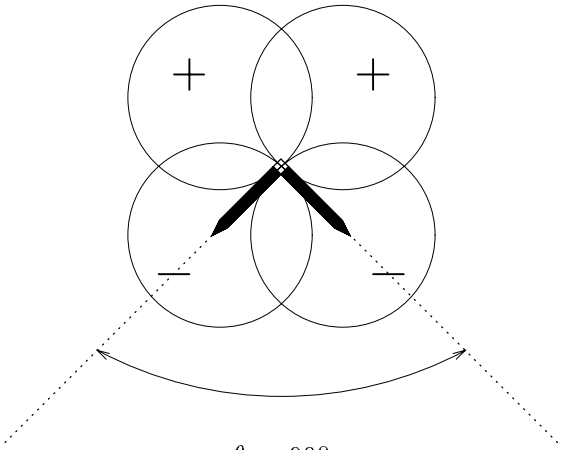


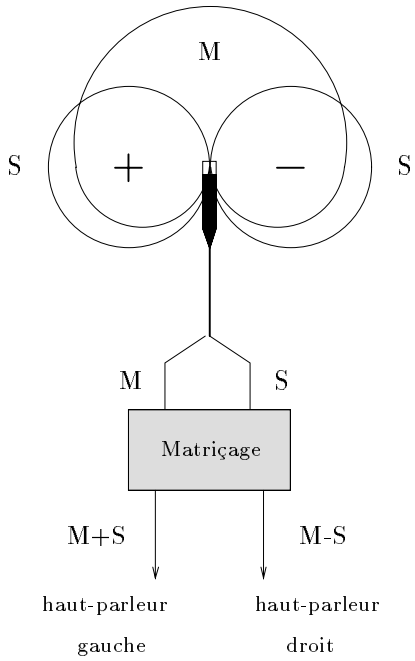
Fig. 2.8 - Stéréophonie de temps: Couple AB omni (microphones omnidirectifs non coïncidents)



(a) Couple XY (microphones cardioïdes coïncidents)



(b) Couple Stereosonic (microphones bidirectifs coïncidents orientés à 90°)



(c) Système MS – Mitte-Seite – (combinaison d'un microphone cardioïde et d'un microphone bidirectif coïncidents orientés à 90°)

FIG. 2.9 - Stéréophonie d'intensité

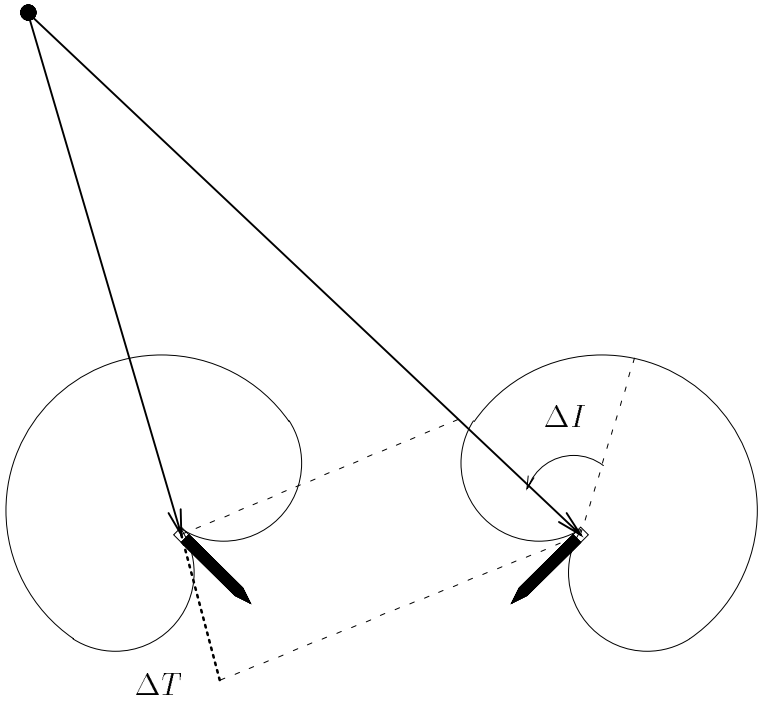


Fig. 2.10 - Stéréophonie mixte: Couple AB (microphones unidirectifs non coïncidents)

2.4.3 Restitution

Le dispositif de restitution stéréophonique conventionnel (cf. Fig. 2.7) possède théoriquement un seul point d'écoute, qui correspond au troisième sommet du triangle équilatéral dont la base est formée par les deux haut-parleurs. Seule cette position garantit une restitution idéale des effets stéréophoniques. Toutefois, en pratique, la localisation des sources virtuelles reste correcte tant que l'auditeur est situé sur l'axe de symétrie des deux enceintes. En se déplaçant le long de cette ligne, il risque cependant de sortir du lobe principal de directivité des haut-parleurs, ce qui peut engendrer des distorsions spectrales.

En revanche, dès que l'auditeur s'écarte de la ligne de symétrie des haut-parleurs, l'image stéréophonique est déformée: l'auditeur continue de percevoir une impression d'espace, mais la position des sources virtuelles est faussée par rapport au point d'écoute central. L'image stéréophonique se décale en effet vers le haut-parleur le plus proche, en même temps qu'elle se rétrécit.

Les contraintes sur la position de l'auditeur imposent donc une considérable limitation au système. Nous allons présenter maintenant plusieurs solutions mises au point pour étendre la zone d'écoute stéréophonique [Kergourlay, 1996].

2.4.4 Restitution stéréophonique étendue à trois points d'écoute: Application à un système de visioconférence [Aoki & Koizumi, 1987]

En 1987, S. Aoki et N. Koizumi ont proposé une méthode pour étendre la zone d'écoute en préservant la localisation des sources virtuelles dans le cadre d'un système de visioconférence avec trois participants. Leur idée a consisté à ajouter aux deux enceintes stéréophoniques, deux paires de haut-parleurs supplémentaires qui, au moyen de retards et d'inversion de phase, vont restituer le champ stéréophonique pour deux positions excentrées.

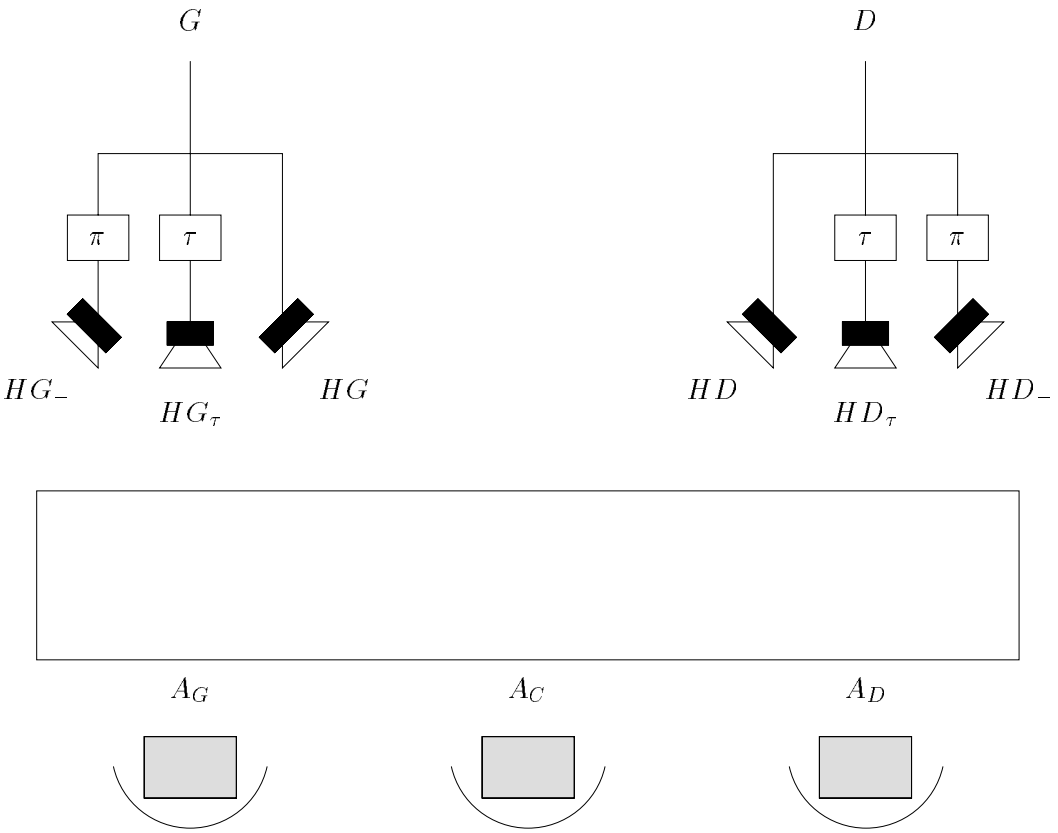


Fig. 2.11 - Dispositif de restitution stéréophonique étendue à trois points d'écoute pour un système de visioconférence [Aoki & Koizumi, 1987]

Le dispositif est illustré sur la figure 2.11. Les trois positions d'écoute sont repérées par les points A_D, A_C et A_G derrière la table de visioconférence. En regard, on distingue trois paires de haut-parleurs :

- la paire stéréophonique conventionnelle qui correspond aux haut-parleurs HD et HG alimentés par les deux signaux stéréophoniques et dont les axes sont croisés devant la zone d'écoute,
- la paire constituée des haut-parleurs HD_τ et HG_τ qui sont alimentés par les signaux stéréophoniques retardés d'une valeur de retard τ et qui pointent respectivement vers les positions A_D et A_G,
- la paire constituée des haut-parleurs HD_- et HG_- qui sont alimentés par les signaux stéréophoniques en opposition de phase et dont les axes sont dirigés vers l'extérieur de la zone d'écoute.

La valeur du retard τ est calculée de façon à ce que le signal émis par le haut-parleur HD_τ parvienne au point A_D en même temps que le signal émis par le haut-parleur HG. De même les signaux issus des haut-parleurs HG_τ et HD doivent arriver simultanément au point A_G.

Par suite, si on fait le bilan des contributions des trois paires de haut-parleurs pour les différents points d'écoute, on constate que:

- au point A_D:

  Les contributions des haut-parleurs HD et HD_- s'annulent puisqu'ils sont en *opposition de phase*. La restitution stéréophonique est assurée par les haut-parleurs HD_τ et HG et elle n'est pas perturbée ni par le haut-parleur HG_τ en raison de l'*effet de précédence*, étant donné qu'il est retardé, ni par le haut-parleur HG_-, du fait de sa *directivité*, attendu qu'il est orienté dans la direction opposée.

- au point A_G:

  En raisonnant par symétrie avec la position A_D, on se rend compte que la restitution stéréophonique est assurée par les haut-parleurs HG_τ et HD.

- au point A_C:

  La restitution stéréophonique est essentiellement dominée par les haut-parleurs HD et HG. Les haut-parleurs HD_τ et HG_τ n'interviennent pas en raison de l'effet de précédence, ni les haut-parleurs HD_- et HG_- à cause de leur directivité.

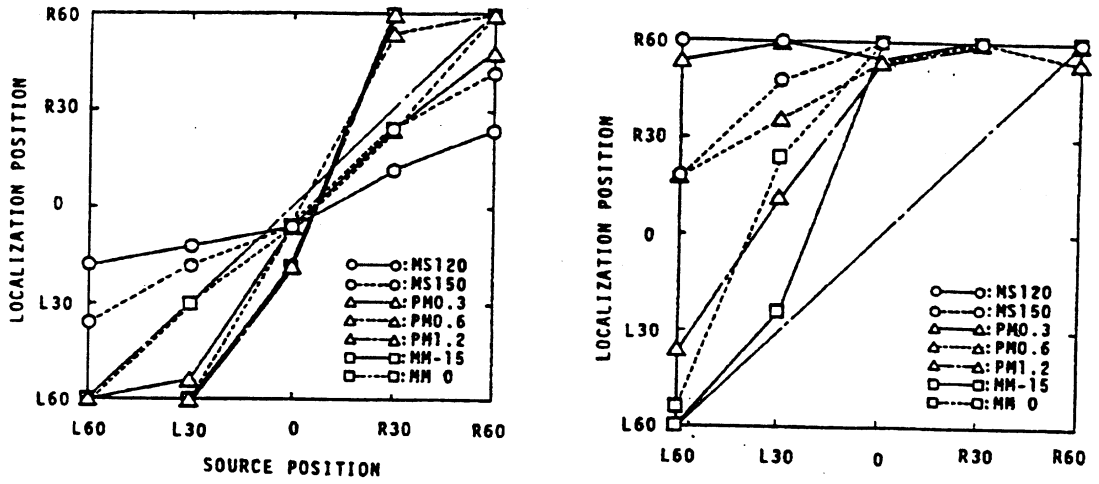
Ainsi, trois auditeurs situés aux points A_D, A_C et A_G perçoivent la même image stéréophonique. Des tests subjectifs ont évalué comment la source virtuelle est localisée par les auditeurs excentrés (cf. Fig. 2.12 & 2.13). Leurs résultats démontrent l'efficacité du système. Cette méthode d'extension de la zone d'écoute stéréophonique s'avère très séduisante. Il reste cependant à déterminer si l'image sonore demeure stable si les auditeurs se déplacent au sein de la zone d'écoute, ce qui est une des conditions du mur de téléprésence (cf. Chapitre 1). Par ailleurs, des effets de coloration spectrale sont à craindre.

On retient l'idée que l'extension de la zone d'écoute passe par l'ajout de sources complémentaires. Dans le cas présent, pour deux positions d'écoute additionnelles, on a recours à deux paires stéréophoniques supplémentaires. Par ailleurs, un travail de réflexion semble devoir être mené sur la manière d'alimenter l'ensemble des sources utilisées. On constate en effet que, dans le dispositif précédent, deux haut-parleurs ne travaillent qu'à annuler le champ émis par les deux autres haut-parleurs sur une partie de la zone d'écoute. Afin d'optimiser l'énergie *utile* du système, il serait plus pertinent de chercher plutôt à faire collaborer de façon plus constructrive l'ensemble des sources à la restitution globale du champ sonore, d'autant que la mise en œuvre d'interférences destructrives risque de générer des instabilités du champ sonore restitué. Ainsi, dans l'holophonie, l'ensemble des sources contrôle de façon globale et homogène, la restitution du champ sonore sur toute la zone d'écoute, sur la base du Principe de Huygens.

2.4.5 Restitution stéréophonique étendue avec des paires d'enceintes croisées: Application à la sonorisation de spectacles [Arnaud, 1996]

Depuis plusieurs années, le laboratoire d'acoustique de l'I.N.A.[11] s'intéresse au problème de la diffusion stéréophonique sur une zone étendue. Inspirés d'une démarche à la fois scientifique et empirique, ces travaux ont l'avantage d'être assortis d'une évaluation du système proposé dans des conditions réelles de fonctionnement pour sonoriser un spectacle [Arnaud, 1996].

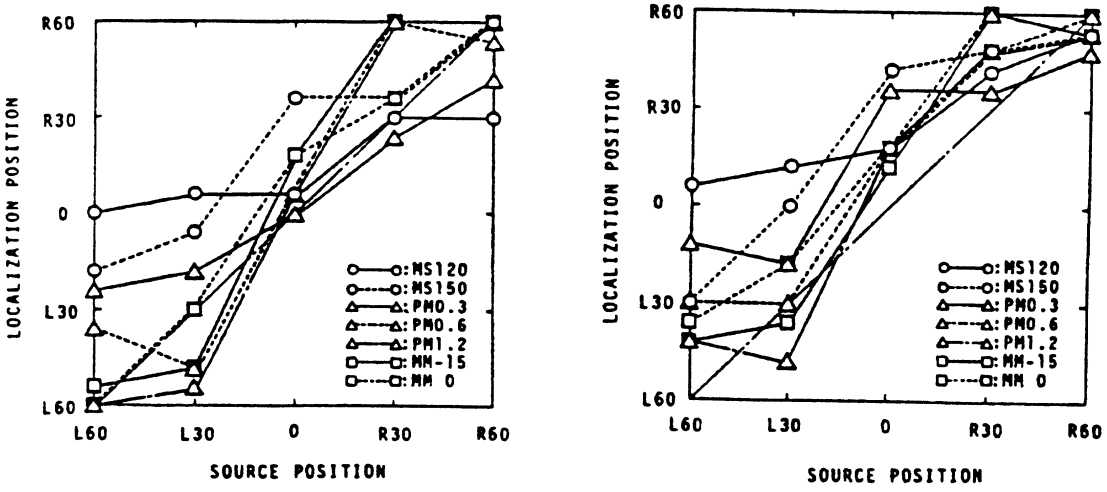
11. Institut National de l'Audiovisuel (Bry-sur-Marne, France)



(a) Position centrale

(b) Position excentrée

FIG. 2.12 - Test de localisation pour un système stéréophonique conventionnel [Aoki & Koizumi, 1987]



(a) Position centrale

(b) Position excentrée

FIG. 2.13 - Test de localisation pour le dispositif de restitution stéréophonique étendue à trois points d'écoute [Aoki & Koizumi, 1987]

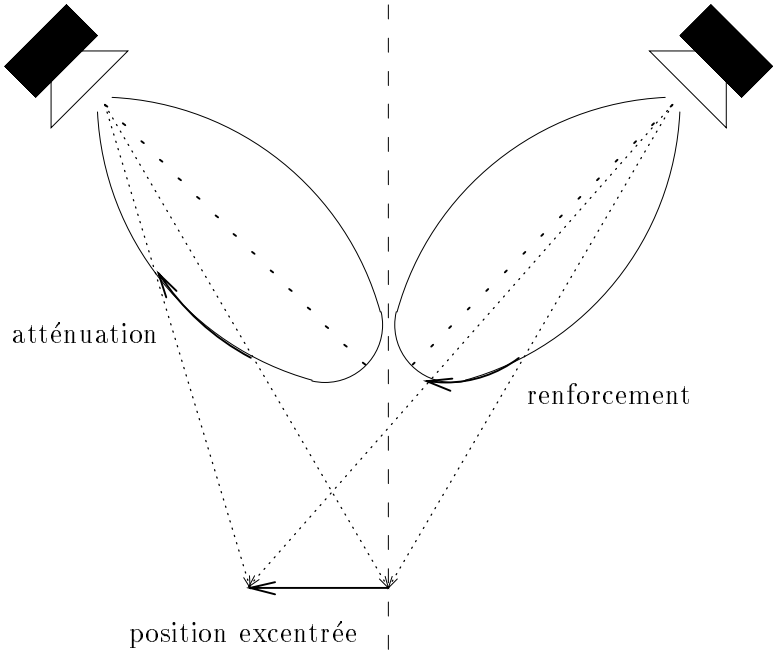


FIG. 2.14 - Utilisation de la directivité des enceintes pour compenser une différence de temps par une différence d'intensité [Arnaud, 1996]

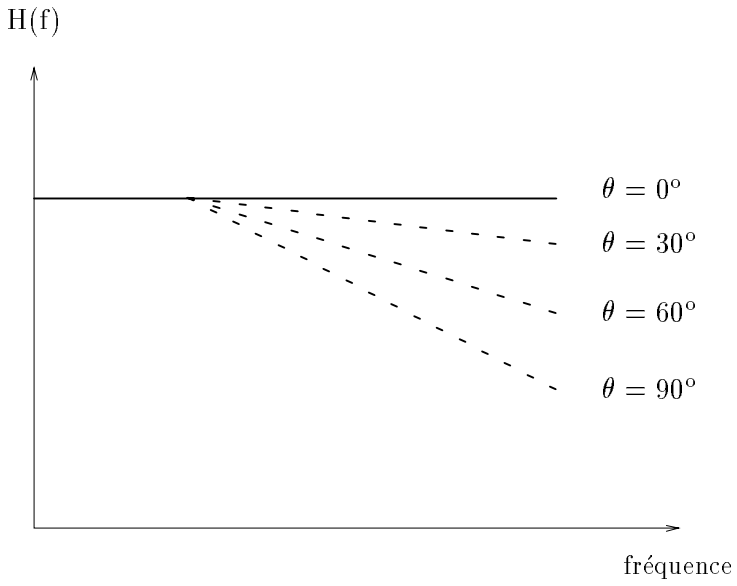
Principe

Lorsque l'auditeur s'écarte de l'axe médian des haut-parleurs (cf. Fig. 2.14), il se rapproche d'une des enceintes, ce qui introduit des ΔI et ΔT additionnelles qui viennent fausser les différences initiales entre les deux canaux stérérophoniques. Lorsque l'auditeur est suffisamment éloigné des enceintes électroacoustiques, l'effet de la différence de temps prédomine, la différence d'intensité devenant négligeable. Or, plusieurs auteurs ont montré qu'il était possible au niveau perceptif de compenser une différence de temps par une différence d'intensité — et inversement — pour repositionner la source virtuelle [Blauert, 1983]. L'idée consiste donc à utiliser la directivité des enceintes électroacoustiques de façon à compenser les ΔT par des ΔI (cf. Fig. 2.14). Dans ce but, les deux enceintes stéréophoniques sont croisées devant la zone d'écoute, de telle sorte que, lorsque l'auditeur s'écarte de l'axe de symétrie des haut-parleurs, il s'éloigne de l'axe de directivité principale de l'enceinte la plus proche tandis qu'il se rapproche de celui de l'enceinte la plus éloignée (cf. Fig. 2.14). Le son perçu en avance est ainsi atténué, ce qui correspond bien à compenser les ΔT dues aux erreurs de positionnement par des ΔI introduites par la directivité des enceintes.

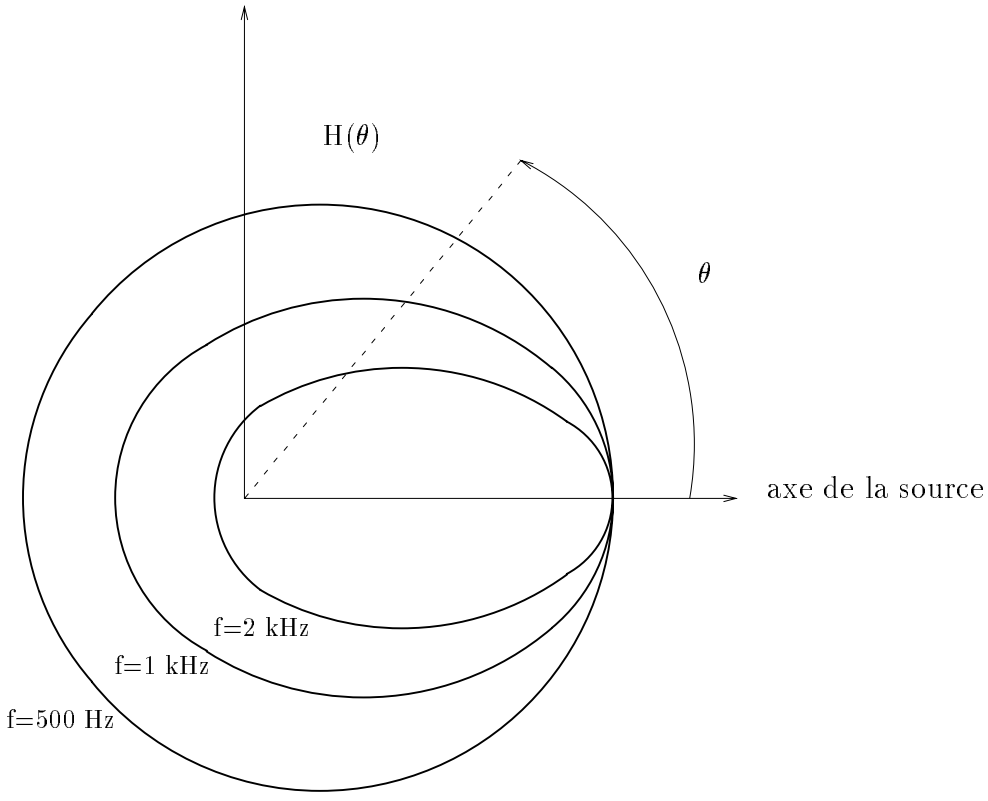
Enceinte à Directivité Contrôlée Croissante (E.D.C.C.)

Cette solution a déjà été proposée par d'autres auteurs (cf. [Bauer, 1960] par exemple), toute l'originalité des travaux menés à l'I.N.A. réside dans le soin porté à sa mise en œuvre. Le système repose sur l'utilisation de paires d'enceintes à *directivité contrôlée croissante* (E.D.C.C.) en fréquence. Le concept des E.D.C.C. garantit la totale maîtrise de la réponse de l'enceinte — en amplitude et en phase — dans toutes les directions, afin de maintenir un rendu cohérent quel que soit l'angle d'écoute. Ce comportement est requis pour la mise en œuvre de paires d'enceintes croisées dont le principe a été décrit au paragraphe précédent. Ce dispositif consiste en effet à croiser les enceintes stéréophoniques devant la zone d'écoute: l'auditeur n'est donc plus situé dans l'axe principal des enceintes et il importe de préserver la qualité de restitution des enceintes en dehors de leur lobe principal de directivité.

On pourrait penser qu'une enceinte à *directivité constante* constitue la meilleure solution. Mais, en pratique, ce type d'enceintes offre une directivité contrôlée sur un angle limité, en dehors duquel la réponse



(a) Réponse en fréquence $H(f)$



(b) Diagramme de directivité $H(\theta)$

FIG. 2.15 - Principe d'une enceinte à directivité contrôlée croissante (E.D.C.C.)

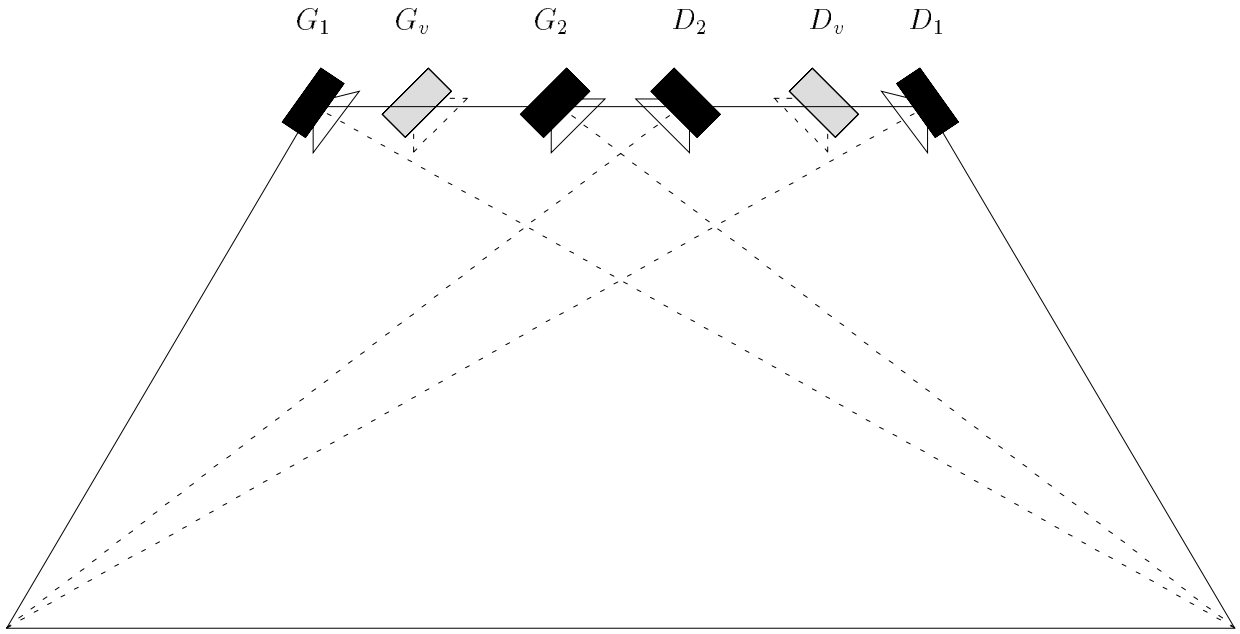


FIG. 2.16 - Dispositif de paires d'enceintes croisées pour une zone de restitution stéréophonique étendue (vue de dessus): L'ensemble des enceintes gauches et droites sont utilisées pour synthétiser deux sources stéréophoniques gauche et droite virtuelles dont la localisation est indépendante de la position de l'auditeur au sein de la zone d'écoute considérée.

devient très irrégulière avec, en particulier, des accidents de phase. Pour obtenir des résultats homogènes dans toutes les directions d'émission, une enceinte dont la directivité croît linéairement en fréquence (cf. Fig. 2.15) est préférable, ce qui définit le concept d'"enceinte à directivité contrôlée croissante".

La principale limitation de ce système provient de la complexité des lois de compensation ΔI / ΔT qui dépendent de la fréquence de façon fortement non linéaire, ce qui en rend le contrôle très difficile. Il est donc important de valider la démarche par des mesures physiques et des tests psychoacoustiques.

Application à la sonorisation de spectacles

Dans le cadre du festival IMAGINA [12], l'I.N.A. a été chargé de sonoriser un chapiteau (Espace Fontvieille à Monaco) pour y projeter les films en compétition. Etant données les dimensions de la zone à sonoriser, le principe des E.D.C.C. a été étendu à plusieurs paires d'enceintes croisées. La figure 2.16 illustre le dispositif pour deux paires d'enceintes croisées. L'idée consiste à synthétiser une paire stéréophonique *virtuelle* en combinant les contributions des différentes paires réparties devant la zone d'écoute. Plus précisément, une source virtuelle gauche est synthétisée en ajoutant les contributions de l'ensemble des sources gauches de chaque paire, de même qu'une source virtuelle droite est obtenue en sommant les contributions de toutes les sources droites. Par exemple, dans le cas de deux paires d'enceintes croisées (G_1,D_1) et (G_2,D_2) (cf. Fig. 2.16), les sources stéréophoniques virtuelles G_v et D_v sont respectivement synthétisées à partir des sources G_1 et G_2 d'une part, D_1 et D_2 d'autre part. La position des sources virtuelles est calculée par une méthode vectorielle (cf. Fig. 2.17): pour chaque enceinte, on trace à partir du point d'écoute considéré, un vecteur qui pointe dans la direction de l'enceinte et dont la norme est proportionnelle à la quantité d'énergie rayonnée par l'enceinte dans cette direction.

12. Le festival IMAGINA est une manifestation annuelle consacrées aux créations liées à l'image de synthèse.

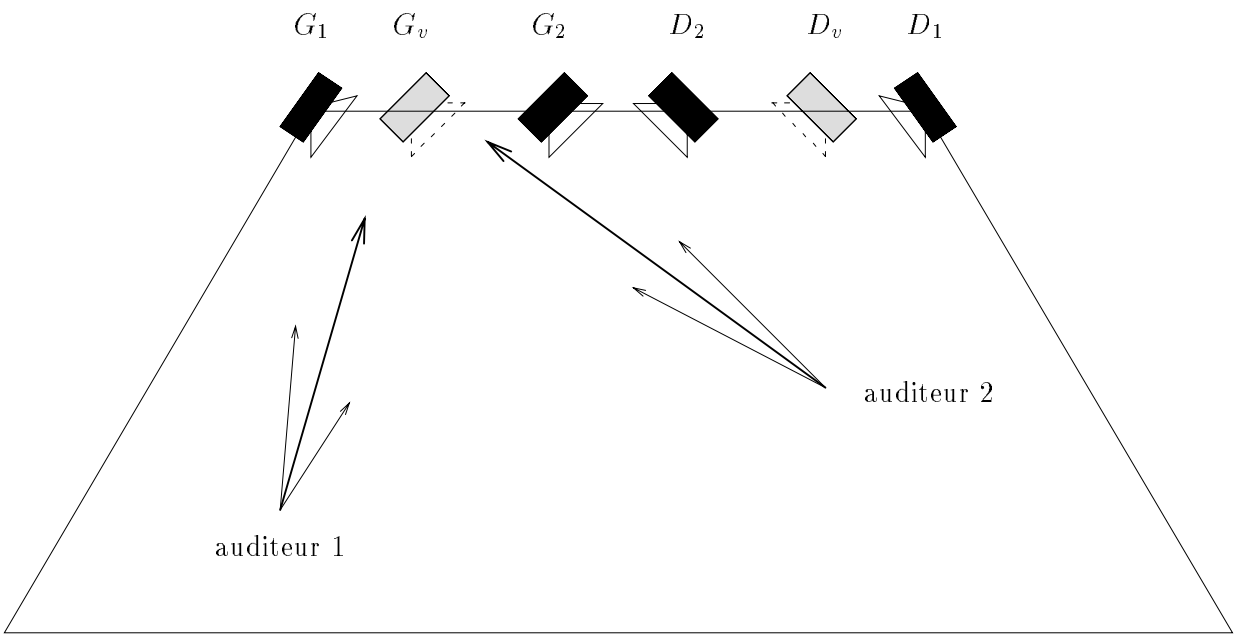
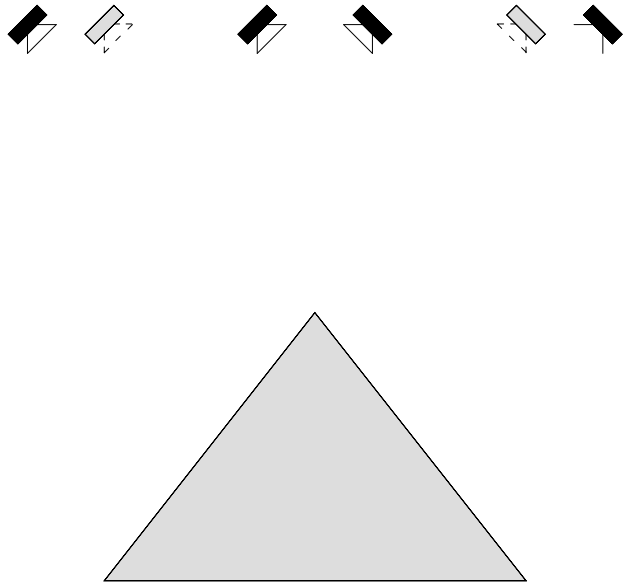
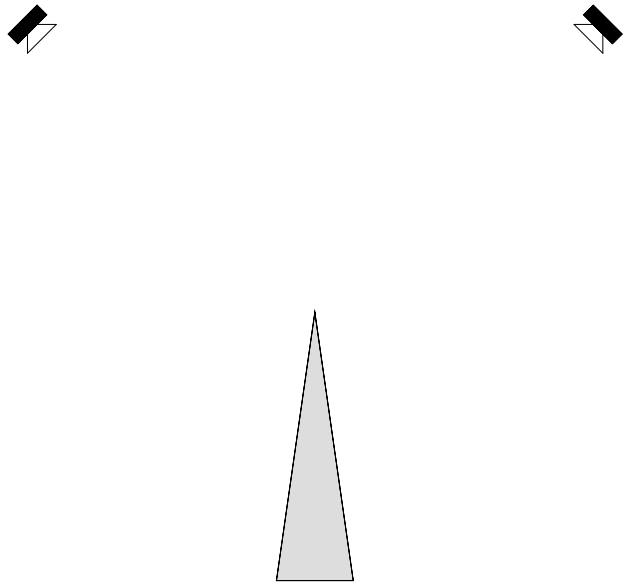


Fig. 2.17 - Méthode vectorielle de construction des sources virtuelles



(a) Système de multidiffusion stéréophonique de l'I.N.A.



(b) Système stéréophonique conventionnel

FIG. 2.18 - Etendue de la zone de restitution stéréophonique stable évaluée sur la base d'un critère de distorsion d'imagerie V (la surface grisée représente la zone sur laquelle le critère V est inférieur ou égal à 10%): comparaison des performances du système de multidiffusion stéréophonique de l'I.N.A. avec un dispositif stéréophonique conventionnel [Arnaud, 1996]

L'orientation des enceintes est déterminée afin de préserver une image stéréophonique homogène et stable sur une zone étendue. Le principe des enceintes croisées devant la zone d'écoute (cf. Fig. 2.14) est alors généralisé à plusieurs paires stéréophoniques. Chaque enceinte est dirigée vers le point extrême situé à l'opposé de la zone d'écoute (cf. Fig. 2.16): de cette façon, elle pointe vers le point le plus éloigné qui va donc bénéficier de son rayonnement maximal. Inversement, pour les positions d'écoute les plus proches de la source, le niveau sonore perçu est atténué par sa directivité de l'enceinte. Ce dispositif peut être appliqué à un nombre quelconque de paires d'enceintes.

La validité de cette approche a pu être testée avec le système mis en place au festival IMAGINA. L'étendue de la zone offrant une restitution stéréophonique stable a notamment été évaluée. Les résultats indiquent que cette zone représente près de 75% de l'espace à sonoriser. Des tests psychoacoustiques réalisés avec un dispositif similaire monté sur un plateau de télévision à l'I.N.A. ont complété cette évaluation. Ils confirment le meilleur comportement du système par rapport au dispositif stéréophonique classique, en ce qui concerne la stabilité de la localisation des sources (cf. Fig. 2.18). Cependant, les performances du système s'avèrent être sensibles au type de prise de son — selon qu'il s'agit de stéréophonie de temps, d'intensité, ou de multimicrophonie[13] —, à la la nature des sources, ainsi qu'à l'environnement acoustique de restitution. De plus, il convient de noter que la stabilité de la localisation n'est évaluée que pour une source virtuelle centrale. On peut se demander si les résultats restent valables pour d'autres positions de sources virtuelles.

Par ailleurs, il faut noter que cette approche, comme la précédente d'ailleurs, ne permet que d'étendre la zone d'écoute: les limitations inhérentes au procédé stéréophonique demeurent. En particulier, la spatialisation sonore reste partielle et ne concerne toujours que la portion d'espace comprise entre les deux sources stéréophoniques.

2.4.6 Stéréophonie dirigée: Panoramique d'intensité

Une stéréophonie artificielle

Dans tout ce qui précède, les deux signaux stéréophoniques sont obtenus à partir d'une prise de son par un couple stéréophonique. En d'autres termes, l'effet de spatialisation est réalisé de façon acoustique et les différences de temps ou d'intensité sont introduites par le jeu du positionnement et des directivités des microphones. Avec la stéréophonie dirigée, deux signaux stéréophoniques sont dérivés à partir d'une prise de son monophonique. L'effet de spatialisation est introduit *artificiellement*, ce qui justifie le terme de *stéréophonie dirigée*, en affectant le même signal sur les deux canaux et en jouant sur le gain relatif entre les deux voies, de façon à produire une différence d'intensité qui va permettre de déplacer la position de la source virtuelle vers l'un des deux haut-parleurs, la source virtuelle étant localisée au milieu de deux haut-parleurs lorsqu'aucune différence de gain n'est appliquée. Le procédé est encore désigné sous le nom de *panoramique d'intensité*.

Loi des Sinus et Loi des Tangentes

La position de la source virtuelle, repérée par son azimut φ (cf. Fig. 2.19), est reliée aux gains g_1 et g_2 des signaux alimentant les deux haut-parleurs par la *loi des sinus* [Pulkki, 1997]:

$$\frac{\sin\varphi}{\sin\varphi_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{2.1}$$

Cette loi est valable pour les basses fréquences, sous réserve que l'auditeur garde la tête immobile en fixant le point au milieu des deux haut-parleurs. Si l'auditeur tourne la tête dans la direction de la source virtuelle, il convient d'appliquer plutôt la loi des tangentes [Pulkki, 1997]:

$$\frac{\tan\varphi}{\tan\varphi_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{2.2}$$

Cependant, les deux lois diffèrent peu en pratique.

13. Une prise de son multimicrophonique consiste à enregistrer chaque source avec un microphone individuel. En postproduction, les différents signaux ainsi obtenus sont sommés pour constituer les deux voies stéréophoniques. Au cours de cette opération, les sources sonores sont positionnées dans l'espace en jouant sur le gain relatif des deux voies selon une procédure de "panoramique d'intensité", aussi considérée comme une *stéréophonie dirigée*, qui va être décrite à la section suivante (cf. Section 2.4.6).

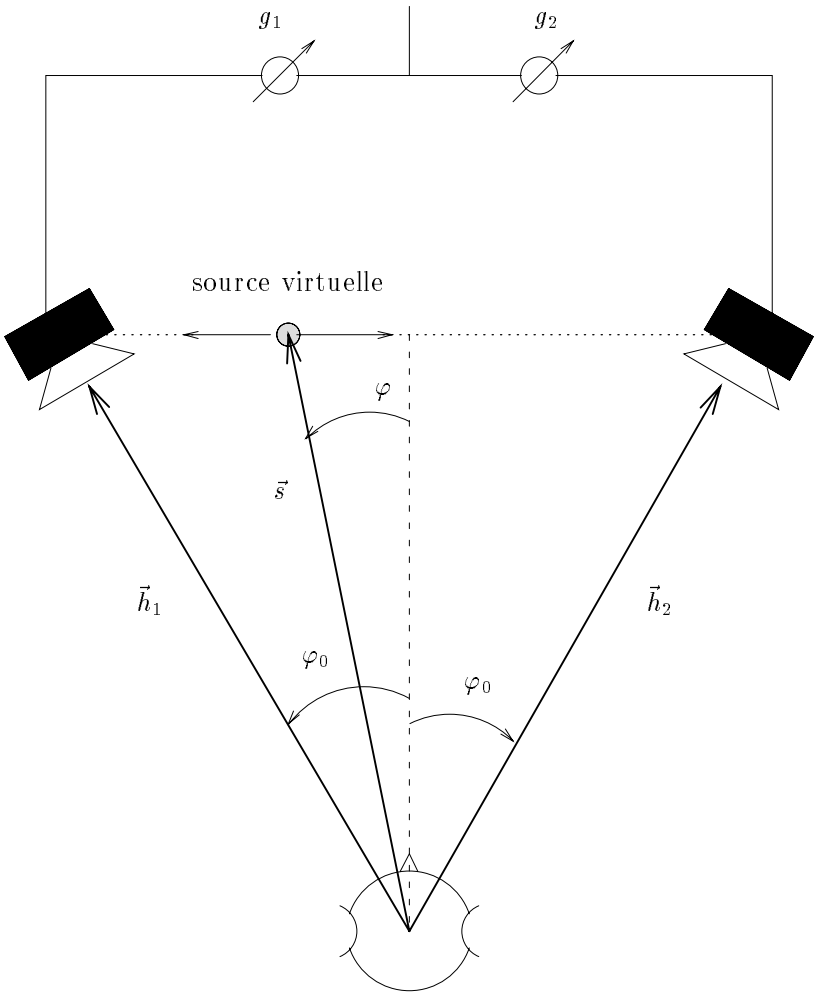


Fig. 2.19 - Principe de la stéréophonie dirigée: Contrôle de la localisation de la source virtuelle par un panoramique d'intensité

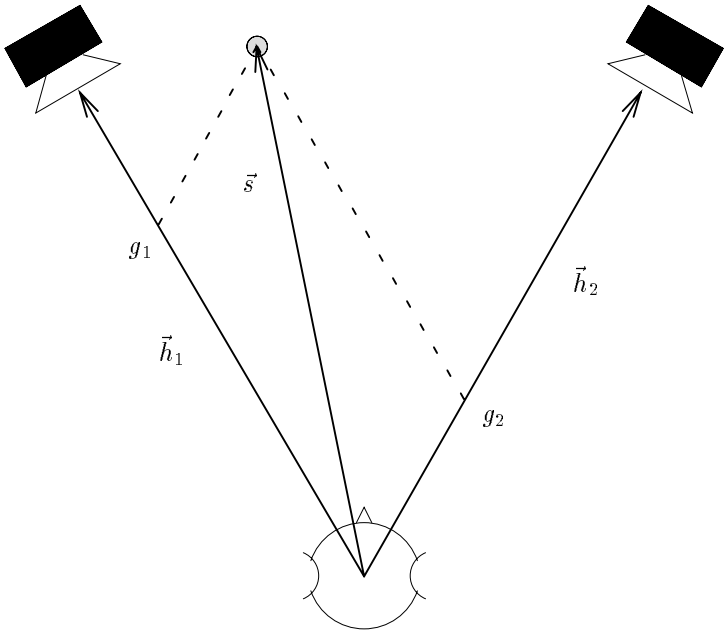


FIG. 2.20 - Méthode V.B.A.P.: Projection de la source virtuelle sur une base vectorielle constituée à partir des haut-parleurs

Méthode V.B.A.P [Pulkki, 1997]

Récemment, la loi des tangentes a été reformulée sous forme vectorielle, pour donner une méthode générale de panoramique d'intensité, désignée sous le nom de méthode V.B.A.P (*Vector Base Amplitude Panning*) [Pulkki, 1997] et qui consiste à projeter la source virtuelle sur la base vectorielle constituée à partir des deux haut-parleurs, l'auditeur représentant l'origine du repère (cf. Fig. 2.20). Ainsi, si les vecteurs $\vec{h}_1[h_{1_x}, h_{1_y}, h_{1_z}]$ et $\vec{h}_2[h_{2_x}, h_{2_y}, h_{2_z}$ repèrent les positions des deux haut-parleurs et le vecteur $\vec{s}[s_x, s_y, s_z]$, la position de la source virtuelle (cf. Fig. 2.20), on exprime le vecteur $\vec{s}$ comme une combinaison linéaire des vecteurs $\vec{h}_1$ et $\vec{h}_2$, les coefficients de la combinaison étant donnés par les valeurs des gains g_1 et g_2 des haut-parleurs:

$$\vec{s} = g_1 \ \vec{h}_1 + g_2 \ \vec{h}_2 \tag{2.3}$$

En d'autres termes, les gains (g_1,g_2) représentent les coordonnées de la source virtuelle dans la base ($\vec{h}_1$,$\vec{h}_2$). L'équation précédente peut se réécrire sous forme matricielle:

$$\mathbf{s} = \mathbf{h}\ \mathbf{g} \tag{2.4}$$

avec:

$$\begin{array}{lcl} \mathbf{s}^T & = & [s_x s_y s_z] \\ \mathbf{g}^T & = & [g_1 g_2] \\ \mathbf{h}^T & = & [\mathbf{h_1 h_2}] \\ \mathbf{h_1}^T & = & [h_{1_x} h_{1_y} h_{1_z}] \\ \mathbf{h_2}^T & = & [h_{2_x} h_{2_y} h_{2_z}] \end{array}$$

Les gains des haut-parleurs sont alors obtenus en résolvant cette équation matricielle dont la solution existe dès que la matrice $\mathbf{h}$ possède une matrice inverse $\mathbf{h}^{-1}$:

$$\mathbf{g} = \mathbf{h}^{-1}\ \mathbf{s} \tag{2.5}$$

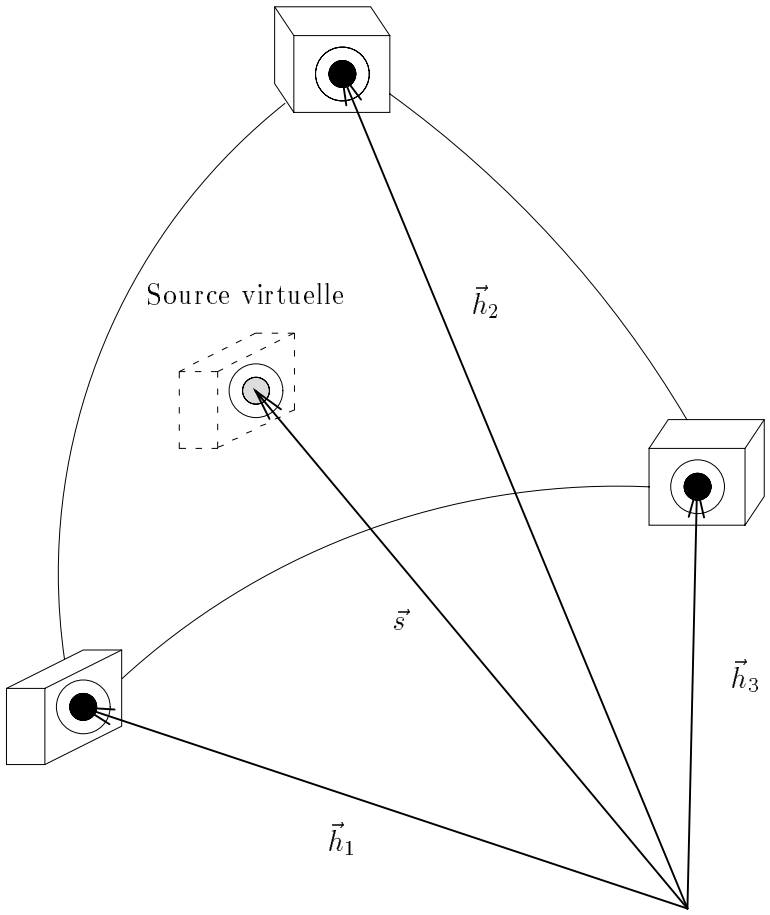


FIG. 2.21 - Méthode V.B.A.P. étendue à trois dimensions

On peut vérifier que cette solution redonne la loi des tangentes [Pulkki, 1997]. Il faut noter que, d'un point de vue psychoacoustique, la méthode V.B.A.P. n'est valable qu'aux basses fréquences. Sur le même principe que la méthode V.B.A.P., mais en raisonnant en termes d'énergie, une autre approche a été dérivée pour les hautes fréquences sous le nom de V.B.I.P. (*Vector Base Intensity Panning*) [Pernaux *et al.*, 1998]. Pour restituer l'ensemble du spectre, il conviendrait donc de coupler les deux approches, V.B.A.P. et V.B.I.P., au moyen d'un filtrage par bande.

Panoramique d'intensité 3D

L'intérêt de la méthode V.B.A.P. est en fait double: son premier intérêt réside dans la généralisation du procédé de panoramique d'intensité, mais, dans un second temps, elle va permettre *d'étendre ce procédé de spatialisation aux trois dimensions de l'espace*. Le principe de la projection de la source virtuelle sur une base vectorielle obtenue à partir des haut-parleurs peut en effet être appliqué à trois haut-parleurs formant avec l'auditeur un trièdre. Il suffit de considérer, dans l'équation 2.4, trois haut-parleurs au lieu de deux:

$$\begin{aligned} \mathbf{h} &= [\mathbf{h_1 h_2 h_3}] \\ \mathbf{g}^T &= [g_1 g_2 g_3] \end{aligned}$$

La source virtuelle peut alors être positionnée en n'importe quel point à l'intérieur du domaine délimité entre les trois haut-parleurs (cf. Fig. 2.21). Par extension, il est possible de synthétiser des sources virtuelles

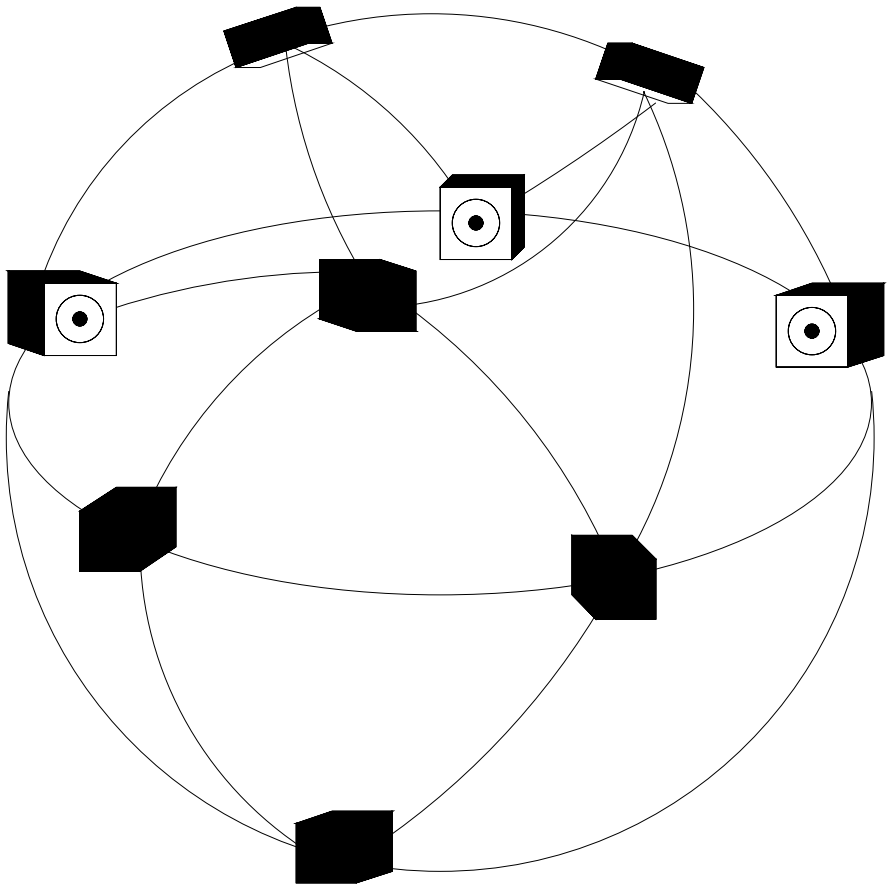


Fig. 2.22 - Panoramique d'intensité 3D avec la méthode V.B.A.P.: L'auditeur est entouré par une sphère de haut-parleurs.

dans n'importe quelle direction, en entourant l'auditeur par une sphère de haut-parleurs (cf. Fig. 2.22). Les haut-parleurs adjacents sont regroupés en triplets qui définissent autant de bases vectorielles sur lesquelles se projettent les sources virtuelles. Pour une position donnée de source virtuelle, un seul triplet de haut-parleurs est activé: il s'agit du triplet défini par les haut-parleurs les plus proches de la position de la source considérée. Les gains des trois haut-parleurs sélectionnés sont calculés à partir de l'équation 2.4.

Avec la méthode V.B.A.P., il est donc possible d'étendre les performances du panoramique d'intensité à une restitution spatialisée 3D par une sphère de haut-parleurs. Cependant, dans son principe, cette approche ne permet pas d'étendre la zone d'écoute: la restitution n'est correcte en théorie que pour un auditeur situé au centre de la sphère. On peut néanmoins penser que le principe d'un *panoramique d'intensité local*, dans la mesure où la source virtuelle est systématiquement synthétisée par les haut-parleurs les plus proches de sa position, offre un rendu plus stable de la spatialisation, notamment lorsque l'auditeur s'écarte de la position d'écoute idéale définie par le centre de la sphère. Dans le même esprit, pour les systèmes de diffusion stéréophoniques pour le cinéma ou la télévision, un haut-parleur central a été ajouté à la paire stéréophonique, afin de stabiliser la localisation des sources virtuelles pour les spectateurs excentrés. On retrouve l'idée sous-jacente d'accroître la taille de la zone d'écoute en augmentant le nombre de sources. En termes de spatialisation sonore, un panoramique d'intensité local possède cependant un inconvénient: lorsque la source virtuelle se déplace, le rendu sonore présente des hétérogénéités plus ou moins marquées [Pernaux *et al.*, 1998]. Par opposition, les approches qui cherchent à étendre le nombre de haut-parleurs impliqués

dans la synthèse d'une source virtuelle, tendent à offrir un rendu plus homogène.

Un autre problème posé par le panoramique d'intensité réside dans la prise de son: alors qu'en stéréophonie conventionnelle, système de de prise de son et système de restitution sont liés de façon totalement indissociable, le panoramique d'intensité ne gère que la restitution. La méthode de prise de son la plus adaptée est la *multimicrophonie*, qui correspond une prise de son par des microphones individuels de proximité, mais la position des sources à synthétiser doit être identifiée par une procédure séparée, puisque les signaux enregistrés ne contiennent aucun codage acoustique de la position des sources sonores. En stéréophonie conventionnelle, ce codage est représenté par les différences Δ_T ou Δ_I.

2.4.7 Conclusion

En dépit de ses performances limitées, la stéréophonie reste aujourd'hui le système de restitution sonore spatialisée le plus utilisé, notamment pour les applications destinées au grand public, ceci pour la simple raison qu'en termes de matériel et de mise en œuvre, il offre un effet de spatialisation à moindre coût. Une prise de son stéréophonique ne requiert en effet qu'un couple de microphones et, pour la restitution, seuld deux haut-parleurs sont nécessaires.

Cependant, en stéréophonie, la spatialisation sonore demeure relativement "embryonnaire", puisque, d'une part, le modèle de spatialisation du champ sonore est grossier et relativement instable et que, d'autre part, une seule dimension de l'espace est spatialisée, la spatialisation dans cette dimension n'étant de surcroît que partielle. L'approche de panoramique d'intensité généralisé aux trois dimensions de l'espace, V.B.A.P. (cf. Section 2.4.6), propose une solution séduisante pour étendre la spatialisation à tout l'espace entourant l'auditeur, mais elle implique d'accroître considérablement le nombre de haut-parleurs pour la restitution.

Le second problème posé par la stéréophonie est sa zone d'écoute limitée. Deux solutions pour l'étendre, dont une développée spécifiquement pour un système de visioconférence, ont été présentées: elles passent toutes par une augmentation du nombre de haut-parleurs. La seconde solution mise au point à l'I.N.A. (cf. Section 2.4.5) ne nécessite qu'un nombre minime de sources supplémentaires, mais, étant donné qu'elle est destinée à des applications de sonorisation de spectacles, on peut craindre que l'effet de spatialisation ne soit pas contrôlé de manière suffisamment fine, du moins du point de vue du contexte de visioconférence. De la première solution (cf. Section 2.4.4), on retiendra qu'il faut compter une paire de haut-parleurs par auditeur supplémentaire.

En dépit de ces tentatives d'extension de la zone d'écoute, les performances de la stéréophonie en termes de spatialisation sonore s'avèrent être à la fois imprécises et limitées. Nous allons donc nous intéresser à une approche plus générale, basée sur un modèle de spatialisation sonore plus précis et plus complet: les techniques binaurales.

2.5 Techniques binaurales

2.5.1 Principe

Dans la stéréophonie, l'effet de spatialisation est basé sur la reproduction des différences d'intensité et de temps du champ sonore perçu par les deux oreilles de l'auditeur. Les techniques binaurales[14] procèdent d'une idée similaire, mais elles reposent sur une modélisation à la fois plus globale et plus rigoureuse. Fondamentalement, elles se proposent de *reproduire le champ sonore induit au niveau des oreilles de l'auditeur*. Ainsi, dans le champ restitué, sont présents non seulement les effets de la propagation entre la source et l'auditeur (atténuation, retard, effet de salle...), mais aussi l'ensemble des phénomènes engendrés par le corps de l'auditeur, tels que la diffraction par la tête, les réflexions sur le haut du corps et le pavillon de l'oreille externe. Les différences de temps et d'intensité restituées par la stéréophonie ne rendent compte de ces phénomènes que de façon incomplète et très approximative. De plus, comme les différences interaurales ne contrôlent que la localisation horizontale, le champ sonore n'est spatialisé que dans le plan horizontal. Au contraire, en visant la reproduction fidèle du champ excitant le tympan, les techniques binaurales restituent l'ensemble des indices qu'utilise l'appareil auditif pour interpréter le champ sonore, qu'il s'agisse des différences interaurales — pour la localisation horizontale — ou des indices spectraux monoraux — pour la localisation verticale —.

14. Pour tout complément d'information sur les aspects théoriques et pratiques des techniques binaurales, l'article de H. Møller [Møller, 1992] constitue une excellente synthèse du sujet.

Le champ sonore est donc spatialisé dans les *trois dimensions* de l'espace, c'est-à-dire que les sources sonores virtuelles peuvent être synthétisées dans tout l'espace entourant l'auditeur.

2.5.2 Prise de son

La prise de son est effectuée au moyen de deux capteurs placés à l'intérieur du canal auditif de chaque oreille. Idéalement, l'enregistrement devrait être réalisé dans les propres oreilles de l'auditeur afin de prendre en compte ses particularités morphologiques. Pour des raisons pratiques évidentes, on a souvent recours à une *tête artificielle* constituée d'un mannequin qui reproduit la moitié supérieure d'un corps humain. Il faut cependant être conscient qu'une tête artificielle ne rend compte que d'une perception moyenne qui, si elle est valable en bonne approximation pour un grand nombre d'individus, n'est exacte pour aucun... On a d'ailleurs observé que les enregistrements sur tête artificielle sont susceptibles d'introduire dans le champ perçu certaines aberrations, telles que des *inversions avant/arrière*, c'est-à-dire qu'une source censée être située devant l'auditeur est perçue derrière lui, ou des phénomènes de *localisation intracrânienne*, c'est-à-dire que les sons sont perçus comme si les sources étaient situées à l'intérieur de la tête [Møller, 1992]. Ce problème constitue un des points délicats des techniques binaurales.

2.5.3 Restitution

Restitution sur casque

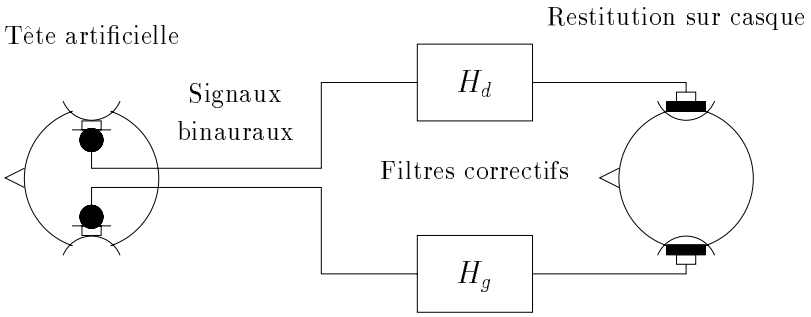
Pour garantir un système de reproduction totalement transparent, le champ enregistré en un point du canal auditif devrait être restitué au même point. Une restitution sur *casque* représente donc la solution la mieux adaptée. A l'origine, les enregistrements binauraux étaient exclusivement destinés à une diffusion sur écouteurs. Le système binaural conventionnel est illustré sur la figure 2.23a. Avant d'être restitués par le casque, les deux signaux binauraux sont filtrés afin de compenser la réponse des transducteurs électroacoustiques, à savoir les microphones de prise de son et le casque de restitution. Le filtre corrige également les éventuels écarts de position entre le point d'enregistrement — en général à l'intérieur du canal auditif — et le point de restitution — à l'entrée du canal auditif —. Toutes ces corrections sont déduites de mesures de *fonctions de transfert* électroacoustiques.

Restitution sur haut-parleur: Système *transaural*

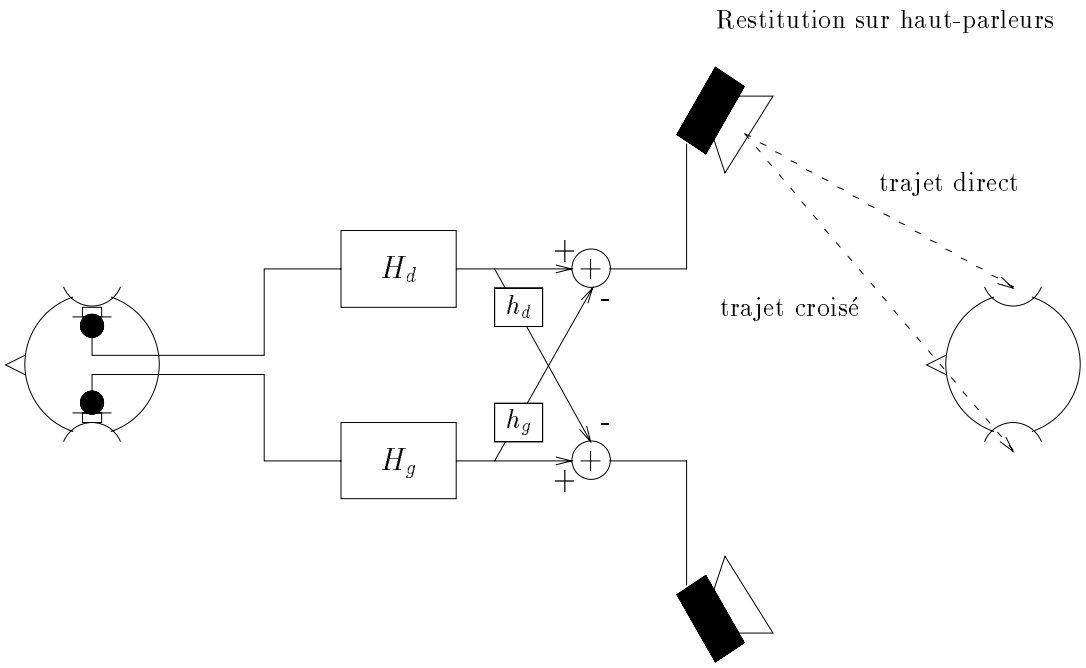
Il est néanmoins possible de restituer un enregistrement binaural sur deux *haut-parleurs*. On préfère alors parler de *système transaural* [Møller, 1992]. Dans le cas d'une diffusion sur haut-parleurs, il faut remarquer que chaque oreille perçoit à la fois le son émis par l'enceinte ipsilatérale — trajet *direct* — et par l'enceinte contralatérale — trajet *croisé* —. Pour obtenir une qualité de restitution équivalente à une diffusion sur casque, il faut donc éliminer les trajets croisés: dans ce but, la contribution "parasite" de l'enceinte contralatérale est soustraite au préalable au signal alimentant chacun des haut-parleurs. Outre l'égalisation de la réponse des transducteurs, la diffusion sur haut-parleurs nécessite donc de compenser la propagation entre les enceintes et les oreilles de l'auditeur, à la fois les trajets directs — reliant chaque oreille à l'enceinte ipsilatérale — et les trajets croisés — reliant chaque oreille à l'enceinte contralatérale —, de façon à restituer au niveau de chaque tympan, indépendamment du système de restitution, le signal de pression correspondant au signal enregistré dans le canal auditif. Cette compensation est réalisée en traitant les signaux binauraux par des filtres dont les réponses inversent les différentes fonctions de transfert électroacosutiques reliant chaque enceinte à chaque oreille. La figure 2.23b décrit un dispositif transaural avec son circuit de correction.

2.5.4 Performances

Les techniques binaurales constituent un outil de reproduction sonore 3D très performant, dès lors que le procédé est parfaitement maîtrisé, mais elles souffrent de problèmes liés à leur mise en œuvre délicate. La qualité du rendu est en effet très sensible à un certain nombre de paramètres pratiques, tels que la précision du positionnement ou l'égalisation des transducteurs [Møller, 1992]. De plus, la complexité des filtres binauraux requiert une forte puissance de calcul.



(a) Dispositif de prise et restitution binaurales conventionnelles: Les modules H_g et H_d représentent les filtres de correction destinés à l'égalisation des transducteurs et à la compensation de la propagation dans le canal auditif.



(b) Dispositif de prise et restitution transaurales: Outre les filtres de correction H_g et H_d, le circuit de traitement comprend des filtres destinés à compenser les trajets croisés entre les enceintes et l'auditeur.

FIG. 2.23 - Techniques binaurales: Système binaural conventionnel et Système transaural (d'après [Møller, 1992])

Par ailleurs, dans leur essence, les techniques binaurales sont fortement *individualistes* et, par la même, ne se prêtent pas à une diffusion pour une large audience. Deux raisons fondamentales l'expliquent:

- Premièrement, la qualité de la reproduction — et notamment l'effet de spatialisation — est d'autant plus précise et fidèle que les signaux binauraux prennent en compte, aussi bien à la prise de son qu'à la diffusion, les *spécificités individuelles* de l'auditeur. Dans le dispositif idéal, la prise de son est effectuée avec des microphones placés dans les propres oreilles de l'auditeur et les filtres de correction sont calculés à partir de mesures réalisées également dans ses oreilles, mais en ce cas la reproduction n'est correcte que pour cet individu. Dès que l'on s'écarte de ces conditions — par exemple en utilisant une tête artificielle —, la qualité de la restitution se dégrade. On doit donc réaliser un compromis entre la qualité de la reproduction et le degré de généralité des signaux binauraux. Or, si, dans le souci de rendre la restitution accessible à un plus grand nombre d'auditeurs, on réduit la part des codages spécifiques à l'individu, la reproduction sonore 3D perd tout son relief pour se ramener aux performances d'un système stéréophonique conventionnel.

- Deuxièmement, comme une restitution sur casque est exclue en contexte de visisoconférence de groupe, la diffusion se peut se faire que sur des haut-parleurs. En ce cas, les filtres de correction, qui sont destinés à compenser la propagation entre les enceintes électroacoustiques et l'auditeur, sont calculés pour une position donnée de l'auditeur et les corrections appliquées ne sont donc valables que pour cette position d'écoute. Par conséquent, un système transaural n'admet qu'*un point d'écoute correcte* qui est assorti en outre de contraintes très strictes sur le positionnement de l'auditeur et plus précisément de ses oreilles[15]. Théoriquement l'auditeur ne peut pas tourner la tête.

Par suite, *une restitution binaurale n'est valable en théorie que pour un individu spécifique et une position d'écoute unique*. Cependant, J. Bauck et D.H. Cooper ont récemment proposé de généraliser le système transaural à plusieurs auditeurs [Bauck & Cooper, 1996]. Leur idée est présentée dans le paragraphe qui suit.

2.5.5 Système transaural généralisé

Dans son principe, le système transaural est basé sur un ensemble de traitements destinés à compenser la propagation entre les enceintes électroacoustiques et les oreilles de l'auditeur. Il s'agit d'étendre ce système à un nombre quelconque de haut-parleurs et d'auditeurs. Mathématiquement, le problème se pose de la façon suivante. Soient M haut-parleurs et N auditeurs. Désignons par[16]:

$$\mathbf{s}^T = [s_1 s_2] \tag{2.6}$$

$$\mathbf{t}^T = [t_1 t_2 ... t_M] \tag{2.7}$$

$$\mathbf{u}^T = [u_1 u_2 ... u_{2N}] \tag{2.8}$$

les vecteurs représentant respectivement les signaux binauraux enregistrés, les signaux alimentant les M haut-parleurs et les signaux de pression induits au niveau de chaque oreille des N auditeurs (2N signaux au total). La matrice X définit les fonctions de transfert des trajets acoustiques entre les haut-parleurs et les auditeurs:

$$\mathbf{u} = X\mathbf{t} \ , \tag{2.9}$$

tandis que la matrice Y traduit les traitements appliqués aux signaux avant leur diffusion (cf. Fig. 2.24):

$$\mathbf{t} = Y\mathbf{s} \ . \tag{2.10}$$

15. Dans son article sur le système transaural généralisé [Bauck & Cooper, 1996], J. Bauck prétend — sans en apporter la preuve — qu'au contraire, la restitution transaurale est plus robuste aux erreurs de positionnement de l'auditeur que la stéréophonie conventionnelle. Ce résultat nous paraît d'autant plus sujet à caution que la sensibilité des techniques binaurales aux défauts d'égalisation — réponse des transducteurs ou compensation de la propagation acoustique — est reconnue et constitue d'ailleurs une des limitations du procédé [Møller, 1992].

16. Dans tout le raisonnement, on se placera indifféremment dans le domaine temporel ou fréquentiel.

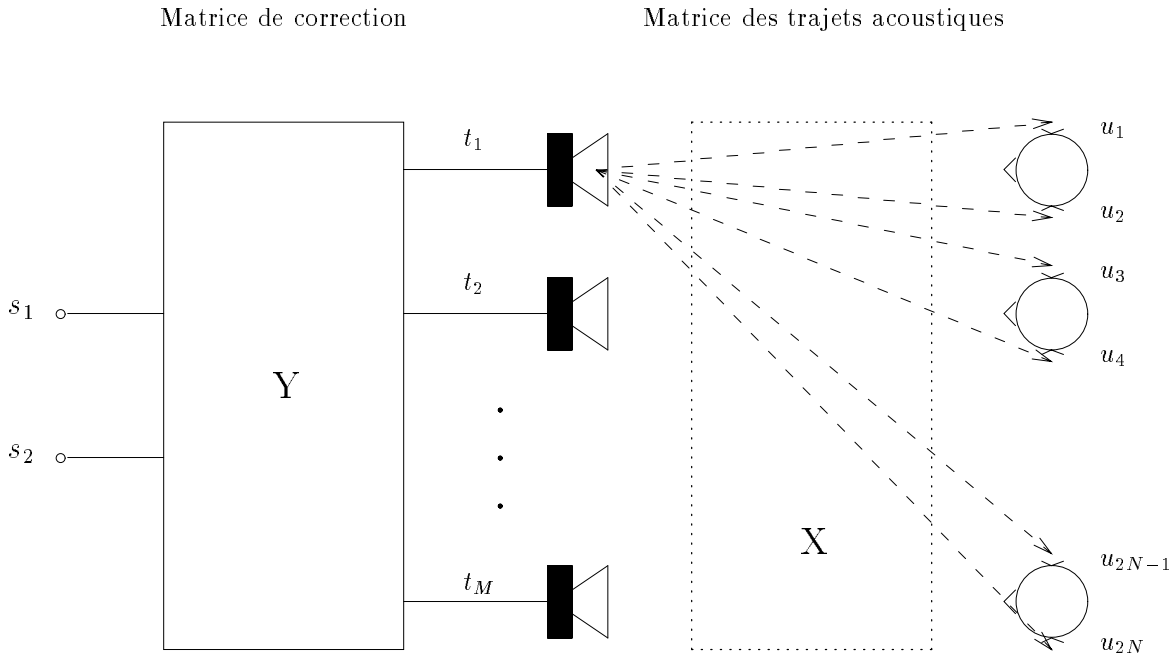


FIG. 2.24 - Principe du système transaural généralisé

En définitive, on veut reproduire les signaux binauraux sur les deux oreilles de chaque auditeur, les vecteurs **s** et **u** sont donc reliés par une matrice Z de la forme:

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{bmatrix} , \tag{2.11}$$

telle que:

$$\mathbf{u} = Z\ \mathbf{s} \tag{2.12}$$

Or, d'après les équations 2.9 et 2.10:

$$\mathbf{u} = X\ Y\ \mathbf{s} \tag{2.13}$$

Par suite, il vient:

$$XY = Z\ . \tag{2.14}$$

Ainsi, le problème consiste à déterminer la matrice Y connaissant les matrices X et Z. On est amené à *inverser la matrice* X, à condition bien entendu qu'elle soit inversible. J. Bauck et D.H. Cooper suggèrent d'exprimer Y à partir de la matrice *pseudo-inverse*[17] de X, dénotée X^+:

$$Y = X^+ Z\ . \tag{2.15}$$

17. Pour rappel, la matrice pseudo-inverse d'une matrice X est définie par: .(inverse au sens de Moore-Penrose [Bauck & Cooper, 1996])

Cette solution est à *norme minimale*, c'est-à-dire que, d'un point de vue électroacoustique, elle minimise la puissance délivrée à chaque haut-parleur [Bauck & Cooper, 1996]. De la même façon, les excursions des éventuels pics d'amplitude des signaux sont limitées. Par ailleurs, les auteurs montrent comment il est possible de simplifier le travail d'implémentation des filtres en procédant à certaines manipulations algébriques.

La principale limitation de la méthode résulte du problème d'inversibilité des fonctions de transfert acoustiques, inversibilité qui devient très délicate en présence d'un effet de salle. Très séduisante sur le plan théorique, la méthode transaurale généralisée pose de sérieuses difficultés dans sa mise en œuvre et il est dommage que cette question ne soit pas abordée par les auteurs. On note que, dans son principe, cette approche permet de prendre en compte les fonctions de transfert (H.R.T.F) propres à chaque auditeur.

Un exemple particulier d'application mérite cependant d'être signalé. Il s'agit d'un système de restitution transaurale étendue constitué d'un monopôle placé devant la zone d'écoute et associé à plusieurs dipôles qui sont disposés derrière chaque auditeur (cf. Fig. 2.25). Le monopôle est alimenté par le signal Σ qui représente la somme des deux signaux binauraux (signaux G et D, respectivement pour les oreilles gauche et droite):

$$\Sigma = G + D \ ,$$

tandis que chaque dipôle reçoit le signal Δ qui correspond à leur différence:

$$\Delta = G - D \ .$$

Le rayonnement dipolaire est utilisé de telle sorte qu'au niveau de l'oreille gauche se superposent le signal Σ émis par le monopôle et le signal Δ émis par le lobe gauche du dipôle:

$$\Sigma + \Delta = 2G \ ,$$

Le signal résultant correspond bien au signal binaural gauche, moyennant un facteur d'amplitude 2. En revanche, au niveau de l'oreille droite, c'est le signal en opposition de phase $-\Delta$ émis par le lobe droit du monopôle, qui vient s'ajouter au signal Σ:

$$\Sigma - \Delta = 2D \ ,$$

de façon à reconstituer le signal binaural droit D. Chaque auditeur étant plongé dans le champ proche immédiat de son dipôle, il ne risque pas d'être gêné par les dipôles des autres auditeurs, d'autant que les sources dipolaires sont caractérisées par un rayonnement de faible puissance.

2.5.6 Conclusion

Par rapport à la stéréophonie, les techniques binaurales constituent une avancée majeure, en ce sens qu'elles permettent véritablement de reproduire un champ sonore 3D, dans lequel l'intégralité des indices de localisation auditive sont restitués.

Néanmoins, l'approche binaurale s'avère assez inadaptée au contexte de visioconférence de groupe, dans la mesure où elle est fondamentalement *individualiste*:

- d'abord parce qu'elle fait intervenir les *codages acoustiques spatiaux spécifiques à l'individu*, ce qui pose de sérieuses difficultés pour une application destinée à une large audience, car cela nécessiterait d'enregistrer au préalable les fonctions de transfert de chaque individu,

- ensuite parce que *la zone de restitution correcte, stricto sensu, se limite à deux points* correspondant aux deux oreilles de l'auditeur en excluant toute possibilité de se déplacer.

Il existe des systèmes — systèmes de "Headtracker" — qui permettent de suivre les mouvements de la tête de l'auditeur pour les compenser dans le traitement des signaux restitués. De tels systèmes pourraient être utilisés pour offrir une plus grande latitude de mouvements à l'auditeur, mais ils permettraient de corriger seulement les rotations de la tête et, en tout état de cause, ils ne fonctionneraient efficacement que dans la limite de petits mouvements. En outre, les systèmes de Headtracker imposeraient d'équiper l'auditeur d'un appareillage contraignant, ce qui est inacceptable du point de vue d'un mur de téléprésence (cf. Chapitre 1).

Une solution pour étendre la zone d'écoute consiste à généraliser le principe d'une restitution transaurale à plusieurs auditeurs, mais elle implique d'inverser des réponses impulsionnelles de salles, ce qui s'avère assez

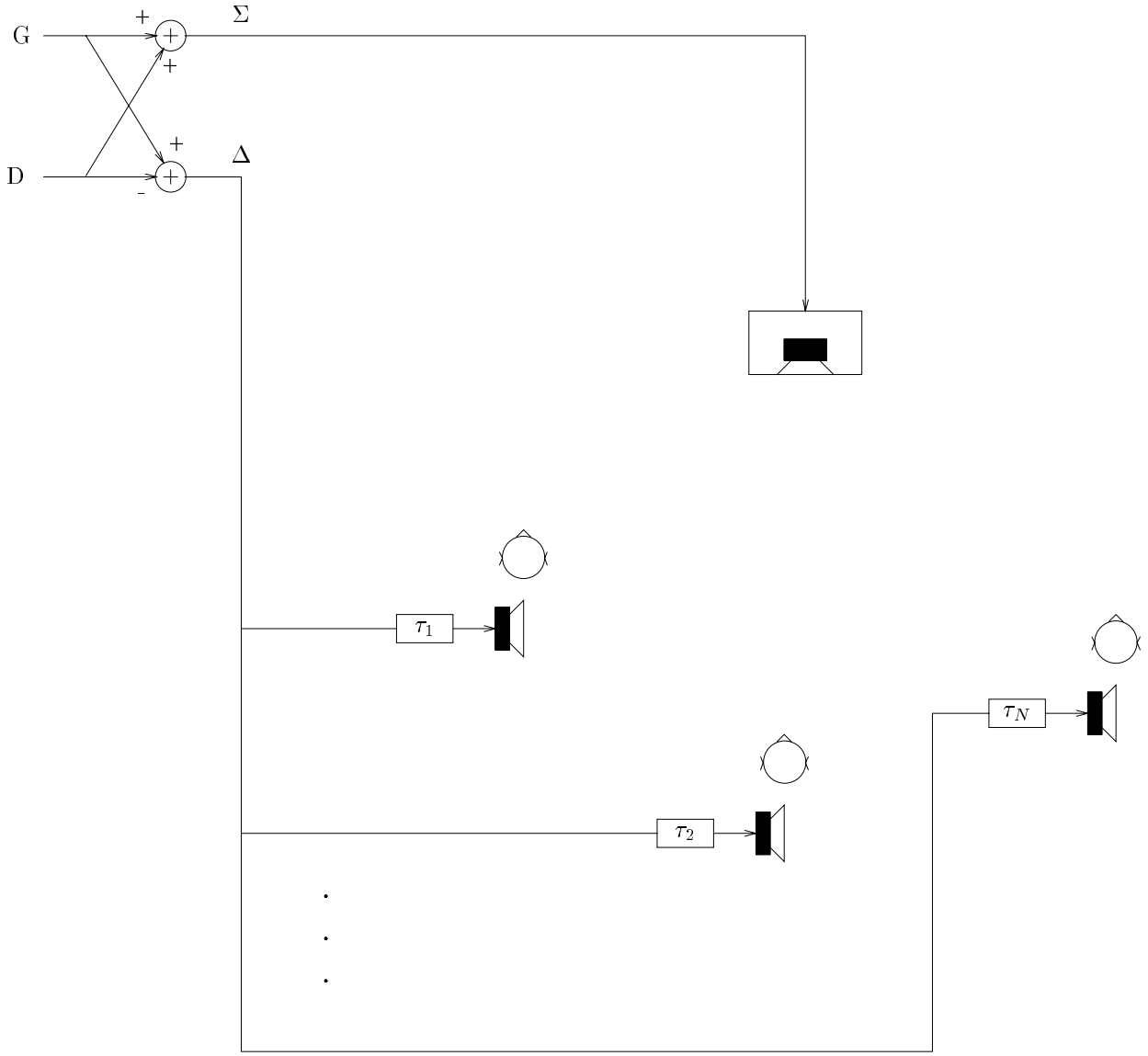


Fig. 2.25 - Exemple de système de restitution transaurale généralisée: Combinaison d'un monopôle et de plusieurs dipôles disposés derrière chaque auditeur

problématique en pratique. De plus, dans cette approche, les auditeurs ne peuvent toujours pas se déplacer au sein de la zone d'écoute. On note aussi qu'avec le système transaural, il faut toujours compter deux haut-parleurs par auditeur.

Il reste qu'actuellement, de nombreuses études cherchent à lever les différentes limitations des techniques binaurales. Les travaux récemment publiés s'intéressent notamment aux points suivants:

- la mise au point de jeux de HRTF universelles [Dudouet & Martin, 1999],
- la prise en compte des mouvements de la tête [Wenzel, 1999],
- l'inversion de l'effet de salle [Lopez *et al.*, 1999],
- la réduction de la complexité des filtres binauraux [Larcher & Jot, 1997].

2.6 Système ambisonique

Le procédé ambisonique est basé sur une approche de spatialisation sonore qui a vu le jour dans le courant des années 70. Son inventeur, M.A. Gerzon, était, à l'origine, titulaire d'une thèse en Mathématiques, ce qui confère une orientation spécifique à son approche. Il faut néanmoins reconnaître qu'il ne s'est pas cantonné à une analyse théorique du problème, mais a su mettre en œuvre ses résultats dans un système concret.

En dépit de son intérêt, le système ambisonique est longtemps resté une *méthode marginale* peu connue des professionnels des techniques audiovisuelles, à l'exception des studios de la B.B.C.. Depuis quelques années, elle bénéficie d'un renouveau d'intérêt, en particulier de la part du monde scientifique, en raison de la montée en puissance des domaines de la spatialisation sonore et de la restitution sonore multicanale, notamment pour le cinéma et la télévision.

Parmi l'ensemble des techniques de reproduction sonore spatialisée, la technique ambisonique constitue une approche marginale non seulement par son utilisation, mais aussi par la personnalité de son inventeur qui semble avoir cherché à entourer sa méthode d'un "flou scientifique" qui la rend difficile, voire décourageante, à appréhender. Récemment, plusieurs articles ont contribué à la démythifier [Bamford, 1995] [Daniel *et al.*, 1998] [Nicol & Emerit, 1999a][18] , néanmoins il faut garder présent à l'esprit que, moins qu'une méthode parfaitement définie, le système ambisonique représente plutôt un *concept*, voire un *ensemble de règles relatives à la mise en œuvre d'un système de restitution sonore spatialisée*[19].

Fondamentalement, le procédé ambisonique est basé sur une *approche hybride* qui combine le principe de *reconstruction physique* du champ sonore et la prise en compte des mécanismes *psychoacoustiques*. De façon simplifiée, il consiste à reproduire le champ acoustique en un point correspondant au centre de la tête de l'auditeur, mais la reconstruction proprement physique n'opère qu'aux basses fréquences. Pour les hautes fréquences, la procédure de reconstruction obéit à une approche énergétique des phénomènes. Cette idée est intéressante, car on comprend intuitivement qu'une reconstruction physique aux hautes fréquences s'avère rapidement coûteuse. De plus, cette reconstruction n'est pas forcément nécessaire à cause des limitations de la perception auditive. Par ailleurs, on peut penser que, dès lors qu'une partie du spectre est correctement restituée, la perception globale est préservée, d'autant qu'il s'agit des basses fréquences dont le rôle déterminant pour la localisation auditive a été établi. Nous allons à présent détailler le principe de la prise et la restitution du son selon le procédé ambisonique.

2.6.1 Prise de son

Une extension du système Stereosonic

Une prise de son ambisonique consiste à *enregistrer en un point à la fois la pression acoustique (W) et les trois composantes (X,Y,Z) du vecteur de vitesse particulaire selon les trois directions de l'espace* [Farrar, 1979a]. Un *microphone de pression* (microphone omnidirectif) permet d'enregistrer le signal de pression,

18. De manière plus informelle, mais tout aussi déterminante, les nombreuses discussions que j'ai eu à l'Ircam avec Jean-Marc Jot à propos du système ambisonique m'ont été d'un grand secours pour en comprendre l'esprit et les mécanismes. Elles sont largement responsables de mon intérêt pour cette méthode et un certain nombre des idées qui vont être présentées dans ce document sont issues de ces entretiens.

19. M.A. Gerzon a d'ailleurs su entretenir l'ambiguïté dans ses écrits.

tandis que les composantes du vecteur de vitesse particulaire peuvent être enregistrées par trois *microphones à gradient de pression* (microphones bidirectifs), chaque microphone pointant dans une des directions de l'espace (cf. Fig. 2.26). Ainsi une prise de son ambisonique rappelle le principe du couple Stereosonic mis au point dans les années 30 par A.D. Blumlein et qui est constitué de deux microphones bidirectifs orientés perpendiculairement (cf. Fig. 2.9b) [Blumlein, 1934]. Dans le cas d'une onde plane, d'amplitude a et de vecteur d'onde[20] $\vec{k}$, décrite par:

$$p(\vec{r}) = a\ e^{j\vec{k}.\vec{r}} \tag{2.16}$$

où le vecteur d'onde, dont la direction est repérée par les angles (φ_0, θ_0) (cf. Fig. 2.27), est donné par:

$$\vec{k} = k \left| \begin{array}{l} \sin\theta_0\ \cos\varphi_0 \\ \sin\theta_0\ \sin\varphi_0 \\ \cos\theta_0 \end{array} \right. \tag{2.17}$$

les signaux (W,X,Y,Z) s'expriment:

$$\left\{ \begin{array}{lcl} W & = & a \\ X & = & a\ \sqrt{2}\ \sin\theta_0\ \cos\varphi_0 \\ Y & = & a\ \sqrt{2}\ \sin\theta_0\ \sin\varphi_0 \\ Z & = & a\ \sqrt{2}\ \cos\theta_0 \end{array} \right. \tag{2.18}$$

Rôles des composantes X, Y et Z: Codage de l'information spatiale

L'opération par laquelle le champ sonore est enregistré sous la forme des quatre signaux (W,X,Y,Z) représente son *encodage*, c'est-à-dire l'opération par laquelle les informations spatiales du champ acoustique sont transcrites en signaux électriques (W,X,Y,Z). En d'autres termes, ces signaux constituent la représentation codée de l'information contenue dans le champ sonore et cette information concerne non seulement ses propriétés temporelles, mais aussi ses propriétés spatiales qui sont exclusivement contenues dans les trois signaux (X,Y,Z). En effet, le signal W est, par essence, omnidirectif, alors que les signaux (X,Y,Z) sont associés à des microphones bidirectifs qui explorent les trois directions de l'espace et expriment donc la répartition spatiale de l'énergie [Farrar, 1979a] [Gerzon, 1985].

Nous allons essayer de préciser ces idées en reprenant l'exemple d'une onde plane (cf. Equ. 2.16). Exprimons la pression acoustique et le vecteur de vitesse particulaire $\vec{v}$ induits par cette onde au point $\vec{r} = \vec{0}$:

$$\left\{ \begin{array}{lcl} p(\vec{0}) & = & a \\ \vec{v}(\vec{0}) & = & \frac{a}{\rho c}\ \vec{k} \end{array} \right. \tag{2.19}$$

où les termes ρ et c dénotent respectivement la masse volumique de l'air et la célérité des ondes acoustiques. Des expressions de la pression $p(\vec{0})$ et de la vitesse particulaire $\vec{v}(\vec{0})$, il ressort que le signal de pression n'exprime que l'amplitude de l'onde plane, tandis que la vitesse particulaire correspond, à un facteur multiplicatif près, au vecteur d'onde dont la direction définit l'incidence de l'onde plane et traduit donc son comportement spatial. Par suite, le vecteur de vitesse particulaire décrit bien les propriétés spatiales de l'onde plane considérée.

Plus généralement, dans le cas d'un champ acoustique quelconque, la vitesse particulaire n'est autre que le gradient de la pression, en vertu de l'équation d'Euler [Bruneau, 1983]. Elle exprime donc la dérivée spatiale de la pression acoustique et par la même contient des informations relatives à sa distribution spatiale. De plus, par analogie avec un *développement de Taylor*, on peut aussi considérer que les composantes (X,Y,Z), qui expriment les dérivées spatiales à l'ordre 1 de la pression acoustique, étendent la description du champ sonore au voisinage du point $\vec{r}$. En ajoutant ces trois composantes au signal de pression W, le champ n'est donc plus connu seulement en un point, mais aussi sur tout un volume qui représente le voisinage immédiat de ce

20. Le nombre d'onde est défini par:

$$k = \frac{\omega}{c}$$

où ω désigne la pulsation temporelle et c, la célérité des ondes acoustiques.

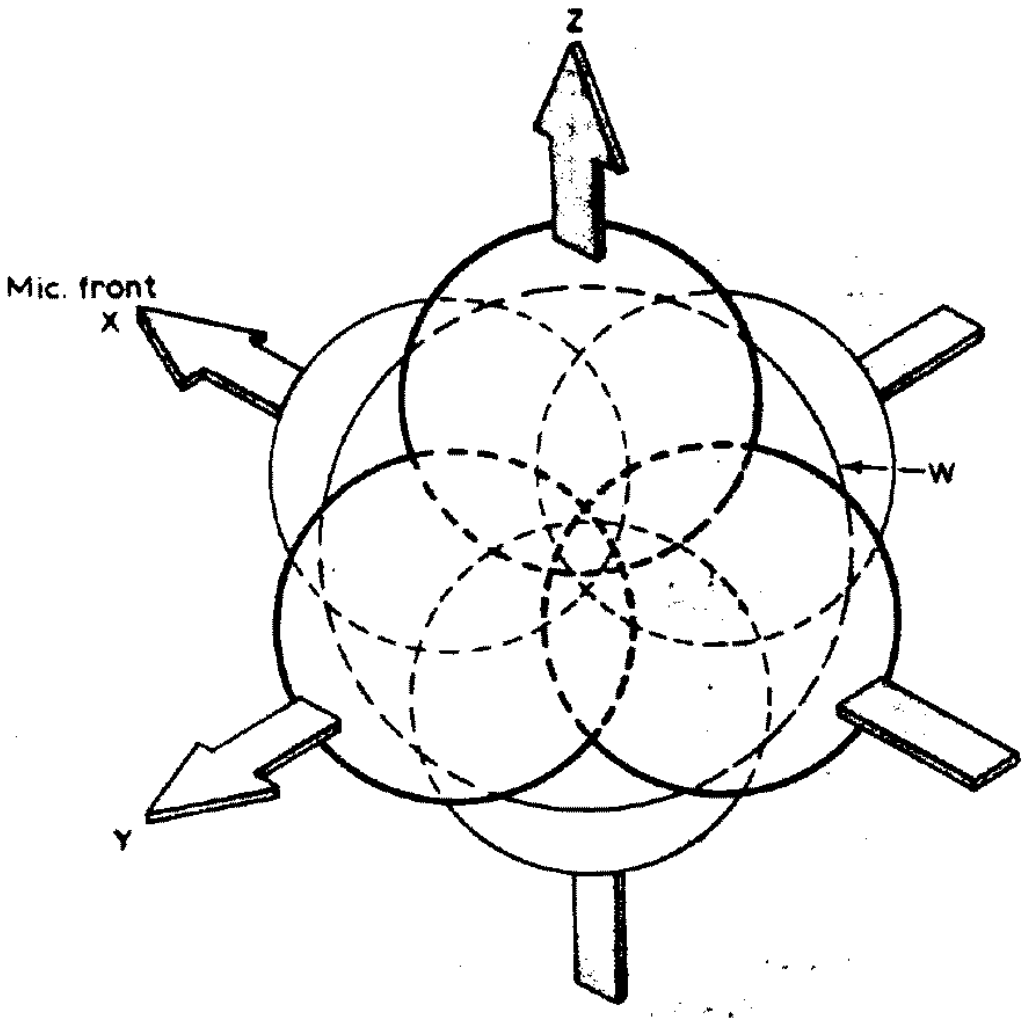


Fig. 2.26 - Prise de son ambisonique: Association d'un microphone omnidirectif (composante W) à trois microphones bidirectifs (composantes X,Y,Z)

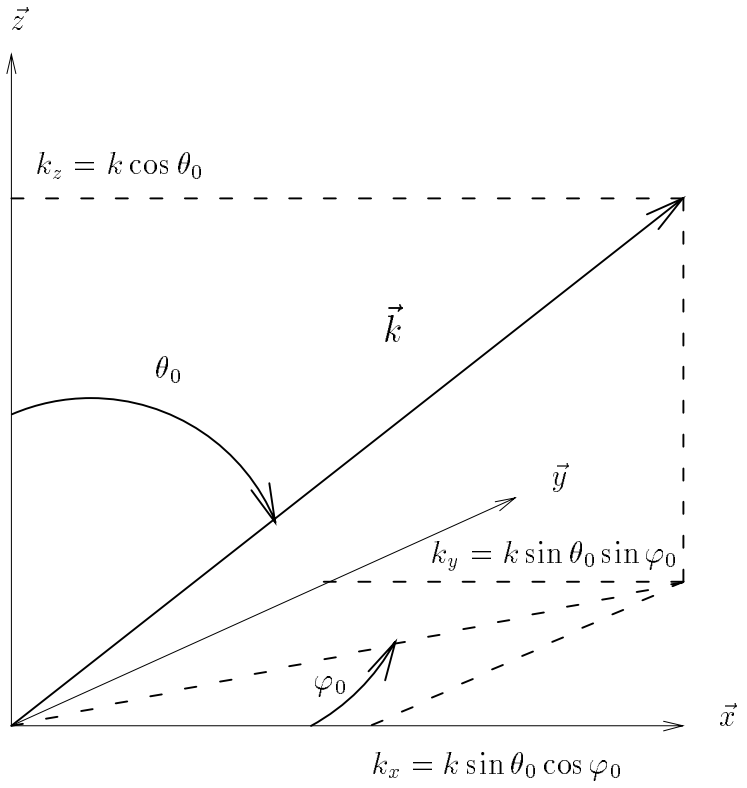
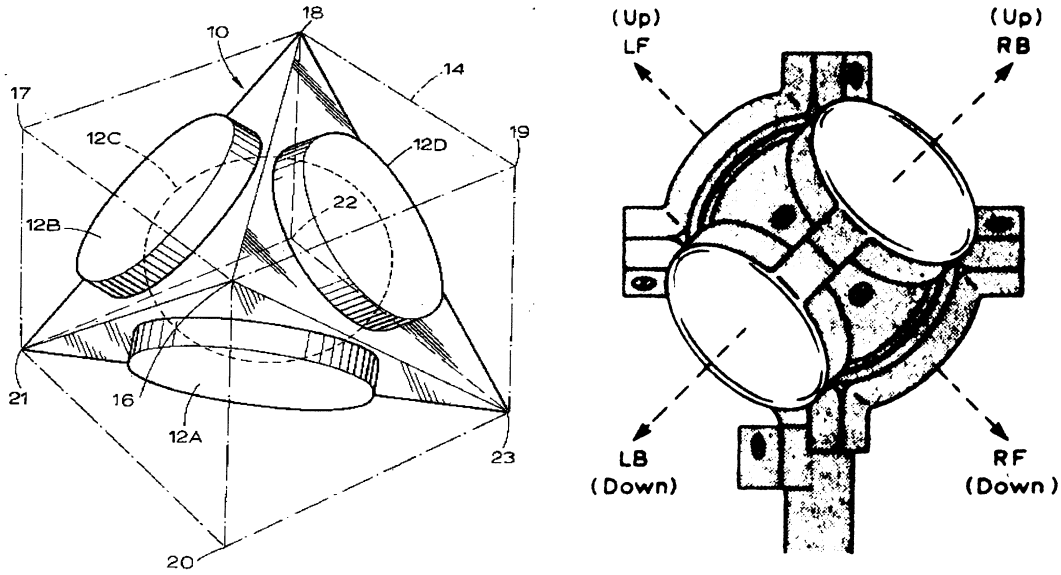


FIG. 2.27 - Coordonnées du vecteur d'onde associé à une onde plane (φ_0: angle d'azimut, θ_0: angle d'élévation)



(a) Tétraèdre régulier [Craven & Gerzon, 1977]

(b) Montage des quatre capsules microphoniques (LF, LB, RF, RB) [Farrar, 1979a]

Fig. 2.28 - Microphone Soundfield: Combinaison de quatre capsules cardioïdes montées sur les faces d'un tétraèdre régulier

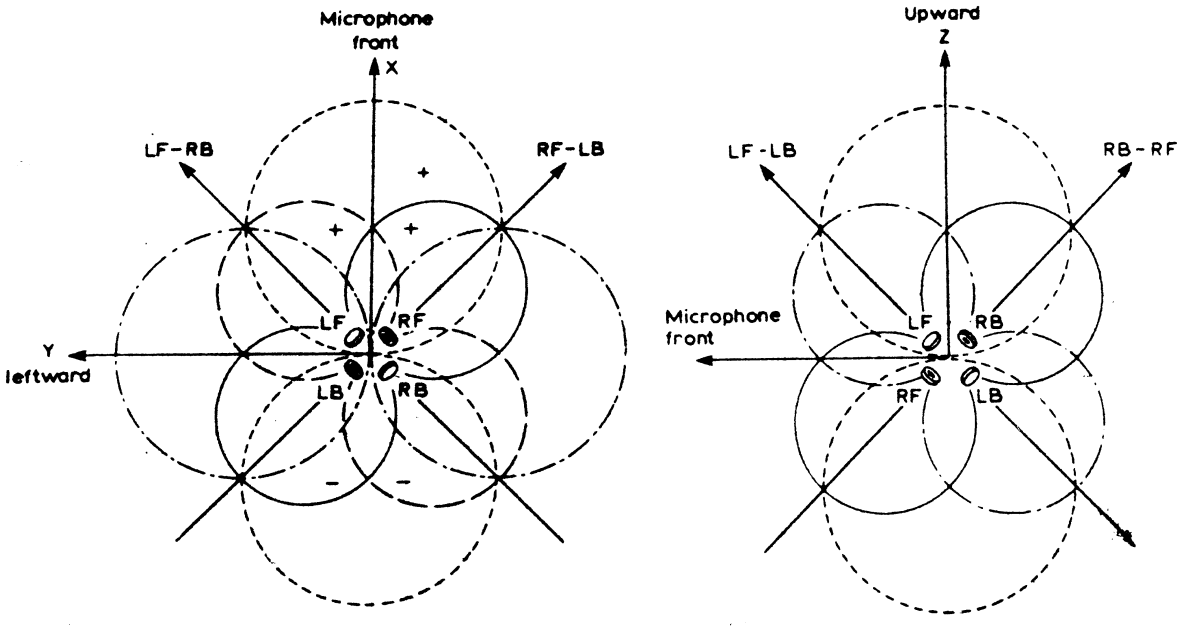
point et qu'on peut figurer par une sphère de rayon s qui est lié à la longueur d'onde. Aux basses fréquences, cette sphère englobe la tête de l'auditeur. Dans l'hypothèse d'une reconstruction physique à l'identique, les oreilles de l'auditeur sont donc bien plongées dans un champ acoustique identique à celui qu'il aurait perçu en présence des sources réelles et dont l'ensemble des propriétés temporelles et spatiales est restitué. En revanche, lorsque la fréquence augmente, le rayon de la sphère se réduit à quelques centimètres, voire quelques millimètres: par suite, les signaux (X,Y,Z) ne permettent plus une reproduction fidèle du champ sonore au niveau des oreilles de l'auditeur.

Microphone Soundfield

Une prise de son ambisonique requiert en théorie un microphone de pression et trois microphones à gradient de pression. Cependant, le microphone qui est dédié aux enregistrements ambisoniques et qui est connu sous le nom de microphone *Soundfield* [Craven & Gerzon, 1977] [Farrar, 1979a] [Farrar, 1979b] est constitué en réalité de quatre capsules cardioïdes distribuées sur les faces d'un tétraèdre régulier (cf. Fig. 2.28). Les quatre microphones sont ainsi positionnés sur la surface d'une sphère. Ce dispositif est une solution élégante pour pallier l'impossibilité matérielle de placer quatre microphones au même point. Il présente aussi l'avantage d'utiliser *quatre capsules microphoniques aux caractéristiques identiques*, ce qui simplifie le problème de l'appariement et de l'alignement entre les différents microphones.

Les capsules cardioïdes sont référencées en fonction de leur position et de leur orientation (cf. Fig. 2.29):

- LF (*Left Front*): capsule gauche orientée vers l'avant,
- RF (*Right Front*): capsule droite orientée vers l'avant,
- LB (*Left Back*): capsule gauche orientée vers l'arrière,
- RB (*Right Back*): capsule droite orientée vers l'arrière.



(a) Projection sur le plan horizontal $(\vec{x}, \vec{y})$

(b) Projection dans le plan vertical $(\vec{x}, \vec{z})$

FIG. 2.29 - Microphone Soundfield: Dérivation des signaux (W,X,Y,Z) à partir des signaux (LF, LB, RF, RB) enregistrés par les capsules cardioïdes [Farrar, 1979a]

Un examen attentif de leur géométrie met en évidence des relations simples entre leurs signaux de sortie (LF,LR,RF,RB) et les signaux d'encodage ambisonique (W,X,Y,Z) [Farrar, 1979a]:

$$\begin{cases} W &= LF + LB + RF + RB \\ X &= LF - LB + RF - RB \\ Y &= LF + LB - RF - RB \\ Z &= LF - LB - RF - RB \end{cases} \tag{2.20}$$

Différents formats d'encodage des signaux ambisoniques

Les signaux (LF,LR,RF,RB) et (W,X,Y,Z) constituent en fait deux formats d'encodage ambisonique et sont respectivement désignés comme le *format A* qui correspond aux signaux de sortie du microphone Soundfield et le *format B* qui définit les signaux "physiques" identifiés à la pression et à la vitesse particulaire [Gerzon, 1992a]. Dans le format B, l'information spatiale a été extraite pour la rendre directement accessible en vue de la restitution.

Il existe un autre format: le *format UHJ* qui est principalement utilisé pour la transmission des signaux. Il est basé sur quatre signaux (Σ,Δ,T,Q) qui sont reliés aux signaux (W,X,Y,Z) de la façon suivante:

$$\begin{cases} \Sigma &= 0.9397\ W + 0.1856\ X \\ \Delta &= j(-0.342\ W + 0.5099 X) + 0.655\ Y \\ T &= j(-0.1432\ W + 0.6512 X) - 0.7071\ Y \\ Q &= 0.9772 Z \end{cases} \tag{2.21}$$

Ce format est *compatible avec une diffusion stéréophonique*[21]. Les signaux stéréophoniques gauche et droite

21. C'est pourquoi il est utilisé en transmission où le système de diffusion n'est pas connu a priori.

(G,D) se déduisent en effet simplement des signaux (Σ,Δ):

$$\begin{cases} G &= \frac{\Sigma+\Delta}{2} \\ D &= \frac{\Sigma-\Delta}{2} \end{cases} \tag{2.22}$$

Pour cette raison, les signaux (Σ,Δ) sont respectivement référencés comme le signal "somme" et le signal "différence".

2.6.2 Restitution

Décodage de l'information spatiale

La restitution repose sur l'opération de *décodage* des signaux (W,X,Y,Z). Opération inverse de l'encodage, le décodage consiste à restituer sous forme acoustique les informations spatiales codées dans les signaux (W,X,Y,Z). La restituion se fait au moyen d'un dispositif de N haut-parleursalimentés par N signaux qui se déduisent des quatre signaux enregistrés par une procédure de matriçage:

$$\mathbf{g} = \mathbf{M}_d \ \mathbf{b} \tag{2.23}$$

où les vecteurs $\mathbf{b}$ et $\mathbf{g}$ décrivent respectivement les signaux d'enregistrement au format B et les signaux d'alimentation des haut-parleurs g_i (i=1,...,N):

$$\begin{aligned} \mathbf{b}^T &= [W \ X \ Y \ Z] \\ \mathbf{g}^T &= [g_1 \ g_2 \ g_3 \ \cdots \ g_N] \end{aligned}$$

La matrice $\mathbf{M}_d$ réalise le matriçage des signaux.

Ce matriçage constitue l'étape essentielle du décodage et il est conçu de façon à assurer une reproduction fidèle du champ sonore au niveau de l'auditeur. Le principal paramètre du décodage est la géométrie du dispositif de haut-parleurs: la matrice de décodage $\mathbf{M}_d$ ne dépend donc que des positions de haut-parleurs. Il convient de noter que la géométrie du dispositif de haut-parleurs est totalement libre: en effet, dès lors que la position des haut-parleurs est prise en compte dans la matrice de décodage et, par suite, peut être compensée, n'importe quelle géométrie est admissible. Il importe seulement d'appliquer la matrice de décodage associée à la géométrie du dispositif choisi. Il est toutefois recommandé de placer l'ensemble des haut-parleurs à égale distance de l'auditeur, encore que d'éventuelles différences de distance peuvent être corrigées par l'introduction de retards.

Une fois choisie une disposition de haut-parleurs, la matrice de décodage $\mathbf{M}_d$ est calculée en cherchant à optimiser la reproduction du champ sonore à la position de l'auditeur [Gerzon, 1992d] [Gerzon, 1992c]. La qualité de la reproduction est évaluée au moyen de deux critères définis sous les noms de *vecteur Vélocité* $\vec{r}_V$ et de *vecteur Energie* $\vec{r}_E$ [Gerzon, 1992b], qui s'attachent à décrire la localisation auditive d'un point de vue à la fois quantitatif (direction perçue) et qualitatif (stabilité de l'image sonore et confort d'écoute notamment) [Gerzon, 1974] [Gerzon, 1977].

Vecteur *Vélocité*

Le vecteur *Vélocité* est donné par [Gerzon, 1992b]:

$$\vec{r}_V = \sum_{i=1}^{N} \Re \left[\frac{g_i}{\sum_{i=1}^{N} g_i} \right] \ \vec{h}_{0_i} \quad (g_i \in \mathbb{C}) \tag{2.24}$$

Les vecteurs $\vec{h}_{0_i}$ repèrent les positions des haut-parleurs, plus exactement leur direction, puisqu'il s'agit de vecteurs *unitaires* obtenus en normalisant les vecteurs $\vec{h}_i$ qui indiquent les coordonnées des haut-parleurs dans un repère dont l'auditeur figure l'origine (cf. Fig. 2.30):

$$\vec{h}_{0_i} = \frac{\vec{h}_i}{|\vec{h}_i|} \quad \forall i = 1, 2, \cdots, N \tag{2.25}$$

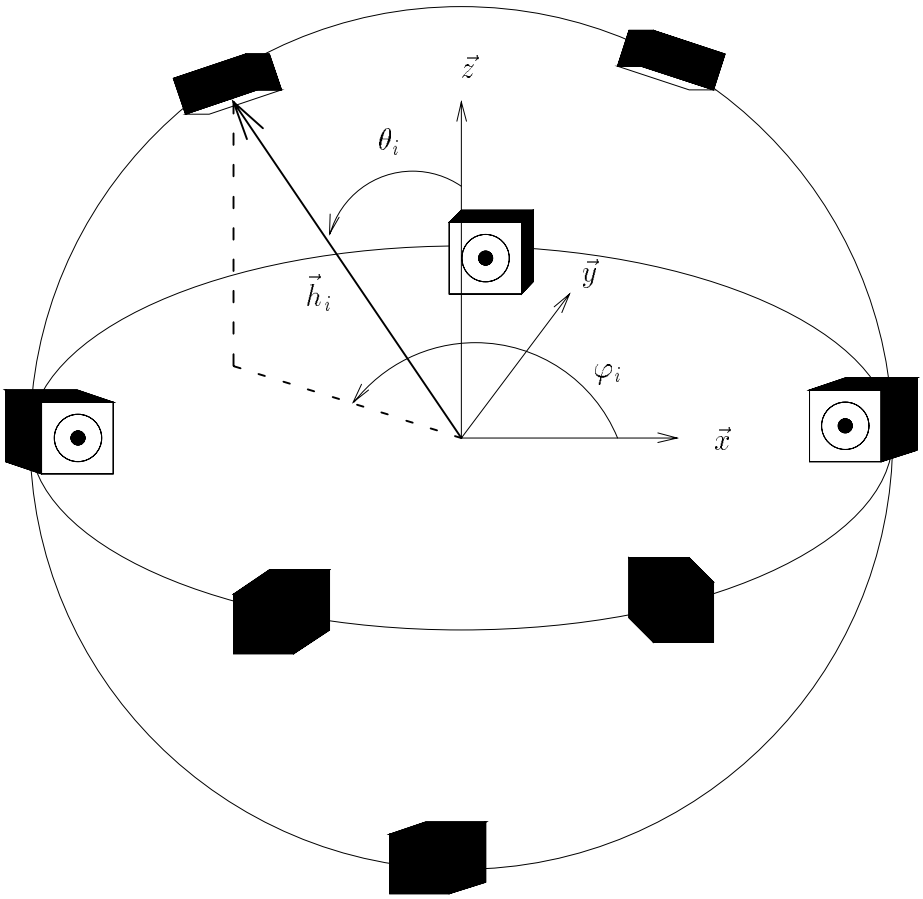


Fig. 2.30 - Système de restitution ambisonique: Distribution de haut-parleurs sur la surface d'une sphère centrée sur l'auditeur

Ce critère s'applique uniquement au domaine des *basses fréquences*, dont la limite est en général fixée à 700 Hz, cette valeur étant identifiée par les études psychoacoustiques sur la localisation auditive, comme une frontière entre deux modes distincts de perception [Gerzon, 1992b] [Blauert, 1983]. Dans le vecteur *Vélocité* $\vec{r}_V$, deux éléments sont à analyser:

- sa direction:

$$\vec{r}_{0_V} = \frac{\vec{r}_V}{|\vec{r}_V|}$$

- et sa norme (ou *phasiness*):

$$r_V = |\vec{r}_V|$$

Le vecteur $\vec{r}_{0_V}$ repère la direction dans laquelle sera localisée la source virtuelle synthétisée par l'ensemble des N haut-parleurs. Pour une restitution optimale du champ sonore, cette direction doit s'identifier à celle de la source réelle qu'on veut reproduire, ce qui définit la *première règle de décodage (B.F.: Basses Fréquences)*:

$$\vec{r}_{0_V} = \frac{\vec{r}_s}{|\vec{r}_s|} \tag{2.26}$$

Dans les écrits de M.A. Gerzon, l'origine de ce critère est rattachée à la théorie de la localisation de Makita [Makita, 1962]. Cette théorie concerne les phénomènes de localisation dans les systèmes stéréophoniques: en analysant les propriétés du champ sonore induit au niveau de l'auditeur, Y. Makita a cherché à exprimer la direction apparente de la source virtuelle, c'est-à-dire la direction dans laquelle l'auditeur va localiser la source virtuelle, en fonction des signaux alimentant les haut-parleurs. Dans son calcul, il identifie la direction apparente de la source sonore à la *normale au front d'onde évaluée à la position de l'auditeur*, c'est-à-dire la direction vers laquelle l'auditeur doit tourner la tête pour annuler les différences de phase perçues entre ses deux oreilles. Or, on peut montrer que le vecteur $\vec{r}_{0_V}$ correspond à la normale au front d'onde au point $\vec{r} = \vec{0}$, ce qui justifie la référence à la théorie de Makita. De manière plus intuitive, le vecteur $\vec{r}_{0_V}$ définit la "direction moyenne" de provenance des ondes sonores, dans la mesure où l'équation 2.24 peut s'interpréter comme la moyenne algébrique des incidences des ondes émises par l'ensemble des haut-parleurs, chaque incidence étant pondérée par le gain affecté au haut-parleur qui lui est associé. Il s'agit d'une méthode usuelle de construction vectorielle de source virtuelle synthétisée à partir d'un nombre donné de haut-parleurs (cf. Section 2.4.5: Fig. 2.17).

Alors que le critère de direction $\vec{r}_{0_V}$ est de type quantitatif, en ce sens qu'il exprime la direction apparente de la source virtuelle, le critère de norme r_V est purement qualitatif puisqu'il caractérise la qualité globale de l'image sonore restituée. Ce critère est notamment associé à la stabilité de l'image lorsque l'auditeur tourne la tête et plus généralement à son confort d'écoute, les défauts de restitution se traduisant souvent par une sensation de fatigue auditive [Gerzon, 1992b]. En fait, on montre que la norme r_V est liée à la vitesse apparente de propagation de l'onde synthétique, c'est-à-dire l'onde reproduite par les N haut-parleurs [Daniel *et al.*, 1998] [Makita, 1962]. Cette onde s'identifie en effet localement à une onde plane dont le nombre d'onde n'est plus k, mais k' qui se déduit de k moyennant un facteur multiplicatif qui n'est autre que r_V:

$$k' = r_V \; k \tag{2.27}$$

ce qui correspond à définir une *vitesse apparente de propagation de l'onde* c' qui vaut:

$$c' = \frac{c}{r_V} \tag{2.28}$$

La norme r_V exprime donc le rapport entre la vitesse de propagation d'une onde *naturelle* et la vitesse de propagation d'une onde *synthétique*. Or, cette altération de la vitesse de propagation de l'onde reproduite affecte les *différences interaurales de phase* perçues par l'auditeur et vient par suite fausser la localisation des sources virtuelles [Daniel *et al.*, 1998]. Ainsi une valeur $r_V < 1$ tend à ramener les sources virtuelles dans le plan médian de l'auditeur, tandis qu'une valeur $r_V > 1$ les en éloigne. Pour une restitution optimale, il convient donc d'imposer la valeur:

$$r_V = 1 \tag{2.29}$$

ce qui définit la *seconde règle de décodage (B.F.)*.

Les deux règles de décodages que constituent les équations 2.26 et 2.29 déterminent la stratégie du décodage aux basses fréquences. On montre que le respect de ses deux règles est équivalent à une reconstruction physique du champ sonore — qui inclut la reconstruction à la fois de la pression et de la vitesse particulaire — au point $\vec{r} = \vec{0}$. On rappelle que la condition de reconstruction exacte de la vitesse particulaire signifie qu'on étend implicitement la zone de reconstruction exacte de la pression au voisinage immédiat du point $\vec{r} = \vec{0}$.

L'ensemble des idées exposées dans ce qui précède ont été mises en évidence à partir de raisonnements basés sur des ondes planes. Un éclairage nouveau sur l'interprétation du vecteur *Vélocité* a été apporté par un article récent qui donne une définition générale de ce critère valable pour n'importe quel champ acoustique [Daniel *et al.*, 1999]. Le vecteur *Vélocité* est exprimé sous la forme d'une admittance acoustique spécifique:

$$\vec{r}_V = \Re\left[\rho c\ \frac{\vec{v}}{p}\right] \tag{2.30}$$

Dans le cas d'un champ constitué d'une superposition de N ondes planes, c'est-à-dire en première approximation le champ synthétisé par N haut-parleurs répartis sur une sphère, on vérifie que cette expression redonne bien la définition de M.A. Gerzon (cf. Equ. 2.24). A partir de cette définition générale, le vecteur *Vélocité* est interprété en termes d'intensité acoustique, et, plus précisément, on montre qu'il est directement relié au vecteur d'intensité active $\vec{I}$:

$$\vec{I}(\vec{r}) = \frac{1}{2}\ k\omega\rho\ p^2(\vec{r})\ \vec{r}_V(\vec{r}) \tag{2.31}$$

Ce résultat confirme clairement que le vecteur *Vélocité* décrit les *transports de l'énergie acoustique*, à la fois en termes de direction de propagation ($\vec{r}_{0_V}$) et de vitesse de propagation (r_V) des ondes. De plus, l'intensité active étant proportionnelle au gradient de la phase ϕ du signal de pression défini par la relation:

$$p(\vec{r}) = |p(\vec{r})|\ e^{j\,\phi(\vec{r})}\ , \tag{2.32}$$

le vecteur *Vélocité* est également lié à ce gradient:

$$\vec{r}_V(\vec{r}) = \frac{1}{k}\ \vec{\nabla}\phi(\vec{r}) \tag{2.33}$$

Or, le gradient $\vec{\nabla}\phi(\vec{r})$ correspond, à un coefficient multiplicatif près, á la normale au front d'onde, le front d'onde étant, par définition, une courbe d'équiphase. Par suite, le vecteur *Vélocité* coïncide bien avec la normale au front d'onde, comme on l'a déjà indiqué.

Vecteur *Energie*

Le vecteur *Energie* $\vec{r}_E$ est défini comme:

$$\vec{r}_E = \frac{\sum_{i=1}^{N} |g_i|^2\ \vec{h}_{0_i}}{\sum_{i=1}^{N} |g_i|^2} \tag{2.34}$$

Ce critère intervient pour évaluer la reproduction du champ sonore aux hautes fréquences (H.F.), correspondant aux fréquences supérieures à 700 Hz. En effet, le système auditif n'est véritablement sensible aux différences de phase qu'aux basses fréquences. Aussi, lorsque la fréquence augmente, préfère-t-on raisonner sur les énergies. En particulier, en présence d'un champ constitué d'une superposition de plusieurs ondes, il convient de sommer non plus les amplitudes, mais les énergies. Le critère $\vec{r}_E$ résulte de cette distinction entre basses et hautes fréquences.

Ses connotations physiques sont cependant moins évidentes que pour le vecteur *Vélocité*. Bien qu'il s'agisse d'une grandeur homogène à une intensité acoustique, puisqu'on y reconnaît la somme des produits entre la pression et la vitesse particulaire des ondes émises par chaque haut-parleur, le vecteur $\vec{r}_E$ ne peut pas être identifié au vecteur d'intensité acoustique associé à l'onde synthétisée par l'ensemble des haut-parleurs. L'intensité acoustique est en effet une grandeur énergétique, c'est-à-dire du second ordre. Par conséquent, l'intensité totale n'est pas égale à la somme des intensités élémentaires.

Comme pour le vecteur *Vélocité*, on distingue:

– sa direction:

$$\vec{r}_{0_E} = \frac{\vec{r}_E}{|\vec{r}_E|}$$

– et sa norme:

$$r_E = |\vec{r}_E|$$

Sans en apporter de véritable justification, M.A. Gerzon considère que le vecteur $\vec{r}_{0_E}$ repère la direction dans laquelle l'auditeur localise la source virtuelle. Même si le vecteur $\vec{r}_{0_E}$ ne peut être rattaché à aucune grandeur physique, on peut cependant l'identifier à la *direction moyenne de provenance de l'énergie*. L'équation 2.34 définit en effet la moyenne des directions d'arrivée des ondes émises par l'ensemble des haut-parleurs, la direction de chaque haut-parleur étant pondérée par son énergie[22]. Par suite, pour une restitution sonore optimale, la direction visée par le vecteur $\vec{r}_{0_E}$ doit cïncider avec la direction originale de la source (*première règle de décodage H.F.*):

$$\vec{r}_{0_E} = \frac{\vec{r}_s}{|\vec{r}_s|} \tag{2.35}$$

Les deux premières règles de décodage B.F. et H.F. (cf. Equ. 2.26 & 2.35) peuvent être fondues en une seule:

$$\vec{r}_{0_V} = \vec{r}_{0_E} = \frac{\vec{r}_s}{|\vec{r}_s|} \tag{2.36}$$

Quant au critère de norme r_E, il s'interprète dans le même esprit comme le *le taux de concentration de l'énergie* dans la direction du vecteur $\vec{r}_{0_E}$. Par analogie avec le critère r_V, il évalue la qualité de restitution de l'image sonore, en termes de stabilité et de confort d'écoute. On montre aisément que sa valeur est toujours inférieure ou égale à 1. Comme pour r_V, il faudrait, pour une restitution sonore optimale, lui imposer d'être égal à 1, mais cette valeur est impossible à atteindre en pratique. Aussi se contente-t-on de chercher à rendre sa valeur aussi grande que possible, c'est-à-dire que la *seconde règle de décodage H.F. consiste à maximiser* r_E.

Mise en œuvre des critères psychoacoustiques

Comment, en pratique, sont appliquées les règles de décodage que l'on vient d'énoncer? La démarche est la suivante. D'abord, on définit la *disposition des haut-parleurs*: ils peuvent être distribués dans tout l'espace, de préférence sur la surface d'une sphère centrée sur l'auditeur, ou uniquement dans le plan horizontal[23] situé à la hauteur de la tête de l'auditeur. Une fois fixées les positions des haut-parleurs, le problème consiste à exprimer la matrice de décodage $\mathbf{M}_d$ qui permet de satisfaire la première règle de décodage (Equ. 2.36) et la seconde règle de décodage B.F. (Equ. 2.29), à savoir:

$$\left\{ \begin{array}{lcl} \vec{r}_{0_V} & = & \vec{r}_{0_E} = \frac{\vec{r}_s}{|\vec{r}_s|} \\ r_V & = & 1 \end{array} \right. \tag{2.37}$$

Lorsque les haut-parleurs sont distribués selon une géométrie "simple", c'est-à-dire obéissant à certains critères de régularité[24], la matrice de décodage est dérivée analytiquement. Ainsi, si les haut-parleurs forment un *polyèdre régulier*, elle est donnée par [Gerzon, 1992b]:

$$\mathbf{M}_d = \frac{1}{N} \begin{bmatrix} 1 & \sqrt{3}h_{0_{1_x}} & \sqrt{3}h_{0_{1_y}} & \sqrt{3}h_{0_{1_z}} \\ 1 & \sqrt{3}h_{0_{2_x}} & \sqrt{3}h_{0_{2_y}} & \sqrt{3}h_{0_{2_z}} \\ 1 & \sqrt{3}h_{0_{3_x}} & \sqrt{3}h_{0_{3_y}} & \sqrt{3}h_{0_{3_z}} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \sqrt{3}h_{0_{N_x}} & \sqrt{3}h_{0_{N_y}} & \sqrt{3}h_{0_{N_z}} \end{bmatrix} \tag{2.38}$$

22. De même, pour le vecteur *Vélocité*, l'équation 2.24 exprime aussi la moyenne des directions d'arrivée des ondes, à la différence que la pondération affectée à chaque haut-parleur est reliée à son *gain en amplitude* au lieu de son gain en énergie.

23. A noter qu'en ce cas, dans les quatre signaux enregistrés par le microphone Soundfield, le signal Z n'est pas pris en compte.

24. On peut entendre régularité au sens matriciel, étant donné que les propriétés de la configuration géométrique des haut-parleurs déterminent le comportement de la matrice de décodage $\mathbf{M}_d$. Cette idée sera précisée ultérieurement, lorsque l'approche ambisonique sera identifiée à une méthode de reconstruction physique de champ acoustique (cf. Section 2.6.3 & Chapitre 7).

où le vecteur $\vec{h}_{0_i}$ repère la position du ième haut-parleur ($i = 1, \cdots, N$) (cf. Equ. 2.25). Dans le cas d'un polygone régulier (plan horizontal), la matrice de décodage devient [Gerzon, 1992b]:

$$\mathbf{M}_d = \frac{1}{N} \begin{bmatrix} 1 & \sqrt{2}h_{0_{1_x}} & \sqrt{2}h_{0_{1_y}} \\ 1 & \sqrt{2}h_{0_{2_x}} & \sqrt{2}h_{0_{2_y}} \\ 1 & \sqrt{2}h_{0_{3_x}} & \sqrt{2}h_{0_{3_y}} \\ \vdots & \vdots & \vdots \\ 1 & \sqrt{2}h_{0_{N_x}} & \sqrt{2}h_{0_{N_y}} \end{bmatrix} \tag{2.39}$$

Dans le cas général où la distribution des haut-parleurs est quelconque, la matrice de décodage est obtenue en résolvant un *problème d'optimisation non linéaire sous contrainte* [Trébuchet, 1997].

La direction de chaque haut-parleur étant repérée en coordonnées sphériques par les angles (φ_i, θ_i) (cf. Fig. 2.30):

$$\vec{h}_{0_i} = \left| \begin{array}{lcl} h_{0_{i_x}} & = & \sin\theta_i \ \cos\varphi_i \\ h_{0_{i_y}} & = & \sin\theta_i \ \sin\varphi_i \\ h_{0_{i_z}} & = & \cos\theta_i \end{array} \right. \tag{2.40}$$

les gains de chacun des haut-parleurs s'expriment, dans le cas d'un polyèdre:

$$\begin{aligned} g_i &= \frac{1}{N} \left(W + X\ \sqrt{2}\ h_{0_{i_x}} + Y\ \sqrt{2}\ h_{0_{i_y}} + Z\ \sqrt{2}\ h_{0_{i_z}} \right) \\ &= \frac{1}{N} \left(W + X\ \sqrt{2}\ \sin\theta_i\ \cos\varphi_i + Y\ \sqrt{2}\ \sin\theta_i\ \sin\varphi_i + Z\ \sqrt{2}\ \cos\theta_i \right) \end{aligned} \tag{2.41}$$

et, dans le cas d'un polygone:

$$\begin{aligned} g_i &= \frac{1}{N} \left(W + X\ \sqrt{2}\ h_{0_{i_x}} + Y\ \sqrt{2}\ h_{0_{i_y}} \right) \\ &= \frac{1}{N} \left(W + X\ \sqrt{2}\ \cos\varphi_i + Y\ \sqrt{2}\ \sin\varphi_i \right) \end{aligned} \tag{2.42}$$

Il reste ensuit á optimiser la matrice de décodage est optimisée du point de vue des hautes fréquences en vertu de la seconde loi de décodage H.F. [Daniel *et al.*, 1998]. Dans ce but, on applique aux signaux (W,X,Y,Z) des gains correctifs c_W, c_X, c_Y et c_Z:

$$\mathbf{g} = \mathbf{M}_d \begin{pmatrix} c_W & W \\ c_X & X \\ c_Y & Y \\ c_Z & Z \end{pmatrix} \tag{2.43}$$

et le problème consiste à déterminer les valeurs de ces gains qui maximisent le critère r_E. Ces gains n'interviennent que dans les hautes fréquences, c'est-à-dire les fréquences supérieures à 700 Hz, ils sont donc appliqués au moyen de filtres, désignés dans la littérature sous le nom de *shelf filter* [Gerzon, 1980] [Gerzon, 1985]. Un schéma général du circuit de décodage est illustré sur la figure 2.31.

2.6.3 Lien avec les harmoniques cylindriques [Bamford, 1995]

Dans la section précédente, on a vu que le décodage B.F., basé sur le critère $\vec{r}_V$, est équivalent à une reconstruction physique du champ acoustique (pression et vitesse particulaire) au point représentant la position de l'auditeur. Cette idée a été précisée analytiquement par le travail de thèse de J.S Bamford qui a établi le lien entre le procédé ambisonique et une décomposition en harmoniques cylindriques [Bamford, 1995].

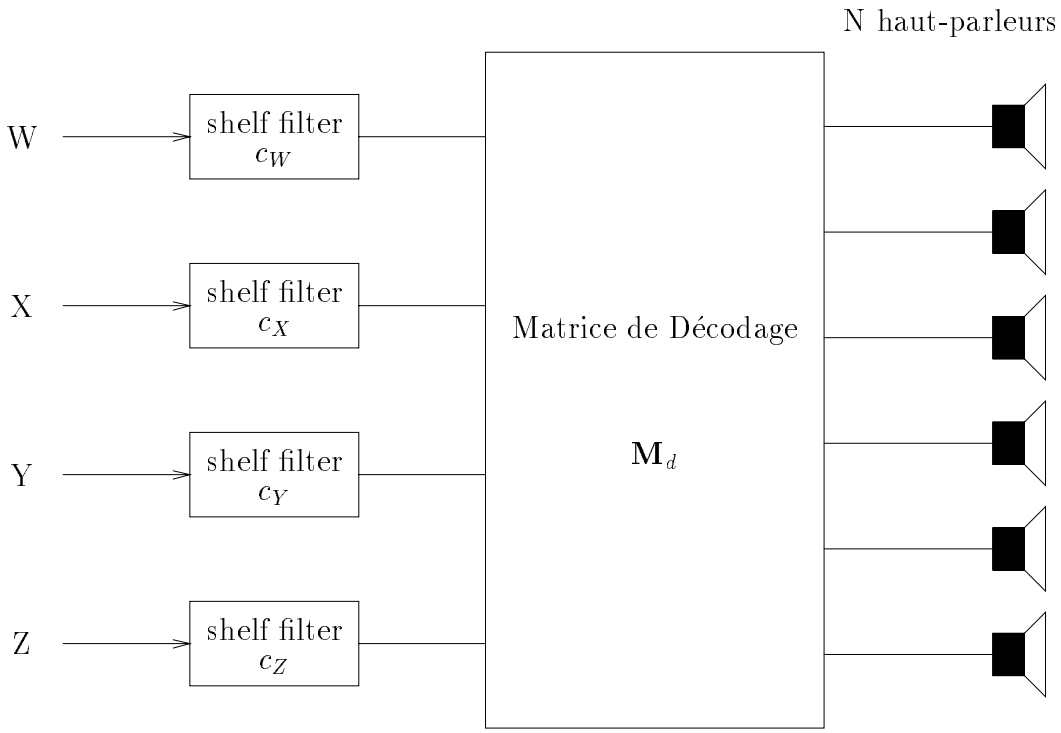


Fig. 2.31 - Synoptique d'un circuit de décodage ambisonique

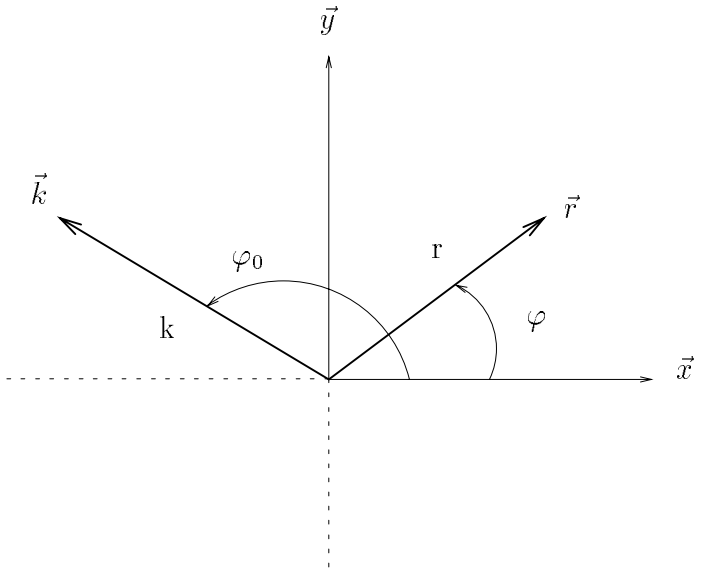


FIG. 2.32 - Onde plane horizontale: Coordonnées cylindriques (vecteur d'onde $\vec{k}$ de l'onde plane et point d'écoute $\vec{r}$)

Décomposition en harmoniques cylindriques

Les travaux de J.S. Bamford sont basés sur la *décomposition d'une onde plane en harmoniques cylindriques* [Morse & Ingard, 1968] [Bruneau, 1983]. Soit une onde plane se propageant parallèlement au plan horizontal $(\vec{x}, \vec{y})$ et dont le vecteur d'onde $\vec{k}$ est donné par (cf. Fig. 2.32):

$$\vec{k} = k \left| \begin{array}{l} \cos\varphi_0 \\ \sin\varphi_0 \\ 0 \end{array} \right. \tag{2.44}$$

on montre que:

$$\begin{aligned} p(r,\varphi) &= a\ e^{j\,kr\cos(\varphi-\varphi_0)} \\ &= a\ J_0(kr) + 2a \sum_{m=1}^{+\infty} j^m J_m(kr)\cos(m\varphi_0)\cos(m\varphi) \\ &\quad +2a \sum_{m=1}^{+\infty} j^m J_m(kr)\sin(m\varphi_0)\sin(m\varphi) \end{aligned} \tag{2.45}$$

Dans cette expression, le point $\vec{r}$ est repéré par ses coordonnées cylindriques $\vec{r}[r, \varphi, z = 0]$ (cf. Fig. 2.32). Cette décomposition est également connu sous le nom de *développement en série de Fourier-Bessel.*

L'équation 2.45 peut être réécrite sous forme matricielle. En introduisant les vecteurs $\mathbf{u}$ et $\mathbf{v}$ définis par:

$$\mathbf{u}^T = [1 \quad \sqrt{2}\cos(\varphi_0) \quad \sqrt{2}\sin(\varphi_0) \quad \cdots \quad \sqrt{2}\cos(m\varphi_0) \quad \sqrt{2}\sin(m\varphi_0) \quad \cdots] \tag{2.46}$$

$$\mathbf{v}^T = [J_0(kr) \quad j\sqrt{2}\cos(\varphi)J_1(kr) \quad j\sqrt{2}\sin(\varphi)J_1(kr) \quad \cdots \quad j^m\sqrt{2}\cos(m\varphi)J_m(kr) \quad j^m\sqrt{2}\sin(m\varphi)J_m(kr) \quad \cdots] \tag{2.47}$$

il vient:

$$p(r,\varphi) = a\ \mathbf{u}^T\ \mathbf{v}(r,\varphi) \tag{2.48}$$

La décomposition en harmoniques cylindriques permet de dissocier les dépendances radiales (en fonction de r) des dépendances azimutales (en fonction de φ) dans le comportement spatial de l'onde plane. Ainsi l'*information directionnelle*, qui est liée à l'angle φ_0, est décrite intégralement par le vecteur $\mathbf{u}$ qui représente les coefficients de la décomposition en harmoniques cylindriques. Or, on reconnaît dans les premiers coefficients du développement u_i les signaux (W,X,Y) d'un enregistrement ambisonique (cf. Equ. 2.18):

$$\left\{ \begin{array}{lcccl} a\ u_0 & = & W & = & a \\ a\ u_1 & = & X & = & a\ \sqrt{2}\ \cos\varphi_0 \\ a\ u_2 & = & Y & = & a\ \sqrt{2}\ \sin\varphi_0 \end{array} \right. \tag{2.49}$$

Un *enregistrement ambisonique* revient donc à une *décomposition en harmoniques cylindriques limitée à l'ordre* $m = M = 1$, où l'entier M représente l'ordre de troncature de la série d'harmoniques (cf. Equ. 2.45). Ce résultat suggère donc un moyen d'enrichir une prise de son ambisonique: il suffit d'étendre l'ordre de troncature à $M = 2$, ou $M = 3$...etc... L'intérêt de cette extension réside dans un *élargissement de la zone d'écoute*. On peut montrer en effet qu'en augmentant l'ordre de troncature de la décomposition, la zone de reconstruction correcte s'étend autour de l'auditeur à un cercle dont le rayon est proportionnel à M, à la manière d'un développement de Taylor [Nicol & Emerit, 1999a]. Une prise de son ambisonique étendue aux ordres supérieurs constituerait donc une amélioration considérable par rapport au système actuel.

A l'ordre M, il faudrait enregistrer $2M + 1$ signaux:

$$\left\{ \begin{array}{lcl} W & = & a \\ X_1 & = & a\ \sqrt{2}\ \cos\varphi_0 \\ Y_1 & = & a\ \sqrt{2}\ \sin\varphi_0 \\ X_2 & = & a\ \sqrt{2}\ \cos(2\varphi_0) \\ Y_2 & = & a\ \sqrt{2}\ \sin(2\varphi_0) \\ \vdots & & \\ X_M & = & a\ \sqrt{2}\ \cos(M\varphi_0) \\ Y_M & = & a\ \sqrt{2}\ \sin(M\varphi_0) \end{array} \right. \tag{2.50}$$

Cette extension pose cependant le problème de la prise de son de tels signaux, car elle implique des microphones à directivité d'ordre supérieur (directivité en $\cos(m\varphi_0)$ pour $m = 2, 3, \cdots, M$) qui n'existent encore qu'à titre expérimental. Cependant, nous verrons au chapitre 7 qu'il est possible de dériver les signaux d'encodage ambisonique d'une prise de son par un réseau circulaire de microphones cardioïdes.

Reconstruction par une superposition d'ondes planes

On a vu qu'une prise de son ambisonique équivaut à une décomposition en harmoniques cylindriques. Etudions maintenant le problème de la reproduction de l'onde plane à partir des signaux enregistrés. On considère N haut-parleurs régulièrement répartis sur un cercle dont le rayon est suffisamment grand pour que l'onde induite au niveau de l'auditeur par chacun des haut-parleurs soit assimilable à une onde plane, ce qui suppose que le rayon du cercle soit grand devant la longueur d'onde. Les haut-parleurs étant positionnés aux angles φ_i définis par:

$$\varphi_i = i\ \frac{2\pi}{N},\ (i = 1, 2, \cdots, N)$$

la pression totale résultant de la superposition des N ondes planes s'écrit:

$$p(r, \phi) = \sum_{i=0}^{N-1} g_i\ e^{jkr\cos(\varphi - \varphi_i)} \tag{2.51}$$

où le facteur g_i désigne le gain appliqué à chaque haut-parleur.

Chacune des ondes planes élémentaires peut alors être décomposée sur la base des harmoniques cylindriques (cf. Equ. 2.45):

$$a_i\ e^{jkr\cos(\varphi - \varphi_i)} = g_i\ \mathbf{u}_i^T\ \mathbf{v}(r, \varphi) \tag{2.52}$$

de sorte que la pression totale (Equ. 2.51) devient:

$$\begin{aligned} p(r,\phi) &= \sum_{i=0}^{N-1} g_i \ \mathbf{u}_i^T \ \mathbf{v}(r,\varphi) \\ &= \mathbf{g}^T \ \mathbf{U} \ \mathbf{v}(r,\varphi) \end{aligned} \tag{2.53}$$

Le vecteur $\mathbf{g}$ représente les gains des N haut-parleurs, tandis que la matrice $\mathbf{U}$ décrit leurs positions définies par leur angle d'incidence φ_i:

$$\begin{aligned} \mathbf{g}^T &= [\ g_1 \ g_2 \cdots g_N\ ] \\ \mathbf{U}^T &= [\ \mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_N\ ] \end{aligned}$$

Les gains g_i des haut-parleurs s'obtiennent en identifiant l'onde synthétique (cf. Equ. 2.53) à l'onde originale (cf. Equ. 2.48), ainsi:

$$p = a \ \mathbf{u}^T \ \mathbf{v} = \mathbf{g}^T \ \mathbf{U} \ \mathbf{v} \tag{2.54}$$

Par suite, les gains g_i sont obtenus en inversant la matrice $\mathbf{U}$, mais cette dernière est de dimensions $N \times (2M+1)$, c'est-à-dire que dans le cas général[25] où $N \neq 2M+1$, elle n'admet pas d'inverse au sens strict, mais elle possède une *pseudo-inverse* qui est donnée par [Daniel *et al.*, 1998]:

$$\mathbf{U}_{pinv} = \mathbf{U}^T \ (\mathbf{U} \ \mathbf{U}^T)^{-1} \tag{2.55}$$

On montre que, dès lors que les haut-parleurs forment un polygone régulier, cette matrice pseudo-inverse prend une forme particulièrement simple[26] [Daniel *et al.*, 1998] [Bamford, 1995]:

$$\mathbf{U}_{pinv} = \frac{1}{N} \ \mathbf{U}^T \tag{2.56}$$

Les gains des haut-parleurs s'expriment alors:

$$\begin{aligned} \mathbf{g} &= a \ \mathbf{U}_{pinv}^T \ \mathbf{u} \\ &= \frac{a}{N} \ \mathbf{U} \ \mathbf{u} \end{aligned} \tag{2.57}$$

On vérifie qu'à l'ordre $M=1$, c'est-à-dire avec $2M+1=3$ signaux enregistrés (W,X,Y) restitués sur un dispositif de N haut-parleurs, on retrouve les gains obtenus par M.A. Gerzon (cf. Equ. 2.42):

$$\begin{aligned} g_i &= \frac{a}{N} \ (1 + 2 \ \cos\varphi_i \ \cos\varphi_0 + 2 \ \sin\varphi_i \ \sin\varphi_0) \\ &= \frac{1}{N} \ \left(W + \sqrt{2} \ \cos\varphi_i \ X + \sqrt{2} \ \sin\varphi_i \ Y \right) \end{aligned} \tag{2.58}$$

Système ambisonique généralisé

Le principal apport des travaux de J.S. Bamford a consisté à replacer le procédé ambisonique, du moins au sens du décodage défini pour les basses fréquences, dans un contexte de reconstruction physique de champ sonore. Il en résulte une meilleure compréhension de la méthode et de ses limitations. En particulier, il apparaît clairement que l'approche ambisonique est fondamentalement basée sur les propriétés des ondes planes, ce qui n'est pas véritablement restrictif, étant donné que tout champ acoustique peut être décomposé

25. Il convient de noter que la situation où $N \neq 2M+1$ signifie:

- si $N < 2M+1$, qu'il s'agit d'un *problème sous-dimensionné*, c'est-à-dire qu'on veut reproduire 2M+1 voies enregistrées avec un nombre N inférieur de canaux de reproduction, ce qui se traduit par une inévitable *perte d'information*,
- si $N > 2M+1$, qu'il s'agit d'un *problème sur-dimensionné*, c'est-à-dire qu'on veut reproduire 2M+1 voies enregistrées avec un nombre N supérieur de canaux de reproduction, ce qui implique une *redondance* entre les haut-parleurs.

Par suite, la situation *optimale*, au sens où elle ne produit ni perte d'information, ni redondance, correspond au cas où le nombre de voies enregistrées est égal au nombre de canaux de reproduction: $N = 2M+1$. Pour un enregistrement ambisonique à l'ordre M, le nombre optimal de haut-parleurs est ainsi $N = 2M+1$ [Poletti, 1996].

26. Ce résultat est d'ailleurs lié aux propriétés d'orthogonalité des harmoniques cylindriques.

en une somme d'ondes planes [Bruneau, 1983] [Morse & Ingard, 1968]. Au delà de ce premier résultat, l'étude de J.S. Bamford met en évidence une possibilité d'améliorer et d'étendre les performances de la méthode ambisonique. On a vu en effet qu'une prise de son ambisonique réalise une décomposition en harmoniques cylindriques qui est tronquée à l'ordre $M = 1$ dans l'approche classique définie par M.A. Gerzon. En étendant l'ordre de troncature aux ordres supérieurs $M \geq 2$, ce qui implique d'augmenter le nombre de signaux enregistrés, il est donc possible de réaliser une prise de son ambisonique plus riche et offrant une zone d'écoute élargie. Cette idée définit l'approche *ambisonique généralisée* qui sera détaillée au chapitre 7, lorsque l'on présentera l'approche unifiée de reconstruction physique de champ sonore, dans laquelle l'holophonie et le procédé ambisonique sont fusionnés.

2.6.4 Conclusion

Le système ambisonique possède plusieurs atouts. D'une part, il séduit par la *simplicité* de sa mise en œuvre, dans la mesure où, fondamentalement, dans le procédé ambisonique, un champ sonore 3D est reproduit à partir de quatre signaux seulement. Du point de vue matériel, l'équipement requis, qui consiste en quatre microphones cardioïdes et quatre haut-parleurs (dispositif minimal), reste aussi très basique. D'autre part, l'idée de *distinguer deux processus différents pour les basses fréquences et hautes fréquences*, en adoptant une *reproduction simplifiée pour les hautes fréquences*, est très intéressante. Inspirée des propriétés psychoacoustiques du système auditif, elle permet de réduire les coûts de reproduction en préservant la qualité du rendu. Cependant, le procédé ambisonique souffre d'un inconvénient majeur dans la perspective de la visioconférence: la zone d'écoute est limitée à un auditeur. Face à ce problème, les travaux de J.S. Bamford suggèrent un élément de solution avec une approche ambisonique généralisée, dans laquelle la zone d'écoute peut être étendue en augmentant le nombre de signaux enregistrés et qui sera développée ultérieurement (cf. Chapitre 7).

2.7 Privilégier une approche physique de reproduction sonore 3D

Ce chapitre aura été l'occasion d'un survol des méthodes de reproduction sonore 3D (stéréophonie, techniques binaurales, système ambisonique), en identifiant les différents façon d'aborder le problème de la spatialisation sonore, à savoir: l'approche physique, l'approche psychoacoustique et l'approche hybride. Cette étude a montré qu'il existe d'ores et déjà des outils performants qui sont capables de reproduire avec succès une image sonore 3D.

Cependant, l'ensemble des méthodes examinées présente un inconvénient rédhibitoire du point de vue d'une application de visioconférence de groupe: la zone d'écoute est si limitée qu'elle n'admet au maximum qu'un auditeur en lui interdisant tout déplacement.

Les méthodes psychoacoustiques ou hybrides proposent des approches simplifiées qui, en contrepartie, renforcent les contraintes sur le nombre et la position des auditeurs. Des solutions pour pallier ce problème ont été développées et nous les avons examinées, mais, quelle que soit l'approche choisie, l'extension de la zone d'écoute passe inéluctablement par un accroissement du nombre de haut-parleurs, et, dans certains cas, cet accroissement comporte le risque de fragiliser la robustesse de la méthode. Or, le principal reproche adressé aux méthodes basées sur une reconstruction physique du champ sonore concerne le nombre de sources nécessaires à la reconstruction. On se rend donc compte qu'en spatialisation sonore, il n'existe pas de solution miracle: *une zone d'écoute étendue implique nécessairement un nombre important de haut-parleurs.* Il est tout aussi évident que plus la taille de la zone à sonoriser est importante, plus le nombre de haut-parleurs requis est élevé.

Aussi, pour la suite de cette étude, a-t-on préféré se tourner vers une méthode plus globale, en privilégiant une approche de reconstruction physique de champ acoustique qui présente l'avantage de rester directement liée aux phénomènes physiques. On a ainsi recherché une compréhension approfondie des processus impliqués, en se réservant la possibilité d'effectuer des simplifications *a posteriori*. On a aussi voulu, dans une seconde étape, analyser l'ensemble des techniques de spatialisation sonore dans le cadre d'un formalisme commun, le plus général possible, qui permet de faire ressortir les divergences et les points communs, les avantages et les inconvénients de chaque méthode, et offre une comparaison directe de leurs performances. Cet objectif a été atteint pour le procédé ambisonique (cf. Chapitre 7).

A présent, nous allons nous intéresser à l'approche de reconstruction physique de champ sonore, en commençant par poser les principes fondamentaux de l'holophonie.

Références Bibliographiques

Aoki S. & Koizumi N. (1987). Expansion of Listening Area with Good Localization in Audio Conferencing. *In: Proc. I.C.A.S.S.P.*

Arnaud Y. (1996). *Etude du système à enceintes croisées: de la pertinence de la localisation stéréophonique en diffusion multipoints.* Rapp. tech. Ecole Nationale Supérieure Louis Lumière.

Bamford J.S. (1995). *An analysis of Ambisonic Sound Systems of First and Second Order.* Ph.D. Thesis, University of Waterloo, Ontario, Canada.

Bauck J. & Cooper D. H. (1996). Generalized Transaural Stereo and Applications. *J. Audio Eng. Soc.*, **44**(9), pp. 683–705.

Bauer B. B. (1960). Broadening the Area of Stereophonic Perception. *J. Audio Eng. Soc.*, **8**(2), pp. 91–94.

Blauert Jens. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization.* The MIT Press, Cambridge, Massachusetts.

Blumlein A.D. (Juin 1934). *U.K. Patent 394,325.*

Bruneau M. (1983). *Introduction aux théories de l'acoustique.* Université du Maine, Le Mans.

Condamines R. (1978). *Stéréophonie: Cours de relief sonore théorique et appliqué.* Masson, Paris.

Craven P.G. & Gerzon M.A. (Août 1977). *U.S. Patent 4,042,779.*

Daniel J., Rault J.-B. & Polack J.-D. (Septembre 1998). Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. *In: Proceedings of the A.E.S. 105th Convention.*

Daniel J., Rault J.-B. & Polack J.-D. (Avril 1999). Acoustic Properties and Perceptive Implications of Stereophonic Phenomena. *In: Proceedings of the A.E.S. 16th International Conference.* pp. 91-102.

Dudouet E. & Martin J. (Mars 1999). A New HRTF Decomposition and Algorithm for Sound Spatialisation. *In: Collected Papers from the Joint Meeting "Berlin 99" (137th Meeting of the Acoustical Society of America / 2nd Convention of the European Acoustics Association).*

Farrar Ken. (1979a). Soundfield Microphone. *Wireless World*, Octobre, pp. 48–50.

Farrar Ken. (1979b). Soundfield Microphone - 2. *Wireless World*, Novembre, pp. 99–103.

Gerzon M.A. (1974). Surround-Sound Psychoacoustics. *Wireless World*, Décembre, pp. 483–486.

Gerzon M.A. (1977). Criteria for Evaluating Surround-Sound Systems. *J. Audio Eng. Soc.*, **25**(6), pp. 400–408.

Gerzon M.A. (1980). Practical Periphony: The Reproduction of Full-Sphere Sound. *In: Proceedings of the A.E.S. 65th Convention.*

Gerzon M.A. (1985). Ambisonics in Multichannel Broadcasting and Video. *J. Audio Eng. Soc.*, **33**(11), pp. 859–871.

GERZON M.A. (1992a). Ambisonic Decoders for HDTV. *In: Proceedings of the A.E.S. 92nd Convention.*

GERZON M.A. (1992b). General Metatheory of Auditory Localisation. *In: Proceedings of the A.E.S. 92nd Convention.*

GERZON M.A. (1992c). Optimum Reproduction Matrices for Multispeaker Stereo. *J. Audio Eng. Soc.*, **40**(7/8), pp. 571–589.

GERZON M.A. (1992d). Panpot Laws for Multispeaker Stereo. *In: Proceedings of the A.E.S. 92nd Convention.*

HERTZ B. F. (1981). 100 Years with Stereo: the Beginning. *Journ. of Audio Eng. Soc.*, **29**(5), pp. 368–372.

HUGONNET C. & WALDER P. (1994). *Théorie et pratique de la prise de son stéréophonique.* Eyrolles, Paris.

JESSEL M. (1973). *Acoustique théorique, propagation et holophonie.* Masson, Paris.

KERGOURLAY G. (1996). *Etude et prédiction de la zone d'écoute stéréo (Rapport de stage, France Telecom C.N.E.T.).* Rapp. tech. Ecole Nationale Supérieure des Télécommunications, Le Mans.

LARCHER V. & JOT J.-M. (Avril 1997). Techniques d'interpolation de filtres audio-numériques. Application à la reproduction spatiale des sons sur écouteurs. *In: Actes du 4e Congrès Français d'Acoustique.* pp. 97-100.

LOPEZ J.J., GONZALEZ A. & ORDUÑA BUSTAMANTE F. (Avril 1999). Measurement of Cross-Talk Cancellation and Equalization Zones in 3-D Sound Reproduction under Real Listening Conditions. *In: Proceedings of the A.E.S. 16th International Conference.* pp. 349-357.

MAKITA Y. (1962). On the Directional Localisation of Sound in the Stereophonic Sound Field. *E.B.U. Review*, Juin, pp. 102–108.

MØLLER H. (1992). Fundamentals of Binaural Technology. *Applied Acoustics*, **36**, pp. 171–218.

MORSE P.M. & INGARD K.U. (1968). *Theoretical Acoustics.* McGraw-Hill, New York.

NICOL R. & EMERIT E. (1999). 3D-Sound Reproduction over an Extensive Listening Area: a Hybrid Method Derived from Holophony and Ambisonic. *In: Proceedings of the A.E.S. 16th International Conference on Spatial Sound Reproduction.* pp. 436-453.

PERNAUX J.-M., BOUSSARD P. & JOT J.-M. (Novembre 1998). Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis. *In: Proceedings of the 98 Digital Audio Effects Workshop (DAFX98).* pp. 76-80.

POLACK J. D. (1995). *Cours de psychoacoustique et d'acoustique des salles.* Université du Maine, Le Mans.

POLETTI MARK. (1996). The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems. *J. Audio Eng. Soc.*, **44**(11), pp. 948–963.

PULKKI VILLE. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *J. Audio Eng. Soc.*, **45**(6), pp. 456–466.

SNOW W. B. (1953). Basic Principles of Stereophonic Sound. *Journal of the SMPTE*, **61**(Novembre), pp. 567–589.

TORICK E. (1998). Highlights in the History of Multichannel Sound. *J. Audio Eng. Soc.*, **46**(1/2), pp. 27–31.

TRÉBUCHET L.-C. (1997). *Etude et mise en œuvre des techniques ambisoniques pour la spatialisation du son.* Rapp. tech. FT.BD.CNET/DSM/RSA/SDA/147/97/LCT. E.N.S.T. / C.C.E.T.T.

WENZEL E.M. (Avril 1999). Effect of Increasing System Latency on Localization of Virtuel Sounds. *In: Proceedings of the A.E.S. 16th International Conference.* pp. 42-50.

Table des Illustrations