BRIQUE PAMU :Module ACOUS

Gaël RICHARD

9 janvier 2003

Table des matières

1. Chapitre partiellement écrit par J.M.Jot

Chapitre VI

Spatialisation des sons [1]

VI.1 Introduction

La *spatialisation* des sons consiste à reproduire un (ou plusieurs) sons dans un espace tridimensionnel (3D) que cela soit à partir de signaux monophoniques ou de signaux multi-canaux. On utilisera souvent dans ce chapitre le terme spatialisation sachant que des termes équivalents tels que *acoustique virtuelle*, *sons 3D*, *audio binaural* peuvent aussi être utilisés ([3]).

Ce chapitre, qui reprend de larges extraits de [13], est organisé de la façon suivante. Dans un premier temps, nous donnons une présentation générale de quelques techniques de codage et de reproduction spatiale des sons. Ensuite, nous présentons plus en détail les approches binaurales qui permettent de reproduire artificiellement (ou encore de simuler) la position d'une source sonore dans un espace 3D à partir d'un enregistrement monophonique.

VI.2 Techniques de codage et de reproduction spatiale des sons

Dans cette partie, nous présentons quelques approches de prise et de reproduction du son permettant de restituer la localisation d'une source dans l'espace sonore subjectif de l'auditeur. Lors de la prise de son, le codage des informations directionnelles est effectué par un dispositif de deux microphones au moins, ce qui implique un stockage ou une transmission du message sonore sur plusieurs canaux. La reproduction s'effectue sur une paire d'écouteurs ou un dispositifs de deux haut-parleurs au moins. Dans un système de spatialisation, il s'agit de réaliser artificiellement le codage du message sonore sur plusieurs canaux à partir d'un signal de prise de son monophonique.

VI.2.1 La stéréophonie conventionnelle

La stéréophonie conventionnelle utilise les différences d'intensité et les décalages temporels entre les deux signaux émis par les haut-parleurs pour fournir les indications relatives aux positions qu'occupaient les sources lors de l'enregistrement. Les deux signaux sont fournis, lors de l'enregistrement, par deux microphones de directivités différentes, orientés différemment ou placés à des positions différentes (on parle alors d'un couple de microphones "non coïncidents"). Le potentiomètre panoramique d'une console de mixage reproduit artificiellement ces différences entre les deux canaux, à partir d'un signal monophonique. Par exemple, un potentiomètre pa-

1. Chapitre partiellement écrit par J.M.Jot

noramique qui permet d'obtenir un signal stéréo (signaux D et G) à partir d'un signal monophonique $x(n)$ peut être réalisé simplement de la façon suivante:

$$D = (1-\alpha)x(n) \quad \text{(VI.1)}$$
$$G = \alpha x(n) \quad \text{(VI.2)}$$

où α varie de 0 à 1 suivant la position désirée de la source.
Un modèle un peu plus réaliste peut être réalisé en posant:

$$c = (1+W)\cos(\frac{\phi}{2}-\frac{\pi}{4}) + (1-W)\sin(\frac{\phi}{2}-\frac{\pi}{4}) \quad \text{(VI.3)}$$
$$d = (1-W)\cos(\frac{\phi}{2}-\frac{\pi}{4}) + (1+W)\sin(\frac{\phi}{2}-\frac{\pi}{4}) \quad \text{(VI.4)}$$

On obtiendra alors les signaux droite et gauche (respectivement D et G):

$$D = |c| \times x(n) \quad \text{(VI.5)}$$
$$G = |d| \times x(n) \quad \text{(VI.6)}$$

où W représente le coefficient de largeur de la stéréo. Une valeur de $W = 1$ annule le canal droit (resp. le canal gauche) pour un angle de -90^o (resp. 90^o).

Notons, que les potentiomètres panoramiques conventionnels n'introduisent, comme celui décrit ci-dessus, qu'une différence d'intensité entre les canaux gauche et droit [22], ce qui simule un enregistrement effectué à l'aide d'un couple de microphones coïncidents. Afin de bénéficier des qualités que de nombreux auteurs reconnaissent aux enregistrements par des couples de microphones non coïncidents, tels que le couple AB "ORTF", une première tentative d'amélioration consisterait à introduire aussi un décalage temporel. Cependant, les positions des sources enregistrées en stéréophonie conventionnelle se trouvent ramenées, lors de la reproduction sur haut-parleurs, à l'intérieur du secteur angulaire horizontal délimité par les directions des deux haut-parleurs (c'est-à-dire entre les azimuts -30^o et $+30^o$). Une écoute au casque modifie considérablement la spatialisation obtenue, et produit invariablement une localisation subjective de la source "à l'intérieur de la tête".

VI.2.2 Les systèmes multi-canaux

le codage et la reproduction des informations directionnelles sur deux canaux seulement conduisent à une restriction de la zone de provenance apparente des sons (frontale en stéréo) et à une forte contrainte sur la position de l'auditeur. Afin de dépasser ces limites, il est nécessaire de coder les informations directionnelles sur un nombre de canaux supérieur à 2, et le nombre de haut-parleurs doit être au moins égal au nombre de canaux d'enregistrement. Deux approches fournissent aujourd'hui des solutions applicables à un auditeur unique ou à un auditoire réduit : l'approche "sound field" (représentée par la norme anglaise "Ambisonics") et l'approche "surround" (représentée par les codages utilisés en cinéma et leurs dérivés).

Le système Ambisonic

Dans ce système, la localisation de la source se fait dans tout le plan horizontal au prix d'un canal de transmission supplémentaire. Le principe de base est le suivant (voir figure VI.1). Une

onde plane dont le plan d'onde est incident avec un angle ψ avec l'axe des x sera perçu par un auditeur situé à une distance r du centre et à un angle θ par rapport à l'axe des x. L'onde plane, représentée par S_ψ, peut alors s'écrire:

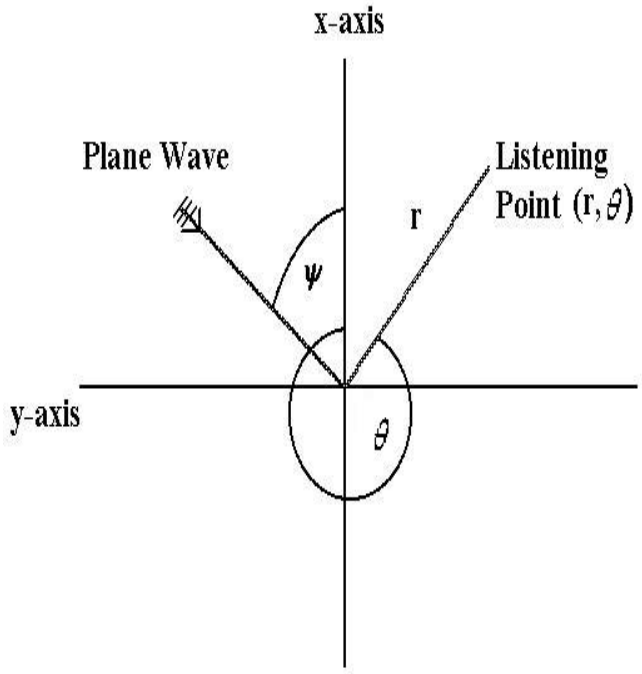

FIG. VI.1 – *Convention d'orientation du plan d'onde*

$$S_\psi = P_\psi e^{ikr\cos(\theta-\psi)} \tag{VI.7}$$

où k est le nombre d'onde et P_ψ l'amplitude du signal. On peut alors montrer que le plan d'onde peut s'écrire en terme d'harmoniques sphériques telles que:

$$S_\psi = P_\psi J_0(kr) + P_\psi \left(\sum_{m=1}^{\infty} 2i^m J_m(kr)[\cos(m\psi)cos(m\theta) + sin(m\psi)sin(m\theta)] \right) \tag{VI.8}$$

où J_0 et J_m sont les fonctions du premier ordre.

Il est alors supposé que le son provenant de chaque haut-parleur est approximativement une onde plane à la position d'écoute. Si cela est le cas, alors chaque haut-parleur produira un signal répondant à l'équation VI.8. La seule différence est que le haut-parleur n'est pas situé à l'angle ψ mais à l'angle ϕ_n, où n est le numéro du haut-parleur.

$$S_n = P_n J_0(kr) + P_n \left(\sum_{m=1}^{\infty} 2i^m J_m(kr)[\cos(m\phi_n)cos(m\theta) + sin(m\phi_n)sin(m\theta)] \right) \tag{VI.9}$$

où P_n est l'amplitude du signal du n^{ieme} haut-parleur. Pour obtenir le signal au niveau de l'auditeur, il suffit alors de sommer les contributions de chaque haut-parleur:

$$S = \sum_{n=1}^{N} P_n J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) \left(\sum_{n=1}^{N} P_n \cos(m\phi_n) cos(m\theta) + \sum_{n=1}^{N} P_n sin(m\phi_n) sin(m\theta) \right) \tag{VI.10}$$

Ainsi, pour reproduire l'onde plane originale avec l'ensemble de haut-parleurs ambisonic, il suffit d'égaler les termes des équations VI.10 et VI.8. On peut alors identifier les équations suivantes:

$$P_\psi = \sum_{n=1}^{N} P_n \tag{VI.11}$$

$$P_\psi \cos(m\psi) = \sum_{n=1}^{N} P_n \cos(m\phi_n) \tag{VI.12}$$

$$P_\psi \sin(m\psi) = \sum_{n=1}^{N} P_n \sin(m\phi_n) \tag{VI.13}$$

En pratique, on restreindra le système ambisonic au premier ou au second ordre ce qui nous donne:

$$W = P_\psi = \sum_{n=1}^{N} P_n \tag{VI.14}$$

$$X = P_\psi \cos(\psi) = \sum_{n=1}^{N} P_n \cos(\phi_n) \tag{VI.15}$$

$$Y = P_\psi \sin(\psi) = \sum_{n=1}^{N} P_n \sin(\phi_n) \tag{VI.16}$$

$$U = P_\psi \cos(2\psi) = \sum_{n=1}^{N} P_n \cos(2\phi_n) \tag{VI.17}$$

$$V = P_\psi \sin(2\psi) = \sum_{n=1}^{N} P_n \sin(2\phi_n) \tag{VI.18}$$

$$\tag{VI.19}$$

Le B-format correspond en fait à un système ambisonic au premier ordre, c'est à dire restreint aux composantes W, X, et Y. On peut alors montrer que le signal P_n qui doit être généré à chaque haut-parleur s'écrit (d'après [25])

$$P_n = (W + 2X\cos(\phi_n) + 2Y\cos(\phi_n)) \tag{VI.20}$$

Si on se place dans un espace 3D, et non plus seulement dans un plan horizontal, on aura alors une quatrième composante, souvent notée Z.

La prise de son format Ambisonics (encore appelé "UHJ" ou "format B") nécessite un microphone composé de 4 capsules, appelé "Sound Field Microphone". A la reproduction, l'auditeur est placé au centre d'une sphère portant 8 à 12 haut-parleurs alimentés par des signaux obtenus par un matriçage des 4 signaux d'origine. Pour une reproduction des directions horizontales uniquement, trois canaux de codage et 4 à 6 haut-parleurs suffisent [11, 10] même si on peut aussi utiliser un nombre plus réduit de hauts parleurs. Le système Ambisonics est cependant lui aussi contraignant en ce qui concerne la position de l'auditeur, mais procure une localisation plus "robuste" que les techniques binaurales, résistant aux rotations de la tête de l'auditeur et aux variations inter-individuelles.

Autres approches

Dans l'approche "surround", l'amélioration visée par rapport à la stéréophonie conventionnelle est la restitution d'une scène sonore peu dépendante de la position de l'auditeur par rapport aux haut-parleurs et incluant des informations d'ambiance et de réverbération provenant des directions latérales et arrières. Cependant, contrairement au système Ambisonics, cette approche privilégie la restitution des événements sonores frontaux, tandis que la reproduction des autres informations spatiales est peu contrôlée. Cette approche est héritée des premières techniques stéréophoniques imaginées pour l'industrie cinématographique à la fin des années 30, qui sont à l'origine de la stéréophonie à deux canaux actuelle. Les systèmes de reproduction cinématographiques comportent encore aujourd'hui, en plus des haut-parleurs droit et gauche, un haut-parleur central placé derrière l'écran (assurant une localisation stable des dialogues pour tous les spectateurs) et éventuellement des haut-parleurs frontaux supplémentaires [15]. Plus tard sont apparus des haut-parleurs placés sur les murs latéraux et arrière de la salle, spécifiquement destinés à diffuser des signaux "d'ambiance" ou d'effets spéciaux. Comme dans le format Ambisonics, le nombre de canaux de transmission est rendu inférieur au nombre de canaux de diffusion par l'utilisation de matrices d'encodage et de décodage qui assurent une compatibilité avec la reproduction stéréo ou mono conventionnelle. L'exemple aujourd'hui le plus répandu de l'approche "surround" est le système "Dolby Stereo" utilisé dans l'industrie cinématographique, dont a été dérivé plus récemment le système "Dolby Surround" visant plus spécifiquement les applications vidéo domestiques. Trois canaux de transmission (gauche, centre et droit) sont employés pour la restitution des sources sonores frontales, tandis qu'un seul canal ("surround") est réservé aux informations d'ambiance.

Les organismes de normalisation internationaux ISO ont récemment standardisé un format de codage "universel" à 5 canaux (MPEG-2 multicanaux) , visant des applications avec ou sans accompagnement visuel (HDTV, vidéo, cinéma, multimedia, mais aussi radiophonie numérique, CD...). La différence essentielle entre ce format et le format "Dolby Surround" réside dans le fait que 2 canaux latéraux (au lieu d'un) sont utilisés pour la transmission des informations d'ambiance. Cela permet une meilleure reproduction des sons diffus (réverbération) et de la localisation des sources latérales. L'avenir de cette norme 3/2 repose sur la possibilité d'encoder ces cinq canaux audio sur un support numérique à capacité limitée, ce qui nécessite l'utilisation de techniques de réduction de débit basées sur des modèles perceptifs (par exemple par l'utilisation du codeur MPEG2/4 AAC). Dans le cadre du format numérique Dolby Surround SR-D, c'est le codeur "AC-3" qui est utilisé.

VI.2.3 Systèmes électroacoustiques dans les grandes salles

Les dispositifs de haut-parleurs recommandés pour la reproduction en format Stéréo 3/2 sont adaptés pour un auditoire de 20 à 30 personnes environ. La reproduction dans une grande

salle nécessite la prise en compte des retards de propagation entre chaque haut-parleur et les diverses zones de l'audience, suivant le principe développé, notamment dans le système "Delta Stéréophonie" [24]. Ce système, éprouvé dans de nombreuses salles depuis la fin des années 1970, réalise l'amplification d'une source sonore acoustique située sur la scène en évitant que les auditeurs aient conscience d'un message sonore provenant des haut-parleurs. Pour cela, les haut-parleurs d'appoint, couvrant uniformément la salle, émettent des versions retardées du signal capté par les microphones sur la scène afin que leur contribution soit fusionnée parmi les échos précoces pour chaque source et tout auditeur. Bien que le principe soit simple, il nécessite une mise en oeuvre complexe (matrice de retards dont les durées doivent être optimisées en fonction de la disposition des haut-parleurs, et automatiquement mises à jour en cas de déplacement d'une source).

VI.2.4 Les techniques binaurales

Pour l'écoute au casque, la localisation "hors de la tête" ne peut être obtenue sans joindre aux différences temporelles et énergétiques des différences spectrales entre les canaux gauche et droit. Ces différences spectrales sont naturellement fournies par un enregistrement binaural, où les microphones sont placés à l'intérieur ou à l'entrée des deux conduits auditifs d'un individu ou d'une tête artificielle.

L'objet de la synthèse binaurale est de reproduire artificiellement, à partir d'un enregistrement monophonique d'une source sonore, les signaux de pression que capteraient les deux oreilles si cette source était située à une position donnée dans l'espace par rapport à l'auditeur. Ces techniques, qui ont vu le jour au début des années 70, développées notamment par Blauert [4], font aujourd'hui l'objet d'un effort de recherche important [12].

La synthèse binaurale repose sur la mesure des HRTF (Head-Related Transfer Functions), c'est-à-dire des deux réponses impulsionnelles du canal acoustique reliant la source aux deux conduits auditifs [4]. Le potentiomètre panoramique, appelé ici filtre binaural, peut être directement réalisé par un double filtrage FIR utilisant ces réponses impulsionnelles. Ceci permet en théorie d'obtenir, lors d'une écoute au casque, une localisation aussi bien en azimut (dans le plan horizontal) qu'en élévation. Cependant, cette technique se heurte à deux inconvénients pratiques et expérimentaux:

- Les HRTF mesurées varient selon les individus. Il semble illusoire d'espérer obtenir un jeu de HRTF procurant une localisation hors de la tête pour n'importe quel auditeur [12, 2].
- le coût de l'implémentation en temps réel: la longueur des réponses impulsionnelles mesurées peut atteindre 10 ms (de sorte que le filtrage binaural implémenté sous forme RIF consomme la puissance de calcul de deux processeurs Motorola DSP56000 pour une fréquence d'échantillonnage de 32 kHz).

La reproduction d'un enregistrement binaural sur haut-parleurs apporte des difficultés supplémentaires, dues fait que, contrairement à la stéréophonie conventionnelle, l'enregistrement binaural est naturellement conçu pour une écoute au casque. Lors d'une écoute sur haut-parleurs, il est nécessaire d'effectuer un traitement supplémentaire afin de compenser les trajets "croisés" (contribution du haut-parleur droit au signal reçu par l'oreille gauche, et inversement). Cette technique, initialement proposée par Schroeder et Atal [20], fait elle-même appel à la connaissance des HRTF associées aux positions des deux haut-parleurs (azimut $\pm 30^o$ pour la disposition stéréophonique conventionnelle). Dans une seconde approche, proposée par Damaske [7], le filtrage compensateur des trajets croisés est déterminé empiriquement.

Ces méthodes permettent théoriquement d'obtenir une localisation dans n'importe quelle direction par rapport à l'auditeur, bien que les deux haut-parleurs soient en réalité placés en face de lui. En pratique, cela comporte deux difficultés:

- la réalisation du filtrage de conversion du signal pour écoute au casque (binaural) en un signal pour écoute sur haut-parleurs (transaural), suivant la technique proposée par Schroeder et Atal, n'est pas aussi immédiate que celle des filtres binauraux [20, 18, 6].
- la contrainte sur la position de l'auditeur: la reproduction transaurale suppose que l'auditeur est dans une position donnée par rapport aux haut-parleurs. S'il s'en écarte ou s'il tourne la tête, le filtre de conversion n'est plus effectif et la position subjective de la source tend à se confondre avec l'un des deux haut-parleurs.

VI.3 Mesure d'une réponse impulsionnelle

Pour obtenir une réponse impulsionnelle, il est possible soit de générer un signal d'excitation impulsionnel soit d'utiliser des séquences pseudo-aléatoires.

La méthode consistant à utiliser un signal d'excitation impulsionnel se heurte aux inconvénients suivants : compte tenu de la faible durée du signal, il est nécessaire pour obtenir un rapport signal/bruit satisfaisant d'utiliser une excitation de grande amplitude, ce qui est délicat lorsqu'on utilise un haut-parleur (introduction de non linéarités), tandis que les sources sonores impulsionnelles (éclateurs, pistolets) posent des problèmes de reproductibilité. Lorsqu'un haut-parleur est utilisé, la solution conventionnelle consiste à effectuer un moyennage de mesures successives afin d'améliorer le rapport signal/bruit de la mesure, ce qui conduit à une procédure de mesure fastidieuse.

Une solution plus avantageuse, appliquée initialement à l'acoustique des salles par Schroeder, consiste à utiliser un signal d'excitation pseudo-aléatoire [21, 5, 14]. Il existe aussi des systèmes utilisant une autre technique aux performances comparables : la spectrométrie à filtrage décalé (time-delay spectrometry), basée sur un balayage linéaire en fréquence [26, 19, 8].

La réponse impulsionnelle est fournie par la fonction d'intercorrélation $C_{xy}(n)$ entre le signal d'entrée et le signal de sortie du système mesuré (fig. VI.2). Il suffit pour cela que le signal aléatoire x_n soit un bruit blanc, car on a dans le cas général, en régime stationnaire:

$$C_{xy}(n) = C_{xx}(n) \star h_n$$

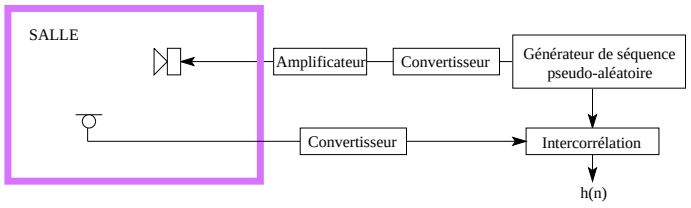

FIG. VI.2 – *Principe de la mesure d'une réponse impulsionnelle par la méthode proposée par Schroeder. [21]*

VI.3.1 Les séquences MLS

Les signaux d'excitation x(n) proposés par Schroeder [21] sont des séquences de longueur maximale (MLS), signaux périodiques de période $L = 2^k - 1$. Ces signaux, qui ne prennent

que les valeurs instantanées +1 et −1, sont aisément produits au moyen d'un calculateur ou de registres à décalage rebouclés [17].
La propriété intéressante de ces séquences est que leur fonction d'autocorrélation est presque exactement une fonction peigne de Dirac de période L. En effet, leur fonction d'autocorrélation cyclique pour la période L, notée ici C_{xx}^L est:

$$C_{xx}^L(n) = \frac{1}{L}\sum_{m=0}^{L-1} x_{m-n}x_m \quad = \quad 1 \quad \text{si} \quad n \equiv 0[mod\ L] \tag{VI.21}$$

$$= \quad \frac{-1}{L} \quad \text{sinon} \tag{VI.22}$$

Ce qui peut s'écrire:

$$C_{xx}^L(n) = (1 + \frac{1}{L})\delta_{[L]}(n) - \frac{1}{L}$$

On vérifie alors que

$$C_{xy}^L(n) = \frac{1}{L}\sum_{m=0}^{L-1} x_{m-n}y_m = (1 + \frac{1}{L})\delta_{[L]}(n) \star h_n - \frac{1}{L}\bar{h}$$

où $\bar{h}$ représente la composante continue de la réponse impulsionnelle et $\star$ l'opération de convolution. Ainsi, à la composante continue près (qui est le plus souvent négligeable), la fonction de corrélation entrée-sortie est la réponse impulsionnelle cherchée, multipliée par $(1 + 1/L)$ et "périodisée" (c'est-à-dire convoluée par le peigne de Dirac $\delta_{[L]}$). La réponse impulsionnelle peut donc être calculée sans approximation sur un horizon d'observation fini (une seule période suffit). Cet avantage provient du fait qu'une séquence MLS est en réalité un signal déterministe: l'élimination du caractère aléatoire du signal d'excitation garantit la reproductibilité au bruit ambiant près [5].
De plus, un algorithme utilisant la transformée de Hadamard permet le calcul rapide de la fonction d'intercorrélation cyclique [5], bien que la période L des signaux ne soit pas une puissance de 2 (ce qui interdit l'utilisation de l'algorithme FFT). Notons que la longueur de la période L doit être suffisante pour inclure la totalité de la réponse impulsionnelle à mesurer, faute de quoi la convolution temporelle ferait apparaître un phénomène d''aliasing temporel'.

VI.3.2 Les codes de Golay

Plus récemment, un autre type de signaux d'excitation a été proposé, les codes de Golay [9]. Leurs avantages sont: 1) la longueur de leur période est une puissance de 2, ce qui permet d'utiliser la FFT pour calculer la corrélation cyclique. 2) L'autocorrélation est *exactement* un peigne de Dirac, il n'y a plus d'erreur due à la composante continue. Néanmoins, ils nécessitent l'acquisition de deux séquences complémentaires, ce qui complique légèrement la mesure.

VI.3.3 Mesure des filtres binauraux (HRTFs)

Le principe de base pour la mesure des fonctions de transfert est montré sur la figure VI.5, cette opération devant bien entendue être répétée pour toutes les positions du haut-parleur. Les filtres binauraux capturent ainsi les indices de localisation provenant de la diffraction de l'onde incidente par le corps (torse, tête, oreille...) répétée pour chaque direction de provenance.

De nombreuses études ont utilisé les HRTFs du domaine publique obtenues à l'aide d'une tête de KEMAR [16]. Comme il a été mentionné précédemment, il n'existe pas un ensemble unique

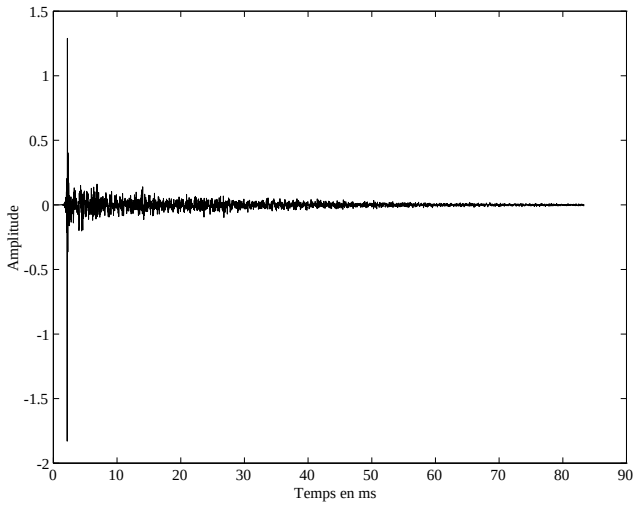

FIG. VI.3 – *Réponse impulsionnelle mesurée à l'aide d'une séquence MLS.*

d'HRTF. En effet, elles varient significativement d'une personne à l'autre en raison notamment des différences morphologiques entre les différentes personnes. Ainsi, un ensemble plus complet de HRTF a été enregistré par le CIPIC et mis partiellement à la disposition de tous ([1]). Ces HRTF ont été obtenus à l'aide d'un dispositif expérimental qui est utilisé aussi bien pour obtenir des HRTF à partir d'une tête KEMAR ou pour des HRTF d'une personne réelle (voir Figure VI.6).

Ce dispositif a permis d'obtenir un ensemble de réponses impulsionneles de filtres de têtes, encore appellées HRIR pour *Head Related Impulse Responses*. Chaque HRIR (ou HRTF) est échantillonnée à 44.1 kHz et est fonction de l'azimuth, de l'élévation et du temps. Les mesures sont faites pour des valeurs discrétisées de ces paramètres (25 azimuths différent compris entre -90^o et $+90^o$, 64 élévations entre -45^o et 230^o par pas de 5.625^o et 200 échantillons temporels correspondant à une durée d'environ 4.5 ms). Ces HRTFs ont été obtenus pour plus de 90 personnes (incluant deux têtes KEMAR avec des pavillons d'oreille différents), mais seulement une partie de cette base est rendue publique. Pour chaque personne, un ensemble de données morphologiques sont aussi mesurées pour permettre d'effectuer des études en vue de relier un ensemble de HRTF à des paramètres morphologiques (voir figure VI.7).

VI.4 Synthèse binaurale

On s'intéresse ici à la réalisation d'un "potentiomètre panoramique binaural", fournissant un signal stéréo à partir d'un signal mono, et permettant le contrôle en temps réel de la direction de provenance perçue par l'auditeur lors d'une écoute au casque, ou bien sur une paire de haut-parleurs (reproduction transaurale) auquel cas, comme on l'a vu précédemment, il faut réaliser un filtre convertisseur "transaural" dont le rôle est de compenser la perturbation apportée par les "trajets croisés". On cherche généralement à modéliser les fonctions de transfert binaurales, afin d'en extraire une représentation plus simple et plus économique pour le filtrage.

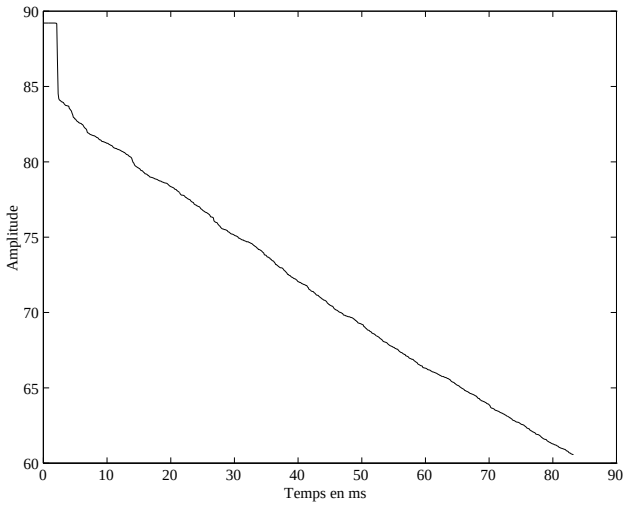

FIG. VI.4 – *Energy Curve Decay de cette réponse.*

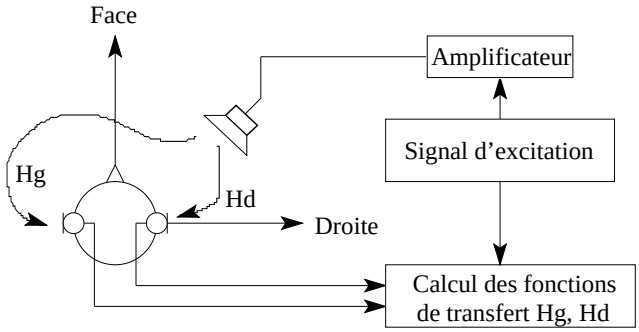

FIG. VI.5 – *Mesure d'un signal binaural.*

VI.4.1 Les filtres binauraux et transauraux

La figure VI.8 montre le principe de synthèse d'un signal binaural à partir d'un signal monophonique.

La figure VI.9 montre le principe du filtrage de conversion binaural → transaural. Ce principe nécessite la mesure des deux fonctions de transfert binaurales correspondant aux positions des haut-parleurs et de l'auditeur dans le lieu d'écoute. Les quatre fonctions de transfert ainsi obtenues forment la matrice de transfert qui caractérise la transformation des signaux Z_g et Z_d émis par les haut-parleurs vers les signaux Y_g et Y_d reçus par les conduits auditifs. Nous utilisons ici une notation omettant la dépendance fréquentielle des fonctions de transfert:

$$\begin{bmatrix} Y_g \\ Y_d \end{bmatrix} = \begin{bmatrix} H_{gg} & H_{dg} \\ H_{gd} & H_{dd} \end{bmatrix} \begin{bmatrix} Z_g \\ Z_d \end{bmatrix} \tag{VI.23}$$

Le filtrage de conversion binaural → transaural réalise la matrice de transfert inverse de la matrice précédente, ce qui conduit naturellement à une implémentation sous la "forme treillis" proposée par Cooper et Bauck [6] et représentée sur la fig. VI.10 :

FIG. VI.6 – *Mesure des HRTFs sur une tête Kemar (haut) et sur une personne réelle (bas) (d'après [1])*

$$\begin{bmatrix} Z_g \\ Z_d \end{bmatrix} = \frac{\begin{bmatrix} H_{dd} & -H_{dg} \\ -H_{gd} & H_{gg} \end{bmatrix}}{H_{gg}H_{dd} - H_{dg}H_{gd}} \begin{bmatrix} Y_g \\ Y_d \end{bmatrix} \tag{VI.24}$$

Nous admettons désormais l'hypothèse de symétrie du dispositif de reproduction et des conduits auditifs, et introduisons les notations suivantes, où l'on prend pour canal de référence le canal gauche:

$$H_g = G \quad ; \quad H_d = D \quad ; \quad H_{gg} = H_{dd} = G_0 \quad ; \quad H_{gd} = H_{dg} = D_0$$

G et D forment la fonction de transfert binaurale pour une position quelconque de la source, tandis que G_0 et D_0 forment la fonction de transfert binaurale pour la position du haut-parleur gauche (le haut-parleur droit étant situé symétriquement par rapport au plan médian de la tête de l'auditeur). Avec ces notations, l'éq. VI.24 s'écrit:

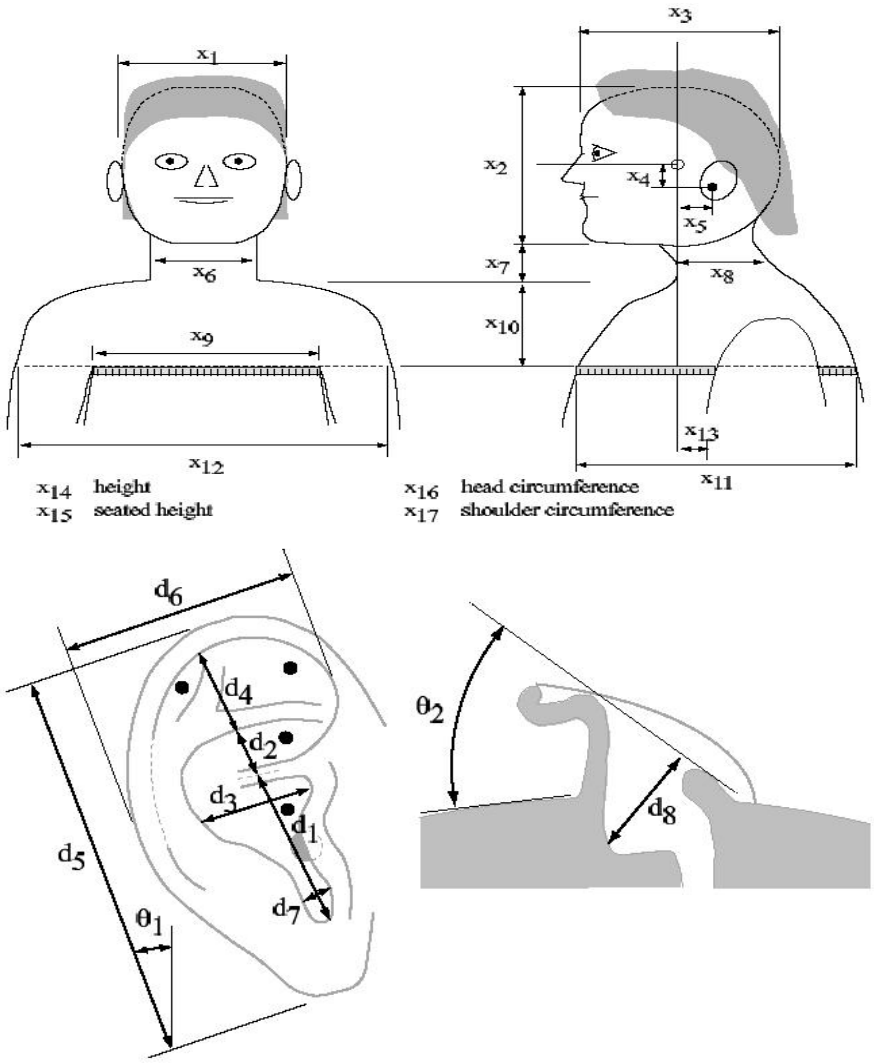

FIG. VI.7 – *Mesure des données morphologiques pour chaque personne (d'après [1])*

$$\begin{bmatrix} Z_g \\ Z_d \end{bmatrix} = \frac{\begin{bmatrix} G_0 & -D_0 \\ -D_0 & G_0 \end{bmatrix}}{G_0^2 - D_0^2} \begin{bmatrix} Y_g \\ Y_d \end{bmatrix} = \frac{1}{2} \left(\frac{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}{G_0 + D_0} + \frac{\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}}{G_0 - D_0} \right) \begin{bmatrix} Y_g \\ Y_d \end{bmatrix} \qquad \text{(VI.25)}$$

Si l'on ignore le coefficient 1/2, cela conduit à l'implémentation du filtre transaural sous la forme "shuffler" proposée par Cooper et Bauck [6] et représentée sur la fig. VI.11 à gauche. Poursuivant cette approche, on constate que tous les traitements de conversion envisageables dans le cadre des techniques binaurales peuvent être réalisés sous une forme nécessitant seulement l'implémentation de deux filtres élémentaires (voir figs. VI.10 et VI.11). Le filtre mono → transaural (fig. VI.10 à droite) est équivalent à un filtre mono → binaural associé en série avec un filtre binaural → transaural, et constitue effectivement un "potentiomètre panoramique" au sens classique (pour reproduction sur haut-parleurs). Le filtre transaural → binaural (fig. VI.11 à droite), inverse du filtre de conversion transaural précédent, permet l'écoute au casque d'un

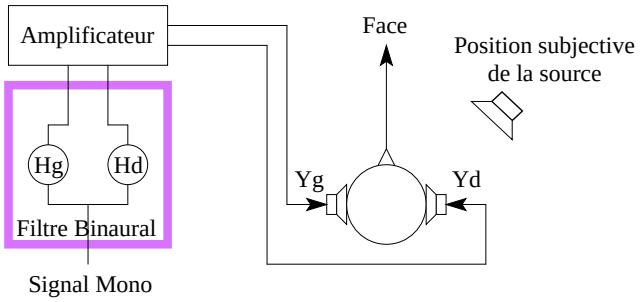

FIG. VI.8 – *Synthèse d'un signal binaural*

signal destiné à une reproduction sur haut-parleurs, c'est-à-dire rétablit artificiellement le trajet de chaque haut-parleur vers l'oreille opposée. Idéalement, l'écoute au casque d'un enregistrement stéréophonique conventionnel devrait toujours s'effectuer au travers d'un tel filtre.

La reproduction transaurale sur une paire de haut-parleurs en disposition stéréophonique conventionnelle fournit une illusion auditive très convaincante mais nécessite, pour la simulation de sources localisées latéralement ou à l'arrière, que l'auditeur soit placé précisément à la position idéale par rapport aux enceintes. Au contraire, la reproduction sur écouteurs d'un signal binaural est très convaincante pour des positions arrières ou latérales de la source, mais plus incertaine pour les positions frontales (souvent perçues à l'intérieur de la tête, à l'arrière ou en élévation). La disparité interindividuelle des HRTF et la réponse en fréquence des écouteurs sont ici des facteurs importants. Pour lever l'incertitude entre les directions correspondant à une valeur donnée du retard interaural ("cône de confusion"), le système auditif ne dispose que d'indices spectraux ou bien cognitifs (informations visuelles, mouvements de la tête).

Dans les systèmes de réalité virtuelle, de développement très récent, le spectateur porte un viseur reproduisant une image de synthèse mise à jour en temps réel en fonction des données fournies par un capteur de position et d'orientation solidaire de la tête ("headtracker"). Ces données de position peuvent être exploitées par un synthétiseur binaural afin que la localisation des événements sonores virtuels soit définie en référence à l'environnement extérieur simulé, indépendamment des mouvements de la tête du spectateur. Cette compensation des mouvements de la tête de l'auditeur, dont on peut attendre qu'elle améliore sensiblement le réalisme de la reproduction auditive, suppose que la synthèse binaurale s'effectue en temps réel et sans retard de traitement trop important.

VI.4.2 Modélisation des fonctions de transfert

Pour diminuer encore le coût de l'implémentation des filtres binauraux ou transauraux, il est nécessaire de faire appel à des techniques de modélisation paramétrique. Ces techniques introduisent nécessairement une approximation de la réponse en fréquence (en phase et en amplitude), et reposent sur la minimisation d'un critère d'erreur dont la pertinence du point de vue perceptif est délicate à vérifier. Il s'agit de trouver les coefficients de deux polynômes en z^{-1} d'ordres donnés, notés $A(z)$ et $B(z)$, minimisant la norme de l'erreur en sortie du filtre de fonction de transfert $B(z)/A(z)$:

$$\epsilon(\omega) = G(e^{j\omega}) - \frac{B(e^{j\omega})}{A(e^{j\omega})}$$

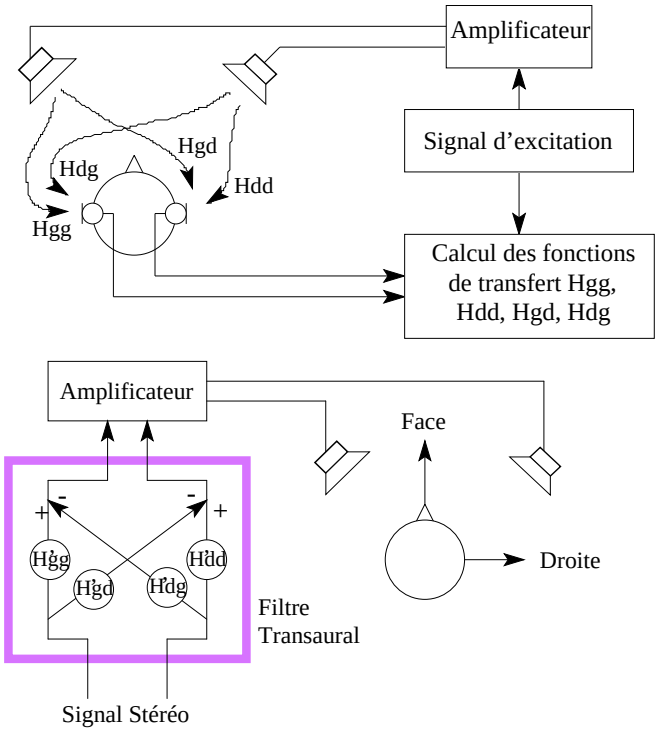

FIG. VI.9 – *Réalisation d'un système transaural adapté aux conditions d'écoute. Le dispositif de reproduction n'est pas forcément symétrique et le filtre transaural compense ("déconvolue") l'effet de l'amplificateur, des enceintes et, au besoin, des échos du lieu d'écoute*

où $G(e^{j\omega})$ représente la réponse en fréquence complexe à approximer. On utilise le plus souvent la norme quadratique, et il s'avère alors plus commode de reformuler le problème sous la forme d'une minimisation de l'erreur d'équation:

$$J = ||G(e^{j\omega})A(e^{j\omega}) - B(e^{j\omega})||$$

On trouvera dans [23] un inventaire très complet des méthodes de modélisation paramétrique.

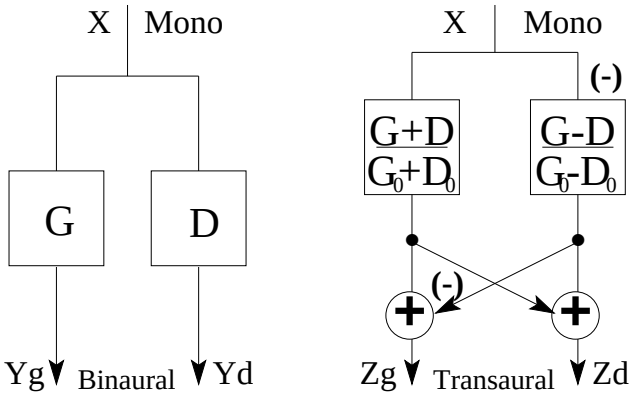

FIG. VI.10 – *Réalisation des divers filtres de conversion utiles dans le cadre des techniques binaurales, à l'aide de la structure "shuffler" proposée par Cooper et Bauck.*

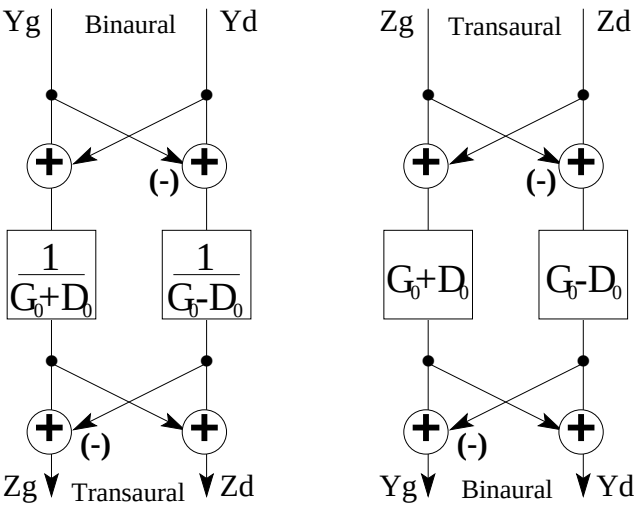

FIG. VI.11 – *Transcodage d'un signal binaural en un signal transaural*

Bibliographie

[1] V. Algazi, R. Duda, D. Thompson, and C. Avendano. The cipic hrtf database. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001. see also http://interface.cipic.ucdavis.edu/CIL_html/CIL_whatis.htm.

[2] D.R. Begault. Challenges to the successful implementation of 3-d sound. *J. Audio Eng. Soc.*, 39(11):864–870, 1991.

[3] D.R. Begault. *3D Sound for Vitural Reality and Multimedia*. AP Professional, 1994.

[4] J. Blauert. *Spatial Hearing*. M.I.T Press, Cambridge, 1983.

[5] J. Borish and J.B. Angel. An efficient algorithm for measuring the impulse response using pseudo-random noise. *J. Audio Eng. Soc.*, 31:478–488, 1983.

[6] D.H. Cooper and J.L. Bauck. Prospects for transaural recording. *J. Audio Eng. Soc.*, 37(1/2):3–19, 1989.

[7] P. Damaske. Head-related two-channel stereophony with loudspeaker reproduction. *J. Acoust. Soc. Am.*, 50(4):1109–1115, 1971.

[8] P. D'Antonio and J.H. Konnert. Complex time-response measurements using time-delay spectrometry. *J. Audio Eng. Soc.*, 37(9):674–690, 1989.

[9] S. Foster. Impulse response measurements using golay codes. *Proc. IEEE ICASSP-86*, pages 929–932, 1986.

[10] R.K. Furness. Ambisonics - an overview. *Proc. 8th A.E.S Int. Conf.*, pages 181–190, 1990.

[11] M. A. Gerzon. Ambisonics in multichannel broadcasting and video. *J. Audio Eng. Soc.*, 33(11), Nov 1985.

[12] D. Griesinger. Binaural techniques for music reproduction. *Proc. 8th A.E.S Int. Conf.*, pages 197–207, 1990.

[13] J.M. Jot. *Réverbération artificielle et spatialisation des sons*, chapter Chap 7. in traitement des signaux audio-fréquences, j. laroche, doc enst edition.

[14] J.P. Jullien, A. Gilloire, and A. Saliou. Caractérisation d'une méthode de mesure de réponse impulsionnelle en acoustique des salles. *11th Int. Conf. Acoust. Paris*, 6:217–220, 1983.

[15] S. Julstrom. A high-performance surround sound process for home video. *J. Audio Eng. Soc.*, 35(7/8), July/August 1987.

[16] D. Gardner & K. Martin. Hrtf measurements of a kemar. *jasa*, 97:3907–3908, 95. see also http://www.sound.media.mit.edu/KEMAR.html.

[17] F.J. McWilliams and N.J.A. Sloane. Pseudo-random sequences and arrays. *Proc. IEEE*, 64(12):1715–1729, Dec 1976.

[18] H. Moller. Cancellation of crosstalk in artificial head recordings reproduced through loudspeakers. *Proc. 84th AES Conv, Paris*, 1988. preprint 2610 (G-7).

[19] M.A. Poletti. Linearly swept frequency measurements, time-delay spectrometry, and the wigner distribution. *J. Audio Eng. Soc.*, 36(6):457–468, 1988.

[20] M.R. Schroeder. Digital simulation of sound transmission in reverberant spaces. *J. Acoust. Soc. Am.*, 47(2):424–431, 1970.

[21] M.R. Schroeder. Integrated impulse method for measuring sound decay without using impulses. *J. Acoust. Soc. Am.*, 66(2):497–500, 1979.

[22] P. Simon. *Le Livre Des Techniques Du Son.* Fréquences, Paris, 1988.

[23] J. O. Smith. *Techniques for Digital Filter Design and System Identification with Application to the Violin.* PhD thesis, Stanford University, Stanford, CA, Jun 1983.

[24] G. Steinke. Delta stereophony - a sound system with true direction and distance perception for large multipurpose halls. *J. Audio Eng. Soc.*, 31(7):500–511, 1983.

[25] John Vanderkooy and Stanley Lipshitz. Anomalies of wavefront reconstruction in stereo and surround-sound reproduction. *Proc. 83rd Convention of the Audio Engineering Society*, Oct. 1987.

[26] J. Vanderkooy. Another approach to time-delay spectrometry. *J. Audio Eng. Soc.*, 34:523–538, 1986.