

Guide pratique des techniques du son

Enregistrement, mixage et spatialisation, appliqués à la musique

Jean-Loup Pecquais

Table des matières

Avant-propos	9
A qui s'adresse cet ouvrage	9
Mise à jour	9
Structures	10
À propos de l'auteur	10
I Généralités	11
1 Oreille et audition	12
1.1 L'oreille externe	12
1.2 L'oreille moyenne	14
1.3 L'oreille interne	14
2 Quantifier et qualifier le son	16
2.1 Phénomène physique	16
2.1.1 Quelques définitions	16
2.1.2 Relation entre temps, distance et fréquence	19
2.2 Perception du son	21
2.2.1 Spectre, timbre et vocabulaire	21
2.2.2 Pression acoustique & niveau sonore	22
2.2.3 Positionnement dans l'espace	23
3 Acoustique des salles	27
3.1 Généralités	27
3.1.1 La réverbération	27
3.1.2 Calcul du temps de réverbération	28
3.1.3 Limite de l'équation de Sabine	29
3.1.4 L'indice de "Speech Clarity" C50	30
3.1.5 Le phénomène d'onde stationnaire	30
3.2 Premières réflexions et filtres en peignes	31
3.3 Traitement acoustique	32
3.3.1 Les types de traitements	35
3.3.2 Considération d'acoustique pour le travail de son	35

4 Notions élémentaires d'électronique	37
4.1 Les grandeurs physiques	37
4.1.1 L'intensité	37
4.1.2 La tension	37
4.1.3 L'impédance	39
4.2 Les composants électroniques	39
4.2.1 Les composants passifs	39
4.2.2 Tubes & semi-conducteurs	41
4.3 L'influence de l'impédance entre différents appareils.	42
5 Description d'une production musicale type	44
5.1 les acteurs de la réalisation d'une œuvre enregistrée	44
5.2 La préproduction	46
5.3 La production	47
5.4 La postproduction	47
II Outils et équipements	48
6 Le chemin du signal	49
7 Les microphones	51
7.1 Petit historique des microphones	51
7.2 Les types et technologies de microphones	52
7.2.1 Les microphones électrostatiques/à condensateur	53
7.2.2 Les microphones à ruban	54
7.2.3 Les microphones dynamiques	54
7.2.4 La taille des membranes	55
7.2.5 Microphones à tubes ou transistors?	55
7.3 Timbre et directivité	55
7.3.1 Capsules à pression	57
7.3.2 Capsules à gradient de pression	57
7.3.3 Et les autres directivités ?	57
7.3.4 Directivités réelles & détimbrage	58
8 Transport de signaux analogiques	60
8.1 Anatomie d'un câble	60
8.1.1 Longueur du câble, son et impédance	62
8.1.2 Les connexions asymétriques	62
8.1.3 Les connexions symétriques	62
8.2 Les fiches & connecteurs	63
8.3 Exemples pratiques	64
8.3.1 Câble jack « mono »	64

8.3.2	Câble jack « stéréo »	64
8.3.3	Câble « Y »	64
8.3.4	Les connexions d'insert	64
8.4	Routage des signaux	65
9	Les préamplificateurs	66
9.1	Informations techniques des préamplificateurs	66
9.2	Critères de choix d'un préamplificateur	68
9.3	Les technologies de préampli	68
9.4	Les réglages d'un préampli	68
10	La conversion analogique numérique	70
10.1	La nécessité de la conversion analogique numérique	70
10.2	Théorie de l'échantillonnage	70
10.2.1	D'un signal continu vers un signal échantillonné	70
10.2.2	La fréquence d'échantillonnage	71
10.2.3	La résolution de quantification	71
10.3	Quelle influence sur le signal ?	72
10.4	La conversion sigma-delta	73
11	Transports de signaux numériques	74
11.1	Les liaisons point à point	74
11.1.1	AES3 ; AES/EBU	75
11.1.2	ADAT	75
11.1.3	MADI	75
11.2	Les réseaux audio numériques	76
12	Introduction à l'informatique musicale	78
12.1	Fonctionnement d'un ordinateur	79
12.2	Les systèmes d'exploitation	79
12.2.1	Linux	79
12.2.2	Apple MacOS	79
12.2.3	Microsoft Windows	79
12.3	Les pilotes audio	79
12.3.1	ASIO	79
12.3.2	CoreAudio	79
12.3.3	ALSA	79
12.3.4	Jack Audio	79
12.4	Les Stations de Travail Audio-Numérique (DAW)	79
12.4.1	Le moteur audio	79
12.4.2	Les fonctionnalités	79
12.5	Les protocoles de transmission d'informations	79
12.5.1	MIDI	79

12.5.2 OSC	79
13 Enceintes et amplificateurs	80
13.1 Anatomie d'un haut-parleur	80
13.2 Amplification et impédance	81
13.3 Puissance et sensibilité	82
13.4 Conseils pratiques	83
13.4.1 Choisir une paire d'écoutes	83
13.4.2 Placer correctement son écoute	83
13.5 L'écoute au casque	84
13.5.1 Casque fermé ou casque ouvert ?	84
III Méthodologie de prise de son	86
14 Ecoute critique : première partie	87
15 Mono et multi-microphonie	88
16 La prise de son au couple	89
16.1 Généralités sur les mécanismes de la localisation du son par l'oreille humaine	89
16.1.1 La localisation par différence de temps	89
16.1.2 La localisation par différence d'intensité	90
16.1.3 Prévalence fréquentielle de ces deux phénomènes	90
16.2 Principes de la prise de son au couple	90
16.2.1 Comment choisir un angle de prise de son.	91
16.2.2 Comment réaliser un angle de prise de son.	91
16.2.3 Privilégier le Δi ou le Δt ?	92
16.3 Les topologies classiques de prise de son au couple	92
16.3.1 Le couple Blumlein / XY	92
16.3.2 Le couple MS	93
16.3.3 Le couple ORTF	93
16.3.4 Les couples AB	93
16.4 Compléter une prise de son au couple par des appooints	95
17 Déphasage et remise en phase	96
17.1 Les effets sonores de déphasage	96
17.2 Approche mathématique	96
17.3 Les sources de déphasage	96
18 En pratiques	97
18.1 Le confort du musicien	97
18.2 Le choix de l'instrument	98
18.3 Le choix de l'acoustique	98

18.4 Placer et choisir son microphone	98
18.4.1 Le rapport à la distance du microphone	99
18.4.2 Quand choisir une prise de son stéréophonique	99
18.4.3 Quand choisir la multi-microphonie	100
18.5 Le choix du préamplificateur	100
18.5.1 L'influence du préampli sur la « couleur » du son	101
IV Outils de mixage	102
19 Anatomie d'une console de mixage	103
20 Manipulation de la phase	104
21 Egalisation et égaliseurs	105
22 Les compresseurs	106
23 Les limiteurs	107
24 Les autres outils de gestion de la dynamique	108
25 Mise en espace et réverbération	109
26 Saturation et distortion du signal	110
27 Effets & modulations	111
V Méthodologie de mixage	112
28 Ecoute critique : deuxième partie	113
29 Approche générale	114
30 Etude des cas d'école	115
31 Pièges et erreurs à éviter	116
VI Introduction à la spatialisation sonore	117
32 Introduction	118
33 Historique de la spatialisation	120
33.1 De la monophonie à la stéréophonie : un besoin d'espace	120

33.2 La quadriphonie : une entreprise infructueuse	124
33.3 La grande aventure du son spatialisé au cinéma	125
33.3.1 Fantasia et le Fantasound	126
33.3.2 Le Dolby Stereo et le son optique matricé	126
33.3.3 Le passage au son numérique	128
33.3.4 Passage au mixage orienté objet et le retour en force de Dolby	130
33.4 Le binaural : la spatialisation sonore pour tous	131
33.5 Les pionniers de la musique spatialisée	134
33.6 L'ambisonie : décorréler l'espace de production et l'espace d'écoute	138
34 Les espaces haut-parlants	141
34.1 Catégorisation des espaces haut-parlants	141
34.2 Les systèmes de diffusions	142
34.3 Une simple analogie image-son	144
34.4 Études des systèmes normés	145
34.4.1 Les systèmes frontaux	145
34.4.2 Les systèmes englobants à une dimension	146
34.4.3 Les systèmes englobants à deux dimensions	151
35 Le binaural	153
35.1 La prise de son binaurale	153
35.2 La synthèse binaurale	153
35.3 Le transaural	153
36 Orienté canal	154
36.1 L'approche perceptive	154
36.1.1 le VBAP	154
36.1.2 Le VBIP	154
36.1.3 Le LBAP	154
36.1.4 Autres variantes	154
36.2 L'approche matricielle	154
36.2.1 Le DBAP	154
36.2.2 Le KNN	154
37 L'ambisonique	155
37.1 L'ambisonie du premier ordre (FOA)	156
37.1.1 Captation du champ sonore	156
37.1.2 Synthèse du champ sonore	159
37.1.3 Restitution du champ sonore	159
37.1.4 Décodage pour arrangement de haut-parleurs irréguliers	161
37.2 L'ambisonie d'ordre plus élevé (HOA)	164
37.2.1 Les outils de captations d'ordre plus élevé	165
37.2.2 Les différents formats ambisoniques	166

37.3 Avantages de l'ambisonie	167
38 Mixage orienté objet	168
38.1 L'Audio Definition Model	169
38.1.1 Présentation	169
38.1.2 Description des métadonnées	169
38.1.3 Les métadonnées de spatialisation	171
38.1.4 Inscription des métadonnées dans un fichier audio	172
38.2 Les profils ADM	173
38.3 L'ADM OSC	174
38.4 L'étape de rendue	174
38.4.1 Description de la loi de panoramique	174
38.4.2 Largeur et décorrélation	175
38.4.3 Décodage ambisonique	176
39 La synthèse de fronts d'ondes (WFS)	177
VII Les outils de spatialisation sonore	178
40 Choisir une station de travail	179
41 Panner binaural	180
42 Panner orienté canaux	181
43 Panner ambisonique	182
44 Gestion des effets	183
45 Le mixage orienté objet	184
45.1 MPEG-H et ADM	184
45.2 Dolby Atmos	184
45.3 IRCAM Spat	184
45.4 FLUX:: Spat Revolution	184

Avant-propos

Ce livre est né de la nécessité d'un support de cours pour la formation professionnelle "Technique de Prise de Son", dispensée par l'auteur. Il intègre donc l'ensemble des notions abordées, expliquées en détail, ainsi que des exemples sonores.

Ce livre est écrit dans la philosophie de l'Open Source. L'intégralité de son contenu est donc disponible gratuitement. Son code source est accessible dans un dépôt GitHub. Ainsi, il est possible à tout à chacun de reporter les éventuelles erreurs ou de proposer des modifications. La grande majorité des outils utilisés pour sa rédaction et la création du contenu sont open source : R & Rmarkdown, Python, FAUST et draw.io.

A qui s'adresse cet ouvrage

Ce livre s'adresse à toutes personnes désireuses d'en apprendre plus sur le son ainsi que sur les métiers de preneur de son et de mixeur. Ainsi, il fait état des principes physiques nécessaires à la bonne appréhension des techniques de travail des métiers susnommés, avec le souci de les rendre accessibles à toutes et tous.

Il pourra donc servir aux musiciens, aux étudiants, et pourquoi pas, à certains professionnels des métiers du son et du divertissement en général.

Mise à jour

La distribution numérique de ce livre permet une mise à jour régulière de son contenu. Cela implique deux choses :

- Certaines sections peuvent être incomplètes, et seront complétées plus tard
- C'est une bonne idée de revenir consulter ce site régulièrement

Pour l'instant, ce livre n'inclut pas encore d'exemples sonores, cela est en cours de création.

Structures

Dans un premier temps, le livre aborde des principes généraux, aussi bien sur la physique que sur l'environnement de production de la musique enregistrée. Est ensuite abordé l'ensemble de la chaîne audio, en y explicitant le rôle et le fonctionnement de chacun de ses composants. L'objectif est de fournir une base technique objective au preneur de son.

Dans un second temps, le livre détaille un ensemble de techniques de prise de son et de mixage, insistant particulièrement sur les mécanismes généraux de la prise et sur l'écoute critique.

La partie dédiée à la pratique du mixage son n'est pas encore disponible en ligne.

À propos de l'auteur

Jean-Loup Pecquais est formateur et consultant dans le monde de l'audio professionnel (FLUX:: Immersive, Whiti Audio, Arkalya). Il est plus particulièrement spécialisé dans les techniques de mixage sonore immersives. Il est diplômé de l'ENS Louis-Lumière en 2019.

partie I

Généralités

1 Oreille et audition

L'oreille est l'organe qui, chez l'Homme, permet d'entendre les sons environnants. Comme nous le détaillerons dans le chapitre suivant, le son est une vibration de l'air. Afin que celle-ci puisse être audible, l'oreille doit d'abord transformer cette vibration de l'air (caractérisée par une variation de pression) en signal électrique finalement transmis à notre cerveau.

On distingue ainsi entendre, perception passive du son, et écouter, qui suppose une participation active de l'auditeur.

Le sens lié à cette faculté de perception du son se nomme **l'ouïe**.

i Note

Nous ne discuterons pas ici du rôle de l'oreille dans l'équilibre.

Cet organe se décompose classiquement en trois parties, l'oreille externe, l'oreille moyenne et l'oreille interne.

1.1 L'oreille externe

Comme son nom l'indique, l'oreille externe est la partie visible de notre organe de l'audition. Elle est caractérisée par ce pavillon à la forme si particulière. Le rôle de ce **pavillon auriculaire** permet de "collecter" le son. Sa forme applique une emprunte fréquentielle sur le son, donnant une information principalement d'ordre spatial. Ce pavillon joue donc un rôle important dans notre capacité à localiser les sons, avec une précision maximale lorsque l'évènement sonore se trouve devant nous.

i Note

Comme nous le verrons plus tard, la forme de notre tête et même de notre torse a également un rôle important dans notre capacité à localiser des évènements sonores dans l'espace.

L'oreille externe comporte aussi le canal auditif, avec à son terme le tympan. Ce dernier est une simple membrane, vibrant de façon homologue à l'onde sonore lui provenant. L'ensemble



Figure 1.1: Schéma de l'oreille

de ce système de “captation” possède un rendement maximal sur les fréquences autour de 3 kHz (fréquences aiguës).

1.2 L'oreille moyenne

Derrière le tympan se trouve l'oreille moyenne, et plus particulièrement une collection de trois os, nommés “marteau”, “enclume” et “étrier”. Le **marteau** est directement relié au tympan et transmet les vibrations du tympan à l’“**enclume**” puis à l’“**étrier**”. Ce dernier permet la transmission du son à l'oreille interne par la fenêtre ovale.

Nous y trouvons aussi la trompe d'eustache, reliée à la gorge, réalisant ainsi un équilibrage de la pression entre celle de l'oreille interne et celle nous environnant.

Le rôle principal de cette oreille moyenne est de réaliser une adaptation d'impédance acoustique entre l'air et le milieu liquide, dans lequel baigne notre oreille interne. En d'autres termes, elle permet un transfert efficace de l'énergie sonore vers l'oreille interne. Sans cette étape, notre sens de l'audition serait grandement amoindri.

Évoquons également le **muscle stapédiens**. Il permet, lorsqu'il est contracté, de limiter l'amplitude de mouvement des trois os évoqués précédemment. Il agit alors comme une protection lorsque nous sommes confrontés à de forts niveaux sonores. Si cette rigidification n'avait pas lieu, notre tympan serait beaucoup plus facilement arraché par des stimulus sonores importants.

Cependant, ce muscle stapédiens échappe à notre contrôle cognitif et est donc mis en action par réflexe. Son temps de mise en action est d'au moins 40 millisecondes (ms) après l'émission d'un son supérieur à 90-100 décibel (dB). Pire, la protection maximale n'est atteinte que 150 ms plus tard. Cela signifie que, si nous sommes exposés à des déflagrations sonores très importantes (armes à feu, explosions, etc.), notre **réflexe stapédiens** n'aura pas le temps de s'activer, et notre système auditif sera sévèrement endommagé.

1.3 L'oreille interne

Nous avons précédemment évoqué l'attachement de l'étrier à la fenêtre ovale. Cette dernière fait elle-même partie de la cochlée, où commence l'oreille interne. Cette cochlée prend une forme de coquille d'escargot et renferme la membrane basilaire et l'organe de Corti, récepteur de l'audition. Son ensemble est immergé dans différentes lymphes (milieu liquide).

La membrane basilaire court le long de la cochlée et près de 30 000 récepteurs ressemblant à de petits cheveux la parcourent. Cette membrane vibre lorsque la lymphe change de pression. Sur l'ensemble de sa longueur est répartie notre sensibilité aux différentes fréquences. La partie proche de l'oreille moyenne est plus sensible aux aiguës, alors que la zone en bout de son

enroulement est plus sensible aux graves. On a donc une correspondance entre fréquence et emplacement sur la membrane basilaire. Lorsque les récepteurs vibrent suffisamment fortement, un signal électrique est émis par l'organe de Cortie dans le système nerveux.

La diminution de sensibilité, voire la perte de certaines fréquences audible dans l'audition, est associée à la mort de ces récepteurs peuplant la membrane basilaire. Ce phénomène est irréversible et peut aboutir à l'apparition d'acouphènes.

Avertissement

L'oreille est un organe fragile, dont l'endommagement est irréversible. Il convient donc d'en prendre soin en limitant :

- son exposition à de forts niveaux sonores
- son exposition au bruit

Dans les cas où l'exposition est contrainte (voire souhaité en concert par exemple), il est vivement conseillé de porter des protections auditives (bouchons d'oreilles).

Il est également bon de rappeler que, dans la législation française, les niveaux sonores de diffusions dans des lieux de spectacles sont normés et ne doivent pas dépasser certains seuils. L'ingénieur du son affecté au mixage est donc **responsable** du respect de ces normes.

En studio, il n'existe pas de norme de niveau de diffusions dans les casques ou les écoutes, mais l'ingénieur du son reste tout de même responsable de ce que les musiciens entendent, et donc de toute perte d'audition de l'un d'eux lors d'une session de travail.

2 Quantifier et qualifier le son

Le son peut s'appréhender de plusieurs façons différentes. Particulièrement, sa description physique et psychoacoustique est très précieuse pour tous les praticiens du son. Il convient donc, afin de pouvoir proposer un dispositif cohérent de prise de son, de comprendre la physique élémentaire du son ainsi que d'être capable de le décrire efficacement.

2.1 Phénomène physique

2.1.1 Quelques définitions

Le son est une vibration mécanique d'un fluide. Dans le cadre de ce cours, nous ne considérerons que l'air comme médium de propagation. Cette onde cause une variation de la pression dans l'espace. Nous, les êtres humains, le percevons grâce à notre ouïe. Il s'agit donc, par définition, d'un phénomène ondulatoire et peut être caractérisé par un nombre d'oscillations par seconde, aussi appelé fréquence. On estime que notre espèce est sensible aux fréquences allant de 20 Hz (très grave) jusqu'à 20 000 Hz (très aigu).

On parlera d'**événement sonore** pour parler généralement de phénomènes physiques produisant une onde sonore.

Les sons composés d'une seule fréquence se nomment **sons purs**. Cependant, de tels signaux n'existent pas dans la nature, et sont souvent utilisés afin de réaliser des mesures ou des tests psychoacoustiques.

Dans notre environnement, les sons sont donc composés de plusieurs fréquences. La fréquence la plus grave d'un son est sa **fréquence fondamentale**. Les autres sont alors appelées **partiels**. Si ces partielles ont pour fréquence un multiple de la fréquence fondamentale, alors on les nomme **harmoniques**.

Plus généralement, on admettra que la composition fréquentielle, ou spectrale, de tout son peut être décomposée par une somme de sinusoïde. L'outil permettant de passer de la représentation temporelle d'un signal à sa représentation fréquentiel s'appelle la **transformée de Fourier**.



Figure 2.1: Onde sinusoïdale et visualisation de son spectre



Figure 2.2: Signal carré et visualisation de son spectre

La fréquence fondamentale donne la **hauteur** du son (sa note en musique par exemple). Les partiels enrichissent cette fréquence fondamentale et créent le **timbre** d'un son. C'est en partie grâce au timbre que l'on peut reconnaître différents instruments de musiques jouant la même note.

Un son se caractérise également par l'évolution de son amplitude au cours du temps. On parle alors de son **enveloppe**. Un modèle courant d'enveloppe est l'ADSR : *Attack*, *Decay*, *Sustain*, *Release*, soit *Attaque*, *Décroissance*, *Maintien* et *Relâchement*.

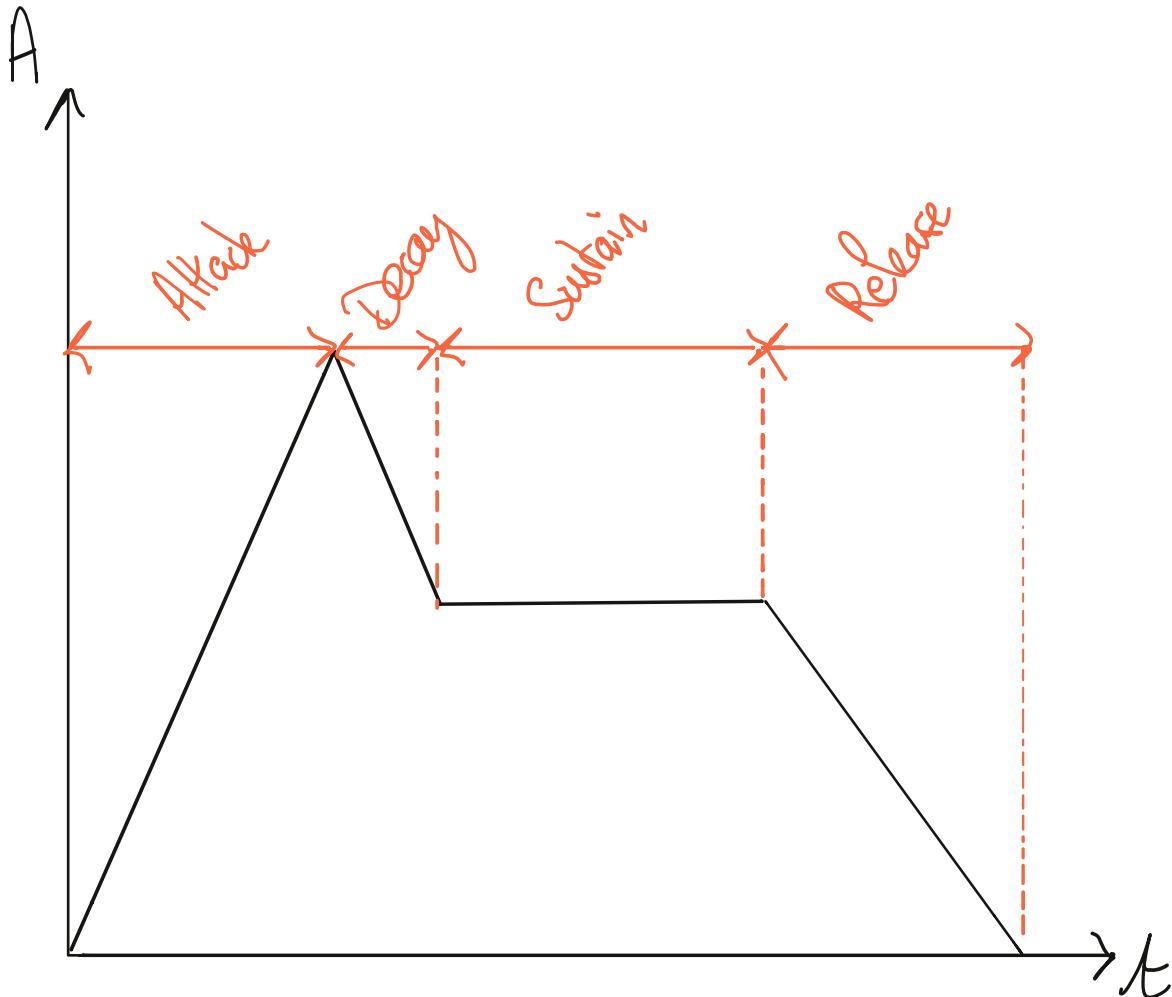


Figure 2.3: Exemple d'enveloppe ADSR

Lorsque son temps est très bref, l'ensemble *attaque* et *décroissance* forme les **transitoires**. Cette partie du signal est responsable de la sensation percussive du son.

2.1.2 Relation entre temps, distance et fréquence

Il est important de garder à l'esprit que les notions de temps, de fréquence et de distance sont étroitement liées. Nous avons vu ci-dessus que tous les sons peuvent être décrits par une somme de sinusoïde. Leur fréquence la plus grave, dite fondamentale, permet de définir la **période**. La période est le temps que met un signal à répéter son motif oscillatoire (voir schémas 3.1 et 3.2). Le lien mathématique entre fréquence et période est très simple, car l'un est l'inverse de l'autre :

$$f = \frac{1}{T}$$

Si nous étudions les fréquences extrêmes, audibles par notre ouïe, nous trouvons que pour $f_{min} = 20 \text{ Hz}$, sa période $T_{f_{min}} = 50 \text{ ms}$. Pour $f_{max} = 20\,000 \text{ Hz}$, $T_{f_{max}} = 0.5 \text{ ms}$.

Une onde sonore est également caractérisée par sa **célérité**. Celle-ci est constante dans un milieu donné. Dans l'air, à une température de 15°C et au niveau de la mer, sa célérité c est de 340 m.s^{-1} . On admettra cette valeur pour réaliser l'ensemble de nos différents calculs.

Comme son unité l'indique, la célérité du son est homogène à une distance divisée par un temps, soit :

$$c = \frac{d}{t}$$

Suivant cette formule, nous pouvons alors calculer la **longueur d'onde** correspondant à une fréquence. La longueur d'onde se note λ .

$$\lambda = cT \iff \lambda = \frac{c}{f}$$

Si nous étudions à nouveau les bornes minimale et maximale de notre audition, nous trouvons que $\lambda_{f_{min}} = 17 \text{ m}$ et $\lambda_{f_{max}} = 17 \text{ mm}$.

Nous pouvons également calculer le temps de propagation du son. En pratique, nous serons souvent intéressés par le temps de propagation séparant deux points dans l'espace (par exemple, le temps séparant deux microphones par rapport à un instrument).

$$t = \frac{d_2 - d_1}{c}$$



Figure 2.4: Distance entre deux microphones.

2.2 Perception du son

Nous avons abordé quelques notions de physique permettant de mieux caractériser le phénomène sonore. Comme indiqué au début de ce chapitre, le son peut également être discuté sous l'angle de notre ouïe, et donc, de notre perception. Cette branche de la science se nomme la psychoacoustique et cherche à étudier la façon dont nous percevons le son.

Notre corps, et a fortiori notre cerveau, sont des machines extrêmement complexes. Nous sommes équipés d'une multitude de capteurs permettant de sentir le contact d'une matière, des odeurs, d'entendre, de goûter, de voir, de positionner nos membres dans l'espace, de ressentir la douleur, etc. Pris indépendamment, chacun de ces sens est déjà un phénomène complexe à décrire, mais il existe en plus une grande interdépendance entre ceux-ci. Par exemple, l'interdépendance entre la vision et l'audition est à l'origine d'un certain nombre de mécanismes biaisant notre écoute.

Nous nous bornerons au fil de ce cours à quelques notions liées à l'ouïe et à son interdépendance à d'autre sens quand cela sera pertinent.

2.2.1 Spectre, timbre et vocabulaire

D'un point de vue perceptif, le spectre d'un évènement sonore est facilement remarquable. Il est, par contre, beaucoup plus difficile à qualifier. Il n'est pas rare de rencontrer les adjectifs "chaud", "brillant", "rond", "aéré", "ouvert", "sombre", voir d'autres encore plus ésotérique, pour tenter de communiquer la sensation ressentie à l'écoute de tel ou tel son.

Cette difficulté liée à l'absence de vocabulaire commun quant à la qualification le son emmène systématiquement la redéfinition de ce vocable en fonction de son interlocuteur. En effet, le mot "rond" ne signifiera pas forcément la même chose selon à qui on s'adresse. Une stratégie possible consiste à questionner son interlocuteur sur l'utilisation de ses adjectifs tout en cherchant à y associer des exemples sonores.

Nous pouvons tout de même nous essayer à cet exercice pour nous permettre d'avoir un vocabulaire commun au fil de ce cours. Vous aurez sans doute compris qu'il n'y aura, dans les termes employés, aucun critère absolu.

Proposition d'association entre bandes de fréquences et sensation.

- **20 Hz — 80 Hz** : Subharmonique, sensation tripale
- **80 Hz — 160 Hz** : Grave, sensation d'assise
- **160 Hz — 380 Hz** : bas-médium, sensation de « chaleur », voir « boueux »
- **380 Hz — 1400 Hz** : Medium, sensation de « boîte » quand trop présent, sonne « creux » quand trop absent
- **1400 Hz — 3200 Hz** : Haut-medium : zone de sensibilité maximale de l'oreille.
- **3200 Hz — 8000 Hz** : Aigu, apporte de la précision voir de l'agressivité



Figure 2.5: Proposition de découpage du spectre

- **8000 Hz — 20 000 Hz** : Air, apporte une sensation d'ouverture voir de finesse

Il est intéressant de former son oreille à reconnaître une plage de fréquence, ainsi que d'y associer son propre vocabulaire et une sensation. Les appellations proposées ci-dessus ne sont à prendre que comme guides et n'ont pas valeur de référence. Cela favorise une écoute critique et analytique.

! Important

Aussi, les fréquences graves ont un effet masquant sur les fréquences plus aiguës. Ce phénomène est dû au fonctionnement de notre oreille, et plus particulièrement de la cochlée.

2.2.2 Pression acoustique & niveau sonore

Nous l'avons abordé plus haut, lorsqu'une onde sonore se déplace dans l'air, on constate la variation de la pression atmosphérique en ce point. Dès lors, il est facile de corrélérer l'amplitude de la variation de la pression avec le niveau sonore entendu (ou mesuré).

L'unité du système international de la pression est le **pascal (Pa)**. Or, il est très rare de parler de la pression acoustique en pascal, car la variation de cette pression exprimée en pascal ne correspond pas à ce que nous percevons. En d'autres termes, si la pression acoustique exprimée en pascal double, nous ne percevons pas un son deux fois plus fort.

Notre oreille fonctionne de façon logarithmique, et non linéairement, face à une variation de pression acoustique. C'est pour cela que l'on parle généralement de **niveau de pression acoustique**, où **SPL** (pour Sound Pressure Level en anglais), qui s'exprimera en **décibel**. La relation entre la variation de pression et le niveau de pression acoustique se fait grâce à la relation :

$$L_p = 20 \log_{10} \left(\frac{p_{eff}}{p_{ref}} \right) \quad p_{ref} = 20\mu Pa$$

i Note

Si la pression acoustique double, on observe une augmentation du niveau sonore de 6 dB SPL. Lorsqu'on ressent un doublement du niveau sonore, on observe une augmentation de 20 dB.

La pression acoustique est divisée par deux à chaque doublement de distance.

La question se complexifie lorsque l'on rajoute la dimension fréquentielle à la question de la perception du niveau sonore. En effet, nous percevons des niveaux sonores différents pour différentes fréquences pourtant émises au même niveau de pression acoustique. Pour inclure cette dépendance fréquentielle, nous avons mis en place une unité de mesure : la **sonie** ou **bruyance** (**loudness** en anglais). Il est donc possible ensuite de définir des courbes d'**isosonie**, c'est-à-dire des courbes indiquant un niveau sonore de perception égale en fonction de la fréquence et du niveau de pression acoustique.

Que conclure de cet abaque ?

- Notre oreille ne perçoit pas les fréquences de manière égale.
- Notre zone de sensibilité maximale se situe dans l'aigu (3k-4k Hz).
- Notre perception d'un matériau sonore en fonction du niveau auquel nous l'écoutons !

2.2.3 Positionnement dans l'espace

Notre système auditif nous permet de situer l'émission d'un son dans l'espace. Cette capacité de localisation repose sur un ensemble de facteurs étroitement liés entre eux.

On qualifie notre écoute de **binaurale**, littéralement, écouter avec deux oreilles. La présence de deux "capteurs de pression" (oserait-on parler de microphones ?) sur les faces latérales de notre crâne et un premier élément expliquant notre capacité de localisation du son.

En effet, l'espacement de nos oreilles (en moyenne 15 cm), créer un décalage temporel entre nos deux canaux d'écoutes. Ce léger retard entendu d'un côté ou de l'autre nous permettra de placer un son plutôt à gauche ou plutôt à notre droite. On appelle cet écart de temps **différence de temps interaural**, ou **ITD** (interaural time difference en anglais) et se note Δt .

On pourrait d'ailleurs, grâce aux formules de ce début de chapitre, calculer le retard maximal moyen entre nos deux oreilles.

$$\Delta t_{max} = \frac{d}{c} = \frac{0,15}{340} = 0.4 ms$$

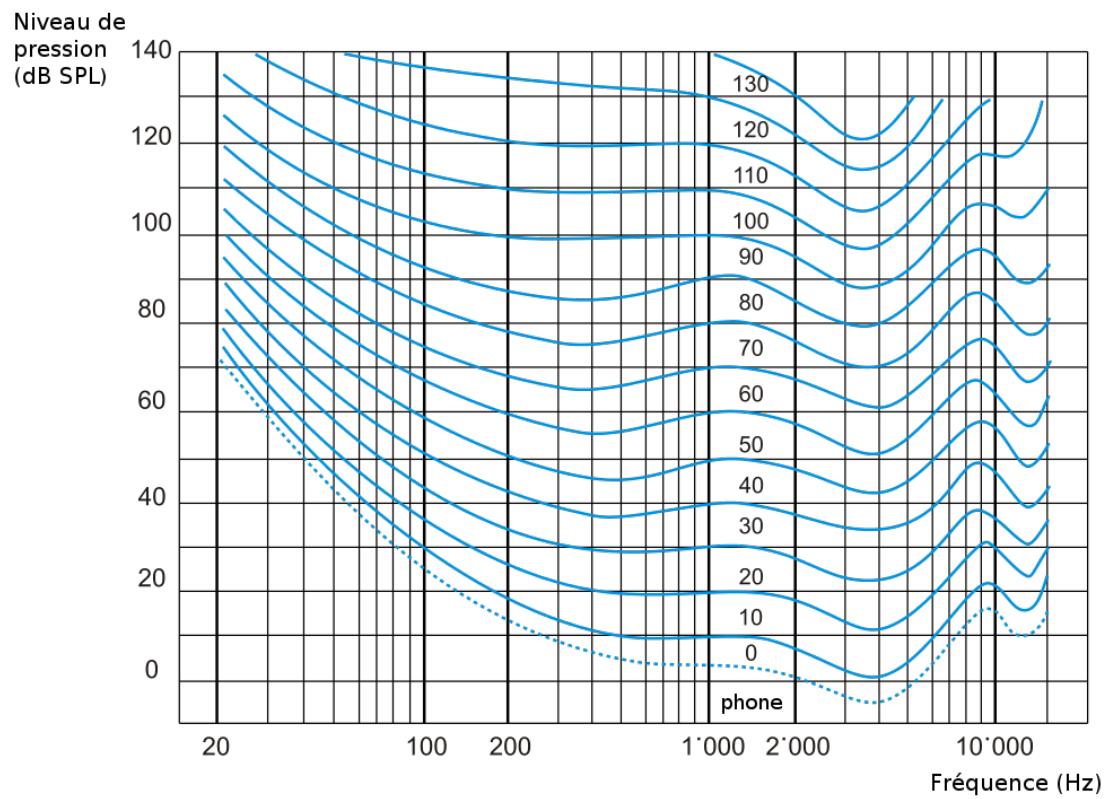


Figure 2.6: Courbes d'isosonie, aussi dites de Fletcher-Munson

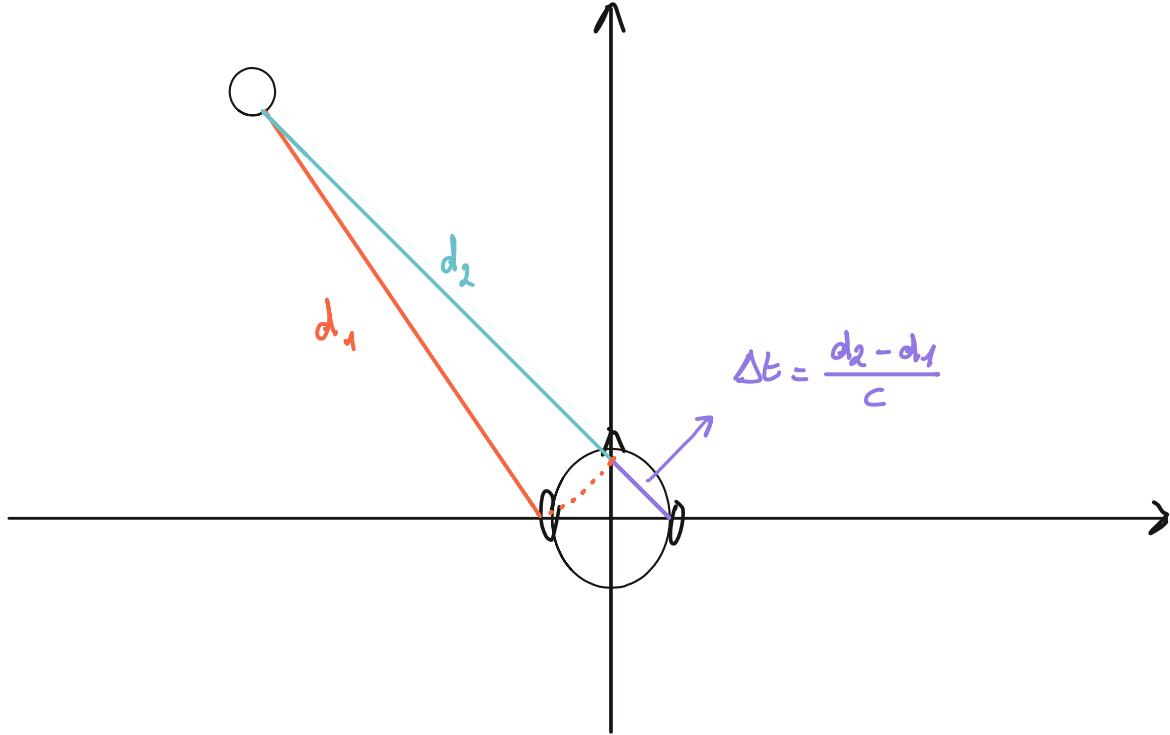


Figure 2.7: Illustration de l'ITD

Si nos oreilles sont espacées de quelques centimètres, notre tête les séparant représente un obstacle acoustique non négligeable. De plus, les pavillons des oreilles imposent également une certaine directivité à notre écoute. En première approximation, on pourra donc considérer que l'ensemble formé par la tête et les pavillons implique une atténuation linéaire des ondes sonores, elle-même fonction de l'angle d'incidence. On appelle cette différence de niveau **différence d'intensité interaural**, ou **ILD** (interaural level difference) et se note Δi . On considère que si la différence de niveau de pression acoustique entre les deux oreilles est supérieure à 20 dB, on entendra l'évènement sonore complètement latéralisé.

! Important

L'ombre acoustique que représentent la tête et le pavillon n'est en réalité pas du tout linéaire en fréquence. La modification du timbre induite par ce système n'est pas perçue par notre cerveau comme une information de couleur, mais bien comme une information de spatialisation. Ainsi, selon l'angle d'incidence de l'évènement sonore, son spectre sera filtré d'une certaine manière qui permettra à notre cerveau de le positionner dans l'espace. La réponse en fréquence d'une tête se nomme **HRTF (Head Related Transfer Function)**.

Enfin, nous sommes également capables de déterminer la distance d'un évènement sonore. La plupart des paramètres permettant d'évaluer cette distance sont relatifs. Cela signifie que l'évènement doit être comparé à un autre pour pouvoir le repositionner dans l'espace. On pourra alors comparer :

- Leurs niveaux sonores : un évènement sonore plus fort paraît plus proche
- Leurs timbres : l'absorption de l'air aura pour effet de diminuer les fréquences aiguës
- La sensation de réverbération associée : plus le signal de l'évènement sonore semblera solliciter la réponse acoustique du lieu, plus celui-ci semblera fort.
- Le temps d'arrivée des premières réflexions : le son direct d'un évènement sonore lointain arrivera quasi simultanément avec ses premières réflexions. Le son direct d'un évènement sonore proche arrivera avant ses premières réflexions.

Le chapitre suivant traitera des notions d'acoustique élémentaire ainsi que de la réverbération.

3 Acoustique des salles

Tout environnement, sollicité par un évènement sonore, produit une réponse acoustique. Cette réponse acoustique est appelée réverbération. Elle est caractéristique d'un lieu et peut, dans certains cas, être une allié précieuse dans notre travail. Dans d'autres, elle est source de problèmes et complexifie grandement notre travail d'écoute analytique.

3.1 Généralités

3.1.1 La réverbération

Afin d'étudier l'acoustique d'une salle, on procède à la mesure de sa réponse impulsionnelle. Pour se faire, on émet dans le lieu à mesurer un signal audio impulsionnel (clappement de main, explosion d'un ballon, émission d'une impulsion de Dirac grâce à un haut-parleur), et l'on enregistre le résultat à l'aide d'un microphone de mesure.

La réponse impulsionnelle d'une salle est généralement décrite en deux temps : le temps des premières réflexions et le temps du champ diffus.

Les premières réflexions sont les premiers rebonds d'une onde sonore sur les parois d'une salle et sont caractéristiques de la signature acoustique du lieu. Ces rebonds reviennent à l'auditeur avec un certain temps. Ce retard se nomme souvent « pré-délai » dans les moteurs de réverbération artificiels. Ce pré-délai est fonction de deux paramètres :

- la taille de la pièce ; plus la pièce est petite, plus les premières réflexions reviendront à l'auditeur rapidement.
- les positions de la source sonore et de l'auditeur ; plus l'auditeur est proche de la source, plus les premières réflexions arriveront après le son direct, plus l'auditeur est loin de la source, plus les premières réflexions arriveront en même temps que le son direct.

Lorsque les premières réflexions elles-mêmes auront rebondi plusieurs fois sur les parois du lieu, le phénomène d'écho des premières réflexions va se muer en champs diffus, par nature plus dense. La longueur du champ diffus se mesure grâce au RT60. Cette méthode de mesure propose de regarder le temps que met la réverbération à perdre 60 dB. Ce temps permettra ensuite de donner une longueur de réverbération.

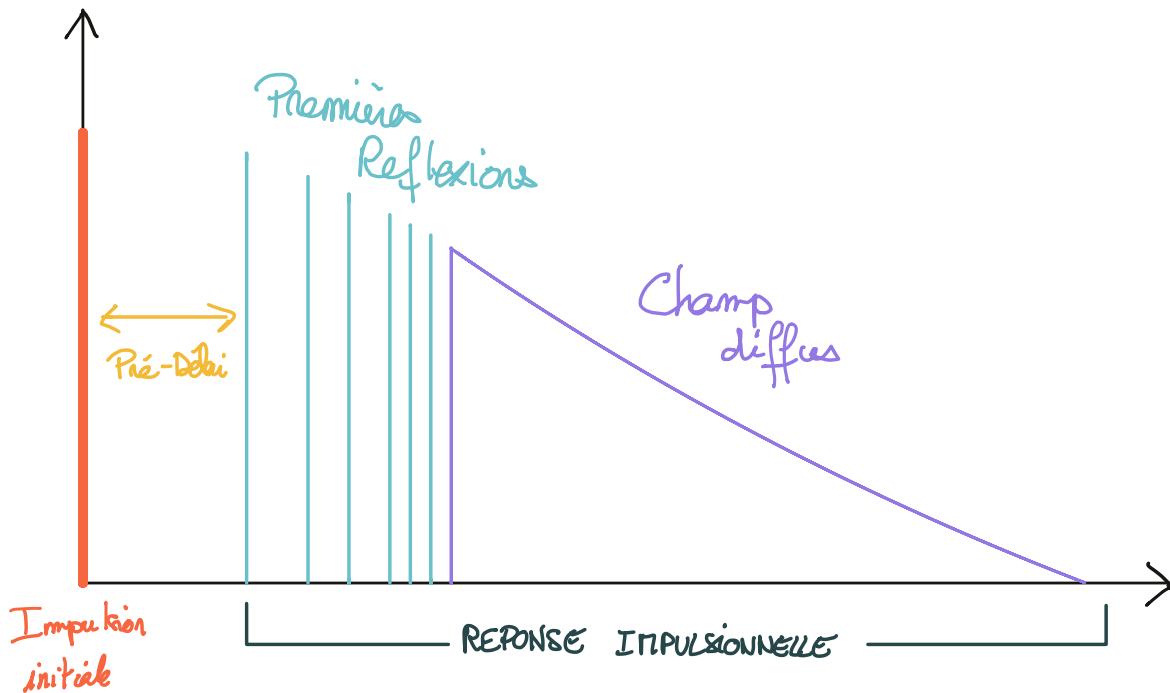


Figure 3.1: Schéma d'une réponse impulsionnelle de réverbération.

3.1.2 Calcul du temps de réverbération

L'équation de Sabine permet de calculer le temps de réverbération d'une salle à partir de son volume et du coefficient d'absorption de ses matériaux.

$$RT_{60} = 0.1611 \times \frac{V}{\sum_{i=0}^k S_i \cdot \alpha_i + \mu V}$$

V s'exprime en m^3 et S en m^2 . α est le coefficient d'absorption du matériau, en sabins. Ce coefficient est compris entre 0 et 1, plus il est important plus le matériau est absorbant. μ est le coefficient d'absorption de l'air (m^{-1}).

Astuce

Pour des fréquences inférieures à 8 kHz, on peut considérer le terme μV comme négligeable

En guise d'exemple sur l'utilisation de la formule ci-dessus, prenons le cas d'une pièce de $25 m^2$ ($5 m$ par $5 m$) et de $2.40 m$ de hauteur. Nous considérons que le sol est en parquet et les murs en plâtre. Nous avons donc $25 m^2$ de parquet et $4 \times (5 \times 2.4) = 48 m^2$. On trouve sur les sites

de fabricant de matériaux que le plâtre peint a un coefficient d'absorption de 0.05 sabins et le bois un coefficient de 0.15 sabins. Notre calcul final.

$$RT_{60} = 0.1611 \times \frac{25 \times 2.4}{25 \times 0.15 + 48 \times 0.05} \approx 1.57 \text{ s}$$

On peut dès lors calculer la **distance critique**, distance à partir de laquelle on entendra autant un évènement sonore que la réponse acoustique de la salle à son stimulus.

$$d_c \approx 0.057 \times \sqrt{\frac{V}{RT_{60}}}$$

Dans notre exemple $d_c \approx 0.35 \text{ m}$.

Il est souvent considéré que la taille de la pièce joue un rôle déterminant sur la longueur de réverbération. L'équation de Sabine indique bien que le coefficient d'absorption des matériaux y joue un rôle beaucoup plus important. Le modèle de réverbération de l'IRCAM va jusqu'à complètement décorrérer la taille de la pièce simulée du temps de réverbération. Au final, la taille de l'espace joue davantage sur la structure temporelle des échos, et donc, principalement sur les premières réflexions.

3.1.3 Limite de l'équation de Sabine

Il convient d'observer plusieurs réserves quant à l'utilisation de l'équation de Sabine. Premièrement, elle ne tient pas compte de l'aspect fréquentiel lié à l'absorption des matériaux. En effet, le temps de réverbération des graves est presque toujours plus long que celui des aigus. Afin de contourner ce problème, on pourra chercher des coefficients d'absorption tenant compte de la fréquence et ainsi résoudre l'équation de Sabine pour certaines plages fréquentielles.

L'équation de Sabine pose également problème pour de petits espaces (régie d'écoute par exemple) en prédisant un temps de réverbération trop long. Dans ce cas, l'équation d'Eyring est plus adaptée.

$$RT_{60} = -0.1611 \times \frac{V}{\sum_{i=0}^k S_i \cdot \ln(1 - \alpha_i) + \mu V}$$

 Avertissement

L'équation d'Eyring n'améliore pas non plus la problématique fréquentielle.

3.1.4 L'indice de “Speech Clarity” C50

L'indice d'intelligibilité (noté C_{50}), ou “Speech Clarity” en anglais, indique la faculté d'une pièce à permettre une bonne compréhension d'une voix parlée. Son principe repose sur la mesure de l'énergie de la réponse impulsionnelle de la pièce avant 50 ms et après 50 ms. On en fait ensuite un rapport logarithmique pour obtenir une valeur en décibel.

$$C_{50} = 10 \times \log \frac{Energie(< 50ms)}{Energie(> 50ms)} dB$$

Plus la valeur du C_{50} est grande, plus la salle concentre la majorité de son énergie avant les 50 ms de propagation de la réverbération. À l'inverse, plus le C_{50} est faible, plus la salle a une énergie prédominante après 50 ms de temps de propagation. Dans ce cas une voix parlée paraîtra moins intelligible, car la réponse acoustique de la pièce engendrera un effet de fusion et de masquage.

3.1.5 Le phénomène d'onde stationnaire

La plupart des pièces de vie sont des salles rectangulaires. Dans ce cas, les surfaces sont toutes parallèles. Ce type de salle est particulièrement propice à l'apparition d'ondes stationnaires. Une onde stationnaire est un phénomène acoustique provoquant l'augmentation de volume de certaines fréquences (ventre) et la disparition d'autres (nœuds).

Nous aborderons ici ce phénomène sous l'angle de l'acoustique des salles, mais il est applicable dans d'autres situations, comme la vibration d'une corde par exemple.

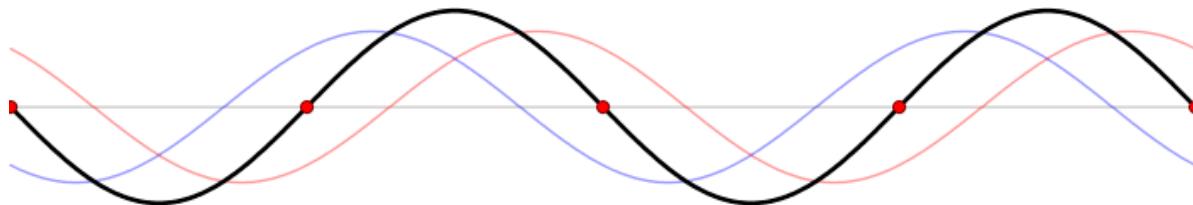


Figure 3.2: Les points rouges représentent les noeuds, les amplitudes maximales sont les ventres.
Infographie par Lucas Vieira

Il est possible de calculer les fréquences d'un mode grâce aux formules vues au chapitre précédent :

$$f(n) = \frac{c}{2L} \cdot n$$

où $c = 340 m.s^{-1}$, L est la longueur considérée de la pièce. Pour $n = 1$ on trouve le **mode propre**. Pour n strictement supérieur à un on trouvera tous les **modes harmoniques**.

Étudions la fréquence du mode propre pour deux cas théoriques : une salle de 16 m^2 (4x4) et une autre de 49 m^2 (7x7). On trouvera donc :

$$f(1)_{L=4m} = 42.5 \text{ Hz} \quad f(1)_{L=7m} = 24 \text{ Hz}$$

On en déduit donc que, plus la pièce est grande, plus la fréquence des modes propres sera grave. Il convient également de considérer la distance de chaque surface parallèle, car les pièces sont rarement cubiques. Cela implique donc la présence de trois modes propres, plus leurs modes harmoniques, pour une seule et même salle.

Au-delà d'une certaine fréquence, dite de *Schroeder*, le recouplement entre les différents modes et leurs harmoniques est tel qu'ils ne sont plus distinguables. Le phénomène de réverbération s'apparente donc à un processus aléatoire gaussien.

$$F_{Schroeder} \approx 2000 \sqrt{\frac{T_{60}}{V}} (\text{Hz})$$

3.2 Premières réflexions et filtres en peignes

Nous avons vu que la réponse acoustique, ou réverbération, d'une salle se décompose généralement en deux parties, la première étant les premières réflexions. Ces premières réflexions sont donc, comme leur nom l'indique, les premiers rebonds que nous entendons suite à un évènement sonore.

Dans de petites pièces, les premières réflexions peuvent être entendues si proche du son direct que cela génère un type de filtrage bien particulier appelé **filtre en peigne**.

Toujours en utilisant les formules définies au premier chapitre, on établit la relation suivante :

$$fc = \frac{1}{2t} = \frac{c}{2d}$$

Où fc correspond à la fréquence d'annulation la plus grave du filtre en peigne. Les autres fréquences se calculent grâce à la relation $f(n) = fc \times n$. Le phénomène de filtre en peigne est donc également harmonique.

Ainsi, on peut calculer les filtres en peignes présents au point d'écoute d'une régie de mixage ou de prise de son grâce à la mesure du chemin des premières réflexions.

La réflexion du son sur une paroi est tout à fait comparable à de l'optique géométrique. Une onde sonore arrivant avec un angle d'incidence α sur une surface sera réfléchie avec le même angle. Ainsi, il est souvent conseillé d'utiliser un miroir lorsque l'on positionne des traitements

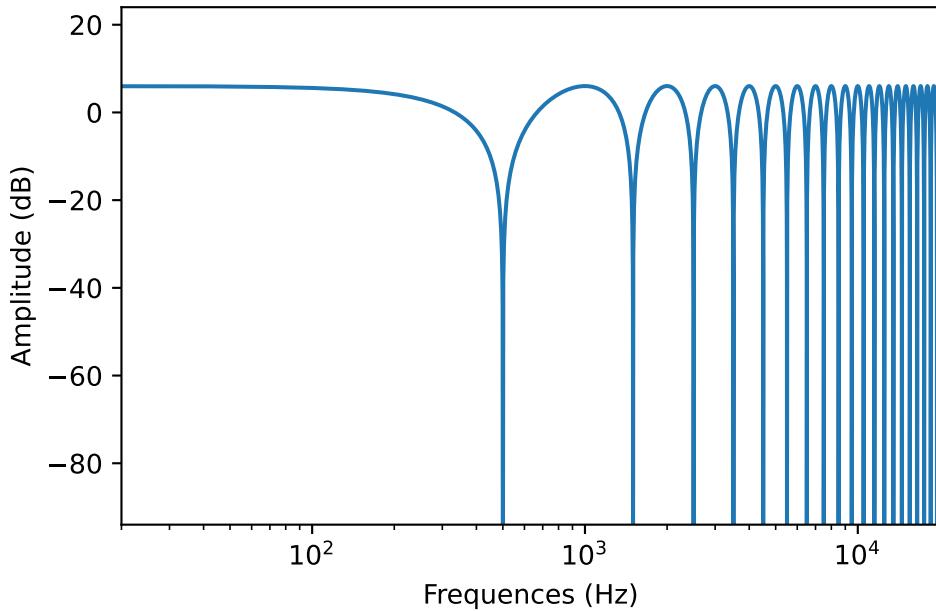


Figure 3.3: Filtre en peigne correspondant à un retard d'une milliseconde

acoustiques. Lorsque la personne assise au point d'écoute voit une enceinte dans un miroir placé sur un mur, on sait alors qu'il faudra placer le panneau à la place du miroir.

On se rend donc compte que l'influence des filtres en peigne générés par les premières réflexions est très importante. Ce phénomène à lui seul explique l'intérêt d'une grande régie d'écoute. En effet, plus une pièce est grande, plus l'écart de temps entre le son direct et les premières réflexions est important. Cela implique deux choses :

- Notre cerveau favorisera le son direct plus facilement (effet de précédence)
- À partir d'une certaine taille, l'effet du filtre en peigne se mue en information d'acoustique pour notre cerveau. Au-delà de 40 ms (trajet d'une première réflexion d'environ 14 m), l'écart entre le son direct et les premières réflexions est tel que nous entendons un écho (effet Haas).

Afin de réduire au maximum les effets des filtres en peignes, il est recommandé de placer des traitements aux points de réflexion critique par rapport à la position d'écoute (voir schéma ci-dessus).

3.3 Traitement acoustique

Grâce aux différents points abordés ci-dessus, nous avons maintenant bien l'idée que l'acoustique d'un lieu est un des facteurs les plus déterminants sur le rendu sonore. Mais c'est

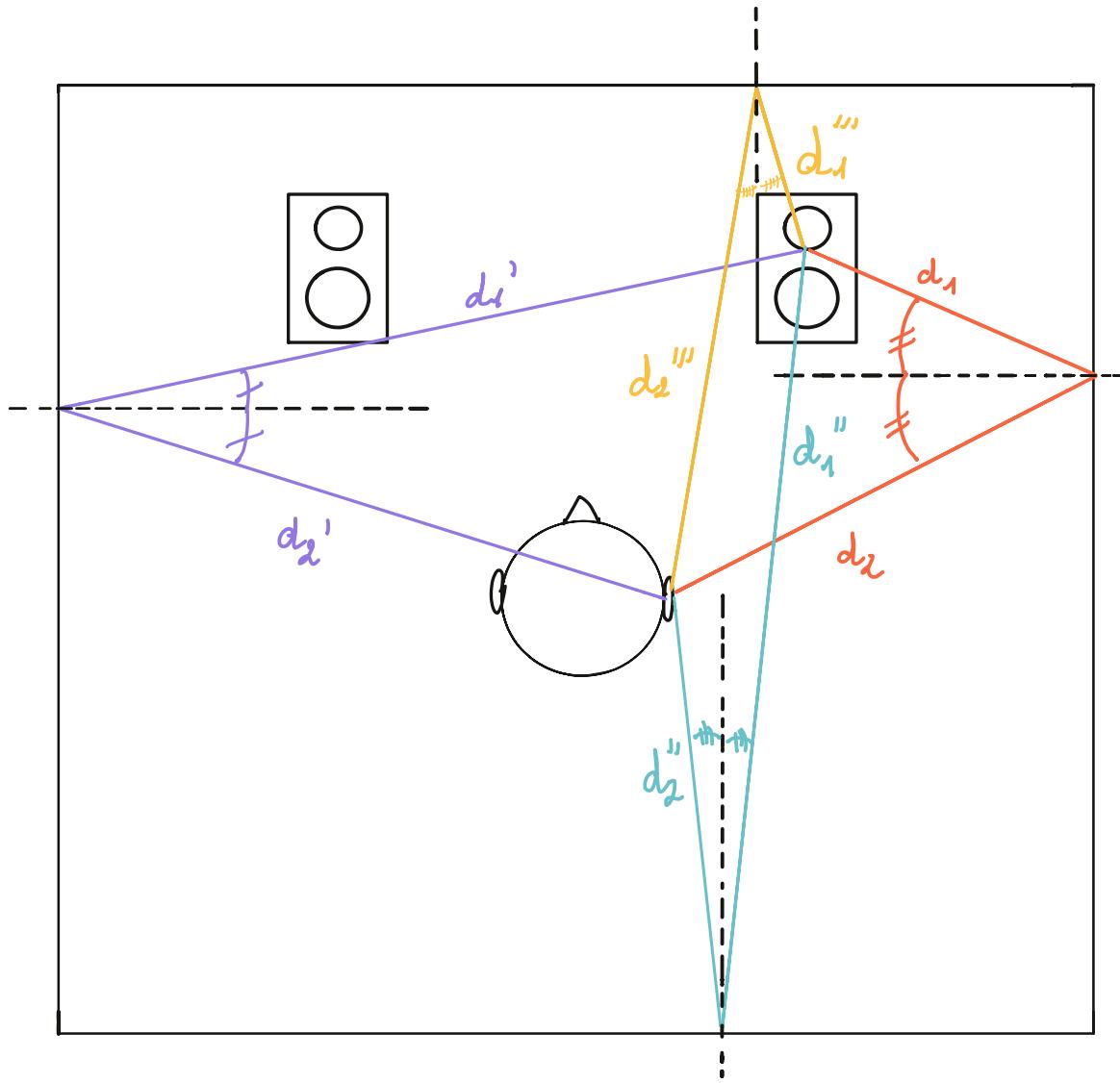


Figure 3.4: Ensemble des premières reflexions entendues par une oreille pour une enceinte (hors plafond et plancher/bureau).

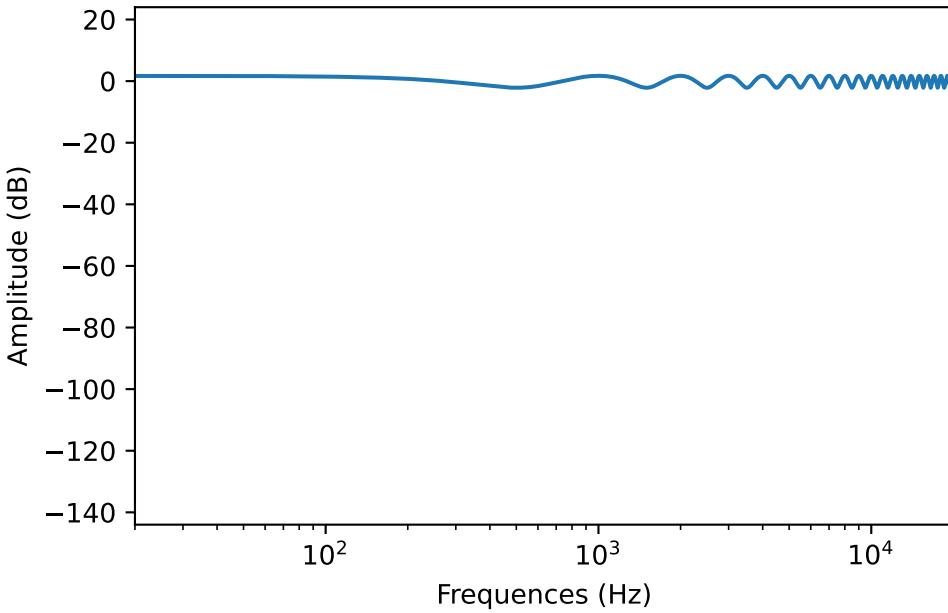


Figure 3.5: Même filtre en peigne, avec une atténuation de 20 dB sur la reflexion

aussi celui sur lequel il est plus difficile et technique d'intervenir.

On favorisera au maximum une architecture optimisée pour l'acoustique. Dans ce but, il convient de n'avoir aucune surface parallèle, cela permettant de grandement limiter l'apparition d'ondes stationnaires. On choisira également des matériaux avec des propriétés acoustiques intéressantes (plâtre et carrelage sont à proscrire, au profit du bois par exemple).

On se posera ensuite la question des endroits de la pièce les plus propices pour y positionner un évènement sonore (enceinte, musicien, etc.). On cherchera donc un point où la contribution des différents modes semble équilibrée. Pour cela, il suffit de se munir d'une enceinte et d'y diffuser une musique ou un signal test qui nous est familier. En déplaçant l'enceinte, on pourra évaluer la contribution acoustique de la pièce en différents points.

Une fois ces considérations prises en compte, on pourra alors aborder le traitement de l'acoustique.

⚠ Avertissement

Il ne faut pas confondre isolation acoustique et traitement acoustique. Dans le premier cas, on cherche à limiter la contribution sonore d'un lieu sur son environnement, dans l'autre on cherche à améliorer la propagation du son dans un espace donné. Une isolation acoustique satisfaisante nécessite de lourds travaux, voire l'aménagement d'une "boîte dans une boîte". Ces notions d'acoustiques dépassent le cadre de ce cours.

3.3.1 Les types de traitements

On trouve, en général, deux types de traitements :

- Les absorbeurs, qui réduisent l'énergie d'une onde sonore à son impact.
- Les diffuseurs, qui répartissent l'énergie d'une onde sonore dans l'espace.

Dans un lieu où la quantité de réverbération est jugée trop importante, on utilisera des absorbeurs. À l'inverse, dans un lieu où l'on souhaite préserver la quantité de réverbération, mais en évitant les phénomènes de modes ou de filtre en peignes, on utilisera des diffuseurs.

Dans de petits lieux, l'usage de diffuseur semble contre-productif, la priorité étant d'absorber au maximum les premières réflexions, celle-ci arrivant très rapidement après l'émission du son direct.

3.3.2 Considération d'acoustique pour le travail de son

Il est vivement recommandé d'installer un studio, de prise de son ou de monitoring, dans un lieu plutôt grand. En effet, plus le lieu est grand, plus il sera facile de positionner un point de prise de son ou d'écoute suffisamment éloigné des parois afin de minimiser l'influence des premières réflexions. Aussi, plus le lieu est grand, plus l'espace y sera suffisant pour installer des traitements acoustiques. Certains types de traitements, comme les basstraps, peuvent prendre une place bien trop importante pour être installée dans des pièces de dimension habituelle (chambres, bureau, etc.). On se rappellera aussi de choisir une pièce de travail avec le minimum de surface parallèle, afin de limiter les ondes stationnaires.

En ce qui concerne les traitements en eux-mêmes, il est vivement recommandé de traiter en priorité le bas du spectre. L'ajout de basstrap est donc prioritaire sur le reste des traitements. Plus la longueur d'onde à traiter est grande (donc la fréquence grave), plus la taille des matériaux devra être importante. On retrouve donc le point abordé précédemment : traiter une pièce correctement, demande un certain espace. Par ailleurs, il est important que les traitements appliqués à un lieu soient linéaires en fréquence, c'est-à-dire qu'il ne se concentre pas sur une seule zone du spectre. Cela arrive souvent avec les kits de mousses peu onéreux, mais n'ayant une réelle efficacité que dans les médiums et hautes fréquences.

Pour une régie d'écoute, on sera tenté de privilégier des traitements d'absorption. En effet, une réverbération trop longue dans une régie de monitoring risque fort de fausser certaines prises de décisions (distance des microphones à la source, quantité de réverbération, etc.). À l'inverse, une pièce avec un temps de réverbération trop court pourra créer un sentiment d'inconfort, voire de malaise.

Pour une salle de prise de son, l'idéal est de disposer d'un grand espace avec un traitement acoustique principalement basé sur de la diffusion, pour ensuite disposer de traitements absorbants amovibles permettant de sculpter le rendu acoustique en fonction de la prise de son

à réaliser. Pour des petits lieux (- de 25 m²), on cherchera à absorber au maximum afin de limiter les effets de filtre en peigne.

4 Notions élémentaires d'électronique

Les chapitres précédents nous ont permis d'aborder un certain nombre de notions fondamentales sur le son ainsi que sur l'acoustique des salles. Nous allons maintenant aborder quelques notions d'électricité et d'électronique. Le but n'est pas de savoir lire un schéma électronique, ou de comprendre comment réaliser tel ou tel circuit, mais bien d'aborder les quelques notions indispensables pour le travail du son.

Durant tout son trajet dans le milieu analogique, le signal sonore est représenté par un courant électrique. Il est donc régi par les mêmes règles que n'importe quel autre courant, même s'il possède une certaine spécificité, comme son oscillation. Un courant électrique se caractérise par le déplacement d'électrons dans un matériau conducteur (le métal par exemple). Un matériau, comme le plastique, qui ne permet pas aux électrons de se déplacer, est qualifié d'**isolant**.

Les électrons font partie des composants de l'atome. Ils sont chargés négativement et se déplacent donc dans le sens inverse du courant.

4.1 Les grandeurs physiques

Commençons par aborder les grandeurs physiques liées à l'électricité.

4.1.1 L'intensité

L'intensité électrique, notée **I** et exprimée en **Ampère (A)**, est une grandeur permettant de mesurer le débit du courant électrique. Ceci est parfaitement analogue à un débit d'eau. Si un robinet est faiblement ouvert, l'écoulement de l'eau sera faible, s'il est complètement ouvert, le débit sera fort.

4.1.2 La tension

La tension, généralement notée **U** et exprimée en **Volt (V)**, désigne une différence de potentiel entre deux points d'un circuit. Imaginons deux réservoirs d'eau, remplis d'un volume différent et connectés par une valve. Dans ce cas, la différence de potentiel serait la différence du volume d'eau entre les deux réservoirs. En d'autres termes, s'il n'y a pas de tension, il n'y a pas de débit.

On choisit, en général, la masse, valant zéro volt, comme point de référence.

Dans le cas de l'audio, la tension électrique du signal sonore est homologue à la variation de pression.

Tout comme la pression acoustique, il est possible de rendre compte d'une variation de tension électrique à un niveau sonore en décibel. La relation liant la tension et le niveau est :

$$L_{dB} = 20 \log\left(\frac{U}{U_{ref}}\right)$$

Il existe plusieurs valeurs pour U_{ref} , donnant lieu à différentes unités de mesures :

- **dBm**, définie à l'apparition du téléphone, propose $U_{ref} = 0.775V$ pour une impédance de 600ω . Cette impédance correspond à celle des lignes téléphoniques.
- **dBu / dBv**, qui ne tient plus compte de la charge d'impédance, $U_{ref} = 0.775V$.
- **dBV**, où $U_{ref} = 1V$

Lorsque la tension double, le niveau augmente de six décibels. Lorsque la tension est multiplié par dix, le niveau augmente de vingt décibels.

On peut également définir l'augmentation du niveau sonore par rapport à la puissance du signal. On admet que :

$$P = \frac{U^2}{Z} U = \sqrt{P \times Z}$$

Où P est la puissance. En remplaçant dans l'équation précédente, on trouve :

$$L_{dB} = 20 \log\left(\frac{\sqrt{P}}{\sqrt{P_{ref}}}\right) = 10 \log\left(\frac{P}{P_{ref}}\right)$$

Lorsque la puissance double, le niveau augmente de trois décibels. Lorsque la puissance est multiplié par dix, le niveau augmente de dix décibels.

On utilisant la loi d'ohm (voir ci-dessous) et la relation entre la puissance, la tension et l'impédance, on trouve également que :

$$P = U \times I$$

4.1.3 L'impédance

Nous connaissons, en général, la **loi d'Ohm**. Celle-ci permet de donner une relation entre l'intensité du courant et sa tension, aux bornes d'un composant d'un circuit (aussi appelé dipôle).

$$U = R \times I$$

Où R est la résistance du dipôle. Elle traduit la facilité d'un courant à se déplacer dans le dipôle. Pour reprendre les analogies ci-dessus, la résistance correspondrait à une valve. À tension constante, si la résistance tend vers zéro, le débit est très important. Si la résistance tend vers l'infini, le débit sera très faible. Si elle est nulle, alors nous sommes dans le cas d'un court-circuit (interrupteur fermé). Si elle est infinie, cela traduit une absence de connexion entre deux points d'un circuit (interrupteur ouvert). L'unité de cette résistance est l'**ohm**.

L'impédance traduit elle aussi l'opposition d'un circuit au passage d'un courant électrique, mais dans le cas d'une **tension oscillante**. Dès lors, l'impédance englobe les effets de résistance, de capacitance et d'inductance (voir ci-dessus).

4.2 Les composants électroniques

4.2.1 Les composants passifs

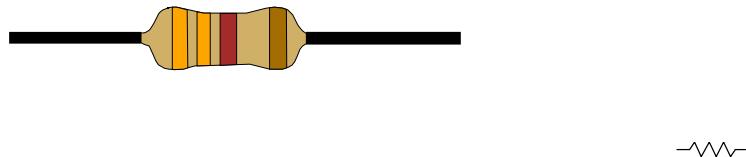


Figure 4.1: Représentation d'une résistance et de son symbole

Étudions maintenant les composants électroniques les plus communs. Nous avons en premier les **résistances**. Ce sont des dipôles purement résistifs. Leur valeur s'exprime en **ohm**. Une résistance s'oppose donc au passage du courant. Pour rappel, la tension a ses bornes est $U = R \times i$.

Viennent ensuite les **condensateurs**. Ils sont constitués de matériaux conducteurs séparés par une couche isolante. La relation entre tension et intensité à ses bornes en régime oscillant est :

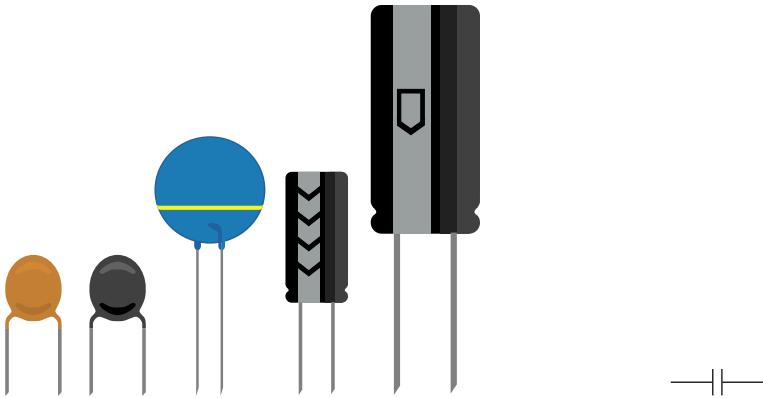


Figure 4.2: Représentation d'un condensateur et de son symbole

$$U = Z_c \times I$$

Où Z_c est l'impédance d'un condensateur idéal. Nous pouvons ici la même analyse que plus haut, quand Z_c tend vers l'infini le courant ne passe plus, quand Z_c tend vers 0, le débit est important. L'impédance d'un condensateur est fonction de sa **capacité** (noté **C**, et s'exprime en **farads**).

$$Z_c = \frac{1}{jC\omega} = 2\pi f$$

Si la fréquence f tend vers l'infini, Z_c tend vers zéro, si la fréquence tend vers zéro, Z_c tend vers l'infini. On constate donc que l'impédance d'un condensateur varie en fonction de sa fréquence. On peut assimiler un condensateur à un interrupteur ouvert en basse fréquence et à un interrupteur fermé en haute fréquence.



Figure 4.3: Symbole d'une bobine

Terminons sur les bobines. Ces composants sont constitués d'un enroulement de câble en cuivre et possède une **inductance** notée **L** et s'exprimant en **henrys**. Étudions à nouveau la relation entre tension et intensité, aux bornes d'une bobine :

$$U = Z_L \times I$$

Où Z_L est l'impédance d'une bobine idéale. Cette impédance se calcule grâce à la relation suivante :

$$Z_L = j\omega L = 2\pi f$$

Si la fréquence f tend vers l'infini, Z_L tend vers l'infini. Si f tend vers zéro, alors Z_L tend vers zéro. On observe donc le comportement inverse du condensateur. Une bobine se comporte comme un court-circuit en basse fréquence et comme un interrupteur ouvert en haut fréquence.

On admet j comme un outil mathématique permettant de simplifier certaines écritures et certains calculs. On l'appelle le nombre complexe, tel que $j^2 = -1$. Dans nos applications, sa présence dans les relations des impédances de condensateur et de bobine implique un déphasage de $-\frac{\pi}{2}$ pour un condensateur, et, de $\frac{\pi}{2}$ pour une bobine.

L'association de résistances, de condensateurs et de bobines donne des circuits RL, RC où RLC, permettant de réaliser des **opérations de filtrage** sur le signal.

4.2.2 Tubes & semi-conducteurs

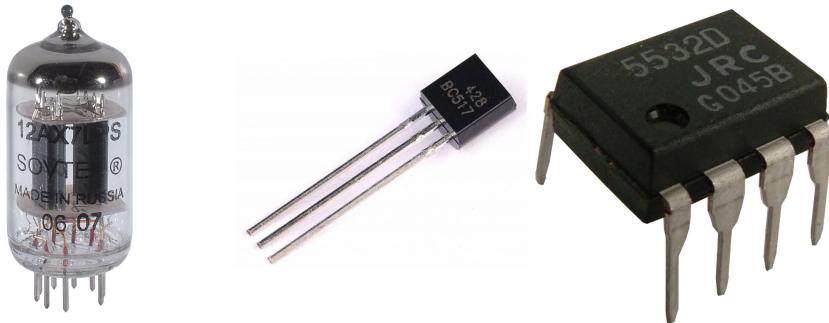


Figure 4.4: Tubes, transistor et circuits intégrés

Les **tubes**, tubes à vide, ou parfois, lampes, sont historiquement les premiers composants permettant d'amplifier le signal, contre une certaine tension d'alimentation. On les retrouve donc dans les préamplificateurs, égaliseurs, compresseurs, et autres amplificateurs jusque dans les années soixante. Ils sont alors progressivement remplacés par les **transistors**, composants appelés **semi-conducteurs**. Ces transistors permettent de réaliser la même amplification du signal qu'une lampe, mais sont beaucoup plus petits et demandent aussi moins de puissance électrique pour réaliser le même facteur d'amplification (aussi appelé **gain**). Peu de temps après la mise au point des transistors, les **circuits intégrés** sont inventés. Ces petites boîtes renferment plusieurs transistors, et peuvent également servir à l'amplification de signaux.

Il est très important de savoir que l'invention du transistor et des circuits intégrés est sans doute l'avancée technologique la plus importante du siècle dernier. Elle a permis le développement exponentiel de l'industrie informatique grâce à la miniaturisation des composants.

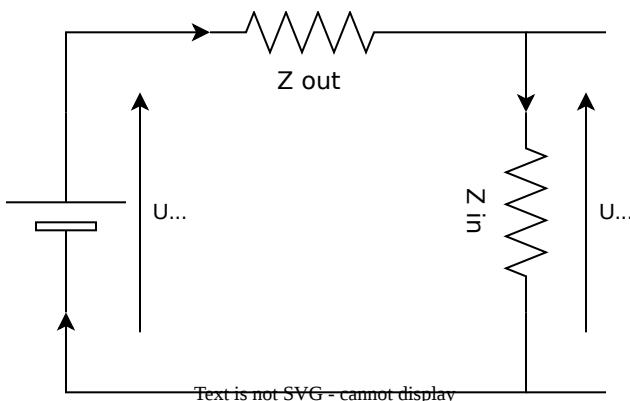
L'utilisation de tubes, de transistors ou de circuits intégrés au sein des machines audio, est souvent associée à une certaine « couleur ». Il y aurait donc un son des tubes, un son des transistors et un son des circuits intégrés. Les différences entre ces dipôles apparaissent principalement dans les zones de **non-linéarité** des composants, typiquement dans leur zone de saturation. La saturation apparaît lorsque la tension du signal amplifié dépasse la tension d'alimentation du composant responsable de cette amplification. On observe alors l'apparition de certaines harmoniques. La distribution des harmoniques générées est différente en fonction du dipôle utilisé.

Il est compliqué d'attribuer une couleur sonore particulière à un composant. En effet, le comportement d'un composant est fondamentalement conditionné par la topologie du circuit dans lequel il est utilisé ainsi que par les autres composants qui l'entourent. Il convient donc, à l'humble avis de l'auteur, d'être relativement prudent sur des expressions telles que « son des tubes » ou « son des transistors », particulièrement quand il s'agit de dire que l'une des technologies « sonnerait mieux » que l'autre. L'histoire de l'électronique musicale regorge d'exemples et de contre-exemples pour chacune de ces affirmations.

4.3 L'influence de l'impédance entre différents appareils.

Sur la fiche technique des appareils, on trouve des valeurs pour son impédance d'entrée et son impédance de sortie. Imaginons que nous connectons un appareil A dans un appareil B. En pratique, nous faisons en sorte que l'impédance de sortie de l'appareil A soit dix fois inférieure à l'impédance d'entrée de l'appareil B. À partir du moment où ces impédances sont proches, voire que l'impédance de sortie de A soit plus grande que celle d'entrée de B, nous allons atténuer le signal transitant entre les deux appareils. Étudions de plus près ce phénomène.

Soit le schéma électronique ci-dessous. On appelle U_{out} la tension de sortie de l'appareil A et Z_{out} son impédance de sortie. De façon similaire, on appelle U_{in} la tension d'entrée de l'appareil B et Z_{in} son impédance d'entrée.



Dans ce circuit, $Z_{eq} = Z_{in} + Z_{out}$ et $U_{out} = Z_{eq} \times i$. Alors, $i = \frac{U_{out}}{Z_{eq}} = \frac{U_{out}}{Z_{in} + Z_{out}}$. Toujours grâce au circuit ci-dessus, on peut dire que $U_{in} = Z_{in} \times I$. En remplaçant dans l'expression précédente on trouve : $\frac{U_{in}}{Z_{in}} = \frac{U_{out}}{Z_{in} + Z_{out}}$

$$\frac{U_{in}}{U_{out}} = \frac{Z_{in}}{Z_{in} + Z_{out}} = \frac{1}{1 + \frac{Z_{out}}{Z_{in}}}$$

Dès lors, si Z_{in} est très grand devant Z_{out} , alors U_{in} tend vers U_{out} . Si Z_{out} est très grand devant Z_{in} , alors U_{in} tend vers 0.

Cela nous amène à démontrer l'affirmation ci-dessus. Maintenant, nous savons que dans un circuit, l'impédance varie en fonction de la fréquence. Dès lors, une mauvaise adaptation d'impédance ne fera pas que diminuer l'amplitude du signal, mais filtrera aussi une partie de spectre, généralement les hautes fréquences.

On considère aussi l'adaptation d'**impédance en tension**. Lorsque que nous considérons la **puissance** les relations sont différentes (cf section sur les hautparleurs).

5 Description d'une production musicale type

Afin de comprendre quels vont être les enjeux du preneur de son, il convient de comprendre dans quel contexte il intervient. Certes, il est le premier métier du son à rentrer en scène, mais l'œuvre à enregistrer a déjà très probablement eu une longue vie. Elle a été composée, arrangée, peut-être même déjà interprétée au cours de concerts.

À ce stade, le preneur de son aura un regard neuf sur la matière. Il aura donc le potentiel de permettre aux créateurs de prendre du recul sur leur travail. Il convient d'ailleurs de rappeler qu'un preneur de son, aussi talentueux et créatif soit-il, est un **assistant de création**. Cela signifie qu'il met à disposition une compétence technique à un d'artiste pour lui permettre d'avancer sur son projet. Cela implique également que celui ou celle qui a le mot final sur le choix des orientations esthétiques est l'artiste en question. Il convient donc, en tant que preneur de son, d'être force de proposition, tout en sachant respecter le choix (qu'ils soient bons ou mauvais) des artistes.

D'un point de vue sonore, le travail de prise de son est absolument critique. Ce sera à ce moment que va se jouer la majorité des choix esthétiques. Il convient donc de réunir les conditions optimales pour :

- offrir aux musiciens et musiciennes la chance de donner leur meilleure interprétation possible
- réaliser une prise de son en adéquation avec l'orientation esthétique du projet

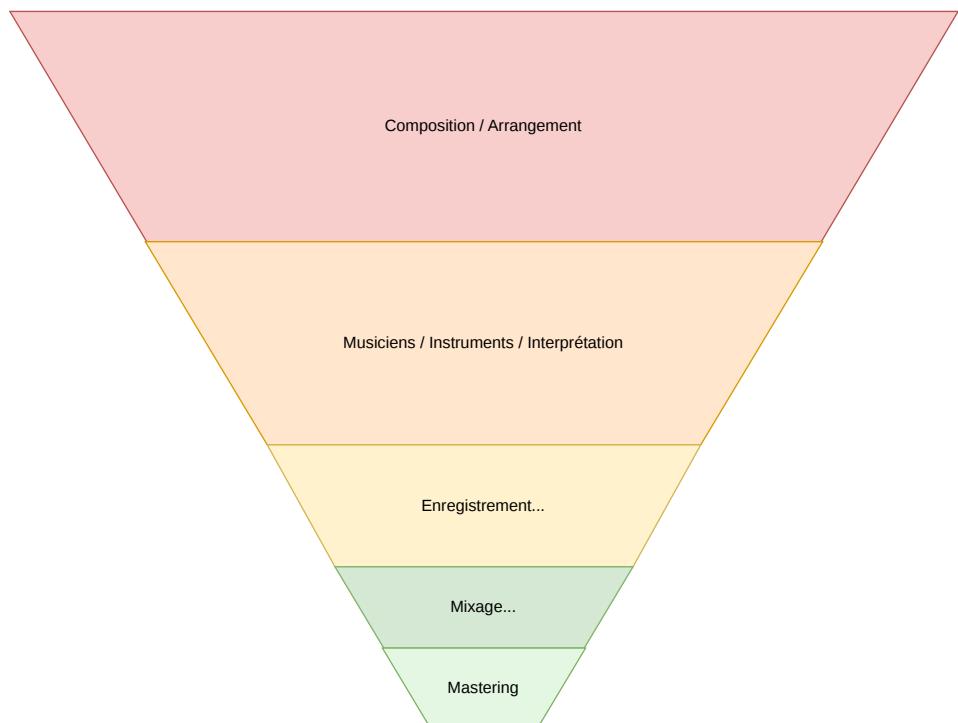
La plupart des choix faits à la prise de son ne pourront pas être renégociés a posteriori. Il convient donc de mettre d'accord les artistes, le directeur artistique et le preneur de son sur les moyens à mettre en œuvre.

5.1 les acteurs de la réalisation d'une œuvre enregistrée

Nous allons ici rapidement discuter des différents rôles apparaissant dans la production d'une œuvre musicale enregistrée. Ceux-ci sont volontairement très séparés, bien que dans les cas pratiques, une personne puisse en incarner plusieurs.

Le compositeur est la personne qui a composé la mélodie et l'harmonie de l'œuvre.

L'arrangeur est chargé de l'orchestration (choix des instruments) et l'écriture des différentes partitions.



Text is not SVG - cannot display

Figure 5.1: Entonnoir de la production musicale

L'interprète a la responsabilité de retranscrire une partition le plus justement possible, à la fois dans sa dimension technique et sensible.

Le directeur artistique (ou DA) supervise l'ensemble de l'enregistrement. Il aura, par exemple, à choisir le preneur de son, le mixeur ou dans quel studio enregistrer. Lors de la session d'enregistrement, il aura à diriger les musiciens (comme un réalisateur dirige ses acteurs au cinéma) afin de leur faire jouer la meilleure interprétation possible pour l'œuvre. Lors du mixage, il sera le principal interlocuteur du mixeur. Pour faire court : il est le garant de l'orientation esthétique du projet.

Le producteur finance l'ensemble de projets. C'est donc un investisseur qui attend un retour sur investissement.

L'appellation abusive de « producteur » pour parler du directeur artistique vient d'un anglicisme du mot « producer ». Le producteur est donc bien l'équivalent du directeur artistique dans les pays anglo-saxons. Si le DA a besoin d'un certain talent, le producteur a surtout besoin d'argent.

Le preneur de son est chargé d'enregistrer les musiciens et musiciennes. Il a donc un rôle premier très technique : il doit inscrire sur un support les ondes sonores produites par ces musiciens. Il a également un rôle esthétique très important, d'un point de vue sonore, car le choix du dispositif de prise de son aura un fort effet sur la suite de la vie de l'œuvre.

Le mixeur intervient après la prise de son et doit réaliser une sommation de l'ensemble des points de captations (microphone) vers un format écoutable par le grand public (mono, stéréo, 5.1, Ambisonique, Dolby Atmos, etc.). Son rôle esthétique est fortement contraint par le travail de prise de son. Si celle-ci est réussie, il pourra amplifier et bonifier les choix de production. Dans le cas contraire, il devra lutter pour essayer de faire sortir le meilleur d'une matière imparfaite.

Le technicien de mastering est le dernier maillon de la chaîne. Son rôle premier sera de préparer le travail de mixage à aux supports de diffusion. Il se devra également d'offrir une oreille nouvelle sur le travail réalisé au mixage.

5.2 La préproduction

La préproduction concerne toutes les étapes d'une œuvre enregistrée qui ont lieu avant ledit enregistrement. On parlera donc en premier lieu de la composition et particulièrement de l'arrangement.

La qualité d'un arrangement aura une influence énorme sur la facilité à mixer une œuvre. Si les instruments sont astucieusement répartis sur l'ensemble du spectre sonore, cela sera une difficulté de moins à gérer au mixage par exemple.

Il est aussi courant pour des artistes de réaliser des « démos ». Celles-ci sont souvent des enregistrements réalisés en home studio afin de définir un cap esthétique pour la suite de la production sonore. C'est un atout extrêmement précieux pour un preneur de son, cela permet de rapidement identifier quel est le projet esthétique de l'œuvre.

5.3 La production

C'est ici que le travail du preneur de son commence. L'étape de production consiste à fixer les interprétations définitives. Le premier objectif est donc de s'assurer du bon enregistrement de tous les canaux prévus. Bien sûr, l'enjeu n'est pas seulement technique, mais aussi esthétique. Et il n'est pas moindre, les choix pris lors de la prise de son seront des carcans impossibles à outrepasser lors de la phase de mixage. Enfin, l'élément le plus déterminant de cette étape est d'obtenir des musiciens leurs meilleures interprétations. La présence d'un directeur artistique est d'une aide précieuse afin de diriger et d'orienter les musiciens. Il permet aussi de faire le lien entre les artistes et l'équipe technique, en exprimant les besoins des uns aux autres.

Sur les projets les plus modestes, le poste de directeur artistique est souvent sacrifié. Il en va donc à l'ingénieur du son de, parfois, remplir ce rôle.

5.4 La postproduction

Arrivé à ce stade, la majorité du travail est déjà accompli, il ne reste que le mixage et le mastering. Classiquement, chacune de ces tâches incombe à un technicien différent. Le travail du mixeur consistera à réaliser la sommation, généralement en stéréo, de l'ensemble des canaux enregistrés lors de la prise de son. Afin de faire cohabiter tous ces signaux, il est commun d'utiliser des traitements pour les répartir sur l'ensemble du spectre et de gérer leur dynamique. Parfois, ces traitements remplissent un rôle esthétique, en déformant le signal d'origine pour aboutir à une nouvelle matière.

Une fois le travail du mixeur terminé, le mastering commence. Le but est d'homogénéiser l'ensemble des titres d'un disque, en volume, en dynamique et en couleur. Ensuite, il convient aussi de définir le niveau de sortie général du disque. La dernière étape consistera à monter l'ordre des morceaux pour le disque, d'y inscrire les métadonnées (nom de l'artiste, des titres, genre musical, etc.) et de générer le fichier final, dédié à l'exploitation.

partie II

Outils et équipements

6 Le chemin du signal

La première mission d'un preneur de son est d'assurer l'arrivée à bon port des signaux dans l'enregistreur. En effet, toute notion de mise en scène sonore et d'esthétique devient très secondaire si le contenu n'a pas été enregistré.

Le diagramme ci-dessous reprend les principaux étages rencontrés par un signal audio dans un contexte de production numérique. Il est essentiel d'être le plus familier possible avec ces différents composants.

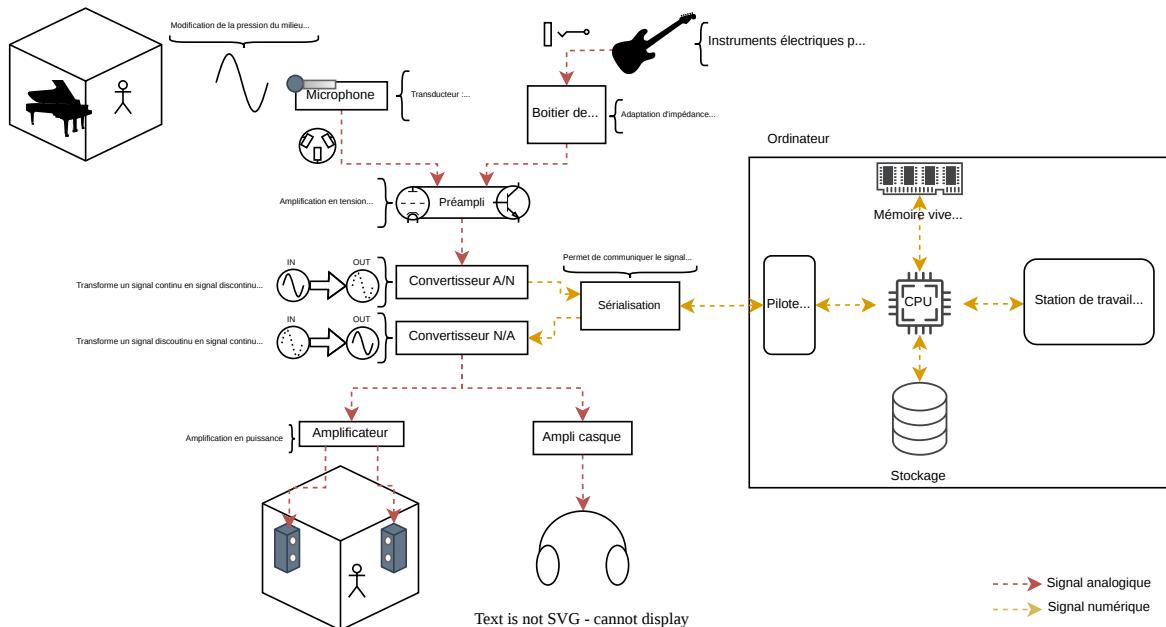


Figure 6.1: Le chemin du signal. Elle peut-être agrandie en ouvrant l'image dans un nouvel onglet.

Nous pourrions catégoriser à partir de ce schéma différents « milieux ». Tout d'abord, nous avons le **milieu acoustique**, où nous trouverons toutes sortes d'instruments de musique, les différents lieux dans lesquels ils pourront s'y trouver. C'est donc le domaine de l'onde sonore mécanique.

On trouve ensuite le **milieu analogique**, où l'onde sonore est représentée, de façon analogue, par des grandeurs électriques. Celles d'un signal sonore dans un circuit analogique sont fonction,

par exemple, de la variation de la pression atmosphérique provoquée en un point par une onde sonore. Les éléments clefs du milieu analogique sont les préamplificateurs et les amplificateurs, mais on trouve aussi certains traitements, comme les égaliseurs et les compresseurs. On définira « analogique » comme une représentation dans laquelle les grandeurs (tension, courant, etc.) qui entrent dans les calculs sont représentées par des grandeurs analogues et qui varient de manière identique (définition du [CNRTL](#)).

Pour passer du milieu acoustique au milieu analogique, et vice-versa, on utilise des microphones et des haut-parleurs. Tous deux sont des **transducteurs**, permettant de transformer une énergie en une autre. Le microphone transforme une énergie mécanique en énergie électrique. Le haut-parleur réalise l'opération inverse.

On en vient ensuite au **milieu numérique**. Fondamentalement, le signal est toujours de nature électrique, mais il a subi une opération très importante nommée **échantillonnage**. On a donc mesuré à intervalle régulier la tension électrique générée par l'onde sonore. Le passage par le numérique permet une myriade de traitements sur le signal, beaucoup plus complexes que ceux permis par l'électronique analogique. L'audio numérique permet aussi un stockage de l'information à moindre coût et l'acheminement d'un grand nombre de voies (canaux) grâce à un faible nombre de modulations (câble).

L'appareil permettant de passer du milieu analogique au milieu numérique est le **convertisseur analogique/numérique**. Il ne s'agit pas d'un transducteur, car les signaux d'entrées et de sorties sont de même nature électrique. Pour opérer l'opération inverse, on utilise un **convertisseur numérique/analogique**.

Le **milieu informatique** nous permet d'utiliser des applications relatives aux traitements du son par le biais d'ordinateurs. Il s'agit aujourd'hui indubitablement de notre outil de travail principal. Nous y réalisons la grande majorité des traitements audio, ainsi que l'enregistrement et le routage des sources.

Le lien entre un ordinateur et un convertisseur A/N/A se fait grâce à un **bus de sérialisation** associé à un **pilote** (ou **driver**). L'ensemble des deux permet de mettre en forme la donnée numérique et de la rendre compréhensible à l'ordinateur.

Chaque élément évoqué ci-dessus sera abordé dans des sections dédiées dans la suite de ce livre.

7 Les microphones

Un microphone est un **transducteur** permettant de transformer une onde acoustique en signal électrique. Cette opération est réalisée par une membrane. Selon la nature du microphone, cette membrane pourra être constituée d'une feuille métallique d'un condensateur ou encore être rattachée à une bobine.

Le microphone est l'outil principal du preneur de son. Le choix du modèle et sa position dans l'espace est déterminants sur le rendu sonore d'une captation. Ces deux paramètres ont par ailleurs une certaine interdépendance : une position souhaitée du microphone pouvant influencer le choix du modèle et vice-versa.

7.1 Petit historique des microphones

Sans vouloir rentrer dans un récit exhaustif sur l'invention et l'évolution des microphones, relater les moments clefs de cette technologie permet d'avoir une vision globale du marché d'aujourd'hui.

La nécessité de capter un évènement sonore grâce à un microphone provient de trois besoins :

- le transmettre (télécommunication)
- l'amplifier (concert, spectacle vivant)
- l'enregistrer (industrie du disque)

En 1876, Alexandre Graham Bell propose un système à base liquide, permettant de transformer une onde sonore en tension électrique. Le système ne fut jamais réellement exploité, car le rendu sonore était jugé trop peu satisfaisant.

Le premier type de microphone utilisé industriellement est le **microphone à charbon** (au UK, par David Edward Hugues, aux US par Emile Berliner et Thomas Edison. Le brevet sera d'ailleurs disputé, avec un gain de cause pour Edison malgré des démonstrations publiques de Hugues antérieur aux publications d'Edison). En raison de sa faible bande passante et de son niveau de bruit élevé, il se révèle de piétres qualités pour l'enregistrement et la transmission de la musique. Il aura, par contre, une place de choix dans les téléphones durant de longues décennies.

Un premier brevet, peut-être même le premier, pour un **microphone dynamique** à bobine mobile est attribué à l'ingénieur et industriel allemand Ernst Werner von Siemens en 1877. La

technologie de la bobine mobile a été mise en pratique dans les années 1920 lorsque la Marconi-Sykes Company a créé le magnétophone pour la radio britannique. Ils s'imposent d'abord dans le monde du concert, dès les années 40, pour leur grande résistance mécanique.

Viennent ensuite les **microphones à condensateur**, dont les premiers modèles remontent à 1916, par le chercheur Edward Wente. Ces microphones sont tout d'abord réputés assez capricieux, leurs réponses en fréquences pouvant varier significativement en fonction de l'humidité de l'air et de la température.

À cause de ces variations sonores présentes dans les premiers microphones à condensateur, on leur préférera un temps les **microphones à ruban**. Ils sont inventés en 1923 par Harry Olson. Ils sont par contre d'une grande fragilité mécanique.

George Neumann est un des noms à connaître dans cette histoire des microphones. On lui doit, entre autres, la stabilisation des microphones statiques. Il sera aussi le premier à produire un microphone (U87) utilisant un transistor en lieu et place des traditionnels tubes.

À partir des années 1970, les microphones dynamiques sont adoptés en studio d'enregistrement, notamment portés par la marque Shure. Ces microphones ont la grande qualité d'être très robustes, et remplaceront leurs homologues à ruban dans bien des cas.

Depuis, les principales améliorations ont concerné la robustesse d'une part, et la miniaturisation des dispositifs d'autre part, menant ainsi au développement des capsules MEMS.

7.2 Les types et technologies de microphones

Avant d'aborder en détail certaines constructions de microphones, il convient de faire attention à certains raccourcis associant des méthodes de fabrications à un niveau présumé de qualité. Par exemple, il est commun d'associer les microphones à électret à une construction « bas de gamme ». Or, c'est oublier que la série 4000 de chez DPA, considérée par beaucoup comme une référence indétrônable de la prise de son, ne contient que des microphones à électret. Les MEMS souffrent du même biais, ceux-ci se retrouvent pourtant de plus en plus souvent sur des microphones ambisoniques, comme le Zyla ou le SPC mic.

Nous allons maintenant aborder les types de microphones suivants :

- Les microphones électrostatiques/à condensateur
- Les microphones à ruban
- Les microphones dynamiques



Figure 7.1: Neumann U87, Schoeps CMC64, Line Audio CM4

7.2.1 Les microphones électrostatiques/à condensateur

Ce sont, historiquement, les premiers microphones à permettre une captation du spectre audible satisfaisante. Ils sont cependant très sensibles aux conditions de température et d'humidité et il fallut attendre les années trente pour que ce problème cesse. Ils nécessitent une alimentation externe, appelée alimentation fantôme, normalisée à +48V. Il existe deux familles de microphones électrostatiques, les **condensateurs à hautes fréquences** et **condensateur polarisés en courant continu**.

Les microphones à condensateur polarisés en courant continu ont le fonctionnement le plus commun. Un courant continu vient polariser la capsule/condensateur. Lorsqu'une onde sonore rencontre la capsule, une de ses armatures se déforme et génère une variation de tension analogue à la variation de pression.

Les microphones à condensateur à haute fréquence proposent une approche différente. Un oscillateur est intégré dans le microphone et la variation de pression enregistrée par le condensateur vient moduler la fréquence de cet oscillateur. Le signal est ensuite démodulé dans la plage audible. Cette méthode de construction offre une impédance de sortie plus faible et une plus grande résistance aux variations de conditions climatiques.

Concernant leurs caractéristiques, ces microphones possèdent des réponses en fréquence souvent très linéaire et une excellente réponse en transitoire. Leur niveau de sortie (sensibilité) est élevé. Leur impédance de sortie est basse.

Exemples : Neumann U87/AKG C414/Shoeps CMC4/Série 4000 DPA/Série MKH Sennheiser



Figure 7.2: Royer R121, Cascade Vinjet, Coles 4038

7.2.2 Les microphones à ruban

Les microphones à ruban souvent préférés à leurs homologues statiques dans les débuts de la musique enregistrée. Leur fonctionnement repose sur l'utilisation d'une feuille métallique placée entre deux aimants. Lorsqu'une onde sonore rencontre cette feuille (le ruban), celle-ci vibre et perturbe le champ électromagnétique créé par les aimants et génère une tension analogue à la variation de pression.

D'un point de vue sonore, les microphones à ruban ont souvent un bas du spectre assez généreux et une réponse plutôt douce pour les hautes fréquences. Ils sont aussi connus pour avoir une impédance de sortie assez élevée et un niveau de sortie faible. Attention à l'alimentation fantôme (+48V), elle peut endommager le microphone.

Exemples : Royer R121/Cohles/Beyerdynamic M160

7.2.3 Les microphones dynamiques

Les microphones dynamiques sont conçus pour des conditions d'utilisation rudes, où les niveaux sonores sont élevés et où le risque de chute est important. Ils sont donc monnaie courante en sonorisation. Leur membrane est attachée à une bobine entourant un aimant. Lorsqu'une onde sonore la met en vibration, la bobine se déplace autour de l'aimant, et, par perturbation du champ électromagnétique, génère une tension de sortie analogue à la variation de pression.

Leur réponse en fréquence est souvent accidentée, particulièrement dans le haut du spectre. Cela peut être vu comme un inconvénient ou comme un outil de « coloration » du son. Comme leurs homologues à ruban, ils possèdent un niveau de sortie faible et une impédance de sortie élevée.

Exemples : Shure SM57/Electrovoice RE20/Sennheiser MD441



Figure 7.3: Shure SM57, Electro-Voice RE20, Sennheiser MD441

7.2.4 La taille des membranes

La taille des membranes influe sur la captation du son. Plus la capsule est grande, plus les fréquences aiguës seront diffractées et donc atténuées dans la prise de son. Un microphone à petite membrane est donc techniquement un microphone plus « juste ». Cependant, l'emploi de large membrane permet aussi d'adoucir un surplus d'énergie dans le haut du spectre.

7.2.5 Microphones à tubes ou transistors?

Historiquement, les tubes ont été les premiers composants électroniques à permettre l'amplification du signal. Le transistor est apparu à la fin des années 40 et a permis de remplir les mêmes fonctions qu'un tube, par une consommation moindre et avec un encombrement beaucoup plus faible.

Certains microphones continuent à être fabriqués avec des tubes, préférant leur comportement vis-à-vis du son. Une écrasante majorité est cependant fabriquée avec des transistors.

Le choix entre un microphone à tube et un microphone à transistor semble cependant anecdotique par rapport à son type, à son placement et à sa directivité.

7.3 Timbre et directivité

La directivité d'un microphone permet de décrire sa capacité à réaliser une « écoute » sélective de son environnement. On rencontre les directivités suivantes :

- Omnidirectionnel : capte l'ensemble du champ sonore de façon indifférenciée.
- Hypercardioïde : compromis entre Omnidirectionnel et cardioïde.
- Cardioïde : capte à l'avant, mais rejette à l'arrière du microphone.

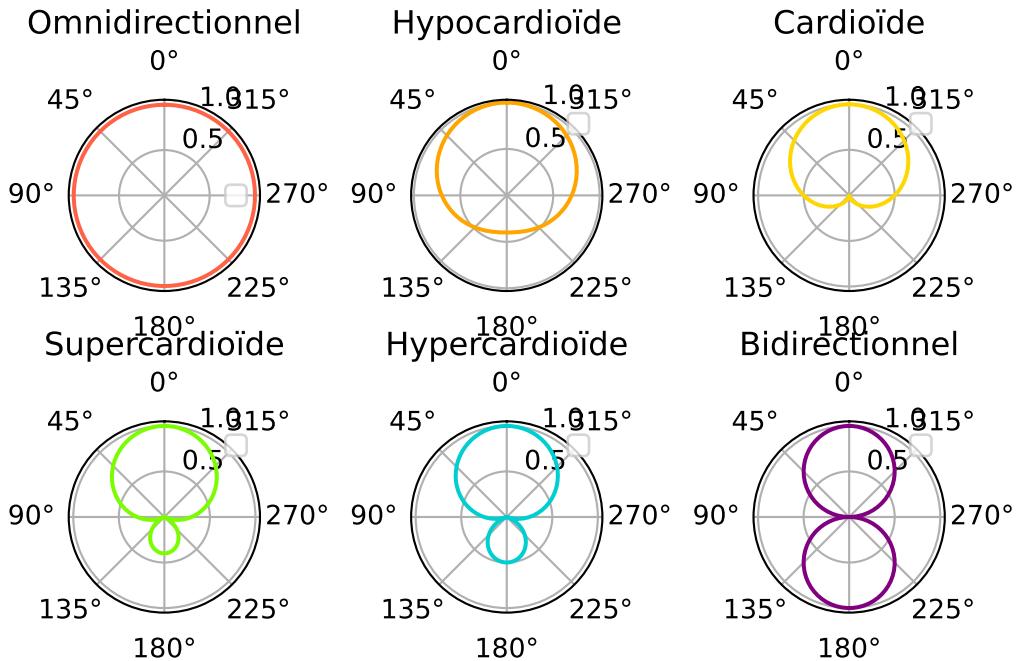


Figure 7.4: Représentation des six grandes directivités de microphones

- Supercardioïde : resserre la zone d'écoute avant au prix de l'apparition d'une résurgence arrière.
- Hypercardioïde : resserre davantage la zone d'écoute et augmente d'autant plus la résurgence arrière.
- Bidirectionnel : capte à l'avant et à l'arrière, mais selon un lobe plus resserré qu'en cardioïde.

Plus la directivité d'un microphone est large, plus la contribution de l'acoustique est apparente. Le timbre est également aussi linéaire que possible. À l'inverse, plus la directivité tant à être étroite, plus le microphone aura une capacité à échantillonner seulement une zone de l'espace. Le timbre est, par contre, amoindri dans le bas du spectre. Les microphones omnidirectionnels sont donc les plus larges et les plus « neutres », tandis que les microphones bidirectionnels sont les plus focalisés et ont la plus importante perte dans le bas du spectre.

Le preneur de son choisit donc une directivité en fonction de la tâche à accomplir. Les microphones directifs ont l'avantage de limiter la contribution d'évènements sonores que l'on ne souhaite pas capter. Les microphones omnidirectionnels ont la faculté d'être un dispositif de prise de son plus transparent, mais seront beaucoup plus sensibles à une acoustique moins optimale, ainsi qu'au bruit environnant.

Nous allons par la suite nous intéresser au cœur du microphone : sa capsule. Il existe deux

familles de capsules, celles dites « à pression », et celles dites à « gradient de pression ».

7.3.1 Capsules à pression

Une capsule sensible à la pression est omnidirectionnelle : elle capte les fluctuations de pressions en un point. Mathématiquement, cette relation s'exprime, en coordonnées polaires, par :

$$\theta = 1$$

L'angle d'incidence de l'onde sonore par rapport au microphone importe donc peu.

Pour réaliser une capsule à pression, on enferme une partie de la membrane dans un milieu acoustique à pression constante.

7.3.2 Capsules à gradient de pression

Une capsule à gradient de pression est sensible à la **variation** du champ de pression. Ces capsules ne sont plus omnidirectionnelles, mais bidirectionnelles : elles captent devant et derrière elles. Mathématiquement, une telle directivité s'exprime par la relation (en coordonnées polaires) :

$$\theta = \cos(\alpha)$$

Où α est l'angle d'incidence d'un son par rapport à la capsule.

Pour réaliser une capsule à gradient de pression, il suffit de laisser exposer les deux faces de la membrane aux variations de pressions.

7.3.3 Et les autres directivités ?

Il est possible, à partir des deux équations ci-dessus, de retrouver toutes les autres directivités. Par exemple, un microphone cardioïde a une équation de directivité polaire tel que :

$$\theta(\alpha) = \frac{1}{2}(1 + \cos[\alpha])$$

Elles découlent donc des deux directivités primaires : omnidirectionnelle et bidirectionnelle. Pour obtenir une directivité particulière, il suffit de « doser » l'influence de ces deux directivités. Par exemple, un microphone cardioïde possède une contribution égale de chacune d'elles. Plus on augmente la proportion de la directive bidirectionnelle, plus on tend vers un microphone

supercardioïde, voire hypercardioïde. À l'inverse, augmenter la proportion de la directivité omnidirectionnelle fait tendre le microphone vers une directivité hypocardioïde.

Il existe deux solutions pour agir sur la contribution des directivités primaires. La première consiste à utiliser un labyrinthe acoustique pour changer le milieu acoustique d'une des faces de la membrane. L'autre consiste à avoir une capsule omnidirectionnelle et une seconde bidirectionnelle et de sommer leur tension de sortie.

Les microphones à multidirectivité permettent à l'utilisateur d'influer, soit sur le labyrinthe acoustique, soit sur la sommation des deux capsules. Il n'est pas rare que ces microphones soient moins performants qu'un microphone spécifiquement dédié à une seule directivité.

On retiendra donc :

$$\theta(\alpha) = A + B \cos(\alpha) \text{ ou } A + B = 1$$

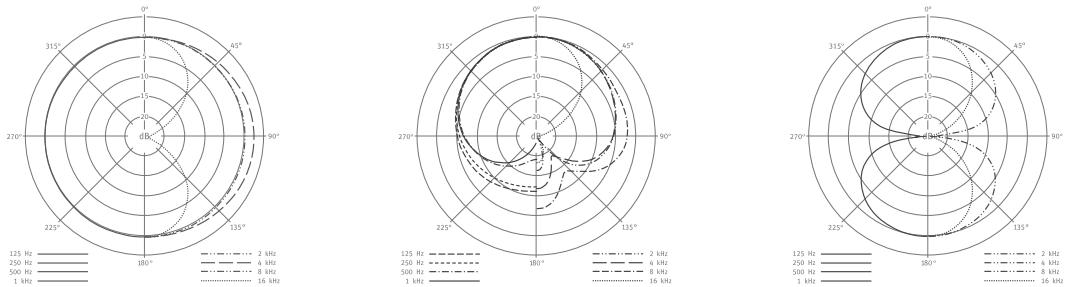


Figure 7.5: Directivités réelles du microphone (U 87)

7.3.4 Directivités réelles & détimbrage

Nous avons jusque là considéré que la directivité d'un microphone était un phénomène indépendant de la fréquence. Or, cela n'est pas vrai. En d'autres termes, la directivité d'un microphone n'est pas la même en fonction de la fréquence de l'onde sonore lui arrivant. Typiquement, un microphone tendra vers une directivité plus resserrée dans le haut du spectre, et vers une directivité plus large dans le bas du spectre.

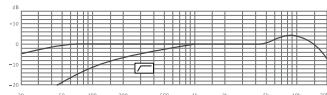


Figure 7.6: Réponse en fréquence du microphone Neumann U87 (omnidirectionnel)

Cela signifie donc que positionner un microphone hors axe face à un évènement sonore n'aura pas seulement un effet sur le niveau du signal en sortie du microphone, mais également sur le

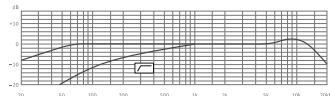


Figure 7.7: Réponse en fréquence du microphone Neumann U87 (cardioïde)

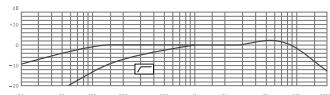


Figure 7.8: Réponse en fréquence du microphone Neumann U87 (figure en huit)

timbre. On appelle alors « **timbré** », un évènement sonore capté plein axe par un microphone, et **détimbré** un évènement sonore capté hors axe.

Ce phénomène est un outil précieux pour les preneurs de son. Par exemple, lorsqu'on enregistre une voix, certains sons sont exagérés par le microphone, particulièrement les « s ». En tournant légèrement le microphone pour placer la voix hors axe, on peut déjà grandement améliorer les problèmes de sifflante avant même de penser à un éventuel traitement ultérieur.

8 Transport de signaux analogiques

Les câbles assurent le transport de signaux électriques. Ces signaux peuvent représenter une onde sonore (précédemment captée par un microphone), une valeur de contrôle pour piloter un appareil (pédale d'expression) ou encore une information numérique (câble USB, ethernet, etc.). Nous aborderons ici les câbles dédiés aux modulations analogiques, et tout particulièrement à l'audio.

Le principe de transport d'un signal analogique est assez simple. Pour une modulation, il faudra au moins un conducteur pour véhiculer le signal, et un conducteur pour la référence de tension, la masse. Pour transporter un signal supplémentaire, on conservera le même conducteur pour la référence et l'on en ajoutera un pour véhiculer le signal supplémentaire, faisant ainsi un total de trois conducteurs.

8.1 Anatomie d'un câble

Les câbles véhiculant un seul signal sont généralement constitués de quatre à cinq composants :

- d'un cœur composé d'un filament (souvent multibrin) en un métal conducteur, véhiculant le signal électrique.
- d'une gaine isolante (plastique) protégeant le cœur
- d'une tresse en cuivre connectée à la masse
- Parfois, d'une feuille de cuivre, enrobant la tresse, permettant de réaliser une cage de faraday et de protéger le cœur des ondes électromagnétiques.
- Enfin, d'une dernière gaine isolante, permettant de protéger l'ensemble du câble.

Ceux permettant de transporter plus de signaux rajouteront des coeurs multibrins entourés de leur gaine isolante. La plupart du temps, les câbles véhiculent un ou deux signaux à la fois, mais certains permettent d'en acheminer beaucoup plus (multipaire, Sub-D). On appelle un câble en fonction de ses connecteurs (ou fiches).

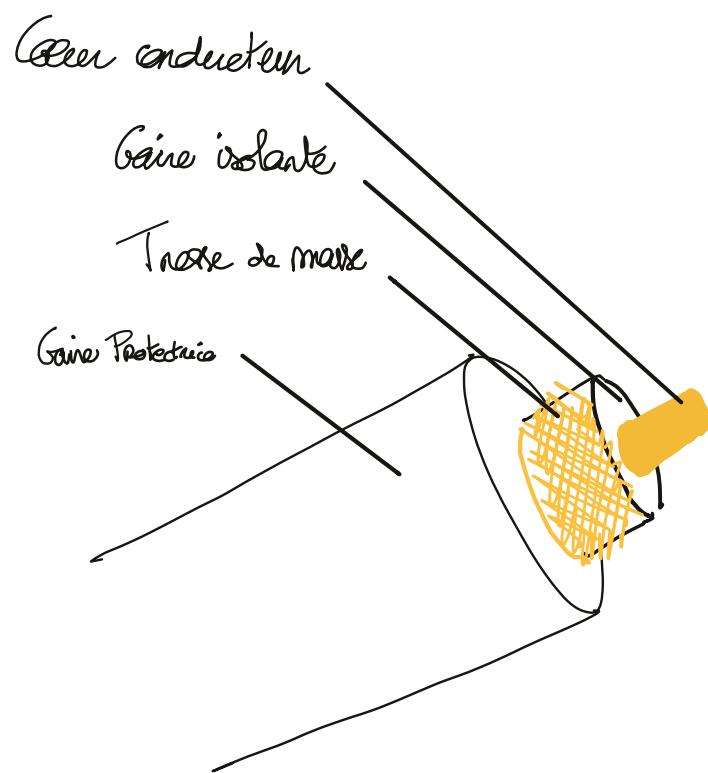


Figure 8.1: Coupe d'un câble

8.1.1 Longueur du câble, son et impédance

Étudions la section d'un câble à modulation unique. Nous pouvons faire plusieurs observations. Le fil conducteur du signal, généralement en cuivre, est d'une longueur non négligeable. Plus cette longueur est importante, plus ce fil aura une **résistance** importante. De plus, ce fil est séparé de la tresse de masse (élément également conducteur) par un isolant. Il existe donc un effet de **capacitance** entre le point chaud et la masse. Enfin, la tresse de masse peut être comparée à une bobine, et possède donc une **inductance**. Nous pouvons donc assimiler un câble à un circuit **RLC** filtrant le haut du spectre audio.

Physiquement, notre description précédente est valide, mais elle est à corrérer à l'**impédance de sortie** de la source. Typiquement, lors de l'utilisation d'un microphone statique, son impédance est suffisamment faible pour que la longueur du câble soit totalement transparente. Certains microphones dynamiques ou à ruban, possédant une impédance plus élevée, peuvent très légèrement souffrir de la longueur du câble.

Ce phénomène d'altération du timbre à cause de la longueur d'un câble a surtout lieu avec les instruments électriques **passifs** (guitares et basses). L'impédance de sorte de ces instruments est tellement grande que l'on peut aisément entendre la différence de son entre deux câbles.

Il est amusant de constater que certains musiciens utilisent ce phénomène, et jouent avec des câbles volontairement trop longs, pour atténuer le haut du spectre. Brian May et Eric Johnson en sont deux exemples.

8.1.2 Les connexions asymétriques

Les connexions asymétriques permettent de transporter un signal entre une source et un récepteur. Il s'agit de la façon la plus simple de connecter deux appareils devant échanger des signaux. Cependant, sur de longue distance, le câble peut se comporter comme une antenne et induire sur le signal certaines ondes électromagnétiques (comme la radio). Ces connexions ne nécessitent qu'un fil conducteur par modulation.

8.1.3 Les connexions symétriques

Le but de ces connexions est de palier au problème des connexions asymétriques. Dans l'appareil émetteur, le signal à transporter est dupliqué, et ce duplicata est inversé en phase. Cette étape s'appelle la **symétrisation**.

C'est deux signaux qui sont appelés point chaud (signal d'origine) et point froid (signal opposé en phase). Sur le trajet du câble, lorsqu'une perturbation électromagnétique est induite sur le signal, celle-ci s'inscrit en phase sur les deux conducteurs (point chaud et point froid). À l'arrivée, l'appareil récepteur inverse la phase du point froid et le somme avec le point chaud.

Cette étape se nomme la **dé-symétrisation**. Ainsi le signal d'origine est sommé en phase, tandis que les interférences sont sommées hors phase et s'annulent.

Les connexions symétriques nécessitent deux fils conducteurs pour chaque modulation.

8.2 Les fiches & connecteurs



Figure 8.2: Jack TS, Jack TRS, Jack Bantam, XLR, Sub-D 25

On appelle « fiche » les éléments mécaniques situés aux extrémités d'un câble et permettant sa connexion à un équipement. Celle-ci nous permet de facilement identifier le type de câble que nous avons entre les mains.

Contrairement aux dires de certains mythes, majoritairement reliée à l'audiophilie, la différence de matériau utilisé pour le contact de la fiche n'a pas d'incidence sur le son.

On rencontre principalement les connectiques jack TS, jack TRS, jack bantam, XLR et Sub-D. Les fiches jack TS ont deux points de connexion. Les fiches jack TRS, XLR et bantam en ont deux. Les Sub-D 25 en possèdent vingt-cinq.

Les fiches jack TS sont souvent utilisées sur les instruments électriques et électroniques (guitare, basse, synthétiseurs, etc.). Ces connexions sont nécessairement symétriques. Les jack TRS sont un peu plus rares et se trouvent généralement sur des instruments avec des sorties stéréophoniques, ou sur des équipements audio possédant des entrées/sorties symétriques. La connectique XLR remplit fondamentalement le même usage qu'un jack TRS, mais offre un verrouillage mécanique, permettant de sécuriser la connexion. On la trouve principalement sur les microphones et sur préamplificateurs. L'avantage du jack TRS est son plus faible encombrement mécanique. On le préfère donc sur les appareils possédant un grand nombre d'entrées/sorties. Le jack bantam se trouve sur les boîtiers de patch. Leur petite taille permet une grande densité de point de connexion. Les patchbay prennent ainsi moins de place. On trouve aussi les connectiques Sub-D 25, principalement pour remplacer des connexions XLR. En effet, une seule connectique Sub-D 25 permet de remplacer huit câbles XLR.

8.3 Exemples pratiques

8.3.1 Câble jack « mono »

Ce type de câble est souvent utilisé sur les instruments électriques. On l'appelle « mono » car il ne véhicule qu'un seul signal.

8.3.2 Câble jack « stéréo »

L'appellation de ce câble est ambiguë. Le terme stéréo fait référence à ses deux voies de connexion, cependant, le nom « stéréo » impliquerait qu'une des voies est destinée à alimenter l'enceinte gauche, et l'autre, le canal droit. Or, ce type de câble peut également convenir pour des connexions symétriques.

On notera qu'il est possible de brancher un câble jack TS dans une fiche TRS. Un seul des canaux sera alors acheminé.

8.3.3 Câble « Y »

Ces câbles possèdent trois connecteurs, et sont le plus souvent équipés de jack TS. Ils permettent de récupérer un signal pour le transmettre sur deux appareils. Attention, la duplication du signal étant passive, on risque un problème d'impédance si l'impédance d'entrée des deux appareils est trop différente. Le cas d'école est souvent rencontré lorsqu'on branche deux casques sur le même amplificateur. Si l'impédance des deux casques est trop différente, un des deux aura presque l'intégralité du niveau du signal alors que l'autre sous-modulera.

8.3.4 Les connexions d'insert

Ces câbles ont la particularité d'avoir une fiche jack TRS et deux fiches jack TS. En studio, on les rencontre très souvent pour insérer un périphérique de traitement dans la chaîne audio. Le « tip » du jack TRS est connecté au « tip » d'un des jack TS qui est connectés sur l'entrée du périphérique. Le deuxième jack TS est raccordé à la sortie de l'appareil et son « tip » est connecté au « ring » du jack TRS.

Ces câbles sont aussi utilisés pour séparer une sortie dite « stéréo » en deux voies « mono ».

8.4 Routage des signaux

En studio d'enregistrement, il n'est pas rare de rencontrer plusieurs cabines de prise de son, chacune équipée de boîtier de patch. Ces boîtiers sont constitués d'un certain nombre d'entrées XLR. On achemine ensuite les sorties de ces boîtiers via des multipaires jusqu'à une « patchbay ». Le rôle de la « patchbay » est de permettre de connecter n'importe quelle entrée (signal provenant d'un microphone) vers n'importe quel préamplificateur.

On trouve évidemment beaucoup d'usage à ces « patchbay ». Elles sont à considérer comme la matrice de routage d'un studio d'enregistrement.

9 Les préamplificateurs

Le rôle d'un préampli est de réaliser une amplification en tension du signal ainsi que de diminuer son impédance. Cette opération est indispensable pour permettre à notre signal de traverser le reste de la chaîne du traitement audio.

Dans les débuts de l'enregistrement audio, les préamplificateurs de microphone étaient des appareils passifs rudimentaires utilisant généralement un couplage par transformateur pour assurer l'isolation et adapter l'impédance du microphone au reste du système audio.

Au fur et à mesure que les techniques d'enregistrement progressaient, les préamplis sont devenus de plus en plus sophistiqués et ont commencé à intégrer des composants électroniques actifs (tube à vide, transistors, circuits intégrés). Ces préamplis actifs offrent une plus grande flexibilité et de meilleures performances.

L'une des principales évolutions de la technologie des préamplis a été l'avènement de l'électronique à semi-conducteurs, remplaçant ainsi la technologie des tubes à vide. Ces préamplis à transistors sont plus petits, plus légers et plus fiables que leurs homologues à tubes, et ils offrent généralement un bruit plus faible et une gamme dynamique plus étendue.

Il n'est pas rare que les préamplificateurs soient choisis pour leur « couleur » sur le signal. Nous commencerons par aborder ces outils d'un point de vue pratique pour enfin déboucher sur cette question.

9.1 Informations techniques des préamplificateurs

Le **gain** (quantité d'amplification) : Les préamplificateurs pour microphones fournissent généralement un gain compris entre 20 et 60 dB, qui est nécessaire pour amplifier les signaux de faible niveau provenant d'un microphone pour atteindre un niveau utilisable. Le gain d'un préampli est généralement réglable, ce qui permet à l'utilisateur de définir le niveau optimal pour un microphone et une application donnée.

L'**impédance** d'entrée d'un préamplificateur de microphone est un paramètre critique, car elle détermine la charge imposée par le préamplificateur au microphone. Une mauvaise adaptation d'impédance peut provoquer des interactions indésirables entre le microphone et le préampli, entraînant, entre autre, des modifications de la réponse en fréquence.

La **réponse en fréquence** d'un préamplificateur est une considération importante, car elle détermine la manière dont le préamplificateur va affecter le son du microphone. Une réponse en fréquence plate est généralement souhaitable.

Le **plancher de bruit** d'un préamplificateur est une mesure du bruit résiduel ajouté au signal par le préamplificateur lui-même. Un plancher de bruit faible est nécessaire pour maintenir un bon rapport signal/bruit dans le système audio. Cette donnée est évidemment d'autant plus critique que le niveau du signal d'entrée est faible. Par extension, on désigne la plage dynamique d'un préamplificateur de microphone comme la gamme comprise entre le niveau maximal du signal que le préamplificateur peut accepter et le plancher de bruit.

La distorsion est une mesure des modifications indésirables du signal audio qui sont introduites par le préampli. De faibles niveaux de distorsion sont souhaitables, car ils garantissent que le son du microphone est capturé et amplifié avec précision. On évoquera alors la **distortion harmonique**, qui, comme son nom l'indique, augmente le signal d'entrée d'harmoniques supplémentaires. On rencontre aussi la **distortion d'intermodulation**. Admettons que nous envoyons en entrée d'un équipement audio deux sons purs de fréquences F1 et F2. La distortion d'intermodulation engendre deux nouvelles fréquences F1+F2 et F1-F2. On cherche donc à maintenir la contribution de ces deux résurgences le plus bas possible.

Ci-dessous, en guise d'exemple, on trouve le tableau des spécifications du préampli Rupert Neve Design 511 :

Measured at Main Output, un-weighted, 22 Hz - 22 kHz, source impedance 150 Ohm balanced. Noise performance can vary depending on the 500 series and / or interference from stray magnetic fields. **NOISE** Unity Gain: Better than -103 dBV Gain @ +66 dB: Better than -60 dBV Equivalent Input Noise: -125 dB

FREQUENCY RESPONSE Main output, no load. +/- 0.1 dBu from 10 Hz to 31.5 kHz -2.6 dB @ 120 kHz

MAXIMUM OUTPUT LEVEL +23 dBu

TOTAL HARMONIC DISTORTION AND NOISE, NO SILK @ 1 kHz, +20 dBu output level, no load: Better than 0.0025% @ 20 Hz, +20 dBu output level, no load: 0.025% Typical (2nd and 3rd harmonic)

TOTAL HARMONIC DISTORTION AND NOISE WITH SILK ENGAGED @ 100 Hz, +20 dBu input level, no load. **TEXTURE** @ min: 0.015%, mostly 3rd harmonic typical **TEXTURE** @ max: 2%, mostly 2nd harmonic typical

GAIN Unity up to +66 dB in 6 dB steps. Trim continuously adjustable from -6 dB to +6 dB.

PHANTOM POWER Supplied by the 500 series rack power supply. Switch selectable on faceplate.

HIGH PASS FILTER Continuously variable swept frequency from 20 Hz to 250 Hz. Slope: 12 dB/Octave

POWER REQUIREMENTS @ +/-16VDC, 100mA

9.2 Critères de choix d'un préamplificateur

Le critère de première importance dans le choix d'un préamplificateur est son gain maximal. Plus l'amplification disponible est grande, plus le préampli sera capable de répondre à des situations exigeantes, telles que l'enregistrement d'un évènement sonore à faible niveau, ou l'emploi d'un microphone à faible sensibilité.

Le second critère important dans le choix d'un préampli est sa réponse en fréquence. Théoriquement, celle-ci doit être la plus neutre possible. Une certaine coloration peut être acceptée (voire souhaitée), mais celle-ci doit rester raisonnable pour répondre à des critères d'utilisations professionnelles.

La réponse en transitoire est un autre élément important. Certains préamplis auront tendance à adoucir la sensation d'attaque des sources. Cet effet n'est pas souhaitable.

Enfin le rapport signal sur bruit doit être le plus grand possible. Nous cherchons toujours à rajouter le moins de bruit possible sur le chemin de notre signal.

9.3 Les technologies de préampli

Nous avons vu dans le chapitre trois qu'il existe trois familles de composants électroniques permettant d'amplifier le signal : les tubes, les transistors et les circuits intégrés. On retrouve donc des topologies de circuit de préamplificateurs utilisant chacun de ces composants.

Chacune de ces topologies offre de très légère variation de son lorsque les préamplis sont poussés dans leur retranchement (seuil de saturation). Comme pour les microphones, il est délicat de parler de son « à tube » ou « à transistor ». De plus, l'influence sur le son d'un préamplificateur apparaît en pratique comme très marginale par rapport au positionnement du microphone.

9.4 Les réglages d'un préampli

Un préampli propose souvent les réglages suivants :

- Un potentiomètre de gain (qui est souvent remplacé par un sélecteur cranté, plus précis, pour les modèles haut de gamme).

- Un bouton activant l'alimentation fantôme. En effet, c'est bien le préampli qui génère cette tension d'alimentation pour les microphones statiques.
- Un bouton d'inversion de phase.
- Un coupe-bas.

10 La conversion analogique numérique

10.1 La nécessité de la conversion analogique numérique

Durant toute la période de l'audio analogique, le support de stockage de prédilection fut la bande magnétique. Cependant, celle-ci n'offre pas un rapport signal sur bruit très satisfaisant, limitant alors la dynamique musicale enregistrable. De plus, la bande a également un coût non négligeable. On a donc cherché à remplacer ce support afin de résoudre ces deux problèmes. Le stockage numérique offre, sous certaines conditions, une dynamique bien supérieure à celle des supports analogiques.

Une représentation numérique de l'audio permet aussi la réalisation de traitement délicat, voire impossible, en analogique. On pense, par exemple, aux algorithmes de réverbération, d'écho et de « pitch shifting » (modification de la hauteur d'un son).

Enfin, nos principaux outils de manipulation du son sont aujourd'hui informatiques. Dès lors, une représentation numérique des signaux est toute indiquée pour les manipuler grâce à nos ordinateurs. Il en découle donc une nécessité de bien maîtriser les principes entourant la numérisation des signaux.

En français, le « digital » est un anglicisme. Le mot correct est donc bien « numérique », et non « digital », qui qualifie ce qui a rapport au doigt.

10.2 Théorie de l'échantillonnage

10.2.1 D'un signal continu vers un signal échantillonné

Une des caractéristiques principales d'un signal analogique est qu'il est continu. Une fonction, en mathématique, est dite continue si elle est définie en n'importe quel instant. Afin d'être numérisé, un signal doit donc être dénombré. En effet, la notion d'infini imposé par la continuité du signal n'a pas d'existence en numérique.

La numérisation du signal est comparable à l'utilisation d'un multimètre pour mesurer une tension. Un convertisseur va prélever la valeur du signal, de façon régulière, au cours du temps.

Afin de correctement numériser un signal, il convient de définir deux paramètres :

- la vitesse de prélèvement, ou **fréquence d'échantillonnage**
- la plage de valeur permise pour le signal, ou **résolution de quantification**

10.2.2 La fréquence d'échantillonnage

Cette fréquence définit le nombre de prélèvements par seconde. Par exemple, un morceau édité sur un CD audio a une fréquence d'échantillonnage de 44 100 Hz (44,1 kHz), cela signifie que le signal est mesuré 44 100 fois par seconde.

La fréquence de travail la plus courante est 48 kHz, mais l'on rencontre parfois des valeurs supérieures, multiple de celle-ci : 96 kHz, 192 kHz, etc. Cette augmentation proportionnelle de la fréquence d'échantillonnage s'appelle **suréchantillonnage**. Certains techniciens espèrent ainsi améliorer la qualité de l'enregistrement. Ce suréchantillonnage à un coût en ressource CPU et en espace de stockage. Un flux audio échantillonné à 96 kHz demande deux fois plus de ressource et d'espace qu'un flux échantillonné à 48 kHz. Cette valeur initiale de 44 100 Hz (ou 48 kHz) n'a pas été choisie au hasard. Pour la comprendre, il faut revenir au phénomène physique que nous cherchons à numériser.

Rappelons que le son est une onde mécanique, et nous l'entendons lorsqu'elle oscille dans une plage de fréquence comprise entre 20 Hz (très grave) et 20 000 Hz (très aigu). Il faut donc que notre système de numérisation soit capable de reproduire une fréquence maximale allant jusqu'à 20 000 Hz. Pour cela, nous utilisons les résultats des travaux des chercheurs Harry Nyquist et Claude Shannon (tous deux ayant travaillé aux laboratoires Bell).

Le **théorème de Shannon-Nyquist** stipule que, pour être capable d'échantillonner un signal de fréquence f , la fréquence d'échantillonnage doit au moins être de $2f$. Ainsi, un ensemble de points généré par une fréquence inférieure à f ne peut correspondre qu'à cette seule et unique fréquence. Notre plage d'écoute étant limitée à 20 kHz, la fréquence d'échantillonnage minimale dont nous avons besoin est de 40 kHz.

Que se passe-t-il si la fréquence du signal dépasse la moitié de la fréquence d'échantillonnage ? Dans ce cas, la vitesse de prélèvement n'est plus suffisante et nous observons l'apparition de nouvelles fréquences ne provenant pas du signal original. Ce phénomène se nomme **repliement spectral**.

10.2.3 La résolution de quantification

La résolution de quantification permet de définir la plage de valeur dynamique permise dans le système numérique. Celle-ci s'exprime en bit. Par exemple, si nous prenons un convertisseur travaillant en 8 bit. Le nombre de valeurs que peut prendre un signal numérisé par un tel convertisseur est de $2^8 - 1 = 255$ en base 10. Admettons que ce convertisseur accepte des signaux ayant une tension en entrée variant entre +15V et -15 V, celles-ci seront **échelonnées**

sur 255 valeurs. Si maintenant, ce convertisseur travaille en 16 bit, il y aura 65 535 échelons. La précision de mesure de la dynamique du signal n'est donc pas du tout la même.

En pratique, augmenter la résolution de quantification permet principalement de définir le niveau de bruit du convertisseur. Plus la résolution est élevée, plus le bruit se retrouvera faible. En 8 bit, l'écart entre le niveau maximal d'un signal et le bruit est de 48 dB, en 16 bit cet écart est de 96 dB, en 24 bit, 144 dB. On peut approximativement calculer cette dynamique par la relation suivante :

$$\Delta_L \approx 6 \times N_{bits}$$

La résolution de quantification standard en enregistrement est 24 bit. La plage dynamique est telle qu'elle rend le travail d'enregistrement beaucoup plus souple sur les niveaux d'acquisition des différentes sources.

10.3 Quelle influence sur le signal ?

Le son numérique a longtemps eu la réputation d'être « dur », particulièrement dans le haut du spectre. Cela s'explique assez facilement par le fonctionnement des premiers convertisseurs.

En effet, toute la difficulté de fabrication d'un convertisseur réside dans la réalisation d'un filtre anti-repliement, pour prévenir le repliement spectral. Ce filtre doit enlever toutes les fréquences au-dessus de la moitié de la fréquence d'échantillonnage, sans pour autant affecter le spectre audible. Ce type de filtre est extrêmement délicat à réaliser en analogique. Cependant, ce problème est résolu grâce à une méthode d'échantillonnage appelée « sigma-delta » (voir ci-dessous).

Le repliement spectral n'apparaît pas seulement lors de la conversion. Il peut également survenir lors de l'utilisation de certains traitements (saturation, simulation analogique, compresseurs). Lorsqu'il devient audible, le repliement spectral se caractérise par l'apparition de fréquences **non harmoniques** souvent qualifiées de « dures » et désagréables. Il est cependant bon de rappeler que ce phénomène, certes bien réel, apparaît dans des conditions de saturation du signal importante et sur des sources sonores riches en hautes fréquences.

Malgré la dure vie que mène parfois la réputation du son numérique, il est important de rappeler qu'il a apporté un grand nombre d'avantages sur le son analogique, **y compris sur des questions de rendus sonores**. Par exemple, la dynamique est bien plus importante, la distorsion involontaire du signal infime et l'ajout de bruit inexistant.

10.4 La conversion sigma-delta

Aujourd’hui, les convertisseurs ne travaillent pas directement à 44,1 kHz/16 bit ou 48 kHz/24 bit. Ils utilisent à la place un procédé appelé échantillonnage sigma-delta. Le principe est d’utiliser une fréquence d’échantillonnage très rapide (384 kHz) et de coder la dynamique du signal, en relatif, sur un seul bit (ce bit prend une valeur de 1 si le nouvel échantillon est plus fort que l’ancien, 0 pour le cas inverse). Les formats de travail que nous utilisons sont générés après cette première étape.

L’intérêt de cette méthode est double :

- Le signal est suréchantillonné dès l’enregistrement
- Les filtres permettant d’éviter le repliement spectral sont donc très simples à réaliser

11 Transports de signaux numériques

Après être passé au travers d'un convertisseur, l'audio est représenté par un ensemble d'échantillons. En raison de cette représentation particulière, on trouve un certain nombre de protocoles de communication de signaux numériques entre appareils. Il ne s'agit plus ici de faire transiter une tension analogue à celle de la variation de pression d'une onde sonore, mais plutôt une tension représentant des mots binaires (bits).

La variation de tension de ces signaux s'apparente à une onde carrée. Ces signaux périodiques possèdent un état « haut » et un état « bas », très pratique pour communiquer des nombres binaires.

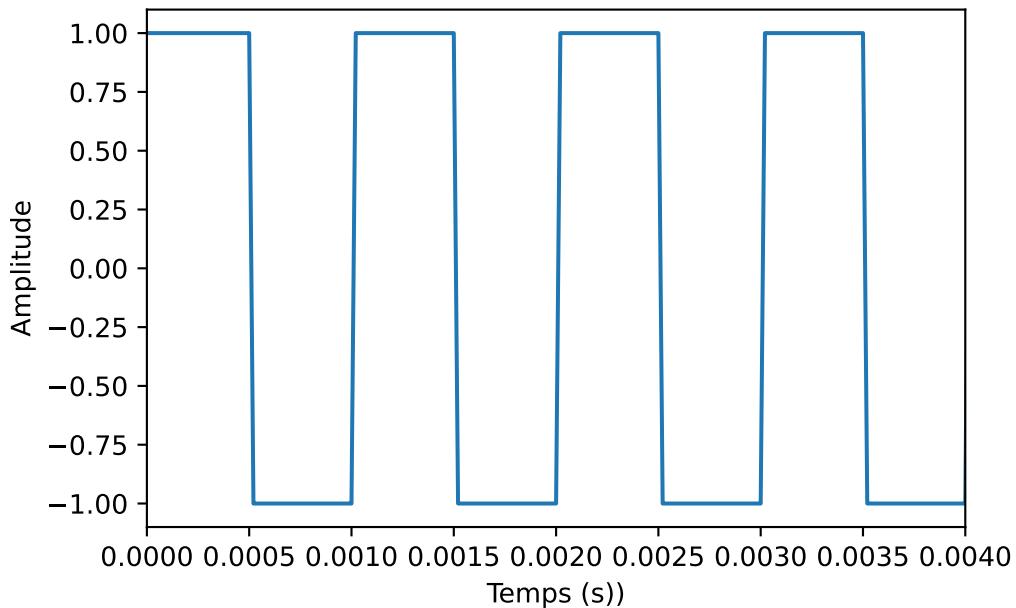


Figure 11.1: Exemples de signal numérique (Non Return to Zero)

11.1 Les liaisons point à point

Les protocoles ci-dessous permettent de transmettre un ou plusieurs signaux numériques entre deux appareils.

11.1.1 AES3 ; AES/EBU

L'AES3 est un protocole défini par l'*Audio Engineering Society* et par l'*European Union Broadcast*. Il est principalement destiné aux appareils audio dits « professionnels ». Il permet de véhiculer deux canaux audio, à une fréquence d'échantillonnage maximal de 48 kHz, via une fiche XLR ou BNC (coaxial).

La S/PDIF est relativement proche de l'AES3, plutôt utilisé dans les équipements grand public, utilisant des câbles coaxiaux (sur fiches RCA) ou optiques (fiche toslink).

11.1.2 ADAT

L'*ADAT lightpipe*, souvent abrégé ADAT, est un autre protocole de transmission de signaux numériques. Il a été développé par Alesis pour fonctionner avec les magnétophones à bandes numériques de la même marque. ADAT signifie en fait « Alesis Digital Audio Tape ». On retrouve ce protocole sur un grand nombre d'appareils, notamment les interfaces audio, afin d'augmenter le nombre d'entrées/sorties accessibles.

L'ADAT peut transporter jusqu'à huit canaux à 44.1/48 kHz, quatre canaux à 88.2/96 kHz et deux canaux à 176.4/192 kHz. Le débit d'information est donc constant, doubler la fréquence d'échantillonnage divise par deux le nombre de canaux.

La connectique la plus courante pour l'ADAT est la fibre optique avec fiches toslink.



Figure 11.2: Câble Toslink

11.1.3 MADI

Le MADI, ou AES10, est un protocole permettant d'acheminer un grand nombre de canaux. On peut donc récupérer soixante-quatre canaux audio à une fréquence de 44.1/48 kHz. Comme pour l'ADAT, le nombre de canaux est divisé par deux à chaque doublement de la fréquence d'échantillonnage.

Ce protocole se retrouve fréquemment dans le monde de l'audio professionnel. Les connexions entre appareils supportant le MADI peuvent se faire soit avec des fibres optiques, soit sur câble coaxial (fiches BNC). Certains constructeurs, comme DIGICO, ont choisi les câbles RJ45 comme support d'acheminement.



Figure 11.3: ?(caption)

Fibre optique et câble BNC

Figure 11.4: ?(caption)

11.2 Les réseaux audio numériques

Aujourd'hui, dans le monde du spectacle vivant, la plupart des salles de spectacle sont équipées avec des solutions de transmission des signaux audio sur réseau. Ces solutions se retrouvent aussi de plus en plus dans les studios d'enregistrement et de production audiovisuelle.

Il existe plusieurs protocoles permettant le déploiement de tels dispositifs, mais leur logique fondamentale reste identique. Chaque appareil capable de se connecter au réseau audio peut recevoir et envoyer un flux audio à n'importe quels autres appareils appartenant au même réseau.

Les réseaux audio sont régis par les mêmes règles que les réseaux informatiques. Chaque appareil pouvant être connecté à un réseau est identifiable par une adresse matérielle unique, appelée **adresse MAC**. Lorsqu'un appareil est connecté sur un réseau, il faut lui attribuer une adresse logique appelée **adresse IP**. Il y a ici deux façons de faire. Soit l'utilisateur attribue manuellement une adresse différente à chaque machine (solution préférée en audio, mais fastidieuse lorsque le réseau comprend un grand nombre d'appareils), soit le réseau possède

un **serveur DHCP** qui se chargera d'attribuer une adresse IP unique à chacun des appareils connectés. Cet outil est généralement intégré dans un appareil nommé **routeur**, permettant d'interconnecter plusieurs appareils ainsi que de gérer le routage des flux d'information. Une fois les appareils interconnectés, chaque constructeur de solutions audio sur IP fournit un logiciel de routage de l'audio entre les appareils.

Les principaux acteurs industriels des réseaux audionumériques sont Audinet avec DANTE, ALC NetworX (appartenant à Lawo) avec Ravenna, et les protocoles open source AES67 et AVB.

12 Introduction à l'informatique musicale

(En cours d'écriture)

12.1 Fonctionnement d'un ordinateur

12.2 Les systèmes d'exploitation

12.2.1 Linux

12.2.2 Apple MacOS

12.2.3 Microsoft Windows

12.3 Les pilotes audio

12.3.1 ASIO

12.3.2 CoreAudio

12.3.3 ALSA

12.3.4 Jack Audio

12.4 Les Stations de Travail Audio-Numérique (DAW)

12.4.1 Le moteur audio

12.4.2 Les fonctionnalités

12.5 Les protocoles de transmission d'informations

12.5.1 MIDI

12.5.2 OSC

13 Enceintes et amplificateurs

Un haut-parleur est un appareil permettant de transformer une énergie électrique en énergie mécanique. On l'appelle, plus spécifiquement, un transducteur électroacoustique. Fondamentalement, il s'agit d'un appareil extrêmement proche d'un microphone. D'ailleurs, il est possible d'utiliser un haut-parleur comme microphone, et vice-versa. Les haut-parleurs ont de faibles rendements, il est donc nécessaire d'amplifier les signaux grâce à des amplificateurs de puissance en amont.

De tous les équipements audio nécessaires pour réaliser une prise de son, les enceintes (et les casques) sont certainement les plus importants. En effet, c'est à travers leur prisme que nous pourrons écouter et contrôler notre travail. Il est donc crucial d'utiliser des écoutes regroupant un certain nombre de critères et, surtout, de les connaître sur le bout des doigts.

13.1 Anatomie d'un haut-parleur

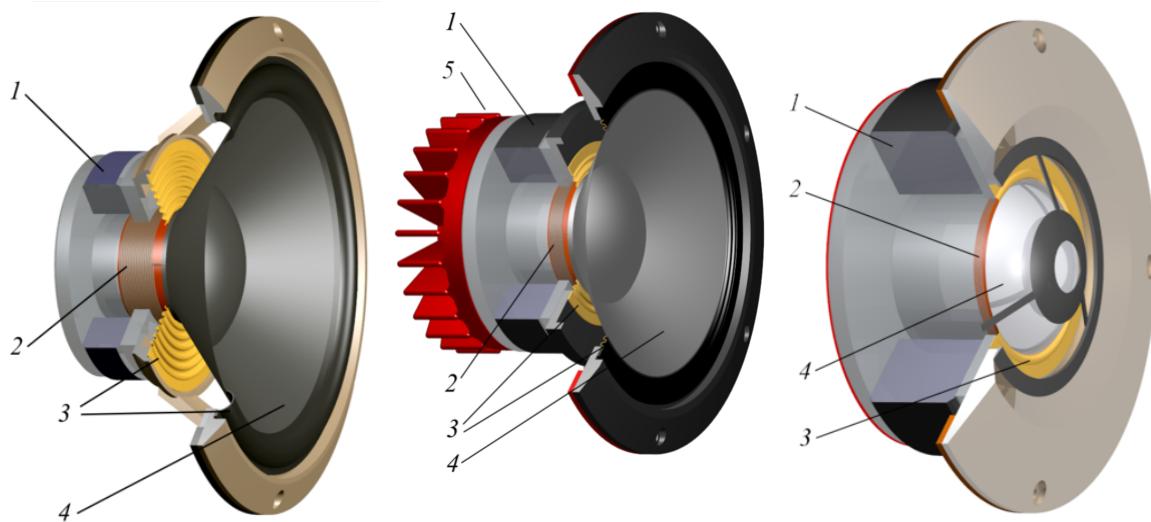


Figure 13.1: Coupe de hautparleur (dans l'ordre, woofer, mid-range, tweeter). Infographie par Svjo, CC BY-SA 3.0

Sur les schémas ci-dessus, nous trouvons les éléments suivants :

1. L'aimant
2. La bobine
3. La suspension
4. La membrane

La plupart des haut-parleurs sont qualifiés de **dynamiques**. La membrane du haut-parleur est reliée à une bobine, elle-même entourée par un système d'aimants. Lorsqu'un courant est appliqué aux bornes de cette bobine, sa position change dû à la modification du champ électromagnétique. Si le courant oscille, la bobine oscille de façon analogue, entraînant la membrane et permet donc la reproduction du son.

On appelle généralement « subwoofer » les haut-parleurs conçus pour retranscrire les fréquences très graves (20-200 Hz), « woofer » les haut-parleurs dédiés aux fréquences graves (50 Hz à 1000 kHz), « mid-range » les haut-parleurs du médium (1 kHz à 6 kHz), et « tweeter », ceux de l'aigu (au-delà de 5 kHz).

Le haut-parleur est sans doute l'appareil audio le plus imparfait qui soit. Il est sujet à de nombreuses sources de distorsion du signal.

Nous ne savons pas fabriquer des haut-parleurs capables de reproduire uniformément toutes les fréquences. Ces derniers sont souvent spécialisés dans une certaine plage de fréquence. La plupart des enceintes de monitoring utilisent 3 voies : deux actives (utilisant des haut-parleurs) pour l'aigu et le médium, et une passive (évent avant ou arrière) pour le grave. L'utilisation de plusieurs voies imposent donc l'utilisation de filtres induisant un déphasage de certaines fréquences.

Également, un haut-parleur peut être approché par un modèle « masse-ressort ». Cela signifie qu'il y a une certaine inertie à sa mise en action et une certaine inertie à sa mise en arrêt. L'enceinte idéale devrait posséder une inertie nulle. Cette inertie est potentiellement responsable d'un adoucissement des transitoires et d'une sensation de flou.

13.2 Amplification et impédance

Nous avons précédemment abordé le préamplificateur, qui permet d'amplifier la tension d'un signal audio analogique et d'en baisser son impédance. On appelle alors « amplificateur » un amplificateur de **puissance**. On rappel que la puissance d'un signal s'exprime par la relation ci-dessous. On cherche donc à augmenter la tension **et** l'intensité du signal.

$$P = U \times I$$

Nous pouvons rapidement aborder la notion de classe d'amplification. En audio, nous n'utilisons que les classes A, AB et D. La classe A utilise un transistor (ou tube) pour amplifier l'ensemble du signal. Elle possède un très mauvais rendement. Cela signifie qu'il

faut fournir beaucoup d'énergie au transistor pour un faible gain sur la puissance du signal. La classe AB utilise deux transistors, un pour les alternances négatives et un pour les alternances positives. Le rendement est meilleur que pour la classe A. Cependant, le point de raccordement entre les deux transistors est assez sensible et peu générée de la distorsion sur le signal (distorsion de croisement). La classe D, aussi appelée à tort « numérique », utilise un transistor afin d'indiquer l'état du signal. On retrouve donc la même idée que dans l'échantillonnage du signal. Ces amplificateurs offrent un **excellent rendement**.

Nous avons également abordé précédemment la notion d'adaptation d'impédance en tension, sans aborder l'adaptation d'impédance en puissance. Pour rappel, afin de préserver la tension entre deux appareils A et B, nous faisons en sorte d'avoir une faible impédance à la sortie de l'appareil A et une grande impédance à l'entrée de l'appareil B. Pour préserver la puissance du signal, ce paradigme ne fonctionne plus. On cherche alors à avoir la même impédance entre la sortie d'un appareil et l'entrée d'un autre. En pratique, on raccordera, sur la sortie « 8 ohms » d'un amplificateur, un haut-parleur ayant une impédance de « 8 ohms ».

On remarque qu'ici, les impédances sont extrêmement faibles. Les impédances typiques des haut-parleurs (et donc des sorties d'amplificateurs en puissance) sont 4, 8 et 16 ohms.

La plupart des enceintes de monitoring ont aujourd'hui une amplification de classe D directement intégré.

13.3 Puissance et sensibilité

On trouve généralement deux mesures de la puissance pour les enceintes. La puissance **crête à crête** (peak) et la **puissance moyenne** (RMS, ou **Root Mean Square**). La puissance se calcule grâce à la formule suivante :

$$P = \frac{U^2}{Z}$$

P est la puissance, U la tension et Z l'impédance. Pour réaliser la mesure de puissance d'un haut-parleur, on le soumet à un signal test, en général un bruit rose, pendant plusieurs heures. La puissance crête à crête se calcule grâce à la tension crête à crête (aussi dite maximale). Par exemple, pour une tension maximale de 100 V, sous une impédance de 8 ohms, on trouve une **puissance crête à crête** d'environ 1200 watts (abrévégé W). On considère généralement que la tension moyenne d'un signal est six décibels plus petite que sa tension maximale. En reprenant notre exemple, pour une tension maximale de 100 V, on a une tension moyenne de 50 V, sous une impédance de 8 ohms, on trouve une **puissance moyenne** d'environ 300 W.

Alors, il faut donc faire très attention sur les spécifications données par les constructeurs, parfois volontairement floues. À défaut, si on lit que la puissance admissible d'un haut-parleur

est de 1000 W, on supposera par défaut qu'il s'agit d'une puissance crête à crête. On ne dépassera donc pas une puissance moyenne d'amplification de 250 W.

La **sensibilité** est une mesure du niveau sonore à un mètre de l'enceinte pour une puissance RMS d'entrée d'un watt. Grâce à cette valeur, on peut calculer le niveau sonore produit par le haut-parleur à diverses distances et en fonction de différentes puissances d'entrées. On comprend aussi que la puissance électrique admissible dans l'enceinte ne donne que peu d'information sur son niveau sonore de sortie.

13.4 Conseils pratiques

13.4.1 Choisir une paire d'écoutes

Choisir une paire d'enceintes peut sembler être un exercice difficile. Il existe énormément de modèles, coûtant de quelques dizaines d'euros à plusieurs dizaines de milliers.

Il y a cependant plusieurs critères assez objectifs pour évaluer la qualité d'une enceinte :

1. La réponse en fréquence : l'enceinte flatte-t-elle particulièrement une zone du spectre ? En délaissant-t-elle une autre ?
2. La réponse en transitoires : les attaques sont-elles respectées ? Retrouve-t-on l'énergie initiale du signal ?
3. La linéarité en fonction du volume : a-t-on une sensation de compression du signal lorsque l'on augmente le niveau envoyé dans l'enceinte ?
4. Le centre fantôme : le centre du système stéréophonique paraît-il stable ? Paraît-il précis ?
5. La couleur sonore de l'enceinte : a-t-on plaisir à écouter du son et de la musique sur ce système ?

13.4.2 Placer correctement son écoute

Afin de satisfaire les critères de la stéréophonie, il convient de respecter les règles suivantes :

- Les deux enceintes doivent être de même marque, de même modèle et appairée. Si une des membranes a dû être changée sur l'une d'elle, l'autre aurait dû recevoir la même opération.
- Les deux enceintes doivent être séparées par un angle de 60°
- L'auditeur doit être placé à équidistance des deux haut-parleurs, et regarder vers le milieu du segment formé par les deux enceintes.

Une fois ces critères respectés, voici quelques conseils sur le placement des enceintes dans une pièce :

On préférera des pièces de grandes tailles, afin de repousser au maximum le temps d'arrivée des premières réflexions. Le système stéréophonique devrait être positionné dans un souci de symétrie : l'enceinte de gauche ne devrait pas être plus proche d'un mur que l'enceinte de droite, par exemple. Dans le cas de petit espace, on préféra coller les enceintes contre un mur. Cela permettra de supprimer l'influence d'une des premières réflexions au prix de l'augmentation du niveau de grave. Il est vivement recommandé de procédé au traitement, même minimal, acoustique de la pièce de travail, à commencer par les zones de réflexions premières et par les angles (ou le grave va s'accumuler). Si le traitement acoustique n'est pas envisageable, il convient de privilégier une écoute à faible niveau et une proximité maximale avec les enceintes.

13.5 L'écoute au casque

Le casque est un outil permettant d'écouter un signal tout en s'extrayant de son environnement (acoustique et/ou bruit). Cependant, de par son mode de fonctionnement, à savoir deux haut-parleurs placés dans une enceinte en contact direct avec les oreilles, il génère un certain nombre de déformations.

Premièrement, la stéréophonie écoutée au casque est hypertrophiée. En effet, dans ces conditions d'écoutes, l'oreille gauche n'entend que le haut-parleur gauche et l'oreille droite n'entend que le haut-parleur droit.

Deuxièmement, il est très difficile de trouver des casques avec une réponse en transitoire satisfaisante. Il convient donc d'être excessivement prudent lorsque l'on mix du contenu percussif sur un casque.

Troisièmement, les casques sont encore moins linéaires en fréquence que les haut-parleurs, il convient là aussi d'être très prudent lors de la réalisation d'un mixage.

Ces défauts peuvent être compensés par l'habitude et la connaissance du système d'écoute, mais la transportabilité d'un mixage (à savoir, sa compatibilité avec d'autres systèmes d'écoute) réalisé au casque est souvent discutable.

13.5.1 Casque fermé ou casque ouvert ?

Le casque fermé, comme son nom l'indique, propose une fabrication enfermant le haut-parleur dans une enceinte close. Cette méthode de fabrication offre l'avantage d'isoler celui qui écoute de l'environnement, mais aussi d'isoler l'environnement de ce qui est diffusé dans le casque. Par contre, ces casques ont souvent une réponse en fréquence très accidentée, et ne sont pas recommandés pour le mixage. Il est par contre vivement recommandé pour les musiciens en session de prise de son.

Le casque ouvert, à l'inverse de son homologue fermé, n'offre aucune isolation acoustique, au prix d'une meilleure réponse en fréquence du casque. Ces casques sont tout indiqués pour le mixage, mais beaucoup moins pour des situations de prise de son.

partie III

Méthodologie de prise de son

14 Ecoute critique : première partie

15 Mono et multi-microphonie

(En cours d'écriture)

16 La prise de son au couple

La prise de son au couple stéréophonique regroupe l'ensemble des techniques de prise de son dédié au système d'écoute stéréophonique (deux enceintes séparées de 60° et orientées vers un auditeur placé à équidistance des deux transducteurs).

Ces techniques permettent une bien meilleure représentation des espaces des prises de son ainsi qu'une localisation des différents éléments enregistrés dans cet espace.

16.1 Généralités sur les mécanismes de la localisation du son par l'oreille humaine

Afin de mieux comprendre comment fonctionne un couple de prise de son, il convient d'étudier rapidement les principes fondamentaux de notre écoute.

Notre capacité à localiser les sons dans l'espace repose principalement sur deux mécanismes :
+ La différence de temps + La différence de niveau

16.1.1 La localisation par différence de temps

Nos oreilles sont espacées, d'environ 15 à 25 cm. Cette distance implique qu'un son émis plus proche de l'oreille droite arrivera également plus tôt qu'à l'oreille gauche. Cet écart de temps, de quelques millisecondes, est suffisant pour donner à notre cerveau un indice sur la localisation du son.

Afin de sentir l'ordre de grandeur en jeu, calculons la différence de temps (Δt) maximale pour un individu possédant un écart d'oreille de 20 cm.

On sait que la célérité du son dans l'air vaut $c = 340m.s^{-1}$, et est invariant en fonction de la fréquence. On sait également que $c = \frac{d}{t}$.

Dès lors, si on pose $d = 20cm$ soit $d = 0.2m$, on peut en déduire que :

$$t = \frac{d}{c} \iff t = \frac{0.2}{340} \approx 0.0006s \approx 0.6ms$$

Afin de mettre en relief ce résultat, il est communément admis que l'oreille humaine commence à faire la différence entre deux répétitions d'un même son à partir de 20ms.

16.1.2 La localisation par différence d'intensité

A priori, l'espace entre nos deux oreilles n'est pas creux. La densité de notre crâne et de son contenu va réfléchir et absorber une partie des fréquences rencontrées.

Également, la partie externe de nos oreilles, appelées pavillon, permet, grâce à sa forme, de donner une directivité à notre écoute.

En d'autres termes, notre tête et le pavillon de nos oreilles se comportent comme un filtre, variant en fonction de l'angle d'incidence de la source. Cette altération du timbre n'est pas perçue comme une coloration, mais bien comme une information de localisation. La modélisation mathématique de ces filtres se retrouve dans la littérature scientifique sous le nom **HRTF**.

Cette atténuation séquentiellement dépendante est décisive dans notre capacité à localiser les sons. On la retrouve communément sous le nom Δi .

16.1.3 Prévalence fréquentielle de ces deux phénomènes

Il est communément admis que le Δt aura une efficacité maximale dans les basses fréquences, et le Δi dans les hautes fréquences.

16.2 Principes de la prise de son au couple

Pour créer son effet stéréophonique, les couples de prise de son utilisent les mêmes mécanismes que notre écoute naturelle :

- La différence de temps
- La différence d'intensité

Il va de soi que, pour fonctionner de façon optimale, les microphones utilisés pour réaliser une prise de son stéréophonique doivent être de même marque, de même modèle et appairée.

L'appairage garantit que les microphones aient des caractéristiques techniques suffisamment proches pour être considérés comme identiques.

Afin de manipuler ces mécanismes, le preneur de son peut jouer sur les paramètres suivant :

- La directivité des microphones
- L'angle entre les capsules
- La distance entre les capsules

Modifier chacun de ses paramètres influe sur l'**angle de prise de son**. Plus l'angle de prise de son est faible, plus l'impression de stéréophonie sera grande. Plus l'angle de prise de son est grand, plus l'impression de stéréophonie sera faible, jusqu'à tendre vers la monophonie.

 Mise en garde

Attention de ne pas confondre l'angle de prise de son avec l'angle entre les capsules.

16.2.1 Comment choisir un angle de prise de son.

L'angle de prise de son est étroitement lié à la distance du couple par rapport à l'évènement sonore à enregistrer. En règle générale, plus le couple est loin des objets sonores à enregistrer, plus son angle de prise de son sera faible. À l'inverse, plus le couple sera proche, plus son angle de prise de son sera grand.

Ensuite, lors de la réalisation d'un couple de prise de son, il est commun d'enregistrer un ensemble d'éléments : plusieurs instruments (batterie), voire plusieurs musiciens (quatuor à corde, orchestre). L'objectif est bien souvent de retrouver une sensation de disposition des éléments dans l'espace proche de la situation réelle. On cherche donc un angle de prise de son suffisamment petit pour que les sources occupent l'intégralité de l'espace stéréophonique, mais également suffisamment grand pour ne pas créer une sensation de trou au centre.

16.2.2 Comment réaliser un angle de prise de son.

Plusieurs outils existent pour aider le preneur de son à choisir son angle de prise de son correctement.

Il est important de commencer par les abaques de Michael Williams, ayant cherché à étudier l'angle de prise de son et ses qualités en fonction des paramètres vues précédemment. Les résultats de ses travaux se trouvent sur le site mmad.info.

On trouve également beaucoup d'application mobile, comme celle du constructeur du microphone Neumann, s'appuyant sur les travaux de Michael Williams pour aider leurs utilisateurs à correctement positionner leurs microphones. Évidemment, et heureusement, rien n'est spécifique à un fabricant de microphones en particulier, l'application d'un constructeur A peut servir pour placer des microphones d'un constructeur B.

Plus récemment, des chercheurs britanniques ont développé une application, nommée **MARRS**, permettant de positionner son couple de prise de son par rapport aux sources via une interface graphique très simple à utiliser. Cette application est disponible sur mobile et sur navigateur internet.

16.2.3 Privilégier le Δi ou le Δt ?

La différence de perception du champ stéréophonique est très différente entre celui produit par le Δi ou par le Δt .

- Un couple reposant sur le Δi aura une sensation de localisation des sources précise. De plus si un tel couple enregistre une source ce déplaçant à vitesse constante, la sensation de déplacement retranscrite par le couple sera, elle aussi, linéaire. Il est également possible de sommer les deux microphones ensemble afin d'obtenir un signal monophonique. Un tel couple est appelé compatible mono.
- Un couple reposant sur le Δt aura une sensation de localisation plus floue, mais apportera un sens de l'espace plus grand et une dimension spacieuse. À l'inverse d'un couple Δi , la sensation d'un déplacement linéaire d'une source n'est pas linéaire. Il n'est pas possible de sommer les deux capsules pour en obtenir une réduction mono sans générer des altérations de timbre sévères.

Chaque couple possède ses avantages et ses inconvénients. Heureusement, nous ne sommes pas limités à l'un où l'autre et nous pouvons à loisir réaliser une combinaison des deux mécanismes.

16.3 Les topologies classiques de prise de son au couple

Le premier ingénieur à se poser la question du son stéréophonique est l'anglais Alan Blumlein en 1929. Il imagine l'entièreté de la chaîne d'enregistrement et de diffusion nécessaire à la stéréophonie. Cependant, la BBC lui impose comme contrainte que toutes ses propositions soient compatibles avec des systèmes monophoniques. Il inventera donc le couple XY et MS.

Plus tard, la plupart des radios européennes développeront des couples de prises de son mêlant Δi et Δt , tel que l'ORTF.

16.3.1 Le couple Blumlein / XY

Les deux microphones sont ici directifs, placés au même point de l'espace et ongulé d'une certaine valeur entre eux.

De par les contraintes technologiques de son époque, Blumlein a décrit ce couple pour une utilisation de deux microphones bidirectionnels. Il est aujourd'hui plus commun de le rencontrer avec deux cardioïdes.

Dans sa version originale, le couple Blumlein comprend donc deux microphones bidirectionnels avec un angle de 90°.

La formulation du couple XY comprend deux microphones cardioïdes avec un angle compris entre 90° et 135° .

16.3.2 Le couple MS

Le couple MS, également inventé par Alan Blumlein, permet de doser la quantité de stéréophonie après l'enregistrement.

Pour se faire, ce couple utilise deux microphones :

- Un omnidirectionnel, historiquement, mais aujourd'hui fréquemment remplacé par un microphone cardioïde.
- Un bidirectionnel

Le microphone omnidirectionnel, ou cardioïde, va rendre compte du centre de la stéréophonie, tandis que le microphone bidirectionnel rendra compte de la latéralité.

Une fois enregistrés, ces deux canaux ont besoin d'être convertis, plus exactement dématricés, vers une paire de canaux stéréophonique. L'opération est très simple :

$$L = M + S$$

$$R = M - S$$

Cette opération peut être réalisée sur une console de mixage, telle que décrite ci-dessous.

16.3.3 Le couple ORTF

Le couple ORTF, inventé par la radio française du même nom, combine l'effet du Δi et du Δt afin de s'approcher de l'écoute humaine.

Sa topologie est précisément définie. Elle propose l'utilisation d'une paire de microphones cardioïde, ouvert à 110° et avec un écart de 17 cm.

16.3.4 Les couples AB

Les couples AB peuvent avoir une définition ambiguë. Une partie de la littérature scientifique considère comme couple AB tout couple non coïncident. À cet égard l'ORTF est considéré comme un couple AB. Pour d'autre, les couples AB ne concernent que des couples constitués de microphones omnidirectionnels.

Ces derniers ont la particularité de n'utiliser que le Δt afin de placer les sources dans l'espace. Le rendu est donc souvent spacieux, au prix d'une certaine instabilité et d'un certain manque de précision de l'image stéréophonique.

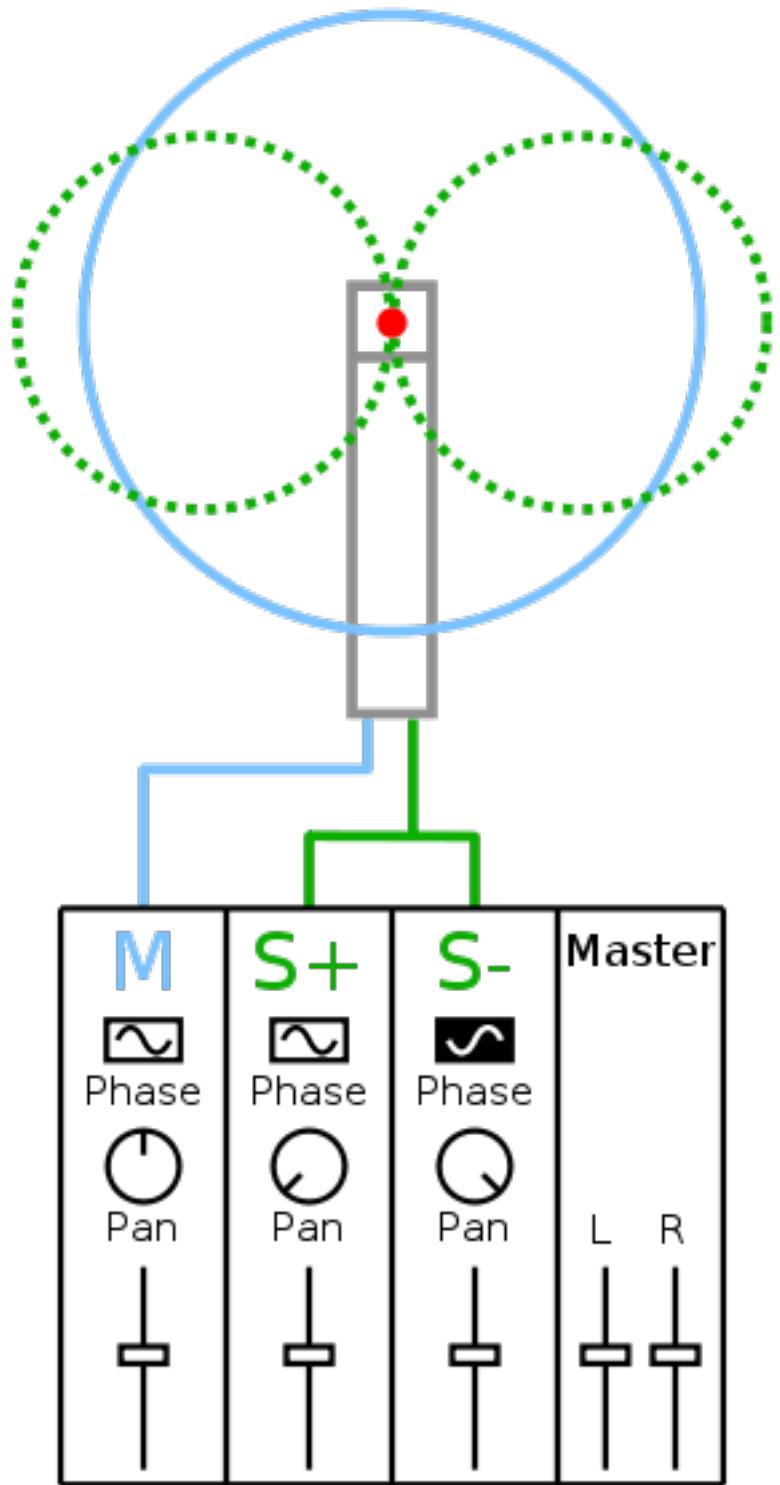


Figure 16.1: Dématriçage MS

16.4 Compléter une prise de son au couple par des appo

Il est commun, lors d'une prise de son au couple, de chercher à obtenir une entière satisfaction sonore à la seule aide du couple. Cependant, cela n'est parfois pas possible, souvent pour des contraintes physiques et acoustiques (un instrument de l'ensemble jouant moins fort que les autres). Dans ces cas, l'utilisation d'appoint, donc de microphone supplémentaire, placé en proximité de la source, va permettre de venir récupérer une précision supplémentaire de l'instrument.

Lors de l'étape de mixage, le couple servira de base principale et l'on viendra ajouter la quantité nécessaire d'appoints pour préciser le propos. Il sera parfois nécessaire de remettre en phase l'appoint et le couple pour améliorer la sommation de l'ensemble.

17 Déphasage et remise en phase

17.1 Les effets sonores de déphasage

Tous les signaux sont caractérisés par une certaine phase. Celle-ci est moins tangible que celles de niveau sonore ou de fréquence. En effet, lorsqu'un signal est écouté seul, celle-ci ne s'entend pas. C'est au moment où plusieurs signaux corrélés (comprendre, enregistrés au même moment, par plusieurs microphones) sont sommés que les différences de phase peuvent s'entendre.

17.2 Approche mathématique

Prenons l'exemple d'un son pur : $\sin(\omega t + \phi)$ où $\omega = 2\pi f$

La phase de ce signal est décrite par $\omega t + \phi$

Les deux paramètres responsables de déphasages audibles sont :

- t , le temps
- ϕ , la phase à l'origine

Attention, pour un son pur, l'effet de la modification de t ou de ϕ semble très similaire. Ce n'est pas le cas pour des signaux pseudo-périodiques, atténués dans le temps.

17.3 Les sources de déphasage

Les causes les plus classiques de déphasages sont :

- Un câble XLR avec une inversion sur le point chaud et le point froid
- Une prise de son avec une différence de distance entre deux microphones
- Une prise de son utilisant deux microphones positionnés de part et d'autre d'une membrane
- Un retard de certaines fréquences lié aux objets rencontrés par les signaux

18 En pratiques

Le bon déroulement de l'enregistrement d'instruments acoustiques dépend de multiple facteur. Trié par ordre d'importance décroissante, nous trouvons :

1. Le confort du musicien
2. La qualité de l'instrument enregistré
3. La qualité de l'acoustique de la pièce où a lieu l'enregistrement
4. Le placement du microphone
5. Le choix du microphone (technologie et directivité)
6. Le choix du préampli

18.1 Le confort du musicien

Même si elle peut sembler triviale, cette « étape » de la chaîne de prise de son est de loin la plus importante. La qualité de l'interprétation donnée par le musicien dépendra grandement de son état moral et psychologique :

- Est-il stressé
- Est-il confiant
- Se sent-il accueilli
- etc.

Nous pourrions considérer qu'un ou une musicienne arrivant dans un studio d'enregistrement se présente avec un taux de confiance maximal envers l'équipe technique. Dès lors l'objectif des différents techniciens est de conserver cette jauge au maximum.

Les premières minutes sont particulièrement importantes et va poser un ressenti fort sur la journée de travail. Il y a donc un équilibre à trouver entre un accueil chaleureux et décontracté et un rapport productiviste et sérieux.

Le système permettant aux musiciens de communiquer entre eux et avec les techniciens est primordial. En pratique, il n'est pas rare de dédier certains microphones du plateau à cette tâche. Du côté régi, le « talkback » est l'outil de communication premier des techniciens présents sur la session. Il convient de l'utiliser avec soin et prudence. Un musicien peut rapidement se sentir isolé, s'il enregistre seul. Il convient de maintenir un contact régulier et précis afin de ne pas l'abandonner dans le seul dans sa cabine. Qui plus est, un quiproquo peut être vite

arrivé avec les systèmes de talkback. Prudence quant à l'état d'ouverture ou de fermeture du microphone.

18.2 Le choix de l'instrument

La plupart des musiciens se présenteront avec leurs instruments. La marge de manœuvre est donc ici quasi nulle.

Cependant il n'est pas rare que le studio possède du « backline », souvent composé de batteries, d'amplificateur guitare et basse, voire de guitares et de basses. Si l'instrument utilisé par le musicien pose problème pour la prise de son, proposer une alternative peut s'avérer être un bon pari. Il convient évidemment de sonder l'ouverture du musicien par rapport à cette proposition, afin de ne pas le braquer.

Il peut également être intéressant de « préparer » les instruments. Cette technique est très courante sur les pianos et les batteries, afin de changer les propriétés acoustiques de l'instrument grâce à l'utilisation de draps, coussins, couvertures disposées dans ou sur l'instrument.

18.3 Le choix de l'acoustique

L'acoustique de la salle d'enregistrement est-elle aussi plus souvent une contrainte qu'une variable d'ajustement.

On préférera souvent de grandes salles afin de limiter l'apparition prématurée de premières réflexions. Plus la salle sera petite, plus celle-ci apportera une forte coloration sur le contenu enregistré. Il convient donc d'être attentif aux petites cabines de studio, celles-ci sont souvent très mates, mais leur apport sur le timbre des instruments qui y sont enregistrés est souvent très important.

Lorsque l'on a la possibilité d'enregistrer dans de grandes salles, il est souvent intéressant de disposer des quelques panneaux acoustiques mobiles, afin de modeler la pièce à sa convenance.

Si l'acoustique imposée est défavorable, on préférera dans ce cas des prises d'hyper proximité, afin de minimiser son effet au maximum.

18.4 Placer et choisir son microphone

En pratique, il est bien difficile de dissocier le choix du microphone de son placement, les deux étant très interdépendants. Cependant, il convient de garder à l'esprit que le positionnement du microphone est, parmi les deux, sans doute le plus déterminant.

La première étape, avant même de choisir un microphone, consiste à écouter l'instrument dans l'acoustique d'enregistrement. Il s'agit ici d'une écoute active. On se déplace autour de l'instrument, on s'en approche, on s'en éloigne, afin de sentir l'interaction entre la source et l'acoustique du lieu. Aussi, il est important de trouver deux zones d'émission particulière de l'instrument : la zone de projection maximale et la zone au timbre le plus favorable. La première peut nous servir à positionner l'instrumentiste par rapport aux autres instruments afin de minimiser les reprises entre microphones. La deuxième zone nous indique l'axe de prise de son.

Cette zone au timbre le plus favorable est relative. Elle dépend de l'instrument, bien sûr, mais aussi du modèle. Elle dépend également du mode de jeu, de l'articulation du joueur et évidemment, de l'esthétique de la musique.

18.4.1 Le rapport a la distance du microphone

La distance de positionnement du microphone est un élément excessivement important sur le rendu esthétique de la prise de son.

En règle générale, plus on prend de distance, plus on approche une prise de son naturaliste, cherchant à reproduire un évènement sonore dans son environnement, tel qu'il aurait été entendu dans la pièce. Plus on se rapproche, plus on fragmente l'événement sonore et plus on l'arrache aussi à son contexte de diffusion.

Afin de déterminer efficacement le placement d'un microphone, il convient d'abord d'en connaître sa distance critique. Celle-ci correspond au point, dans une pièce, où le son provenant directement d'une source est perçu au même niveau sonore que la réponse acoustique à cette source. Cela signifie que si nous plaçons notre microphone au-delà de ce point, nous obtiendrons plus d'acoustique que de son direct de l'instrument.

Il est important aussi de considérer que la directivité du microphone influe sur la distance critique. En effet, plus la directivité du microphone est large (tends vers l'omnidirectionnalité), plus le microphone paraîtra éloigné de la source. À l'inverse, plus la directivité d'un microphone est étroite (tends vers la bidirectionnalité), plus le microphone paraîtra proche.

Dans le cas de l'utilisation de microphone directif, le placement en proximité et hyperproximité va créer une accentuation du contenu basse-fréquence de la source. Cela devient parfois un élément esthétique, comme sur les voix radiophoniques. Cela aussi peut être un défaut, une exagération qu'il conviendra de corriger en postproduction.

18.4.2 Quand choisir une prise de son stéréophonique

La prise de son stéréophonique, comme son nom l'indique, regroupe l'ensemble des techniques de prise de son dédié au système de diffusion stéréophonique (deux enceintes séparées de 60° et orientées vers un auditeur placé à équidistance des deux transducteurs).

L'avantage de tels dispositifs de prises de son est de peupler dès la prise l'espace stéréophonique qui est donné à l'auditeur lors de la diffusion. Ils permettent également de rendre compte de la position de plusieurs évènements sonores ayant lieu dans la même acoustique. Cette dernière est d'ailleurs bien mieux retranscrite par de tels systèmes de prise de son.

Il s'agit à nouveau d'un choix esthétique. Faire le choix d'une prise de son monophonique permet de renforcer la sensation de frontalité et de densité d'une source. À l'inverse, une prise de son stéréophonique donnera une définition spatiale accrue.

18.4.3 Quand choisir la multi-microphonie

La multi-microphonie consiste à enregistrer un instrument via l'utilisation de microphones (principalement) directifs, placés à différents endroits jugés pertinents et en hyperproximité.

Cette approche esthétique de la prise de son est devenue indissociable des « musiques actuelles ». Elle offre l'avantage d'une grande flexibilité de traitement lors de la phase de mixage. Voir, elle implique une certaine partie des traitements.

En effet, une prise d'hyperproximité va systématiquement relever deux défauts :

- un effet de proximité : le grave/bas médium de la source paraît hypertrophié lors de l'emploi de microphones directifs.
- Les dynamiques de jeux sont également hypertrophiées.

L'effet de proximité implique donc bien souvent l'utilisation d'un égaliseur, permettant de corriger cette augmentation artificielle du grave. De même, l'hypertrophie de la dynamique de jeu implique l'usage d'un compresseur afin de corriger ces variations artificielles.

Afin de recréer une sensation de spatialisation, on utilisera principalement deux outils. En premier lieu, le potentiomètre de panoramique afin de diriger ces sons mono dans l'espace stéréophonique, puis les réverbérations artificielles permettra de reconstituer un champ acoustique et de réintégrer ces sources dans une scène sonore.

Si l'approche de la prise au couple pouvait être qualifiée de naturaliste, alors la prise de son en multimicrophone sera son pendant spectaculaire. Évidemment, il convient de ne pas aussi franchement opposer ces deux approches et il existe tout un monde de système de prise de son entre ces deux extrêmes.

18.5 Le choix du préamplificateur

Le rôle du préamplificateur est d'amplifier le signal, le tout en ramenant le minimum de bruit. Un premier élément de choix de préampli va se faire sur le niveau de pression acoustique produit par les sources à enregistrer.

Enregistrer une batterie impose peut de contrainte sur le préampli quand a sa capacité à amplifier sans rajouter beaucoup de bruit sur le signal. À l'inverse, enregistrer des instruments peu sonores, possiblement avec des microphones peu sensibles, implique l'utilisation de préampli avec une excellente réserve de gain et un excellent rapport signal bruit.

18.5.1 L'influence du préampli sur la « couleur » du son

Il est assez connu que le préampli peut également devenir un choix esthétique pour influencer la couleur d'une prise de son. Cette question semble assez complexe. Voici quelques éléments de réponse :

- Le choix du préampli est d'une influence minime par rapport à **tous** les autres choix précédemment fait.
- Les préamplis sont souvent catégorisés, en termes de couleur, via les composants utilisés pour réaliser l'amplification. Attention, un composant électronique dépend toujours du contexte dans lequel il est placé (ici, du circuit électronique). Il est donc difficile de précisément qualifier le son d'un préampli à lampe ou à transistor de façon générique.
- Les impédances d'entrée des préamplis ne sont souvent pas évoquées dans ces discussions. Hors, pour la plus parts des microphones (hors statiques), leur impédance de sortie peut être suffisamment élevée pour engendrer une déperdition en aigu et en transitoire. Cette déperdition peut être heureuse, ou malheureuse, mais surtout bien réelle. Une manière de s'en prémunir peut-être d'utiliser des « booster » de microphones (parfois également appelés préamplis), permettant d'augmenter le niveau de sortie des microphones et aussi d'adapter leur impédance.

partie IV

Outils de mixage

19 Anatomie d'une console de mixage

20 Manipulation de la phase

21 Egalisation et égaliseurs

22 Les compresseurs

23 Les limiteurs

24 Les autres outils de gestion de la dynamique

25 Mise en espace et réverbération

26 Saturation et distortion du signal

27 Effets & modulations

partie V

Méthodologie de mixage

28 Ecoute critique : deuxième partie

29 Approche générale

30 Etude des cas d'école

31 Pièges et erreurs à éviter

partie VI

Introduction à la spatialisation sonore

32 Introduction

Cette section, et les suivantes ont à cœur de traiter le vaste sujet de la spatialisation sonore. Il est assez rare de trouver des ouvrages génériques sur les techniques du son abordant aussi cette discipline. Mais cette lacune s'explique très facilement par la très faible part de marché représenté pour les consommateurs de format audio immersif, comme le fameux 5.1. Ces dernières années, les choses bougent un peu. Dolby est revenu d'une longue traversée du désert avec le « Dolby Atmos », d'abord dans les salles et les auditoriums de cinéma, et aujourd'hui dans le monde de la musique. On retrouve ainsi deux logiciels de mixage phares, Pro Tools d'Avid et Logic Pro X d'Apple, intégrant nativement ce « Dolby Atmos ». Cet intérêt pour les technologies immersives se retrouve aussi chez les constructeurs d'enceintes et de casques, proposant des systèmes de plus en plus simples pour accéder à cette expérience d'écoute. Alors, à la lumière de ces récentes évolutions, il paraît important de se pencher sur les longues et riches histoires de la spatialisation sonore, ainsi qu'aborder les différentes technologies et techniques associées.

Cette partie théorique commence par un historique de la spatialisation sonore, qui nous permettra de réaliser que cette conversation ne date pas d'hier, ni de l'Atmos. Puis, il conviendra de s'attarder sur les systèmes de diffusions et le positionnement des enceintes dans l'espace. On pourra ainsi aborder les différents systèmes normés, amenés par des constructeurs tels que DTS ou Dolby, mais aussi décrire et catégoriser les systèmes d'écoutes qui échappent à ces approches rigides. On différenciera alors les systèmes frontaux, les systèmes englobants à une dimension, les systèmes englobants à deux dimensions (souvent appelés « immersif ») et les systèmes à trois dimensions.

Une fois ces deux blocs très généralistes abordés, nous attaquerons alors une à une les grandes techniques permettant de réaliser des mixages à destination de ces systèmes à multiples enceintes, avec dans l'ordre :

- Le binaural
- L'approche perceptive, ou « orientée canal »
- L'ambisonique
- La WFS, ou synthèse de front d'onde

Nous discuterons aussi du mixage orienté objet, proposé notamment par le Dolby Atmos, mais aussi par d'autres, comme l'IRCAM et son « Spat », également intégré par FLUX:: Immersive dans son Spat Revolution.

i Note

Il n'y aura ici que peu de rappels sur les mécanismes de la perception ainsi que sur notre capacité à localiser les sons dans l'espace. Ces informations sont accessibles dans le chapitre [2](#).

33 Historique de la spatialisation

Avertissement

Section en cours de création.

Les secteurs du jeux-vidéo et de la VR ne sont pas encore traités.

De nos jours (2023 à l'écriture de ce chapitre), le « son immersif » semble appartenir aux nouvelles technologies et la plupart des fabricants se lancent sur ce marché. Il est d'ailleurs très fréquent de voir ce « son immersif » vendu comme quelque chose de nouveau et d'inédit. Pourtant, la question de l'espace sonore, de sa capture et de sa restitution est presque aussi vieille que celle de l'enregistrement.

Nous tâcherons dans ce chapitre de retracer une histoire du son spatialisé, en abordant l'apparition des différentes techniques, mais aussi leurs concepteurs, les chercheurs et les ingénieurs qui y ont participé. Nous traverserons donc près de cent cinquante années d'histoire, en commençant par les premiers essais de son « stéréophonique » jusqu'au mixage orienté objet. D'autre part, nous tâcherons également de parler des compositeurs et des artistes qui se sont accaparé la question de l'espace et sa restitution dans leur œuvre.

Ce chapitre n'a pas pour ambition d'être exhaustif, mais essaye tout de même de donner une image d'ensemble assez juste.

33.1 De la monophonie à la stéréophonie : un besoin d'espace

Pendant près de quatre-vingts ans, la majorité des systèmes de diffusion sonores sont monophoniques. Il faut en effet attendre la fin des années soixante pour que l'écrasante majorité de la musique enregistrée soit produite en stéréophonie. Avant cela, la norme est donc à une écoute ne se constituant que d'un seul haut-parleur.

Pourtant, le besoin d'espace dans la restitution sonore se fait sentir très tôt, car plusieurs ingénieurs se penchent sur cette question dès le début du XXe siècle. **Clément Ader** (1841-1925) semble être le premier à proposer un système de diffusion sur deux canaux. Cet ingénieur français est avant tout connu pour être le premier à avoir fait décoller un engin motorisé plus lourd que l'air en 1890. Il a également participé au déploiement du réseau téléphonique, à Paris, en 1879. C'est à ce moment qu'il a l'idée d'utiliser le réseau de télécommunication pour diffuser l'opéra dans les foyers. Ce dispositif, nommé **Théâtrophone**, est pour la première fois

utilisé en 1881. Les auditeurs peuvent alors écouter la pièce retransmise en direct en plaçant un « écouteurs » sur chacune de leurs oreilles. La diffusion se fait donc sur deux canaux.

À cette époque, les seuls microphones disponibles sont des microphones à charbons, dont on connaît la faible bande passante. La qualité sonore du dispositif est donc médiocre, mais l'engouement du public est réel, et perdurera jusque dans les années 1930.



Figure 33.1: Extrait du journal « Le Gaulois » du 2 avril 1924

La stéréophonie va être pensée, presque simultanément par deux ingénieurs, [Arthur C. Keller](#) (1901-1983) et [Alan D. Blumlein](#) (1903-1942). Le premier travail aux laboratoires Bells, sous la direction d'[Harvey Fletcher](#) (le même Fletcher que les courbes Fletcher-Munson [2.6](#)). Keller se retrouve à travailler sur la diffusion stéréophonique, à la demande du chef d'orchestre [Leopold Stokowski](#), alors chargé de produire des concerts sur les ondes de la NBC avec l'orchestre de Philadelphie. Stokowski est alors très insatisfait de la qualité sonore de ce type de dispositif. Il contacte donc les laboratoires Bell, leur demandant d'y apporter une solution. De cette collaboration naît une collection d'enregistrements dont le plus vieux semble être une interprétation du [Carnaval romain \(Ouverture\)](#) de Berlioz, le cinq décembre 1931.

De l'autre côté de l'océan, Alan Blumlein travaille pour la [Colombia Graphophone Company](#). L'histoire veut qu'il soit frustré de son expérience du son au cinéma, ne comprenant pas pourquoi les voix sont toutes au centre alors que les acteurs se déplacent de part et d'autre de l'écran. Blumlein propose donc, pêle-mêle dans un seul et même brevet, les techniques de prise de son coïncidente XY/[Blumlein](#) et MS (pour garantir la rétrocompatibilité sur les systèmes mono), un système de gravure à deux canaux sur disques microsillons et la description du système d'enceinte que l'on utilise encore aujourd'hui sous le nom de stéréophonie. Il est

amusant de constater que dans ses écrits, Blumlein ne parle pas de son stéréophonique, mais de son binaural. Ceci est d'autant plus surprenant qu'il est le premier à revendiquer l'idée que les haut-parleurs, dans un système stéréophonique, ne correspondent pas aux oreilles, et qu'il convient donc d'étudier ce système avec une approche psychoacoustique. Blumlein invente également un système, qu'il appelle « shuffling », permettant de convertir une différence de phase en différence d'intensité. Pour lui, ce dispositif fait partie intégrante d'un système de prise de son stéréophonique. Cependant, ce « shuffler » ne sera pas du tout conservé par le reste de l'industrie.

i Note

En 1931, La Colombia Graphophone Company et la [Gramophone Company](#) fusionnent pour devenir [EMI](#).

Il existe encore quelques enregistrements du travail d'Alan Blumlein, par exemple, cette démonstration du [mouvement en stéréophonie](#) (dans un style tout à fait anglais), ou encore cette interprétation de la [Chevauchée des Valkyries pour trois pianos](#) enregistrée à Abbey Road. Ces deux enregistrements datent de 1931.

i Note

Les procédés de gravure stéréophonique sur disque microsillon de Keller et Blumlein sont très proches, cependant, il semblerait qu'aucun des deux n'ait eu connaissance des travaux de l'autre. Keller a développé son système avant Blumlein, mais ce dernier a publié son brevet en premier. Les deux approches reposent sur l'idée de graver les deux canaux à 45° de la verticale. Ce procédé semble être tombé dans l'oubli jusqu'à sa « [réinvention](#) » dans les années cinquante par la société Westrex.

Blumlein décède en plein milieu de la seconde guerre mondiale suite au crash d'un bombardier Halifax, dans lequel il effectuait des tests sur le système de [radar H2S](#).

Malgré toutes ces avancées, l'adoption de la stéréophonie traîne, et ce pour deux raisons principales. La première est que les systèmes de gravures sur disques proposés par Blumlein ou Keller ne fonctionnent pas sur les gramophones, à cause du poids de la tête de lecture. Il faut donc attendre leurs remplacements par les disques microsillons au milieu des années cinquante. Le second frein est, comme on le rencontrera pour les systèmes dits « surround », le coût pour s'équiper d'un deuxième amplificateur et d'une deuxième enceinte.

La situation se débloque dans les années soixante. La radio devient stéréophonique, en partie grâce au passage à la modulation FM. Le coût des équipements devient aussi plus raisonnable. À la fin des années soixante, on peut considérer que l'ensemble de la production musicale est passée à la stéréophonie. Il est d'ailleurs intéressant de noter que cette transition vers la stéréophonie n'a pas eu lieu au même rythme selon les genres musicaux. Dans les musiques dites « classiques », la monophonie limite à ce point la représentation de l'espace acoustique que le

LETTERS TO THE EDITOR

Stereophonic Broadcasting: B.B.C.'s Reply

IN your March issue, Major Jeffery takes us to task for not keeping listeners sufficiently informed on our engineering policy and developments. I hope there are not many of your readers who think this, because we always give as much information as we can as soon as it can be given. But when changes or new developments are of an experimental nature we must take care that we do not mislead listeners.

The change of location of the Third Programme transmitter serving the London area was announced in the Press three days before it took place. In addition, microphone announcements were made in both the Third and Home Service programmes before the change was made.

There has been no secret about the new v.h.f. station at Wrotham. As far back as March, 1948, we announced the location of this station, which has been built for experimental frequency modulated and amplitude modulated transmissions. Now that the construction of the station has been completed, we shall shortly issue a statement concerning test transmissions.

We do not propose to radiate

binaural transmissions from Wrotham. We first studied this system of transmission before the war and made some experimental transmissions both by radio and over short-distance wire circuits. The idea is an interesting one, but the advantages are doubtful and the objections are overwhelming. As far as I am aware, there is not now and never has been a regular service of binaural broadcasting in any part of the world. The system requires that the programme chain must be duplicated from the studio right through the low-frequency and high-frequency chain to the aerial. Moreover, there must be duplication of receiving equipment at the listeners' end. The cost would be very high and double the number of carrier frequencies would be required for transmission. It is no exaggeration to say that, in all the wave-bands now used or projected for sound broadcasting, it would be quite impossible to find sufficient channels to enable this to be done.

Finally, there is no doubt that such a system would interest only a relatively small number of listeners, and for this reason, if for no other, it is doubtful whether we should be justified in spending so much money to give a specialized service to so few people.

H. BISHOP,
Chief Engineer, B.B.C.

Figure 33.2: Sur la diffusion stéréophonique — 1950

passage à la stéréophonie permet un réel gain de qualité. Dans les musiques dites « populaires » ou « amplifiées », la nécessité de la stéréophonie est moins évidente, elle d'ailleurs parfois perçue comme superficielle. Par exemple, il faudra attendre le dernier album des Beatles enregistré (Abbey Road - 1969) pour avoir une version stéréo officielle, approuvée par George Martin. Certains ont vu dans la stéréophonie une passade un peu exubérante, mais qui ne durerait pas. L'histoire leur aura prouvé le contraire. Il ne faut pas non plus leur jeter la pierre. En effet, les premiers enregistrements stéréophoniques n'exploitent souvent pas les deux canaux de mixages comme une opportunité de mieux représenter l'espace, mais plutôt pour **démasquer** les éléments entre eux. On trouve donc beaucoup de **mixages** avec les instruments placés dans l'enceinte gauche ou droite, et la voix en plein centre.

Devant la généralisation de la stéréophonie, certains constructeurs, labels et artistes vont alors tenter la courte aventure de la quadriphonie.

33.2 La quadriphonie : une entreprise infructueuse

La **quadriphonie** (quadraphonic, quadrasonic ou encore quadrophonic en anglais) est le premier format dit « surround » accessible au grand public. Les premières tentatives de productions musicales en quadriphonie commencent en 1969. Le directeur artistique Thomas Mowrey a alors réalisé un grand nombre de disques quadriphoniques, notamment pour le label **Deutsche Grammophon**.

Il a existé trois formats de quadriphonie. Ils sont notés de la façon suivante : K-M-N. K indique le nombre de canaux de mixage, M le nombre de canaux du support de diffusion et N le nombre de canaux de restitution (ici, N est donc toujours égal à quatre).

- Le 4-4-4, que l'on pourrait qualifier de « vrai » quadriphonie. Cela impose un équipement tout particulier pour le support de diffusion. En effet, l'écrasante majorité des magnétophones à bandes et des disques microsillons sont stéréo. Certains constructeurs ont donc vendu des magnétophones « quadriphoniques », tandis que d'autres ont tenté de stocker ces quatre canaux sur disques microsillons grâce à des techniques de modulation d'amplitude. Cependant, ces procédés imposent tout de même d'équiper son lecteur vinyle d'une pointe en diamant spécifique.
- Le 4-2-4 cherche à simplifier les choses pour les auditeurs. Grâce à un jeu de matriçage, on encode le signal quadriphonique sur deux canaux, que l'on espère récupérer intact par dématriçage. Si ce matriçage fonctionne parfaitement pour des sons purs, il n'en est rien pour des sons complexes, qui peuvent alors induire une corrélation importante entre les différents canaux. Le matriçage le plus convaincant fut proposé par **Peter Scheiber** et perfectionné par **Benjamin Bauer (SQ Quadriphonic)**, mais même celui-là posait un problème sur la corrélation des canaux de diffusions. En effet le niveau de séparation entre deux enceintes adjacentes n'était que de trois décibels. Ce jeu d'encodage et de décodage distord l'espace sonore et ne produit pas des résultats satisfaisants.

Note

Il « inspirera » pourtant les laboratoires de Dolby pour élaborer leur fort mal nommé « Dolby Stéréo »

- Reste le 2-2-4, que l'on pourrait qualifier du « pire des deux mondes ». Le principe consiste à « inventer » des canaux arrière à un mixage stéréophonique, pour produire un effet englobant. Le résultat est souvent flou et insatisfaisant.

La plupart des enregistrements multicanaux produits par la Deutsche Grammophon ne sont jamais sortis qu'en stéréo. En effet, le label n'a pas été convaincu par la qualité du rendu sonore quadriphonique, notamment à cause du matriçage 4-2-4.

La complexité d'installation, et l'existence de nombreux formats (dont certains n'offrant pas une qualité de restitution satisfaisante) expliquent largement l'échec commercial du format quadriphonique.

L'exploitation commerciale du son quadriphonique ne se limite pas à la musique dite « classique ». Un grand nombre d'artistes et de groupes « pop » verront leurs mixages stéréo déclinés en quadriphonie. Les plus connus sont sans doute les **Pink Floyd**. Dès 1967, ils embarquent en concert l'azimuth coordinator, leur permettant de recréer une diffusion quadriphonique. En 1972-1973, leur ingénieur du son de l'époque, Alan Parson propose un mix quadriphonique de leur album alors en production, *The Dark Side of the Moon*.

Malgré tout, le format sera un échec commercial cuisant : trop cher pour équiper un grand nombre de personnes, souvent de qualité douteuse à cause de l'étape de matriçage, existant sous de nombreux formats, etc. On trouve tout même aujourd'hui un certain nombre de mixages quadriphoniques des années soixante-dix édités sur SACD ou DVD-Audio. Le label anglais **Dutton Vocalion** a dans son catalogue plusieurs dizaines de ces **rééditions**, principalement de musique dite « classique ».

33.3 La grande aventure du son spatialisé au cinéma

Le **cinéma**, naissant autour de 1984, est une des industries du divertissement mettant le plus en avant les intérêts d'une diffusion spatialisée. Sa dimension visuelle spectaculaire et immersive (écran géant, salle noire, etc.) a donc motivé un traitement du son similaire. On qualifie parfois le cinéma des premiers temps de « muet », principalement dû au fait de l'absence de son synchrone rattaché au film. Cependant, imaginer le dispositif cinématographique comme un dispositif silencieux est une erreur. Dès ses débuts, il est très souvent accompagné par un musicien et il est alors courant pour le public de parler, de commenter la projection. N'oublions pas qu'à l'origine le cinéma est un spectacle forain. Ce n'est que dans un second temps que le cinéma prend place dans les théâtres puis dans des lieux qui lui sont dédiés.

Le [Vitaphone](#) est, en 1924, le premier système de synchronisation sonore, développé par Westerne Electric Company en collaboration avec les laboratoires Bells. Le premier film partiellement parlant est [Le Chanter de Jazz](#) (1927) et le premier film intégralement parlant est [Lights of New York](#) (1928).

Le cinéma sonore s'appuie alors sur une diffusion monophonique, le système de diffusion est d'ailleurs caché derrière l'écran.

33.3.1 Fantasia et le Fantasound

[Fantasia](#) est un film de [Walter Disney](#) sorti aux États-Unis en 1940. L'ambition de son auteur est de mettre en image des œuvres incontournables de la musique dite « classique » afin de toucher un nouveau public. Pour cela, il s'associe avec le chef d'orchestre Leopold Stokowski (le même personnage que nous avons évoqué plus haut).

L'ambition technique du film est énorme et, sous les exigences de Walt Disney, pousse l'équipe technique à développer le [Fantasound](#). Ce Fantasound a deux objectifs principaux :

- Premièrement, améliorer la dynamique de diffusion sonore (alors limitée par le support)
- Deuxièmement, **permettre de déplacer des sons dans l'espace.**

Est alors développé le premier système de VCA (Voltage Controlled Amplifier), piloté par la fréquence d'un signal sinusoïdal inscrit sur le film. Le deuxième point est solutionné par l'invention du premier « pan pot ». Le Fantasound utilise cinq canaux de diffusions, trois enceintes frontales (gauche, centre et droit) ainsi que deux enceintes arrières (gauche et droite). Fantasia est ainsi le premier film « surround » de l'histoire.

Note

Le Fantasound est l'ancêtre de ce que l'on appelle aujourd'hui le 5.1 (sans le caisson de grave).

Cependant, le coût global du système Fantasound le rend impraticable, pour le film Fantasia lui-même, mais également pour d'autres productions. Il faut, en effet, un petit régiment d'opérateurs pour pouvoir correctement diffuser le film.

33.3.2 Le Dolby Stereo et le son optique matricé

Durant toute la période du son analogique, le plus grand ennemi de la chaîne de production est le bruit. Ce bruit est particulièrement inhérent aux supports magnétiques et optiques alors utilisés. C'est dans ce contexte que [Ray Dolby](#) fonde son entreprise [Dolby Laboratories](#) en 1965, et commercialise le [Dolby NR](#) en 1966.

Dans les années 70, les laboratoires Dolby élaborent également un système de diffusion cinématographique multicanal. Ce dispositif d'enceinte dit LCRS (gauches, centre, droit, et une enceinte arrière dite « surround »), est associé à un matriçage 4-2-4 grandement « inspiré » par les travaux de Peter Scheiber sur le matriçage quadriphonique. Ils auront simplement la mauvaise idée d'appeler ce système « Dolby Stereo ».

! Important

Le mixage est réalisé sur quatre canaux : gauche, centre, droit et arrière. Ces quatre canaux sont alors **matricés** sur deux canaux que l'on nomme *gauche totale* et *droite totale*.

	L	R	C	S
Lt	1	0	$\frac{\sqrt{2}}{2}$	$\frac{j \times \sqrt{2}}{2}$
Rt	0	1	$\frac{\sqrt{2}}{2}$	$\frac{-j \times \sqrt{2}}{2}$

Pour décoder le signal et retrouver quatre canaux de diffusion, on doit alors **dématriciser** les signaux Lt et Rt.

	Lt	Rt
L	1	0
R	0	1
C	1	1
S	1	-1

Le canal central est alors formé par les signaux **en phase** du matriçage bicanal. Le canal arrière est composé des signaux **hors phase** dans le matriçage bicanal.

Les performances de ce système quant à la restitution de l'espace sont assez médiocres, ou en tout cas, déséquilibrées vers la scène frontale. Cela s'explique évidemment par le dispositif cinématographique et par la présence de l'écran géant captant toute l'attention du spectateur. Il reste tout de même difficile de le considérer comme un système englobant, car l'écart entre les enceintes latérales et l'enceinte arrière est tellement important que l'effet de source fantôme ne peut pas être opérant.

Aussi, il est rare que ce canal « surround », ou arrière, ne soit constitué que d'une seule enceinte à la diffusion. Il est courant de rencontrer un dispositif en « U », situé à l'arrière du spectateur. Cela n'arrange malheureusement pas nos histoires de précisions de restitution du champ spatial.

! Important

En ce qui concerne la gestion du canal LFE (Low Frequency Effect), il n'y a pas de canal dédié. C'est donc un simple système de « bass management », où une partie du grave est filtré des « têtes » (enceintes principales) et redirigé vers le subwoofer.

Le tour de force de Dolby est en réalité de contourner la limitation du support optique sans pour autant remettre en cause la chaîne de production cinématographique. Même après l'avènement du son numérique, une version Dolby Stereo était toujours présente sur la pellicule, pour les salles non équipées en système de diffusion numérique, ou en cas de panne de ce dernier.

Un des métrages emblématiques de cette technologie Dolby Stereo est **La Guerre des Etoiles** de [Georges Lucas](#).

33.3.3 Le passage au son numérique

Le matriçage du signal à de grandes conséquences sur l'intelligibilité de l'espace et pause des problèmes importants de flou de localisation. Lorsque le son numérique se démocratise, la nécessité du matriçage disparaît. On peut ainsi stocker plusieurs canaux audio numérisés sur pellicule. D'ailleurs, il n'est plus nécessaire d'utiliser des réducteurs de bruit, la dynamique du signal numérisé est bien plus grande que celle offerte par les supports analogiques.

i Note

Le passage au son numérique fut le début d'une longue traversée du désert pour l'entreprise Dolby.

L'arrivée du son numérique impose aussi une rude concurrence chez les acteurs de cette mutation. On trouvera ainsi la société nipponne [Sony](#), les Californiens de [DTS](#) et encore et toujours Dolby. On arrive d'ailleurs très bien à relire l'histoire de cette course aux nouveaux formats audio numériques en regardant une pellicule de cette époque.

Dolby occupe déjà toute une partie de la pellicule avec le stockage optique analogique du Dolby Stereo. Sony arrive alors le premier avec son système de son numérique sur support optique avec le SDDS (Sony Dynamic Digital Sound). Lorsque Dolby termine l'élaboration de son encodage AC-3 pour le Dolby Digital, il ne reste alors que l'espace entre les perforations. Finalement, DTS est obligé de trouver une solution de contournement et n'inscrit qu'une piste de timecode sur le film. Cette piste permet alors de synchroniser un lecteur de CD-ROM contenant la piste audio du film.

Le Dolby digital et le format DTS sont tous deux pensés pour une diffusion sur un système 5.1. Le SDDS propose une couverture de l'espace frontale plus importante, avec 5 enceintes accompagnées de deux enceintes arrière pour le « surrounds ». En pratique, ces canaux de diffusions

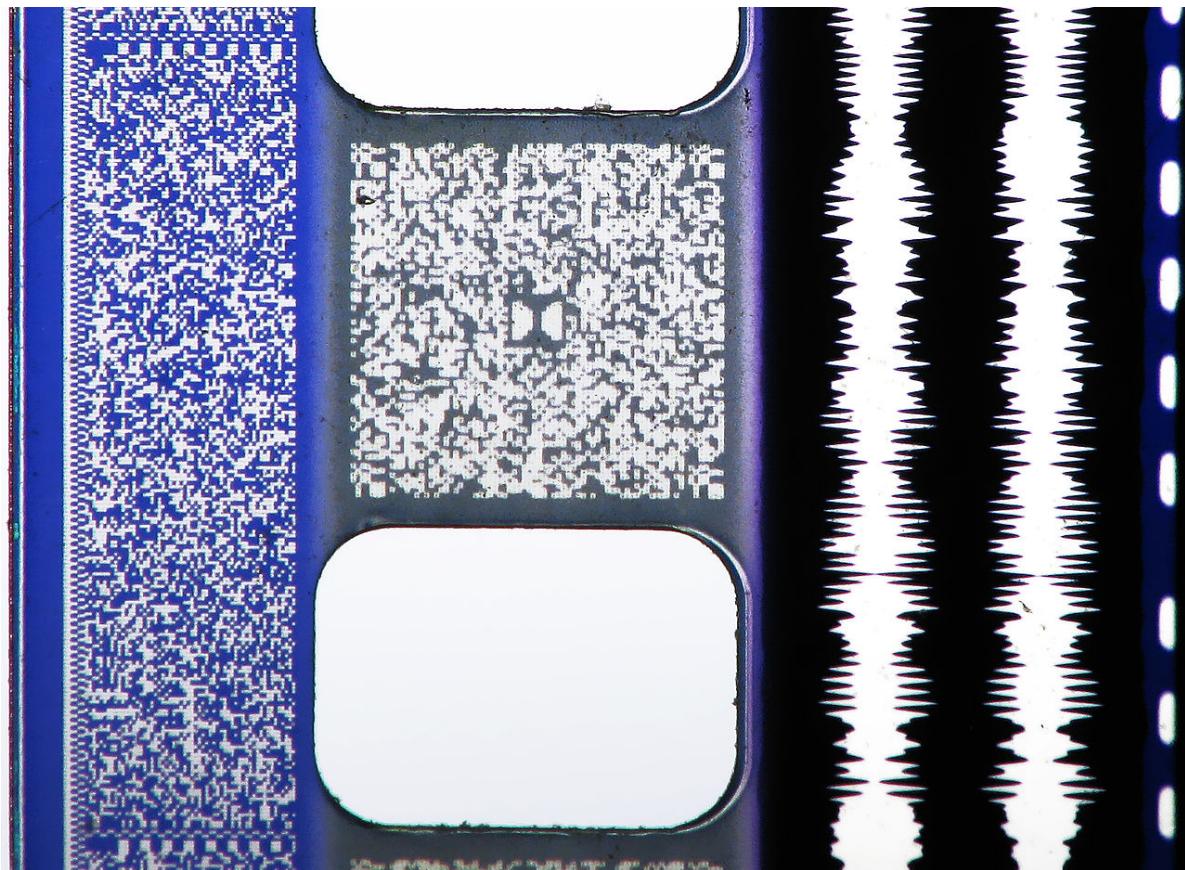


Figure 33.3: Bord d'une copie 35 mm. À droite le Dolby Stereo (analogique), à gauche, le Sony SDDS, au centre, le Dolby Digital, et enfin, tout à droite, la bande de timecode du DTS

supplémentaires ne sont pas exploités lors du mixage, et les mixeurs se content de fournir un mix 5.1 compatible avec l'ensemble des formats. Le premier film mixé en DTS est *Jurassic Park* (1993), de Steven Spielberg, tandis que le premier film utilisant le système Sony est *Last Action Hero* (1993) par John McTiernan.

Le 5.1 n'a pas franchement eu de succès dans son exploitation grand public. Les raisons sont toujours un peu les mêmes : le coût, la place, l'esthétique (certains et certaines n'aiment pas peupler leur séjour d'enceintes). Ce format s'est donc limité à un public de niche, souvent associé à l'audiophilie.

Il y a eu ensuite une multitude de formats dérivés et d'évolutions, comme le Dolby Digital Surround EX, rajoutant une enceinte centrale à l'arrière. Les formats domestiques furent également nombreux, avec les Dolby Surround, Dolby Logic Pro, Dolby Logic Pro II, etc.

L'évolution « majeure » suivante est le passage au 7.1, où l'on considère alors deux enceintes plein gauche et droite et deux enceintes arrière pour compléter le LCR classique. Le premier film mixé en 7.1 est *Toy Story 3* (2010), des studios Pixar.

33.3.4 Passage au mixage orienté objet et le retour en force de Dolby

En 2012, Dolby annonce un tout nouveau format de mixage, le *Dolby Atmos*. C'est une technologie hybride entre mixage orienté canal et mixage orienté objet. Cette approche de mixage orienté objet considère une source sonore (un canal audio mono par exemple) comme un objet, auquel on associe des informations de mixage (volume et position dans l'espace), qui sont ensuite interprétées par un décodeur. Ce dispositif permettait initialement de travailler sur un système d'enceintes dit 7.1.4, soit une 7.1 augmentée de quatre enceintes en élévation. Il s'agit donc d'un des premiers formats cinéma intégrant la composante d'élévation.

L'approche orientée objet permet de solutionner la problématique des « mixdown » des films. En effet, un métrage précédemment mixé en 7.1 doit ensuite être downmixé en 5.1 pour les cinémas moins bien équipés, puis aussi en stéréo pour une exploitation télévisuelle. Ici, il suffit d'indiquer au décodeur le système de haut-parleurs auquel il est connecté et les métadonnées de mixage seront interprétées pour retranscrire au mieux le mixage. On peut même réaliser un mixage dédié à une écoute au casque, en utilisant la synthèse binaurale.

Ce format Dolby Atmos est un succès majeur, replaçant Dolby dans une situation de monopole. Très rapidement, le logiciel *Pro Tools* d'AVID intègre le premier panner Dolby Atmos, et les salles de cinéma s'équipent du décodeur et du système d'enceinte associé.

Ce succès s'infusera aussi dans le monde de la musique. En 2021, Apple annonce le support natif du Dolby Atmos (avec pour la première fois, le moteur de rendu Dolby Atmos intégré) dans son logiciel de musique assistée par ordinateur *Logic Pro*. De plus, son service de streaming, Apple Music, intègre également le support de la lecture de mixages réalisés avec le Dolby Atmos.

33.4 Le binaural : la spatialisation sonore pour tous

Il est bon que nous donnions ici une définition précise du binaural. De nos jours, le binaural rassemble l'ensemble des techniques et des moyens permettant de reproduire l'effet du corps humain (tête, oreilles, torse) sur un signal audio. Un signal binaural comporte donc deux canaux, filtrés par la réponse en fréquence du corps humain. Cependant, dans l'histoire du son, le terme binaural a souvent été utilisé comme synonyme du son stéréophonique.

Note

Par ailleurs, le terme « stéréophonie » se substitue souvent pour désigner tous flux audio comprenant deux canaux. Ce n'est rigoureusement pas juste, car la stéréophonie sous-entendant la diffusion sur un système de haut-parleurs spécifique.

L'[histoire du son binaural](#) s'anime également dans les années 1930, dans les laboratoires Bell. H. Fletcher dirige alors une équipe travaillant sur la mise au point d'un mannequin de cire équipé de microphones sur les joues, nommé Oscar.

Ce premier dispositif permet de réaliser un certain nombre de tests sur notre perception. Il est alors rapidement montré que l'absence d'image lors de l'écoute d'un signal binaural entraîne l'augmentation d'erreurs de localisation (confusion entre l'avant et l'arrière).

En Europe, deux chercheurs de la société néerlandaise [Philips](#), De Boer et Vermeulen, développent le premier mannequin avec une simulation de l'effet du pavillon, grâce à l'intégration des microphones directement dans les oreilles.

Dans les années 40, De Boer dépose un brevet pour un système de prise de son utilisant une simple sphère d'au moins quatorze centimètres de diamètre en lieu et place d'une tête de mannequin. Les microphones sont alors placés de part et d'autre de la sphère.

Toujours aux Pays-Bas, il est alors fait l'expérience de la diffusion de contenus binauraux sur les ondes radio, utilisant sans doute le mannequin de De Boer et Vermeulen.

Durant les années cinquante, plusieurs compagnies développent leurs propres systèmes de prise de son binaural, notamment Schoeps et AKG.

Les années soixante et soixante-dix sont marquées par de nombreuses études sur l'écoute binaurale et l'effet de notre corps sur les signaux. Le système KEMAR devient le premier mannequin de référence. Les mannequins précédents étant souvent récupérés de grands magasins, leurs propriétés acoustiques ne correspondent pas à celle du corps humain. Ce paramètre est alors corrigé par le modèle KEMAR et est ainsi le premier utilisé pour des mesures de prothèses auditives.

En 1973, la société Neumann dévoile sa tête artificielle **KU-80**, équipée de microphones omnidirectionnels **KM83**. L'année suivante, AKG propose elle aussi son système de prise de son binaural **D99c**.

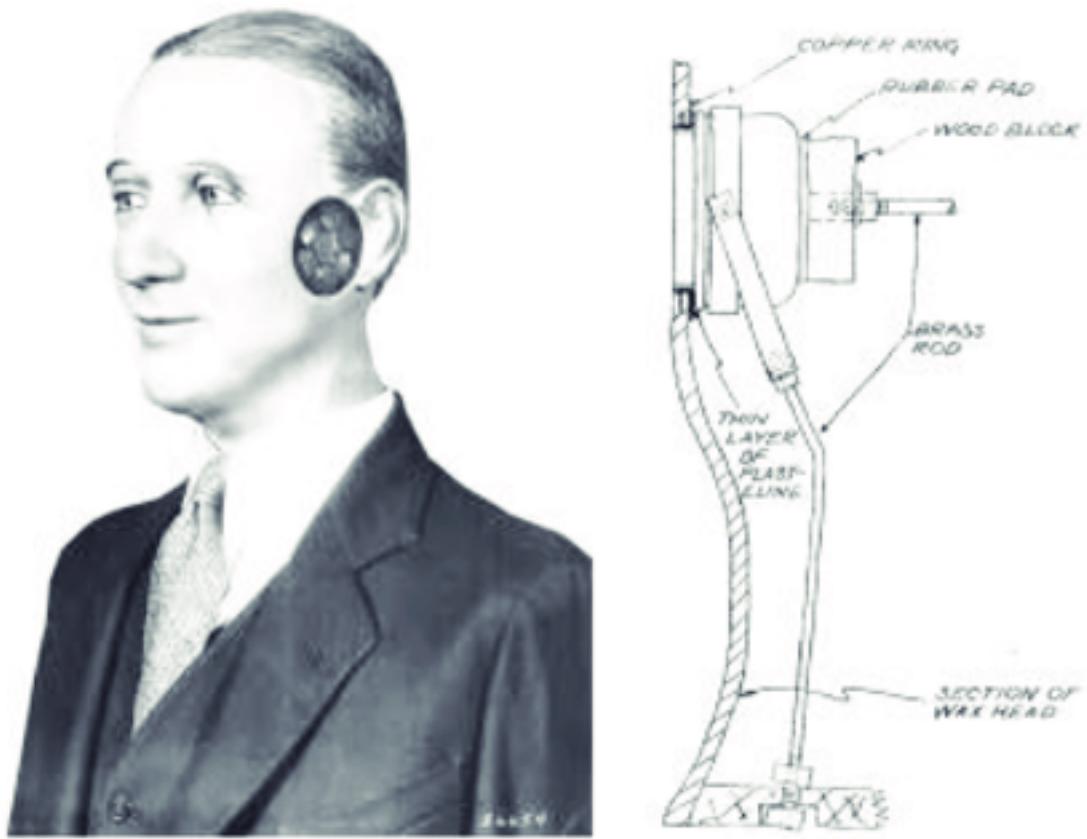


Figure 33.4: Oscar : premier dispositif d'enregistrement binaural

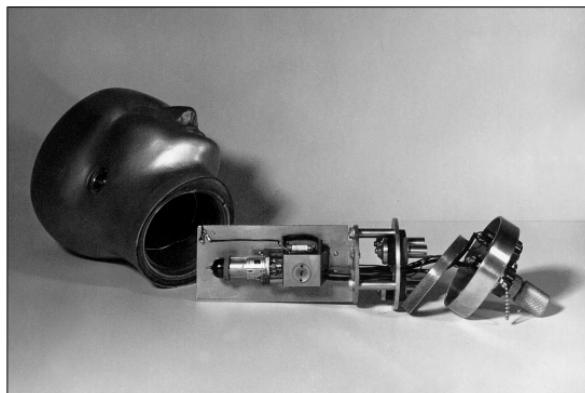


Figure 33.5: Première tête factice d'AKG



Figure 33.6: Sphère de prise de son Schoeps

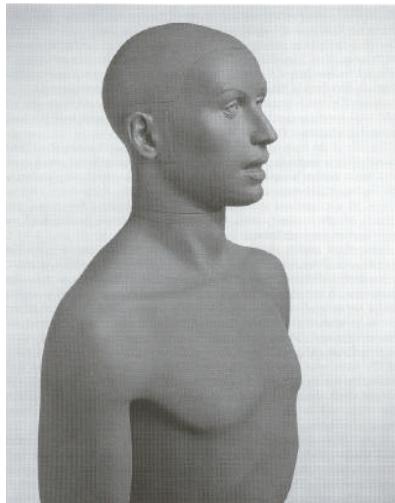


Figure 33.7: KEMAR



Figure 33.8: Neumann KU80



Figure 33.9: AKG d99c

En 1997, un groupe de chercheur publie une [étude comparative des différentes têtes artificielles](#) de l'époque. Il en ressort un problème persistant de trouble de la localisation, particulièrement sur la discrimination de l'avant et de l'arrière. Est alors mise en cause la forme des pavillons des différents mannequins. En partant des résultats de cette étude, plusieurs initiatives similaires apparaissent, consistant à mesurer la réponse en fréquence de la tête d'un grand nombre d'individus, afin de trouver la courbe de réponse correspondant au plus grand nombre. Ce type d'approche donnera, entre autres, naissance à la tête KU-100 de Neumann.

Note

Le résultat de ce type de mesures se nomme **HRTF**, pour Head-Related Transfer Function, ou, fonction de transfert de la tête.

Grâce au passage à l'audionumérique, ainsi qu'à l'augmentation significative de la puissance de calcul des ordinateurs, on voit apparaître dans les années quatre-vingt-dix la technique de la [synthèse binaurale](#) (voir section — [35.2](#)). Il devient alors possible à partir de simples prises de son monophonique de recréer l'effet de la tête sur le signal grâce à une batterie de filtres (toujours ces fameuses HRTF).

Astuce

En France, [Lucie Hardoin](#) est une spécialiste de prise de son et de la postproduction binaurale. Elle met à disposition sur son site web un grand nombre d'extraits pour en découvrir les qualités sonores.

33.5 Les pionniers de la musique spatialisée

S'il est important de couvrir l'histoire technologique, il l'est tout autant de raconter celle de ceux qui ont pensé la spatialisation sonore dans leurs créations. Les mouvements de la [musique concrète](#) et de la [musique électroacoustique](#) ont été particulièrement féconds sur ce sujet.

[Pierre Schaeffer](#) (1910-1995), ingénieur français diplômé de l'école polytechnique, propose les [fondations esthétiques et théoriques](#) de la musique concrète dans les années quarante. Au cœur de ces idées est celle de l'[acousmatique](#). Une musique acousmatique est une musique qui est complètement arrachée à son contexte visuel et n'est plus que son. Elle est grandement permise par les avancées technologiques qui lui sont contemporaines en matière d'enregistrement sonore. L'objectif de l'écoute acousmatique est de permettre à l'auditeur une plus grande imagination à la réponse du stimulus sonore seul. Ce concept est aussi étroitement lié à l'art radiophonique et ses moyens de diffusion. Suite à sa rencontre avec [Pierre Henri](#) (1927-2017), les deux hommes créeront « [Symphonie pour un homme seul](#) » en 1950. Henri collaborera aussi avec [Edgar Varèse](#) en 1954 pour la création de *Déserts* au Théâtre des Champs-Élysées, en tant que spatialisateur. Le concert est aussi diffusé à la radio, et pour la première fois en stéréophonie.

En 1951, Schaeffer créer le « Groupe de Recherche en Musique Concète », avec Pierre Henri à sa tête, qui devient dès 1958 le « [Groupe de Recherche Musique](#) » (GRM). On y retrouve alors, entre autres, [Iannis Xenakis](#), [Michel Chion](#) et [François Bayle](#) (Pierre Henri quitte le projet en 1958, avant la création du GRM).

Note

Pierre Schaeffer fonde également le [service de recherche de la RTF](#) en 1964, dédié à la recherche autour de radio et de la télévision.

Cette branche de la musique a à cœur d'étudier les effets de la captation sonore et de sa diffusion. Si l'on y réfléchit beaucoup sur les façons de transformer le son par la manipulation de son support de stockage, on se pose également la question du moyen de diffusion. En 1974, François Bayle invente l'orchestre de haut-parleurs qu'il nomme [Acousmonium](#). Ce système de diffusion est constitué de multiples haut-parleurs, de taille et qualités sonores différentes, placés sur scène, comme des musiciens. Ces haut-parleurs sont alors « orchestré » par le compositeur qui opère depuis une table de mixage, parfois aussi appelée table de spatialisation. Ce dispositif existe toujours et est régulièrement utilisé en concert.

Astuce

À découvrir, une [archive de 1983](#) sur l'Acousmonium et la musique qu'on lui fait jouer.

Également, est créé en 1970 le Groupe de musique expérimentale de Bourges (**GMEB**) par [Françoise Barrière](#) et [Christian Clozier](#). Il est à l'origine de l'invention d'une table de spatialisation nommée [Gmebaphone](#). Si l'Acousmonium propose un « routage » dynamique des signaux vers les différents haut-parleurs, le Gmebaphone adopte une approche de filtrage, découplant le signal entrant en bandes pour les répartir sur différents haut-parleurs spécialisés.

Le GMEB devient l'**IMEB** en 1997 (Institut de Musique Expérimentale de Bourges) avant d'être fermé définitivement en juillet 2011.

À partir des années 50, un grand nombre de compositeurs (principalement de musique électroacoustique et électronique) se poseront cette question de [l'espace et la spatialisation](#). On peut alors citer : [Iannis Xenakis](#), [Olivier Messaien](#), [Karlheinz Stockhausen](#) et un certain [Pierre Boulez](#) qui fondera l'[IRCAM](#) en 1970.

Astuce

À noter l'excellent site [Electrodoc](#), centre de documentation en musiques électroacoustiques. On y trouve de multiples documentations sur les personnes et les œuvres rattachées à la musique électroacoustique.



Figure 33.10: Une des premières formes de l'Acousmonium



Figure 33.11: Le Gmebaphone

33.6 L'ambisonie : décorrérer l'espace de production et l'espace d'écoute

Depuis le milieu des années soixante, une équipe de chercheurs anglais de la [National Research Development Corporation](#) mènent de nombreuses recherches sur la spatialisation sonore et sur le son « surround ». Nous sommes donc en plein cœur du (bref) boom de la quadriphonie. Parmi eux, on retrouve particulièrement le mathématicien [Micheal Gerzon](#) (1945-1996)

Ensemble, ils élaborent une proposition alternative au son matricé quadriphonique qu'ils nomment « [ambisonie](#) ». Le principe fondamental de cette technique est de décomposer l'espace sonore en plusieurs « directions », ou plus justement, en plusieurs harmoniques sphériques. Alors, nous ne prenons plus le son ni ne le mixons, en tenant compte d'un système de diffusion précis (stéréophonie ou quadriphonie par exemple), mais plutôt en représentant un **espace échantillonné**.

Le premier apport concret des recherches de Michael Gerzon et de ses collègues est le développement d'une technique de prise de son utilisant un tétraèdre de microphones.

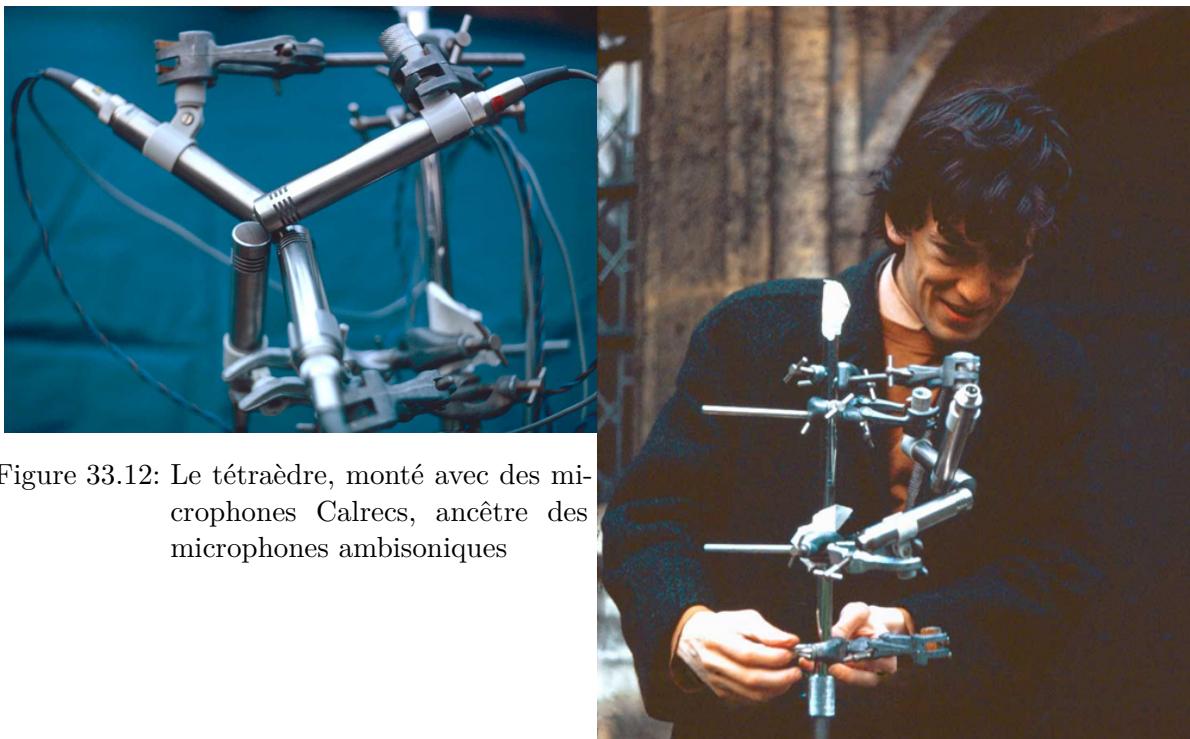


Figure 33.12: Le tétraèdre, monté avec des microphones Calrecs, ancêtre des microphones ambisoniques

Figure 33.13: Michael Gerzon installant le système tétraédrique

De tels dispositifs étant limités par l'encombrement spatial de chaque microphone, il fut ensuite inventé les microphones ambisoniques, regroupant les quatre capsules de façons aussi coïncidentes que possible. Le premier de ce type est le Soundfield, qui fit son baptême en février 1975.

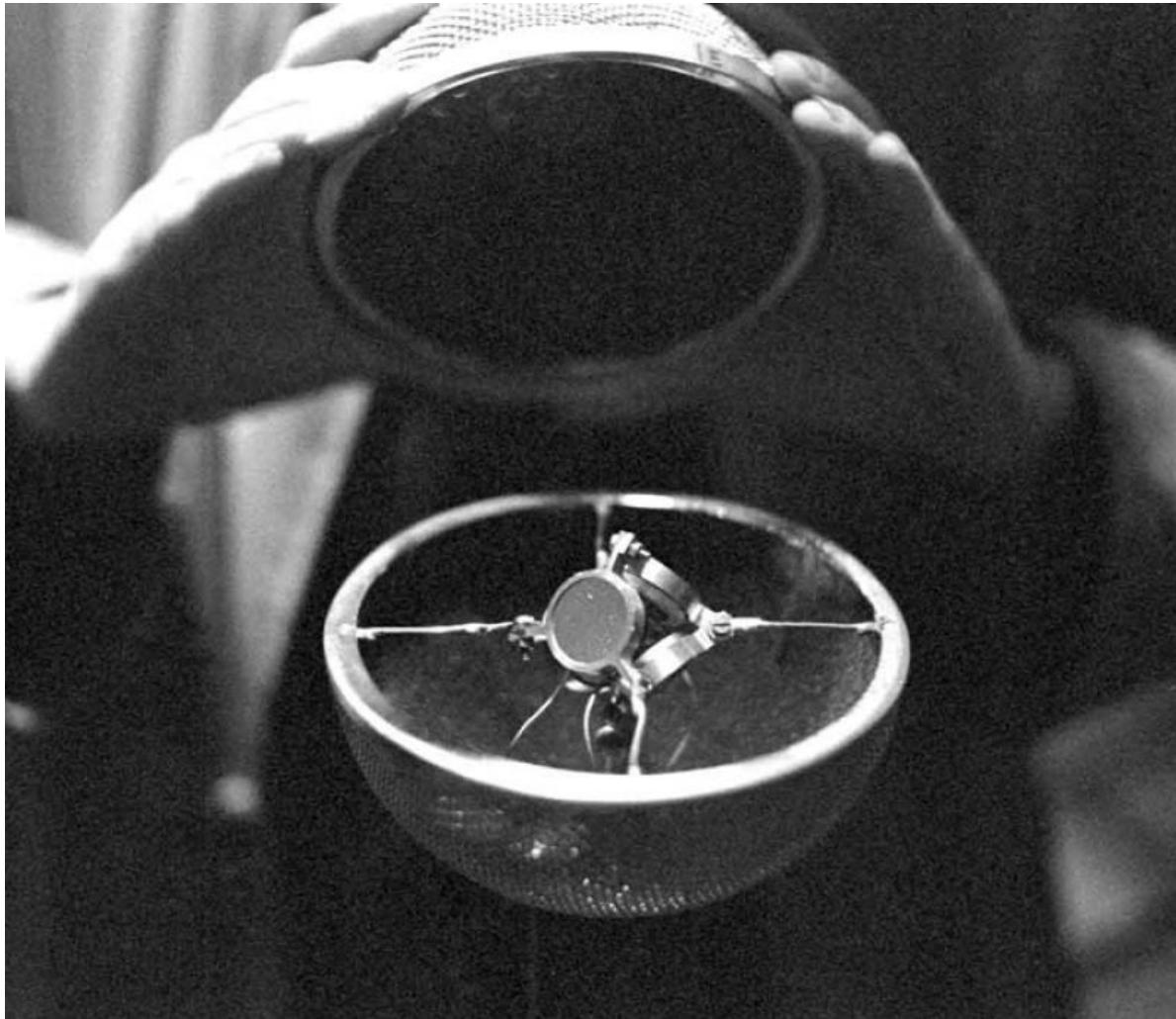


Figure 33.14: Microphone Soundfield

La même année, Michael Gerzon rédige deux articles sur le [principe général](#) de l'ambisonie et son [utilisation en studio](#). On y trouve notamment la définition des différents formats d'ambisonie, principalement:

- Le A-Format, correspondant aux canaux du microphone Soundfield
- Le B-Format, format matricé WXYZ (W est le canal omni, ou commun, les autres canaux sont les harmoniques sphériques permettant d'apposer une direction au son)

L'ambisonie, comme technique de production et de prise de son, ne rencontra malheureusement qu'un intérêt très limité. On notera tout de même les enregistrements du label [Nimbus records](#), réalisés pour la plupart en ambisonie. De plus, lorsque le nombre de haut-parleurs devient trop important, de fortes colorations apparaissent sur le signal. Elle reste alors longtemps une technique restreinte au monde de la recherche.

Note

Michael Gerzon est très influencé par le brevet de 1931 d'Alan Blumlein sur le son stéréophonique. On y retrouve la même idée de matriçage (Stéréo - MS, A-Format — B-Format) et une emphase sur la prise de son coïncidente. Gerzon a d'ailleurs écrit un [article](#) sur le circuit de shuffler d'Alan Blumlein.

L'intérêt autour de l'ambisonie revient dans les années quatre-vingt-dix, notamment grâce aux recherches sur l'Ambisonie d'Ordre Plus Elevé (HOA : Higher Order Ambisonic). Cette stratégie vise à augmenter la finesse d'échantillonnage de l'espace en augmentant le nombre d'harmoniques sphériques utilisés. Cette approche est décrite dans la [thèse de Jérôme Daniels](#). L'ambisonie de Michael Gerzon devient alors l'ambisonie de premier ordre, et chaque nouvel ordre rajoute un certain nombre de nouvelles « directions spatiales ». Cette précision d'échantillonnage permet de meilleurs rendus sur un grand nombre de haut-parleurs, ainsi qu'un décodage ambisonique vers binaural plus qualitatif. En guise d'exemple, les outils de création de vidéos en réalité virtuelle (VR) de Meta supportent l'ambisonie jusqu'au troisième ordre.

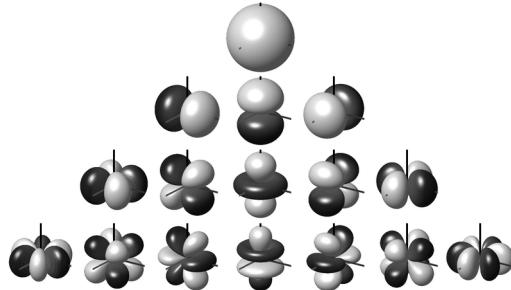


Figure 33.15: Représentation des harmoniques sphériques des ordres 0 à 3

Ce regain d'intérêt et la multiplication des canaux ambisoniques sont à corrélérer avec l'augmentation de la puissance des ordinateurs. En effet, les outils de traitement numériques sont bien plus adéquats que leurs homologues analogiques pour le traitement ambisonique. De plus, l'industrie de la réalité virtuelle, qu'elle soit sous la forme de simples vidéos, ou mieux vidéoludique, a montré un réel intérêt pour les formats audio ambisoniques, notamment pour leurs capacités à être **interactifs**.

34 Les espaces haut-parlants

Une fois n'est pas coutume, nous allons aborder le problème de la spatialisation par sa fin : le système de diffusion. De quoi parle-t-on lorsque l'on parle d'un système « immersif ». À vrai dire, que qualifie réellement ce terme « immersif » ?

Si nous raisonnons simplement, notre expérience naturelle du son est immersive. Nous sommes (plus ou moins) capables d'identifier la position d'un évènement sonore dans l'espace. Dès lors, placer une enceinte dans un lieu est déjà une expérience immersive. En effet, nous pouvons nous déplacer autour, nous en éloigner, nous rapprocher, etc. Aussi, la réverbération naturelle du lieu va aussi créer une sensation d'enveloppement, plus ou moins grande, en fonction de son temps de décroissance et de sa puissance sonore.

Nous commençons donc à pressentir qu'il ne faut pas confondre l'**espace d'écoute**, situé après les enceintes, et l'**espace produit**, situé avant les enceintes. Quand on nous parle d'« immersif », c'est bien à l'espace produit que l'on fait référence. Selon la formule de Jean-Marc Duchenne, on définit ainsi l'**espace haut-parlant** comme ce que l'on donne à entendre.

34.1 Catégorisation des espaces haut-parlants

On peut généralement différencier deux types de spectacle sonores : ceux où la position de l'auditeur est connue et fixe, et ceux où l'on ne peut pas présupposer du positionnement de l'auditeur. Par exemple, lors de l'écoute d'un disque, on suppose que l'auditeur se positionnera au point d'écoute idéal (sweetspot), qui était alors la même place que l'ingénieur du son pendant sa production. À l'inverse, en muséographie, l'auditeur se déplace en permanence et, même si souvent guidés par un sens de visite, ses déplacements sont aléatoires.

Ces deux grandes catégories ont une forte influence sur le choix des technologies de spatialisation.

Note

Le cas du concert amplifié est un peu plus ambigu. Les zones dédiées au public sont connues à l'avance, et l'on suppose que celui-ci ne se déplace pas pendant le spectacle. Cependant, on cherche à tout prix à éviter la création d'un sweetspot. Cela impliquerait que le rendu sonore serait idéal pour quelques personnes et médiocre pour le reste du public. On parle alors de zones de diffusions, que l'on tente de rendre aussi homogènes

que possible. Ce travail est celui du caleur système.

On considère d'ailleurs souvent que les mixages live sont monophoniques, et que le distribue le même mixage sur plusieurs points de diffusion. Cela se nomme la « multi-mono ». Ici, la nécessité de multiples points de diffusions relève de l'homogénéité de la diffusion, et rarement du souci de la spatialisation. Cependant, les techniques de spatialisation permettent aussi d'accroître cette homogénéité, tout en ajoutant une dimension spatiale au spectacle.

Le second cas, où l'auditeur est libre de ses déplacements, donne également la plus grande liberté sur le choix du système à déployer. Les haut-parleurs pourront alors peupler l'espace en n'importe quel point et deviennent alors un élément à part entière de la création sonore. On pourrait même envisager le choix d'un haut-parleur pour une réponse en fréquence particulière, ou pour tout autre défaut qui pourrait prendre sens dans la narratologie d'une œuvre. Cette approche a été un des points de réflexion centraux de la musique concrète et acousmatique. Les espaces haut-parlants supposant un emplacement d'écoute idéal sont les plus courants. Ils concernent la musique enregistrée, le cinéma et une grande majorité du spectacle vivant.

34.2 Les systèmes de diffusions

On peut sous-catégoriser les espaces haut-parlants en caractérisant le système de haut-parleurs utilisé. On distingue ainsi quatre grandes familles :

- Les systèmes frontaux
- Les systèmes englobants à une dimension
- Les systèmes englobants à deux dimensions
- Les systèmes à trois dimensions

! Important

On considère ici un système de coordonnées sphérique, comprenant donc la distance (ρ , sur le schéma ci-dessous), l'azimut (Θ) et l'élévation (δ). Ce repère offre l'avantage d'être au plus proche de notre perception de l'espace.

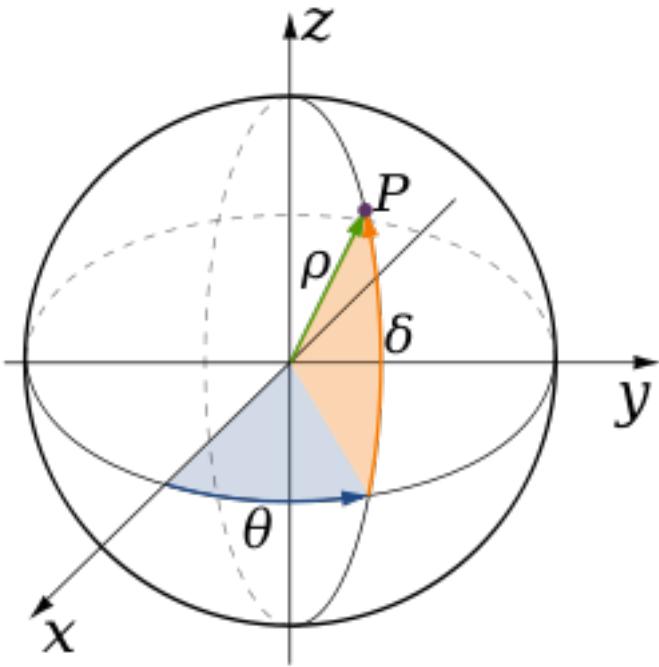


Figure 34.1: Repère de coordonnées sphériques

Dans la première catégorie, on retrouve notamment deux systèmes historiques : la monophonie et la stéréophonie. Le premier se résume par un espace haut-parlant comprenant un unique haut-parleur. Le second est décrit comme un espace haut-parlant comprenant deux haut-parleurs, formant un triangle équilatéral avec l'auditeur. On pourra également évoquer le système LCR, principalement associé au cinéma. En spectacle, on trouve parfois des systèmes en lignes d'enceintes, comme le L-ISA d'L-acoustique, ou encore certains systèmes utilisant la **WFS** (voir le chapitre 39)

Les systèmes englobants à une dimension comprennent l'ensemble des systèmes de haut-parleurs encerclant l'auditeur, offrant ainsi un degré complet de liberté de spatialisation. Le mixeur peut ainsi placer un son tout autour de l'auditeur en changeant son angle d'incidence. On pense alors au système quadriphonique, et à tous les systèmes de cinéma dit « surround » : LC5, 5.1, 7.1, etc.

Les systèmes englobants à deux dimensions rajoutent un nouvel axe de liberté, permettant d'élever un son. Le Dolby Atmos est sans doute le représentant le plus connu de cette famille de système. Ces systèmes sont le plus souvent, pour des raisons physiques évidentes, demi-sphériques, et seule l'élévation positive y est possible. Ces dômes peuvent prendre des formes diverses et variées, avec plusieurs couronnes d'enceintes en hauteur.

Les derniers systèmes, et également les plus rares, sont les systèmes proposant trois axes de liberté. On a donc un **maillage** d'enceinte ponctuant l'espace et permettant ainsi de créer des

effets de profondeurs, non pas par des moyens perceptifs (rapport de volume, réverbération, etc.), mais pas des moyens physiques (enceinte réellement présente au point de diffusion).

Astuce

Afin de simplifier le vocabulaire, on admettra que « surround » est synonyme de systèmes d'enceintes englobant à une dimension, et, que « immersive » et « 3D » son synonyme de système englobant à deux dimensions.

34.3 Une simple analogie image-son

La notion d'échantillonnage d'un signal se retrouve dans de multiples disciplines, notamment celles de l'image et de la vidéo. Par exemple, une image diffusée sur un écran est spatialement échantillonnée. Un peu comme les peintres impressionnistes qui créaient l'illusion d'une image unie à partie de touche de pinceau distincte, on retrouve ici un principe parfaitement analogue. Une image numérique est donc cadre d'un certain nombre de points, et chacun de ces points s'appelle un pixel. C'est le plus petit grain d'une image. Si l'on utilise trop peu de pixels pour décrire une image, ceux-ci deviennent plus gros, et l'on distingue alors ces pixels, et l'illusion d'une image unie est perdue.

La notion intéressante rendue évidente par le traitement de l'image est la notion d'échantillonnage de l'espace. Dans le champ d'application du son, cette « résolution d'espace » existe aussi, et est directement liée au nombre de haut-parleurs. Plus le nombre de haut-parleurs est grand, plus la cohérence de notre espace haut-parlant sera grande. On pourrait alors qualifier les systèmes mono, stéréo, voir 5.1 et 7.1 de systèmes basse résolution.

Note

Gardons bien en tête cette idée d'échantillonnage d'espace, car c'est exactement le sujet de l'ambisonique. La WFS reprend aussi cette d'échantillonnage, mais l'applique au front d'onde de l'onde sonore.

Si l'on prend en exemple le système stéréophonique, nous avons à notre disposition deux enceintes, une à gauche de l'auditeur et une à sa droite. De manière assez évidente, si nous envoyons un signal seulement sur l'enceinte gauche, le son semblera venir de la gauche et vice-versa. Alors, que se passe-t-il lorsque nous envoyons le même signal (amplitude et phase identique) sur les deux enceintes ? Il se créer alors un effet psychoacoustique qui va nous donner l'illusion que le son provient du « centre de gravité » du système (barycentre serait plus juste). Ce phénomène se nomme **centre fantôme**. Si l'on introduit un écart de gain entre les deux enceintes, cette source fantôme se déplace vers l'enceinte la plus forte. Il s'agit alors d'une

forme d'interpolation. On déduit ainsi des points de diffusions physiquement inexistantes grâce aux points de diffusions réels environnants.

Par rapport à notre question de résolution, plus celle d'un système est faible, plus on peut supposer que l'espacement entre les enceintes sera grand. Le poids mis sur cette stratégie d'interpolation est alors plus important, jusqu'au point où l'illusion du centre fantôme cesse d'opérer.

34.4 Études des systèmes normés

34.4.1 Les systèmes frontaux

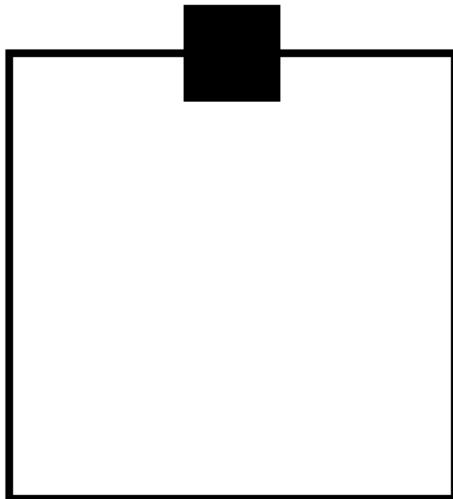


Figure 34.2: Monophonie

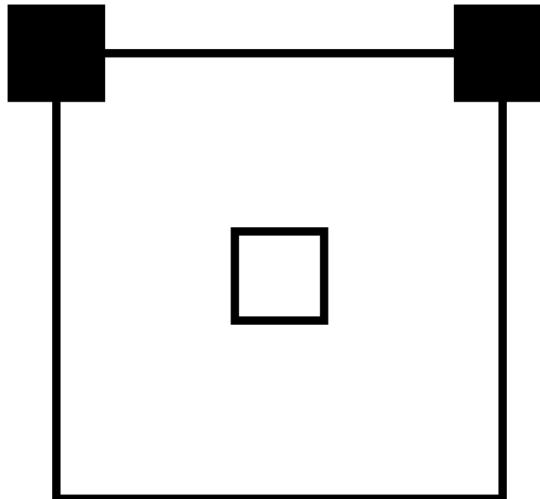


Figure 34.3: Stereophonie

La monophonie n'offre évidemment aucun axe de liberté pour la spatialisation. On pourra éventuellement « tricher » un effet de profondeur en jouant sur le volume des sources sonores et sur le mixage de la réverbération. Depuis la fin des années soixante, la norme de production et d'écoute est la stéréophonie. La stéréophonie repose sur deux enceintes, formant un triangle équilatéral avec l'auditeur. Elle offre alors une scène sonore de 60° face à l'auditeur, et donc un premier axe de liberté dans la spatialisation. Il semble évident que l'on ne peut pas faire jouer un son plus à gauche que l'enceinte gauche, et plus à droite que l'enceinte droite. Dès lors, les axes formés par chacune des enceintes avec le point d'écoute idéal sont des frontières infranchissables par les sons.

Astuce

Le transaural est une technique de spatialisation qui permet, dans une certaine mesure, d'effacer ces frontières. Mais cela vient avec un certain nombre d'inconvénients détaillé dans la chapitre [35.3](#).

Chacune des enceintes est alimentée par un signal audio dédié. On associe alors au système de diffusion stéréophonique un mixage stéréophonique, comptant deux canaux.

Avertissement

Il ne faut pas confondre « stéréophonie » et « deux canaux ». Si la stéréophonie implique un mixage sur deux canaux, un mixage sur deux canaux n'implique pas de la stéréophonie. Par exemple, une écoute au casque n'est pas une écoute stéréophonique. Un mixage binaural repose aussi sur deux canaux, mais n'est pas du tout indiqué pour une écoute sur un système stéréophonique.

Comme nous l'avons vu précédemment (chapitre [33](#)), au cinéma, le besoin d'un canal central, caché derrière l'écran, c'est très rapidement fait sentir, principalement avec l'essor des formats larges-écrans type Cinerama et Cinemascope. On trouve donc des systèmes LCR, soit « left », « center », « right ». Cependant, le support optique de la pellicule ne permet pas d'y stocker plus de deux canaux. On réalise alors un matriçage, permettant de réduire le mixage LCR sur deux canaux (appelés Lt, pour « Left total » et Rt, pour « Right total »), puis, à la diffusion du film, on opère le dématriçage vers le système LCR. Le canal central est alors alimenté par la sommation des canaux L et R.

Note

Le canal central est alors formé par les signaux **en phase** du matriçage bicanal.

Une telle réduction n'est pas indolore sur le signal, et nuit largement à la cohérence d'espace de la diffusion.

34.4.2 Les systèmes englobants à une dimension

La quadriphonie est le premier système de ce type marquant, apparu dans les années soixante-dix. On place alors les quatre enceintes en carré et l'auditeur au centre. On notera ainsi que les angles enceintes-auditeur sont de 90° , soit 30° de plus que la stéréophonie. Nous pouvons déduire que l'on gagne globalement en couverture spatiale, mais l'on perd en précision sur les sources fantômes.

Son homologue dans le monde du cinéma est le LCRS, composé de trois enceintes frontales (gauche, centre, droit) et d'une enceinte arrière. La position exact des enceintes en cinéma

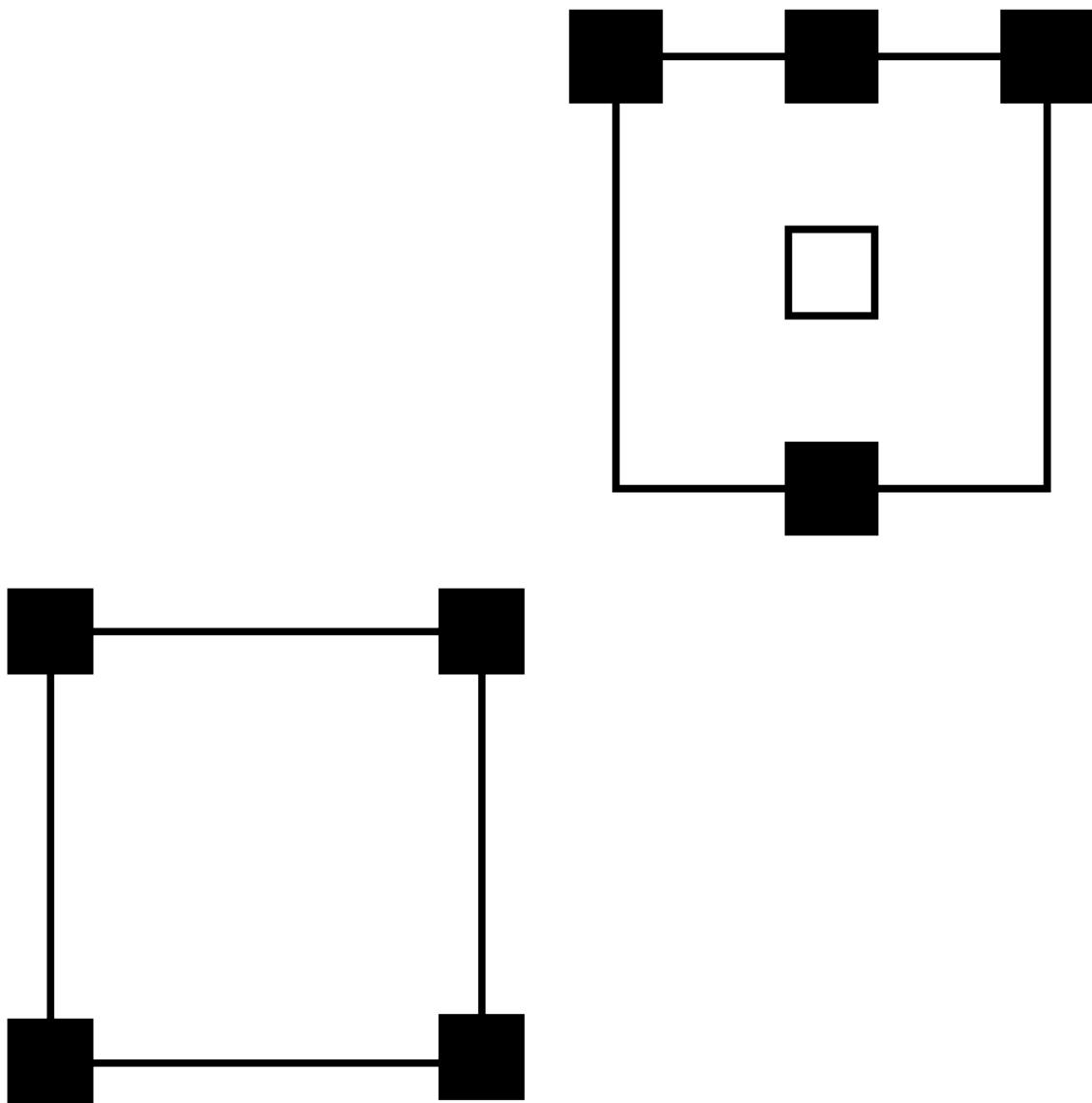


Figure 34.4: Quadriphonie

est toujours un sujet délicat. On considère généralement que les enceintes gauche et droite forme une stéréophonie (plus ou moins 30°) auxquelles ont rajoute une enceinte centrale et une enceinte arrière (respectivement 0° et 180°). Cependant, la diffusion cinématographique n'est pas vraiment tournée vers un point précis. En effet, il convient de convenablement couvrir l'ensemble des spectateurs de la salle. Alors, dans les auditoriums de mixage ainsi que dans les salles, cette preservation d'angles est somme toute assez relative.

Astuce

Le système LCRS est un système historique et n'est plus du tout utilisé aujourd'hui. La quadraphonie quand à elle continue d'avoir une certaine popularité. On peut y avoir deux phénomènes : une nostalgie toujours plus présente pour le "vintage", mais aussi une certaine facilité de prise en main pour les artistes. Il est en effet assez facile d'expliquer la quadraphonie comme quatre espace stéréophoniques (même si ce n'est pas rigoureusement exact).

C'est grâce à la transition du son analogique vers le son numérique que l'on voit apparaître de nouveaux systèmes de diffusions englobants. Toujours dans le cinéma, le premier et le plus connu est le système 5.1, défendu par Dolby et par DTS.

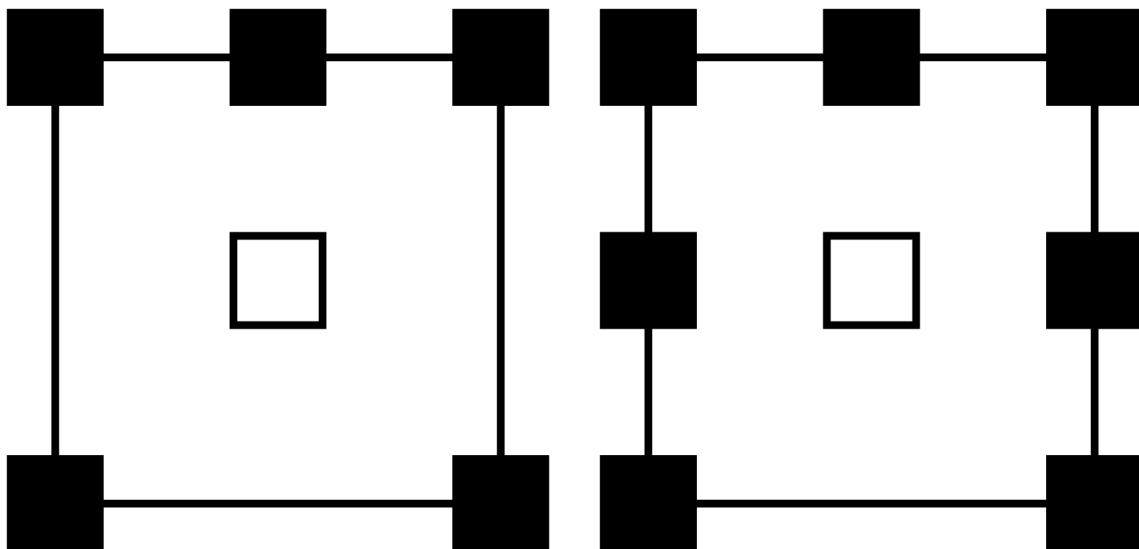


Figure 34.5: 5.1

Figure 34.6: 7.1

Il s'agit au final d'une simple extension du système LCRS, que l'on va augmenter d'un canal de LFE distinct et d'un canal arrière supplémentaire. Alors, le terme arrière est un peu exagéré, car ces enceintes sont placées à plus ou moins 110 degrés de l'enceinte centrale. On a donc plutôt un placement latéral, légèrement décalé vers l'arrière. On notera aussi l'écart important

entre les canaux latéraux avant et ces canaux arrière, grand de 80° . Pire, les enceintes arrières sont séparées d'un angle de 140° . Il est donc délicat d'envisager un placement de sources fantôme à l'arrière de l'auditeur. Aussi, il est plus sage de considérer, lors d'un mixage, qu'un système 5.1 n'offre qu'une scène sonore que de 220° .

Note

Le système quadriphonique est quelque part plus homogène et couvre un espace plus grand que le système 5.1. Par contre, sa définition et sa précision sont moindres sur l'espace frontal.

Le système 5.1 est défini dans la recommandation de l'ITU (International Union Broadcast) n° [BS.775-1](#). On note d'ailleurs que le placement des enceintes possède une marge de tolérance, celles-ci pouvant être placées entre plus ou moins 100° jusqu'à plus ou moins 120° du centre.

Cette recommandation de l'ITU propose également un ordonnancement des canaux. En effet, ce sujet n'est pas normé, et donne lieu parfois à certaines erreurs de rendus. L'ITU propose alors de ranger les canaux dans l'ordre suivant : gauche, droit, centre, LFE, arrière gauche, arrière droit. Cependant, Pro Tools, logiciel de mixage et montage son le plus répandu dans le monde professionnel, adopte l'agencement suivant : gauche, centre, droit, arrière gauche, arrière droit, LFE. Il convient alors d'être excessivement vigilant quant à la diffusion d'un contenu 5.1, et de bien vérifier le bon routage de chaque canal.

Le format marquant suivant est le 7.1 Dolby Surround. On conserve alors le 5.1 que nous avons déjà décrit et y rajoutons deux enceintes, véritable arrière gauche et arrière droit. Il s'est principalement imposé en salle de cinéma.

Nous pouvons également évoquer un système supplémentaire, l'octophonie, mais qui n'a pas eu de réelle exploitation commerciale. En d'autres termes, son utilisation s'est bornée à l'exploitation lors de spectacles vivants, sans intégrer la problématique du stockage, du transport et de la restitution antérieure du programme. Il s'agit cependant de systèmes relativement classiques.

L'octophonie est un nom que l'on rencontre assez fréquemment. Cependant, son nom est souvent rattaché à deux systèmes différents. Dans ce livre, nous appellerons « octophonie » un système de haut-parleurs régulièrement positionnés sur un cercle. Les enceintes sont alors toutes espacées de 45° . On peut voir ce système comme une augmentation de la résolution de la quadriphonie. Cette octophonie offre deux avantages : sa couverture est homogène et sa densité satisfaisante.

L'autre système associé à l'octophonie est le cube, que nous aborderons dans la section suivante.

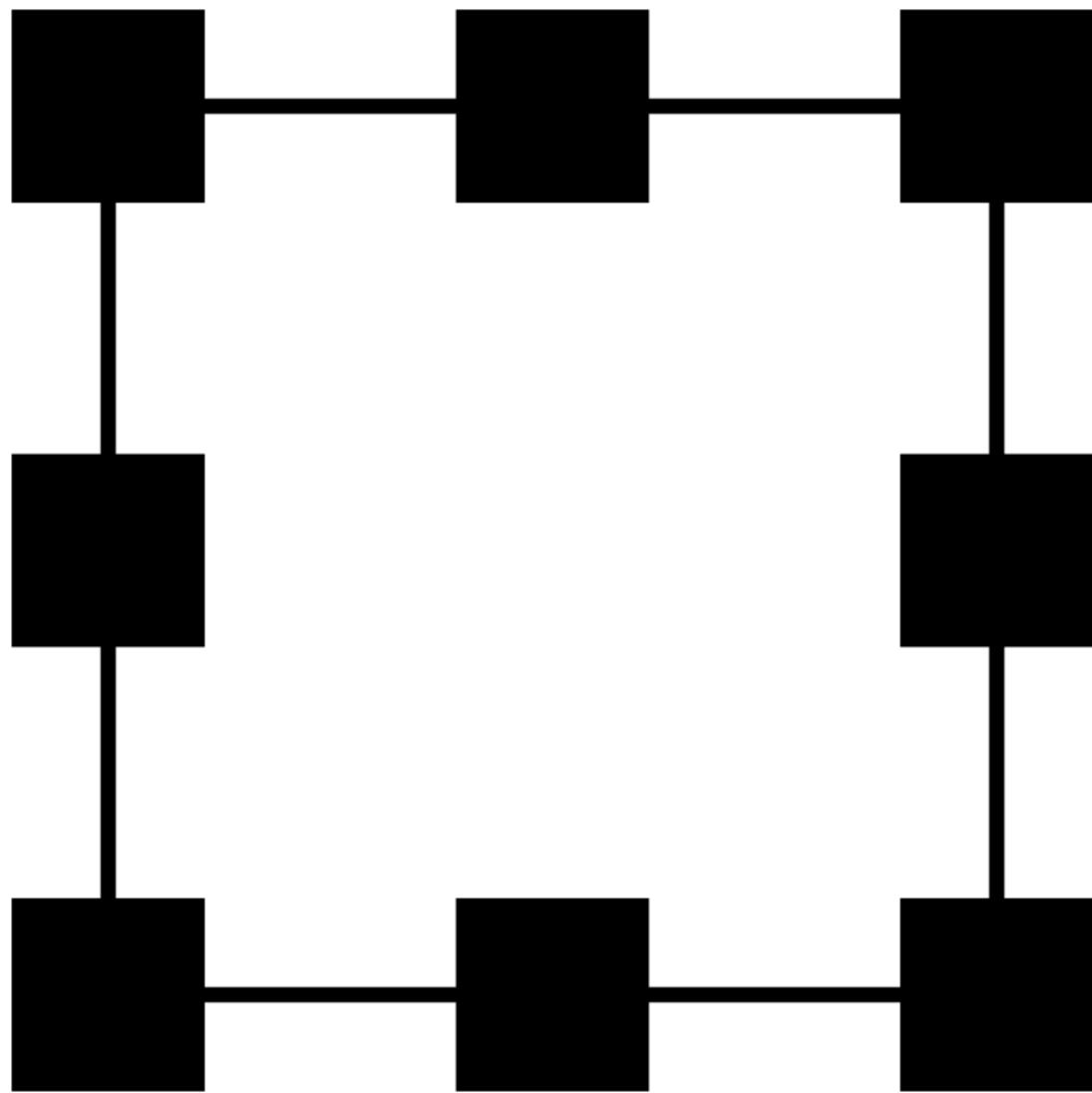


Figure 34.7: Octophonie

34.4.3 Les systèmes englobants à deux dimensions

Avant de nous attarder sur un autre (encore) format de spatialisation de Dolby, nous pouvons faire un détour par le Japon. La société NHK (Nippon Hōsō Kyōkai, ou compagnie de télédifusion japonaise) fait partie du réseau de télédiffusion publique japonaise. Celle-ci possède une branche dédiée à la recherche nommée « NHK Science & Technology Research Laboratories ». Ils ont alors proposé un nouveau format de son « surround » : le 22.2.

Ce système propose trois niveaux d'élévation : un au sol, un à hauteur d'oreille, et un dernier positionné en hauteur. On y ajoute également une enceinte dite de « voice of god », placée au-dessus de la tête de l'auditeur.

- Le niveau inférieur est constitué de trois haut-parleurs frontaux
- Le niveau moyen est composé de dix haut-parleurs. On pourra alors le décomposer comme une octophonie (une enceinte tous les 45°), plus deux enceintes supplémentaires sur la scène frontale.
- Le niveau supérieur comporte huit haut-parleurs, disposés comme une octophonie, plus une enceinte juste au-dessus de l'auditeur.

Ce dispositif offre une couverture relativement homogène de l'espace et surtout plus dense que la plupart de ces concurrents. Malheureusement, hors du Japon, ce système n'est pas du tout répandu.

Note

A consulter, [le document de NHK sur son 22.2](#).

Revenons aussi sur nos questions d'octophonie. Cette appellation est parfois utilisée pour parler d'un système de haut-parleur cube composé de deux plans :

- une quadriphonie au sol
- une quadriphonie en élévation

On couvre ainsi une sphère complète, mais avec une résolution plutôt faible. On met ici beaucoup de pression sur l'effet de source fantôme.

Nous pouvons aussi aborder les systèmes définis algorithmiquement. On indique alors un nombre de haut-parleurs souhaités, et l'on obtient un arrangement de points dans l'espace positionnés homogènement.

Mise en garde

Si le placement de point de façon homogène sur un cercle paraît être une tâche triviale, ce n'est pas du tout le cas pour une sphère.

Retenons deux lois de répartitions importantes :

- Le maillage [t-design](#)
- Le maillage de [Lebedev](#)

Nous reviendrons plus particulièrement sur ces maillages lors de leurs utilisations dans le cadre de l'ambisonique. En dehors de ce cas, ces lois peuvent être utilisées dans la construction de dômes où de sphères de haut-parleurs.

Enfin, nous avons donc les arrangements de haut-parleurs associés au Dolby Atmos. Le plus commun est le 7.1.4, reprenant un 7.1 Dolby et y rajoutant une quadriphonie en élévation. Une première réduction consiste à réduire la couche d'élévation à deux canaux (7.1.2). On trouve également les 5.1.4 et 5.1.2, reprenant la même logique.

Plus intéressant, le Dolby Atmos propose aussi un format 9.1.6, qui apporte enfin une homogénéité correcte, au moins sur le plan principal.

Avertissement

Les techniques de mixage entourant le Dolby Atmos sont très différentes des techniques classiques, reposant sur des lois de panoramiques. Le Dolby Atmos est avant tout un format de mixage orienté objet. Nous abordons cet aspect dans le chapitre [45](#).

Si Dolby possède aujourd'hui la force de frappe la plus importante, on trouve tout de même quelques acteurs pour tenter de lui faire concurrence. DTS propose son propre format de mixage orienté objet, nommé DTS:X. Les systèmes de diffusion sont cependant calqués sur ceux de Dolby. On retrouve aussi Auro 3D, également avec son propre format de mixage orienté objet. Compatible avec les systèmes Dolby Atmos, on retrouve aussi le 13.1 (7.1.4, plus une enceinte centrale sur le plan d'élévation et une "voice of god") et 11.1 (5.1.4, plus une enceinte centrale sur le plan d'élévation et une "voice of god").

35 Le binaural

[Binaural recording](#)

[Binaural recording DPA](#)

35.1 La prise de son binaurale

35.2 La synthèse binaurale

35.3 Le transaural

36 Orienté canal

36.1 L'approche perceptive

36.1.1 Le VBAP

36.1.2 Le VBIP

36.1.3 Le LBAP

36.1.4 Autres variantes

36.2 L'approche matricielle

36.2.1 Le DBAP

36.2.2 Le KNN

37 L'ambisonique



Figure 37.1: Dôme ambisonique du Théâtre de Gennevilliers

L'ambisonique (ou ambisonie) s'attache à décrire le champ acoustique en trois dimensions d'un espace donné en un point. C'est à la fois une technique de prise de son, grâce à l'utilisation de microphones particuliers et une solution de panning en postproduction. De par son indépendance par rapport au dispositif du système de diffusion, on qualifie parfois un mixage ambisonique de « *mixage orienté scène* ». En effet, un signal ambisonique décrit, à l'aide de canaux audio, une scène sonore, et non pas un certain arrangement de haut-parleur (par opposition à l'orienté canal).

L'ambisonie se distingue également des approches plus conventionnelles (approche perceptive du panning usuel) par son approche **physique**. Nous verrons que dans son fonctionnement, l'ambisonie réalise l'**échantillonnage du champ de pression acoustique**. Cet échantillonnage peut varier en précision en fonction de l'**ordre** auquel nous souhaitons travailler.

Nous commencerons par étudier l'ambisonie dit du “premier ordre”, tel que proposé par Michael Gerzon et son équipe dans les années soixante-dix. Seulement ensuite, nous envisagerons

l'ambisonie d'ordre plus élevé, pour pallier aux défauts de l'ambisonie du premier ordre.

37.1 L'ambisonie du premier ordre (FOA)

37.1.1 Captation du champ sonore

L'ambisonie du premier ordre (ou FOA pour « First Order Ambisonic ») voit le jour sous la forme d'une technique de prise de son. Celle-ci permet l'enregistrement d'une scène sonore sur **quatre canaux**, que l'on peut ensuite décoder sur n'importe quel système de haut-parleurs.

Pour capturer le champ acoustique en un point, il faut donc s'intéresser au champ acoustique lui-même. Nous l'avons vu à la Section 2.1, sous sa forme acoustique, une onde sonore se caractérise par la variation locale dans le temps de la pression. Pour mesurer la pression en un point, nous pouvons utiliser un microphone omnidirectionnel (également appelé microphone à pression). Ce microphone omnidirectionnel va donc rendre compte à chaque instant du temps de la valeur de la pression. À ce stade, il n'est pas question de parler de spatialisation, la captation d'une telle capsule étant monophonique.

Il conviendrait donc de mesurer la “direction” du déplacement local des particules d'air. Quelque part, on se demande dans **quel sens et quelle direction varie la pression**. Pour se faire, on utilise un microphone bidirectionnel (aussi appelé à gradient de pression). Ce microphone va donc mesurer, pour chaque instant du temps, la différence entre la pression entre un point de l'espace et un autre très rapproché. On caractérise donc une **variation dans l'espace**.

Une capsule bidirectionnelle nous permet de mesurer sur quel axe se déplacent les molécules d'air (si les molécules oscillent dans l'axe du microphone, on obtient un niveau mesurable, si les molécules oscillent sur l'axe normal (à 90°) de l'axe du microphone, on obtient un niveau négligeable). Cependant, cela ne nous indique pas si la source sonore responsable de la perturbation se trouve plutôt devant ou derrière le microphone. Néanmoins, si nous associons au même point de l'espace une capsule omnidirectionnelle et une capsule bidirectionnelle, nous allons pouvoir lever cette indétermination.

On constate alors que, selon l'angle d'incidence de la source par rapport aux capsules, celle-ci subit une atténuation (et une modification de la phase si la valeur absolue de l'angle d'incidence est ici supérieure à 90°) dans le canal de captation Y, permettant donc de coder sa position autour du microphone. Le canal W capte la source de la même manière, peu importe son angle d'incidence. Nous venons ici de fabriquer un couple MS (voir Section 16.3.2).

Notre dispositif permet ici de capter des ondes sonores selon un seul axe (ici, Y). Si l'on souhaite étendre ce dispositif pour capter l'espace en trois dimensions, nous serions naturellement tentés d'ajouter deux autres capsules bidirectionnelles supplémentaires.

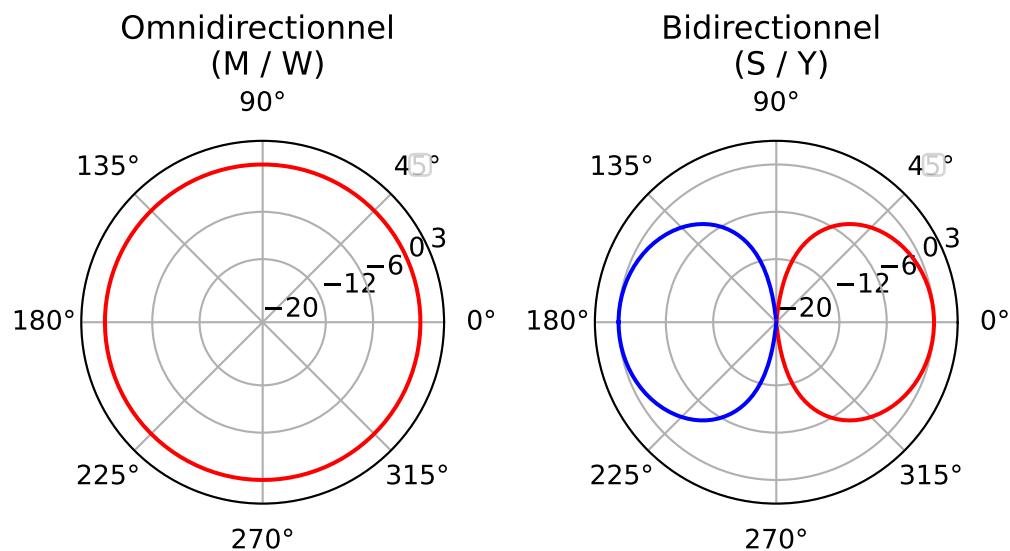


Figure 37.2: Pression et variation de pression sur l'axe Y. Les portions rouges ont une phase positive, les portions bleues ont une phase négative.

Association de deux harmoniques sphériques
(M et S / W et Y)
90°

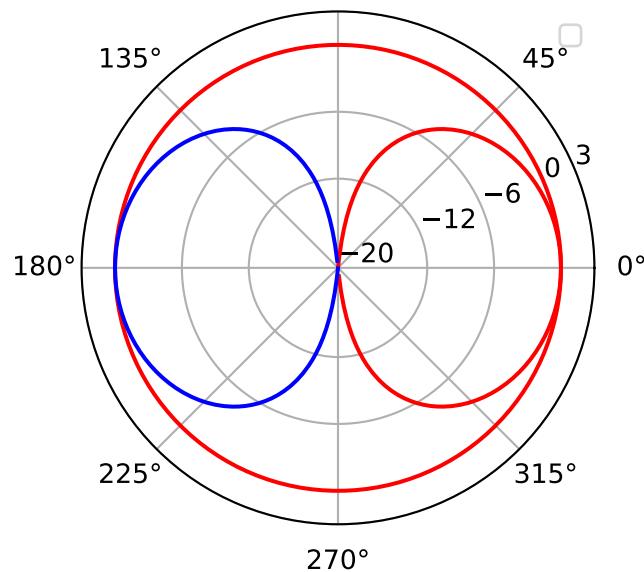


Figure 37.3: Représentation des composantes W et Y sur le même graphique. Les portions rouges ont une phase positive, les portions bleues ont une phase négative.

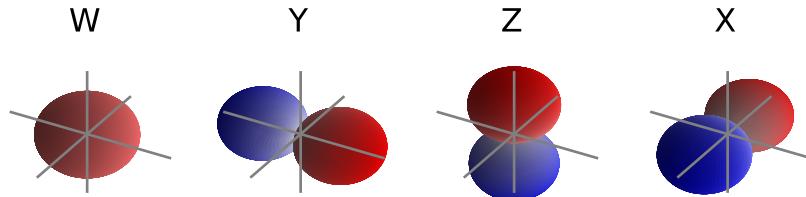


Figure 37.4: Composantes ambisonique du premier ordre. Les surfaces rouges ont une phase positive, les surfaces bleues ont une phase négative.

En pratique, les microphones ambisoniques n'utilisent pas cet arrangement de capsules. En effet, l'encombrement des capsules bidirectionnelles compromet sévèrement la coïncidence du système. La qualité d'un tel microphone dépend largement de sa capacité à positionner les capsules avec la plus grande coïncidence possible, au risque sinon de détériorer la précision de localisation dans le haut du spectre. On utilise alors plutôt quatre capsules cardioïdes, placé sur les surfaces d'un tétraèdre. Cet arrangement est mathématiquement strictement équivalent à l'utilisation d'une capsule omnidirectionnelle et de trois capsules bidirectionnelles. Seulement, l'encombrement moindre des capsules cardioïdes permet une meilleure coïncidence de celles-ci.

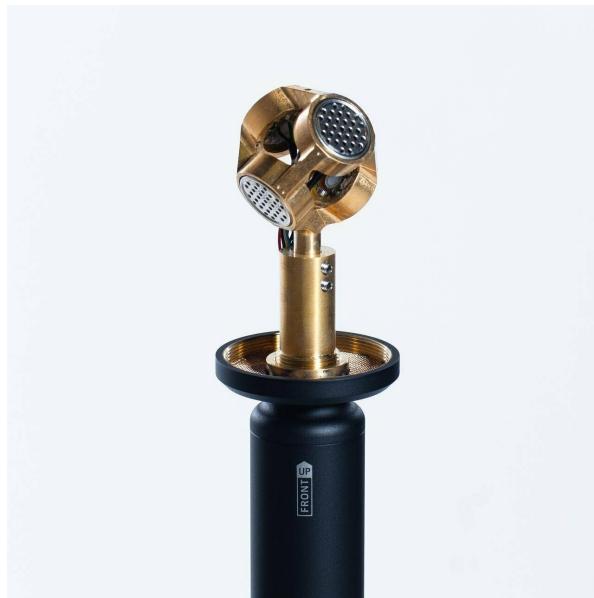


Figure 37.5: Microphone Sennheiser Ambeo

On appelle le flux audio en sortie d'un microphone ambisonique du premier ordre **Format-A**, dans lequel chaque canal correspond à une capsule précise. Par opposition, on appelle généralement **Format-B** un flux ambisonique du premier ordre encodé, où chaque canal correspond

à une des composantes W, X, Y, Z.

Note

L'appellation **Format-A** se retrouve également pour les ordres supérieurs, même si ces microphones sont rares ! Le terme **Format-B** est, quant à lui, discutable pour les ordres supérieurs.

37.1.2 Synthèse du champ sonore

Nous avons vu qu'un microphone ambisonique nous permet de capturer l'espace acoustique entendu d'un point. Nous pouvons également "synthétiser", c'est-à-dire encoder une source sonore monophonique dans un espace acoustique virtuel ambisonique. Pour cela, on utilise simplement des panners ambisoniques.

Pour une incidence donnée de cette source sonore, le panner va affecter une partie de son énergie aux différentes composantes d'un signal ambisonique (W, X, Y, Z) et une phase particulière (positive ou négative). Il est également possible d'encoder des signaux multicanaux (stéréo, quadriphonie, 5.0). Il s'agit d'une technique très efficace pour adapter un mixage réalisé sur un arrangement de haut-parleurs particuliers vers un autre.

Nous ferons état de ces différentes techniques et des différents panners ambisoniques dans le Chapitre 43 .

37.1.3 Restitution du champ sonore

Une fois le signal enregistré, ou encodé en ambisonique, il convient de le décoder sur un arrangement de haut-parleurs particulier. Cette phase de décodage est peut-être la plus complexe à appréhender (et explique très certainement pourquoi l'ambisonique reste à ce jour si peu utilisée).

Illustrons le procédé de décodage en considérant un flux audio ambisonique 2D (W, X, Y) et un système de diffusion quadriphonique.

La figure ci-dessus représente une approche de décodage par **projection** dite **basique**. Pour connaître la quantité de chaque canal à émettre sur chaque enceinte, il suffit de regarder l'axe enceinte-centre, et de lire pour quelle valeur cet axe coupe les harmoniques sphériques. Dans ce cas précis, on trouve :

$$\begin{aligned}L_{front} &= \frac{1}{3}W + X - Y \\R_{front} &= \frac{1}{3}W + X + Y \\L_{back} &= \frac{1}{3}W - X - Y \\R_{back} &= \frac{1}{3}W - X + Y\end{aligned}$$

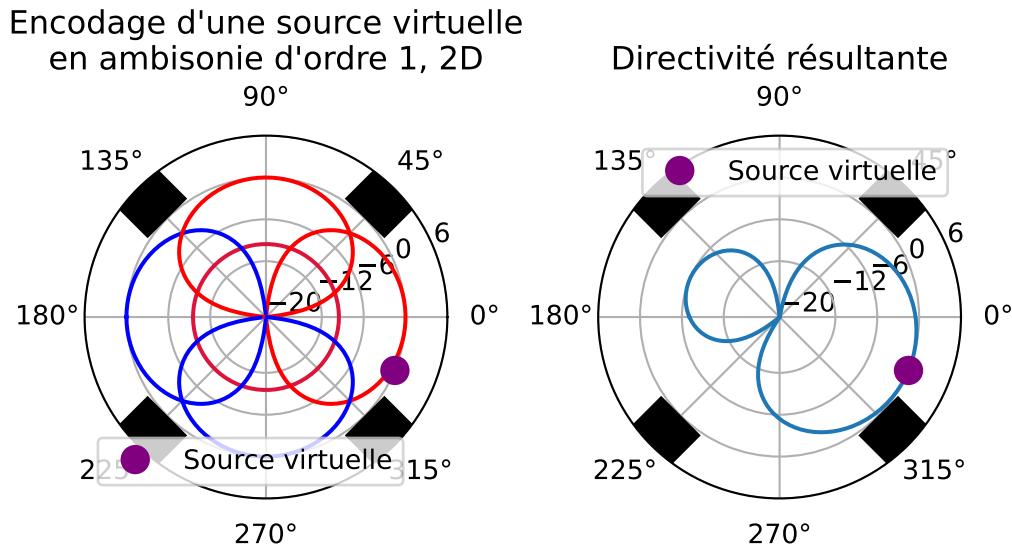


Figure 37.6: Décodage par projection basique.

On peut également étudier la directivité formée pour une source et projeter sur les haut-parleurs pour observer leurs différentes contributions. On remarque dans ce cas (décodage dit basique) la présence d'un lobe arrière. Cela signifie que les enceintes peuvent contribuer à la diffusion d'une source en **opposition de phase** lorsque celles-ci ont certaines positions.

On appelle aussi le décodage par projection “SAD”, pour Sample Ambisonic Decoding. En effet, l’opération consiste à échantillonner la valeur de chaque harmonique sphérique en fonction de l’angle d’incidence de la source.

Il existe d’autres modes de décodage permettant de changer la directivité associée à une source. Ces modes supplémentaires sont “maxRe” et “InPhase”.

Le décodage basique permet un rendu psychoacoustique correct du registre basse fréquence (jusqu’à 500 Hz environ) et offre une précision de localisation maximale au point d’écoute idéal.

Un décodage dit “maxRe” a pour conséquence d’améliorer la localisation dans les hautes fréquences (>700 Hz environ). On constate aussi la réduction des lobes arrière, et donc une contribution moindre des composantes en opposition de phase dans les enceintes.

Lorsqu’un décodeur utilise une optimisation InPhase, la totalité des lobes arrière est supprimée. Cela signifie qu’aucune enceinte ne diffuse de composantes en opposition de phase. Cela a pour conséquence d’améliorer l’homogénéité de la diffusion, d’augmenter la largeur du point d’écoute idéal, au détriment de la précision de localisation.

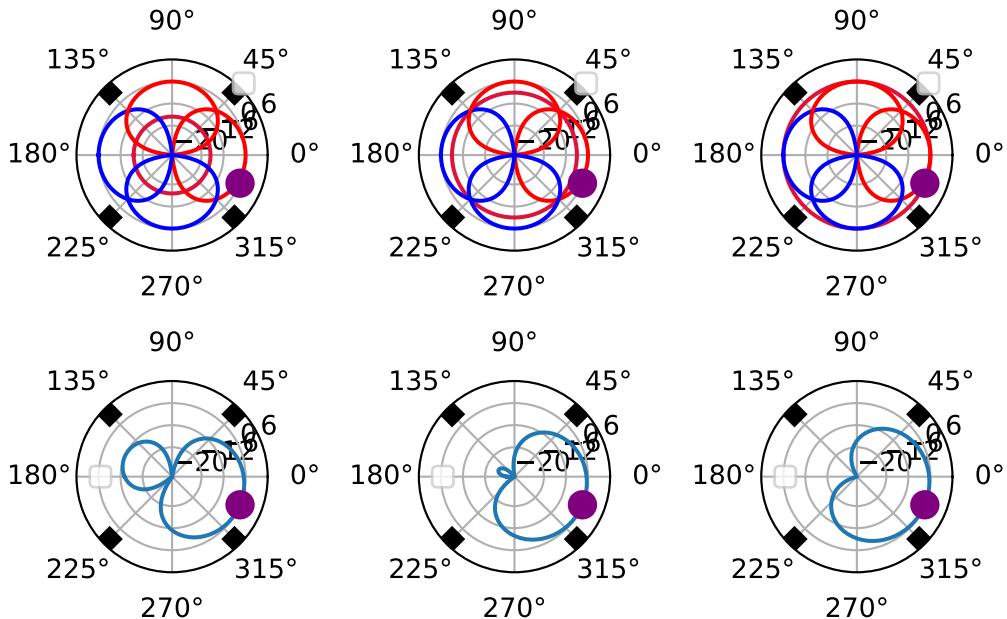


Figure 37.7: Décodage par projection basique, maxRe et inPhase.

i Note

Il n'est pas rare de trouver le mode "maxRe" par défaut dans les décodeurs.

Afin de tenir compte des différences de perception entre les basses fréquences et les hautes fréquences, il est souvent préconisé d'adapter le mode de décodage en fonction de la plage de fréquence. Par exemple, Michael Gerzon propose un décodage "perceptif", en mode basique dans le bas du spectre et maxRe dans les fréquences aiguës. Jérôme Daniel fait également le lien entre zone de couverture, mode de décodage et plage de fréquences.

! Important

Une restitution ambisonique sur un arrangement de haut-parleurs à deux ou trois dimensions n'est pas équivalente. En effet, un décodage 2D offre une meilleure précision de localisation.

37.1.4 Décodage pour arrangement de haut-parleurs irréguliers

Le décodage par projection mentionné précédemment fonctionne bien pour des arrangements de haut-parleurs réguliers dans l'espace. Cependant, la majorité des systèmes utilisés dans

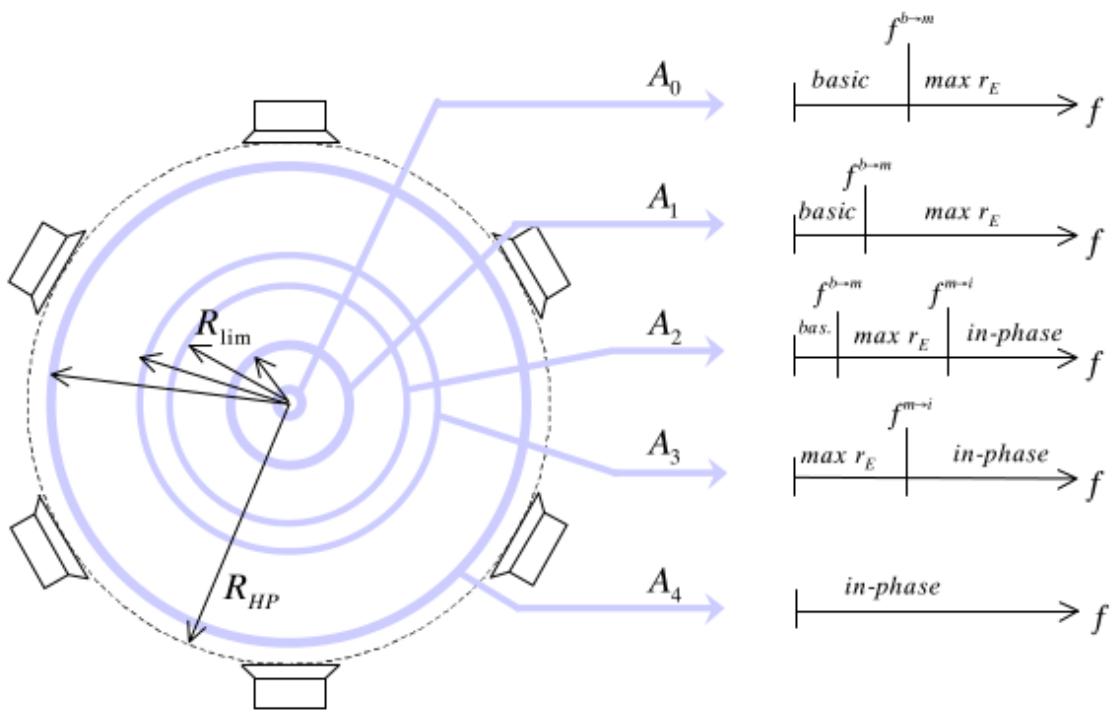


Figure 37.8: Mode de décodage et zone de couverture. J. Daniel (2001).

l'industrie de l'audio ne sont pas réguliers (stéréophonie, 5.1, 7.1.4, etc). Dès lors un décodage par projection ne sera pas homogène en niveau.

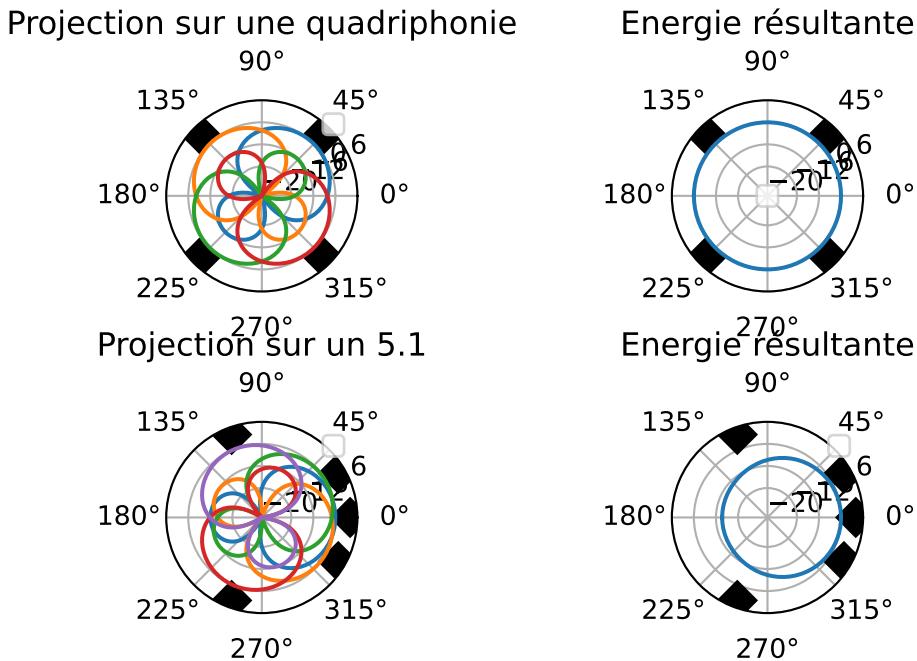


Figure 37.9: Le décodage par projection sur un arrangement 5.0 n'est homogène en énergie.

Il existe alors plusieurs stratégies pour pallier ce problème. La plus utilisée est le décodage **AllRAD** (All-Round Ambisonic Decoder). Celle-ci consiste à décoder le signal ambisonique en deux étapes. Tout d'abord on le décode sur un arrangement de **240 haut-parleurs parfait**, puis on utilise une loi de panoramique VBAP pour réduire ces 240 haut-parleurs à l'arrangement réel.

Une autre stratégie assez connue est l'**EPAD** (Energy Preserving Ambisonic Decoding), qui permet de produire un champ d'énergie invariant en fonction de l'angle d'incidence de la source.

! Important

Peu importe que l'on utilise une stratégie de décodage AllRAD ou EPAD, il convient de **toujours** avoir plus de canaux d'enceintes que de canaux ambisoniques.

37.2 L'ambisonie d'ordre plus élevé (HOA)

L'HOA (pour *Higher Order Ambisonic*) est une extension du système proposé par Michael Gerzon. En effet, l'ambisonie du premier ordre introduit un certain nombre d'effets indésirables lorsqu'on le décode sur un arrangement comprenant un trop grand nombre de haut-parleurs. Il en résulte alors un flou de localisation important, voir l'apparition de problèmes de timbre.

Le FOA reposant sur l'exploitation des harmoniques sphériques aux ordres 0 et 1, il était alors tout indiqué d'utiliser les harmoniques d'ordre supérieur.

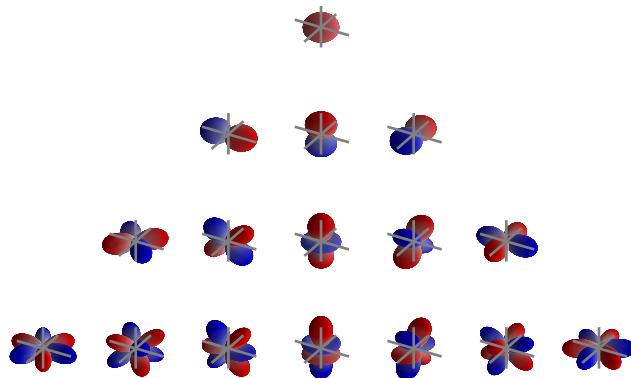


Figure 37.10: Composantes ambisonique jusqu'à l'ordre quatre. Les surfaces rouges ont une phase positive, les surfaces bleues ont une phase négative.

Chaque nouvel ordre ajoute donc un certain nombre de canaux au flux ambisonique. On trouve le nombre de canaux d'un ambisonique d'ordre N avec la formule suivante:

$$N_{ch} = (N + 1)^2$$

On rappelle que lors d'un décodage, on doit toujours avoir plus de canaux d'enceintes que de canaux ambisoniques. Augmenter l'ordre augmente la limite du nombre de haut-parleurs minimum. Cependant, la rétrocompatibilité des ordres inférieurs est garantie par la suppression des canaux correspondant aux ordres supérieurs. Par exemple, un flux ambisonique d'ordre 7 (64 canaux) est trop important pour le décoder sur un système Dolby Atmos 7.1.4 (11 canaux, hors sub). Nous pourrions alors tronquer le signal ambisonique d'ordre 7 pour ne garder que les canaux de 1 à 9. Nous aurions ainsi réalisé une réduction de l'ordre 7 à l'ordre 2, ce dernier étant tout à fait adéquat pour un décodage sur un arrangement de 11 haut-parleurs.

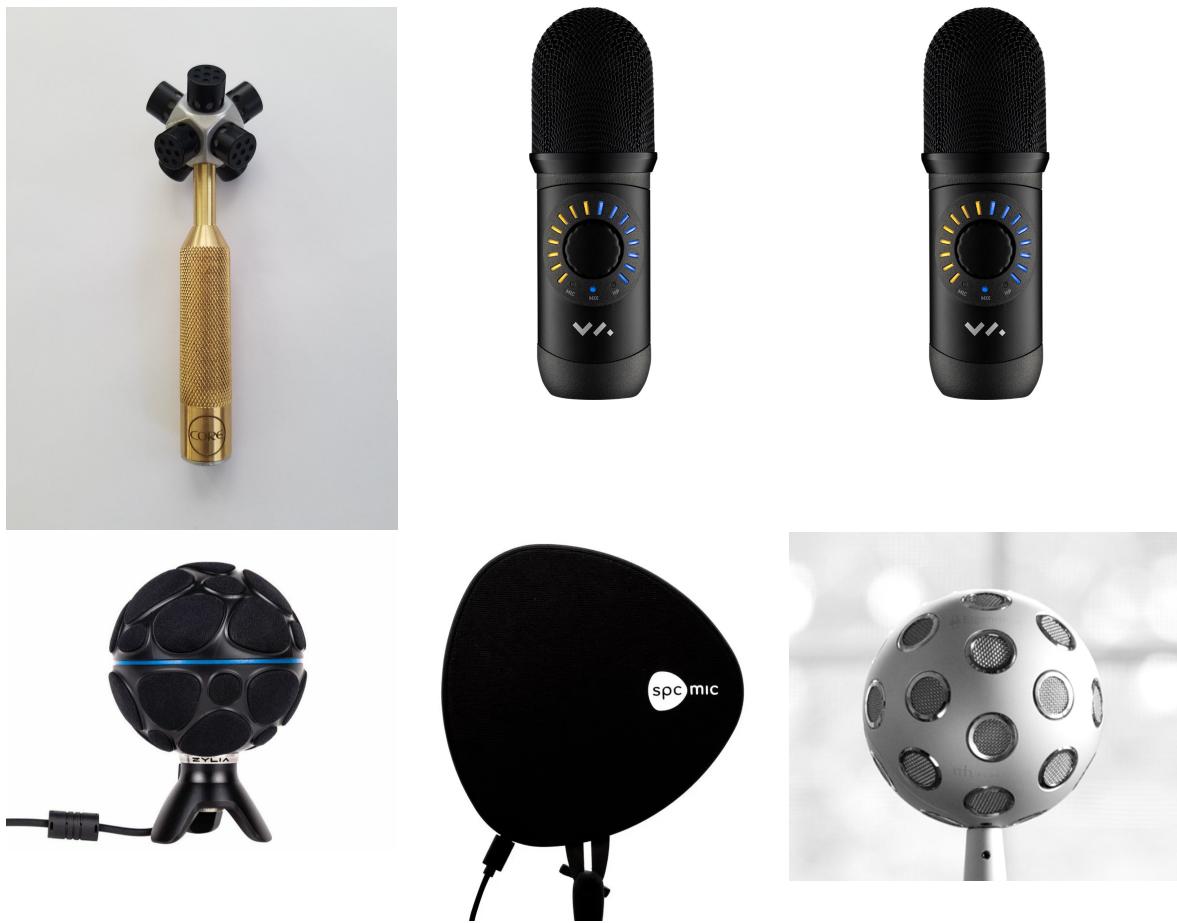
! Important

Ainsi, nous admettrons ces deux règles du décodage ambisonique d'ordre élevé:

- Le nombre de canaux HOA doit toujours être inférieur au nombre de canaux d'enceintes
- Si possible, l'ordre ambisonique doit être le plus élevé possible (tout en respectant la condition précédente)

37.2.1 Les outils de captations d'ordre plus élevé

Il existe un certain nombre de systèmes de prise de son ambisonique d'ordre supérieur. À l'ordre deux, on mentionnera les microphones suivants : [Core Sound OctoMic](#), [Voyage Spatial Mic](#) et l'[Audiotecnica](#). Au troisième ordre, le [Zylia](#), le [Spcmic](#) et l'[Eigenmike](#).



La plupart de ces microphones sont fournis avec des encodeurs spécialisés. En effet, chacun d'entre eux utilise des techniques sensiblement différentes afin de capturer l'espace acoustique. Dès lors, l'encodage du signal des capsules vers de l'ambisonique est bien souvent un procédé propriétaire.

Le nombre de canaux de sortie importants contre indique rapidement l'usage de sorties analogiques. La majorité de ces microphones préfère des sorties USB (fonctionnement d'interface audio) ou audionumériques (Audio sur IP, ADAT).

37.2.2 Les différents formats ambisoniques

Bien que la notion de “format” ambisonique ne soit pas spécifique à l’ambisonie d’ordre élevé, le multiplication du nombre de canaux rend sa compréhension impérative. Deux paramètres sont à prendre en compte : l’ordonnancement des canaux et la normalisation de la composante W par rapport aux autres.

On compte trois principaux ordonnancement ambisoniques, nommés : **Furze-Malham (FuMa)**, **Single Index Designation (SID)** et **Ambisonic Channel Numbering (ACN)**. Le FuMa définit l’ordre des canaux jusqu’au troisième ordre. Le SID est défini par J. Daniel mais reste peu utilisé. L’ACN est aujourd’hui le plus utilisé (format Ambix).

Table 37.1: FuMa

W ₀												0												0											
Y ₂	Z ₃	X ₁				2	3	1				1	2	3				4	5	6	7	8													
V ₈	T ₆	R ₄	S ₅	U ₇		5	7	8	6	4		4	5	6	7	8																			
Q ₁₅	O ₁₃	M ₁₁	K ₉	L ₁₀	N ₁₂	P ₁₄	10	12	14	15	13	11	9	9	10	11	12	13	14	15															

Table 37.2: SID

Table 37.3: ACN

Concernant la normalisation, il existe trois variantes: **maxN**, **N3D** et **SN3D**. La normalisation **maxN** assure que le gain d’une source monophonique pannée n’excède jamais 1 (le panner fonctionne alors seulement en atténuation). La normalisation **N3D** assure que les différentes composantes soient encodées à égale puissance. Enfin, la normalisation **SN3D** est la plus répandue. Elle assure qu’aucune composante ne dépasse les valeurs crêtes du canal W.

! Important

La normalisation n'a aucun effet sur le rendu sonore ambisonique. Il est d'ailleurs possible de passer d'une normalisation à une autre. Il convient seulement d'être vigilant lors de la manipulation de flux ambisonique, et de bien s'assurer de leurs formats d'acquisition, avant de correctement paramétrier ses outils.

Aujourd’hui, le format le plus commun d’ambisonique est l’**Ambix**, utilisant un triage ACN et une normalisation SN3D. Son prédécesseur était ce qu'on nomme couramment le B-Format,

ou parfois FuMa, utilisant un triage FuMa et une normalisation maxN (avec une atténuation de -3 dB).

37.3 Avantages de l'ambisonie

Malgré sa complexité théorique certaine, l'ambisonique offre un bon nombre d'avantages par rapport aux autres techniques de mixage. Avant tout, un mixage ambisonique ne dépend pas du système de restitution. Il peut être décodé après mixage sur, virtuellement n'importe quel arrangement d'enceinte (bien que nous ayons vu certaines contraintes à respecter pour satisfaire les critères de qualité sonore). Travailler à un ordre élevé permet également un décodage vers du binaural cohérent. Il s'agit d'ailleurs d'une approche de la synthèse binaurale courante (Meta 360 et autres casquent de réalité virtuelle).

Le traitement d'un signal ambisonique (égalisation, compression) s'assimile au traitement de multiples canaux. Les outils sont relativement rares, mais cela reste possible. Les traitements impliquant des systèmes linéaire et invariant temporellement (comme les égaliseurs) s'appliquent sans précautions particulières. Attention néanmoins aux traitements dynamiques et non linéaires, ceux-ci peuvent briser la cohérence de la scène sonore. Il est alors recommandé d'utiliser le canal W comme canal de détection afin d'appliquer le traitement de façon indifférenciée sur les autres composantes. On trouve également des traitements spécifiques à l'ambisonique, comme des compresseurs directionnels, permettant de maîtriser l'énergie d'un mixage provenant d'une certaine direction.

L'ambisonie est souvent apprécié pour sa malléabilité et son interactivité. En effet, l'encodage par harmonique sphérique permet d'appliquer des effets en temps réel, même après mixage (et donc sommation). Il est possible de tourner la scène sonore, de l'inverser, de zoomer dedans, etc. Cette propriété de l'ambisonique explique son utilisation courante dans les médias de l'interaction, comme la VR et le jeu-vidéo.

38 Mixage orienté objet

Par définition, un mixage son consiste à mélanger un certain nombre de canaux d'entrées vers un nombre de canaux de sorties plus faible. Le mixeur a pour rôle de réaliser et de maîtriser cette sommation.

Lors d'un mixage orienté objet, la méthode de travail est alors assez différente. Pour chaque canal d'entrée, nous allons y associer des **métadonnées** de mixage, par exemple son gain ou sa position dans l'espace.

Une fois cette association faite entre audio et métadonnées de mixage, on utilise un décodeur qui va, par lecture des métadonnées, adapter ces informations en fonction du système de haut-parleurs qui lui est affecté. Le mixage, c'est-à-dire la sommation des canaux, est alors réellement fait à l'étape de décodage. Quelque part, le mixeur réalise ici un "non-mixage".

L'avantage principal du mixage orienté objet est donc de pouvoir produire différents formats d'écoutes à partir du même mixage (en théorie). On pourrait alors à partir des mêmes métadonnées produire un rendu pour un système stéréophonique, 5.1 et 9.1.6 (ou pourquoi pas, en ambisonique 22e ordre !).

Même si le mixage orienté objet n'implique pas un mixage spatialisé, cette technique de production est souvent employée dans ce contexte pour tenter de simplifier les problématiques de format multiple. Par exemple, on pourra travailler sur un mixage immersif, et pouvoir du même coup obtenir un mixage stéréophonique. On réalise alors un gain de temps évident par rapport à la méthode traditionnelle qui consiste à réaliser "n" mixages si l'on doit produire "n" formats de sorties.

Avertissement

Cette facilité de création de rendus différents n'épargne cependant pas les comportements parfois surprenants d'un mixage d'un système à un autre. Par exemple, un mixage réalisé sur une écoute quadriphonique risque d'être surprenant lors de son passage sur un arrangement 5.1. Il est généralement sage de travailler sur le système le plus grand que l'on souhaite cibler et de contrôler les "downmix". L'inverse peut amener à certaines surprises.

Plusieurs fabricants proposent des formats et des outils de mixage orienté objet. En tête, on retrouve Dolby et son Dolby Atmos, possédant au jour d'aujourd'hui la plus grande part de marché (intégration native dans Pro Tools et Logic Pro). DTS et Auro 3D proposent également leurs propres outils. Enfin, Freinshafer a défini le codec MPEG-H, implémentant

l'Audio Definition Model comme base de métadonnées. L'ADM étant un format ouvert nous servira donc à décortiquer le fonctionnement de l'encodage et le décodage d'un mixage orienté objet.

38.1 L'Audio Definition Model

38.1.1 Présentation

L'[Audio Definition Model](#) est une autre recommandation de l'ITU, dont la dernière édition date de 2019. Ce papier concentre beaucoup d'informations importantes. Principalement, il s'agit de décrire une base commune, espérant devenir une base commune entre les différents acteurs de l'audio orienté objet. Le premier objectif est donc d'assurer l'interopérabilité entre les différentes solutions de mixage orienté objet.

Techniquement, l'ADM est donc un catalogue de descripteurs d'audio. Pour chaque canal audio, un certain nombre d'informations textuelles y sont associées. Elles sont donc **lisibles** par un être humain, et **éditables**.

38.1.2 Description des métadonnées

L'ADM permet la représentation de quatre “types” d'audio :

- Le *channel-based* ou orienté-canal
- L'*HOA* ou Ambisonie d'ordre plus élevé
- Les *Matrix* ou formats matricés
- Les objets

Avertissement

Le mot “objet” a deux sens dans la documentation de l'ADM. Un objet est un format, permettant le rendu dynamique d'un élément sonore (changement de position, de volume, etc.). Cependant, un objet est également un contenu, qui permet de référencer un bloc d'échantillons d'un canal audio.

En ce sens un objet audio en ADM peut représenter le canal d'un flux orienté objet, le canal d'un flux HOA, un élément d'une matrice ou un objet sonore !

L'ADM permet de définir plusieurs programmes, appelés “*audioProgramme*”. Prenons comme exemple le cas d'une série télévisée diffusée en France et en Grande-Bretagne. Nous aurions, dans ce cas, un premier *audioProgramme* contenant les voix françaises et le reste de la bande-son, puis ,un second contenant les voix anglaises et le reste de la bande-son. Nous aurions alors trois **contenus audio** (appelés *audioContent*) : les voix anglaises, les voix françaises et

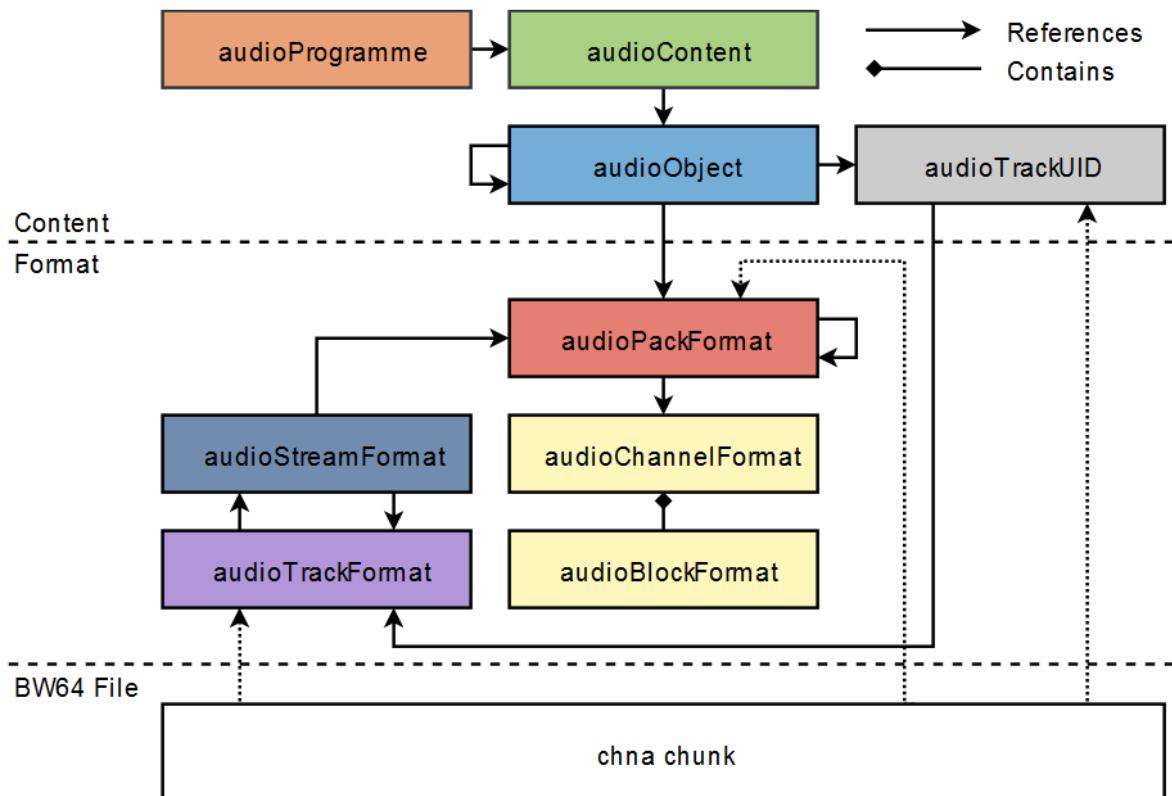


Figure 38.1: Organigramme de l'ADM

le reste de la bande-son (bruitages, montage son, musique, etc.). L'*audioProgramme* “version anglaise” contiendrait le *audioContent* voix anglaise et le reste de la bande son, celui “version française” contiendrait les *audioContent* voix française et le reste de la bande son. Ainsi, le spectateur peut choisir d’écouter l'*audioProgramme* qu’il préfère, de façon interactive.

Chacun de ces *audioContent* contiennent eux-même des objets, appelés *audioObject*. Dans notre exemple, nous aurions différents objets dans l'*audioContent* voix française correspondant aux différents comédiennes et comédiens. Ces *audioObject* permettent d’indiquer quels échantillons audio du fichier BW64 sont concernés. Il convient alors d’indiquer quel est le format audio de ces échantillons. Les *audioObject* pointent vers un *audioPack* qui lui même pointe vers un certain nombre de sous modules permettant de caractériser intégralement le format audio et **où il doit être positionné** dans l'espace.

Si l'*audioObject* est de type “*directSpeaker*”, on l'affecte à un haut-parleur, dont on renseigne une position fixe. Si ce haut-parleur existe dans l'arrangement de haut-parleurs connecté au décodeur ADM, le signal y est routé, sinon on interpole avec les haut-parleurs les plus proches.

Si l'*audioObject* est de type ambisonique, on cherche les autres objets appartenant au même flux, puis on décode l'ensemble, sur le système connecté au décodeur ADM.

Si l'*audioObject* est la composante d'une matrice, on cherche les autres objets appartenant au même flux, puis on dématrice l'ensemble vers nouvel *audioPack* de type “*directSpeaker*”.

Si l'*audioObject* est de type “*object*”, on regarde ses **métadonnées de spatialisation** puis on interpole cette position en utilisant les haut-parleurs les plus proches du point indiqué.

Astuce

Gardons en tête que les *audioObject* “*directSpeaker*” ou ambisonique sont des **mixages**. Les scènes sonores qu'ils représentent sont statiques et immuables. Un *audioObject* de format objet contient un élément sonore (une voix, un instrument, un bruitage), que le décodeur ADM va ensuite mixer. Un objet peut donc être interactif. L'utilisateur pourrait changer son volume, sa position, etc.

38.1.3 Les métadonnées de spatialisation

L'ADM n'intègre qu'un jeu de paramètre très sommaire :

- Le niveau
- La position (en coordonnées cartésiennes et sphériques)
- La taille et diffusion de la source

Le niveau est un simple réglage de gain. La position permet de placer l'objet dans un espace virtuel en trois dimensions et selon deux systèmes de coordonnées. On préfère généralement le système sphérique, plus proche de notre perception sonore, cependant, la description de

certains mouvements est plus simple dans un repère plutôt qu'un autre. Le paramètre de distance permet de placer une source **entre les haut-parleurs et le point d'écoute idéal**.

La largeur permet de jouer sur l'étalement de l'objet dans l'espace. En pratique, accroître ce paramètre augmente le nombre de haut-parleurs contribuant à la diffusion de l'objet sonore. Associé à la taille, le paramètre de diffusion permet d'ajouter de la décorrélation sur les haut-parleurs contribuant à la largeur de restitution de la source.

Il n'y a donc pas de descripteurs perceptifs ou d'espaces. L'approche de la spatialisation y est donc très basique, voire simpliste. Comme nous le verrons dans les sections plus pratiques, cela impose des techniques de mixage relativement complexes à fin de contourner ses limitations.

i Note

Malheureusement la plupart des moteurs de mixage orienté objet souffrent de ces mêmes lacunes, y compris le Dolby Atmos. L'exception principale est l'IRCAM Spat ainsi que le Spat Revolution de FLUX:: Immersive.

38.1.4 Inscription des métadonnées dans un fichier audio

L'ADM décrit la possibilité d'inscrire les métadonnées de mixage directement dans un fichier audio, contenant lui-même l'ensemble des données audio nécessaires. Dans le cadre d'un mixage orienté objet, ce fichier audio peut posséder autant de canaux que d'objets déclarés. Le format retenu pour les stockages de ces informations est le **BW64** (Broadcast Wavefile 64 bit).

i Note

Le **BW64** est une extension du format **BWF**. Tous deux sont également définis dans des recommandations de l'ITU. Ces fichiers prennent la forme de simples fichiers "wave" (monfichier.wav).

Également, on parle d'un en-tête de fichier, pour qualifier l'ensemble des métadonnées (nom, type de fichier, conteneur, etc.) précèdent la donnée utile (ici, les échantillons audio). Attention, les métadonnées ne sont pas toutes relatives à l'ADM. La plupart les fichiers possèdent des métadonnées pour indiquer aux programmes comment exploiter les données qu'ils stockent.

La figure Figure 38.3 fait mention de plusieurs "chunk" (section, bloc) nommé "axml", "bxml", "sxml", "chna". La globalité des informations est stockée sous la forme d'un **XML** dans les *chunks* "axml", "bxml" ou "sxml". Le *chunk* "chna" stocke les informations permettant de faire les liens entre les données audio et leur description dans l'ADM. Il est en général recommandé d'utiliser le *chunk* "axml" (le "bxml" est alors plutôt utilisé pour les surcouches à l'ADM, aussi appelés profiles).

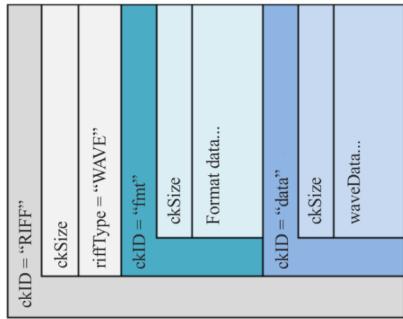


Figure 38.2: Structure d'un fichier wave

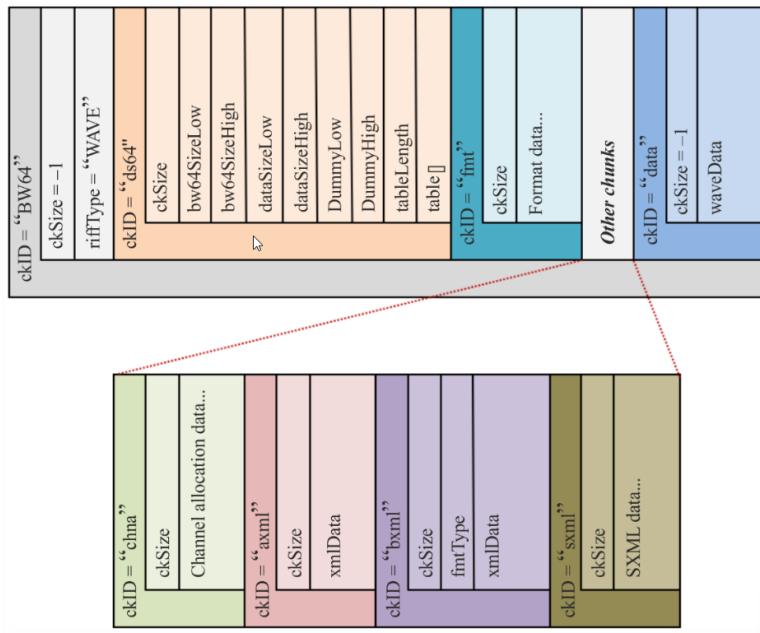


Figure 38.3: Structure d'un fichier bw64

Illustrations extraites du document BS.2088-2 de l'ITU

BS.2088-02

38.2 Les profils ADM

Certains constructeurs ont choisis de proposer des sur-spécification à l'ADM, aussi appelés “profils”. Ces profils ont généralement deux buts :

- Rajouter des métadonnées spécifiques à l'utilisation de codec audio propriétaire.
- Limite les spécifications de l'ADM pour correspondre aux usages de certains domaines

Les deux principaux profils connus sont ceux du [MPEG-H de Fraunhofer](#) et de [Dolby](#)

! Important

Aucun des deux constructeurs n'ont choisi de maintenir le support de l'ambisonique dans leur format.

De plus, il n'existe à ce jour, aucun outil pour convertir un ADM d'un profil quelconque vers un ADM Dolby. La compatibilité des formats n'est donc pas assurée.

L'ADM Sériel (S-ADM)

Pour certaines applications, comme le live ou le streaming, la description des métadonnées sous forme d'un dictionnaire XML n'est pas adéquate. Ici, les flux audio sont en temps réel et ne

dépendent pas nécessairement d'un fichier audio. Il existe donc une norme sérielle de l'ADM, permettant de transmettre les métadonnées dans une connexion numérique, type AES, MADI ou encore à travers de l'audio sur IP.

La description du S-ADM est disponible dans l'[ITU-R BS.2125](#).

38.3 L'ADM OSC

L'ADM OSC est une initiative menée par FLUX:: Immersive, L-Acoustic et Radio France pour proposer une base de communication commune entre les différents moteurs de mixage orienté objet dans le monde du Live. Cela assure donc une interopérabilité minimale entre les principaux acteurs du milieu. À ce jour, les acteurs suivants se sont joints à la conversation : d&b Audiotechnik, DiGiCo, Dolby, Lawo, Magix, Merging Technologies, Meyer Sound, Steinberg.

Les logiciels implémentant l'ADM OSC comptent SPAT Revolution (FLUX::SE), L-ISA Controller (L-Acoustics), Ovation (Merging Technologies), Nuendo (Steinberg), SpaceMap Go (Meyer Sound), QLAB 5 (Figure 53), Space Controller (Sound Particles), Modulo Kinetic (Modulo Pi), Iosono (Barco).

Le code source de ce projet, développé sous une licence open source, est disponible dans ce répository [github](#)

38.4 L'étape de rendue

Tout ce que nous avons vu pour l'instant, au sujet de l'ADM, ne concerne au final que la description d'un mixage, et nous n'avons pour l'instant pas envisagé le "comment". En d'autres termes, nous avons une liste d'ingrédients, mais ne savons pas comment les mélanger et les cuisiner entre eux.

Il existe une [recommandation de l'ITU \(BS-2127\)](#) consacrée au sujet du "renderer", ou processeur de rendu. On fournit donc audio et métadonnées à ce moteur de rendu, qui, grâce à l'analyse de ces métadonnées et au système de haut-parleurs qui lui est connecté, peut calculer le gain de reproduction de chaque source et chaque haut-parleur.

38.4.1 Description de la loi de panoramique

La loi de panoramique recommandée par l'ITU BS-2127 empreinte beaucoup au **LBAP** (voir la section [36.1.3](#)). Nous sommes donc sur une loi de panoramique favorisant un point d'écoute idéal, situé à égale distance des haut-parleurs.

i Note

Ici, la distance ne s'exprime qu'entre les haut-parleurs et le point d'écoute idéal. En d'autres termes, la distance maximale correspond au plan des haut-parleurs, la distance minimale correspond au point d'écoute idéale. La distance ne contribue pas au gain de la source, mais plutôt à une sorte de spreading.

L'algorithme se décrit de la façon suivante :

- On recherche les couches (en élévation) de haut-parleurs au-dessus et en dessous de la source. On calcule alors le gain z en interpolant entre les deux couches.
- Sur chaque couche trouvée, on recherche les deux lignes de haut-parleurs en avant et en arrière de la source. On calcule ainsi le gain y en interpolant entre les deux lignes.
- Enfin, sur chaque ligne trouvée, on recherche la paire de haut-parleurs à gauche et à droit de la source, permettant finalement de calculer le gain x entre les deux haut-parleurs.

i Note

Si une source est positionnée sur un haut-parleur, alors seul celui-ci contribue à la diffusion sonore.

38.4.2 Largeur et décorrélation

Le paramètre de “size” ou taille permet de faire paraître une source plus grande en augmentant le nombre de haut-parleurs contribuant à la diffusion de la source. Ce paramètre “size” est en fait un macro-paramètre qui manipule trois autres paramètres:

- La largeur (étalement horizontal)
- La hauteur (étalement vertical)
- La profondeur (étalement dans la profondeur)

Par principe de source fantôme, augmenter le nombre de haut-parleurs contribuant à la diffusion d'une source n'augmente pas véritablement sa taille apparente dans notre perception de la scène sonore. La source vient plutôt se positionner au [barycentre](#) des haut-parleurs, pondéré par leur énergie de diffusion. Pour réellement créer un effet de “taille”, il est nécessaire d'introduire une différence suffisamment importante pour que notre oreille ait l'illusion d'un étalement dans l'espace. Cette différence doit également être suffisamment faible pour ne pas déformer le signal d'origine. On appelle cela la décorrélation. L'ITU recommande ici l'utilisation de filtres passe-tout sur chaque canal de diffusions.

38.4.3 Décodage ambisonique

L'ITU décrit l'utilisation d'un décodage AllRAD. Voir la [section -@HOA-decode-allrad]

39 La synthèse de fronts d'ondes (WFS)

partie VII

Les outils de spatialisation sonore

40 Choisir une station de travail

41 Panner binaural

42 Panner orienté canaux

43 Panner ambisonique

44 Gestion des effets

45 Le mixage orienté objet

45.1 MPEG-H et ADM

45.2 Dolby Atmos

45.3 IRCAM Spat

45.4 FLUX:: Spat Revolution