

Bergame Périaux • Jean-Luc Ohl • Patrick Thévenot

ina

Le son multicanal

**De la production à la diffusion
du son 5.1, 3D et binaural**

DUNOD

Toutes les marques citées dans cet ouvrage
sont des marques déposées par leurs propriétaires respectifs.

Photo de couverture : © B. Périaux

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements



d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).

© Dunod, 2015

5 rue Laromiguière, 75005 Paris
www.dunod.com

ISBN 978-2-10-074083-3

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Préface

Damn, c'est injuste ! Bergame Périaux, avec la collaboration de Jean-Luc Ohl et Patrick Thévenot, a écrit le livre que je voulais écrire. Après toutes ces années de rencontres et de lectures de bons et de moins bons livres sur le multicanal, j'avais décidé que le seul livre que j'écrirai porterait sur ce sujet. Il traiterait toutes les formes du son multicanal : ses débuts (il y a quelques centaines d'années), ses années de gloire au cinéma, le passage au monde de la musique et du broadcast, le streaming et les téléconférences. La technologie pour l'enregistrement, le mixage et la diffusion d'un programme et plus important encore, l'esthétique et pourquoi nous faisons cela.

Bergame a tout couvert. J'ai cherché une application non traitée, mais je n'ai pas trouvé. Ce livre n'est aucunement une encyclopédie des deux cent premiers résultats sur Google pour la recherche de « son surround » ou « audio multicanal ». La structure de l'ouvrage est très bien pensée, cela ressemble à un dîner français traditionnel avec un amuse-bouche et un apéritif pour ouvrir l'appétit, la salade et le plat principal pour vous emporter, et bien sûr le dessert, sucré et délicieux. Sans mentionner le vin fantastique qui accompagne les plats. Ou bien, on pourrait dire que cela ressemble à une forme de sonate avec son introduction, l'exposition du thème principal (thème secondaire, les transitions de modulation), son développement, la reprise et la coda. Ou plus simplement : c'est une bonne histoire.

Soyez informé, c'est un bagage lourd dans lequel vous devrez trier les informations et vous découvrirez de vieilles connaissances et de nouvelles notions, des ingrédients essentiels et des épices exotiques. L'essentiel est là, les détails aussi, beaucoup de détails, ainsi que les derniers éclairages sur la complexité de ce segment de l'ingénierie audio. Tout ceci est mené avec art et connaissance jumelés avec un grand contenu qui touche l'âme et le cœur.

Damn, c'est un grand livre. Mais bon, maintenant qu'il a été écrit par Bergame, je vais certainement avoir plus de temps libre ! Pour le relire, de bout en bout.

Florian Camerer

Ingénieur du son senior à l'ORF, président de EBU-groupe PLOUD

Table des matières

À PROPOS DES AUTEURS	XVII
REMERCIEMENTS	XIX
SIGLES, TERMES TECHNIQUES, CONVENTIONS	XXI
INTRODUCTION	XXV
CHAPITRE 1 – GÉNÉRALITÉS	1
1.1 Écouter	2
1.2 Historique	2
1.2.1 Cinéma	2
1.2.2 Home-cinéma, DVD, TVHD, radio, jeu vidéo, Internet	5
1.2.3 Écriture, édition et diffusion musicales	9
1.3 Intérêts	10
1.3.1 Une écoute naturelle : réalisme et immersion	10
1.3.2 Un espace : de nouvelles possibilités d'écriture	12
1.3.3 Définition d'image, angle de vision, distance et dispositif sonore	12
1.3.4 Meilleure intelligibilité, extension de la bande passante, interaction avec le corps	13
1.4 Vecteurs	13
1.4.1 Le cinéma : passage au numérique, installations son 3D Barco Auro-3D et Dolby Atmos	13
1.4.2 La TVHD, l'UHDTV et le Super Hi-Vision	14
1.4.3 La radio numérique	14
1.4.4 Les supports Blu-ray, DVD, SACD	14
1.4.5 Le jeu vidéo	15
1.4.6 La salle de spectacle, le théâtre, la retransmission d'opéras au cinéma	15
1.4.7 Internet	15
1.4.8 Le téléphone mobile, la tablette numérique	16
1.5 Problématiques	16
1.5.1 Le système d'écoute	16

1.5.2	L'image sonore multicanal	16
1.5.3	La prise de son spatialisée	17
1.5.4	Des concepts de mixage nouveaux	17
1.5.5	La diffusion, les metadata, le mixage Dolby	17
1.6	Les différentes techniques de spatialisation	17
1.6.1	Stéréophonie ou Multichannel Stereo	17
1.6.2	L'Ambisonie (« scene oriented »)	18
1.6.3	La WFS	18
1.6.4	Le binaural	19
1.6.5	Les objets (« object oriented »)	19
1.7	Formats multicanal	20
1.7.1	Quadriphonie	25
1.7.2	5.1 ITU	26
1.7.3	LCRS Dolby Surround (TV, home-cinéma)	26
1.7.4	5.1 ITU 4 enceintes arrière	27
1.7.5	5.1 amélioré Günther Theile	27
1.7.6	5.1 cinéma	28
1.7.7	5.1 cinéma Dolby Stereo	29
1.7.8	6.1 cinéma	29
1.7.9	6.1, 7.1 IMAX	30
1.7.10	7.1 cinéma SDDS	30
1.7.11	7.1 cinéma image 3D : DTS, Dolby Surround 7.1	31
1.7.12	7.1 Blu-ray Dolby	31
1.7.13	7.1 Blu-ray DTS	32
1.7.14	7.1, 9.1, 11.1 selon recommandation de Mike Williams	34
1.7.15	8.1 Musique acousmatique octophonique	35
1.7.16	7.1, Dolby Surround IIX, 9.1 Dolby Surround IIZ	35
1.7.17	Auro-3D 9.1 ou 10.1 home-cinéma/studio	36
1.7.18	Auro-3D 11.1, 12.1, 13.1 cinéma	37
1.7.19	10.2 Tomlinson Holman	38
1.7.20	11.1 DTS Neo X	39
1.7.21	22.2 NHK (9/10/3.2)	39
1.7.22	Imm Sound 23.1 ou 14.1 (Dolby)	40
1.7.23	Norme ITU-R BS 2051	41
1.7.24	Dolby Atmos	42
1.7.25	WFS	43
1.8	État des lieux des marchés du multicanal	43
1.8.1	Statistiques	43
1.8.2	Premier bilan de la TVHD	44
1.8.3	Perspectives sur la production radiophonique	45
1.8.4	Événements, séminaires, groupes de recherche, formations	46

————— CHAPITRE 2 – LE SYSTÈME D'ÉCOUTE —————		47
2.1	Caractérisation auditive d'une écoute	48
2.1.1	Pourquoi apprendre à écouter ?	48
2.1.2	Critères objectifs nécessaires à la caractérisation d'une écoute	48
2.1.3	Apprentissage d'une méthodologie d'écoute	49
2.2	Finalité d'une écoute monitoring en multicanal	50
2.2.1	Domaine artistique, domaine objectif	50
2.2.2	Problématique spécifique au multicanal	51
2.3	Les différents types de mesures	51
2.3.1	Fréquence glissante	51
2.3.2	Bruit rose	51
2.3.3	MLS, CHIRP, TDS	51
2.3.4	Retard dynamique	52
2.4	L'enceinte acoustique	52
2.4.1	Différents types	52
2.4.2	Mesures et interprétations auditives	53
2.4.3	Conditions de fidélité et choix pour le multicanal	58
2.5	L'influence du local	61
2.5.1	Réponse dans le grave, couplage physique	61
2.5.2	Influence des premières réflexions, couplage psycho-acoustique	63
2.5.3	Influence du champ réverbéré tardif, équilibre tonal et transmissibilité des mixages	66
2.5.4	Le local et l'enceinte, synthèse globale	67
2.6	Les différents types de cabines de mixage	68
2.6.1	Type « Tom Hidley »	68
2.6.2	LEDE	69
2.6.3	Type « à amortissement réparti »	70
2.6.4	Une nouvelle approche : la cabine à « diffusion répartie »	72
2.6.5	Avantages et inconvénients en multicanal	74
2.7	L'influence de l'amplificateur	75
2.7.1	Les différentes technologies	75
2.7.2	Limites des mesures classiques, interactions avec les hauts parleurs et incidence sur l'écoute en 5.1	75
2.8	Normalisation d'une écoute multicanal	76
2.8.1	Normes ITU-R BS.775.1 et 2	76
2.8.2	Normes Cinéma (ISO 2969/1987(E)/ SMPTE ST202-2010)	80
2.8.3	L'écoute au casque multicanal	84
2.9	L'écoute multicanal et l'image	87
2.9.1	Influence sur notre perception	87
2.9.2	Rapports dimensionnels entre les imageries sonores et visuelles	87

2.10	L'écoute domestique en multicanal	88
2.10.1	De la HI-FI au Home-cinéma	88
2.11	Calibrage d'une écoute et optimisation par corrections électroniques	91
2.11.1	Correction ou compensation ?	91
2.11.2	Les différents moyens électroniques, filtres analogiques, filtres numériques IIR, filtres numériques FIR, les optimiseurs de champ acoustique	92
2.11.3	Prise en compte des circuits de production et des cultures d'écoute	97
2.11.4	Diagnostic et test d'un système d'écoute 5.1 par méthode entièrement auditive à l'intention des ingénieurs du son	97
2.12	Perspectives et évolutions	98
2.12.1	L'augmentation du nombre de canaux (7.1 à 22.2 ou 23.1)	98
2.12.2	L'après multicanal ? Le WFS et le Transaural et la 3D	103
<hr style="width: 100%; border: 0.5px solid black; margin: 10px 0;"/> CHAPITRE 3 – LES FACTEURS PERCEPTIFS <hr style="width: 100%; border: 0.5px solid black; margin: 10px 0;"/>		109
3.1	Le standard stéréo	110
3.2	Apport du 5.1	112
3.3	Le dispositif 5.1 et ses contraintes	113
3.4	Localisation	115
3.4.1	Principe de fonctionnement	115
3.4.2	Fonctions de transfert HRTF	118
3.4.3	Précision de localisation	124
3.4.4	Angle minimum audible	126
3.4.5	Perception de la distance	127
3.4.6	Perception des sources latérales et zone d'écoute	129
3.4.7	Effet Haas et multiplication des sources	130
3.4.8	Influence de la vision sur la localisation	131
3.4.9	Internalisation et externalisation	131
3.4.10	Synthèse binaurale	132
3.5	Impression spatiale	135
3.5.1	Définition	135
3.5.2	Enveloppement	138
3.6	Attention auditive	139
3.6.1	Principe général	139
3.6.2	Les causes de la réaction de l'orientation réflexe de l'attention	140
3.6.3	La mémoire visuelle prépare l'identification auditive	141
3.6.4	Comment gérer l'orientation attentionnelle du spectateur ?	141
3.6.5	Quelques règles	142
3.7	Ébauche de solutions pour l'image sonore multicanal	143
3.7.1	Esthétique	143
3.7.2	Localisation	143

3.7.3	Impression spatiale	144
3.7.4	Attention auditive	144
<hr/> CHAPITRE 4 – L'IMAGE SONORE, LES ESTHÉTIQUES, ET LA RÉALISATION <hr/>		145
4.1	Espace de création et espace de reproduction	146
4.2	Présentation des critères objectifs d'analyse	146
4.2.1	Esthétique de l'image, réalisme ou effet	147
4.2.2	Construction de l'image	148
4.2.3	Couleur de l'image	153
4.2.4	Qualité du signal	154
4.3	Application : analyse d'images sonores par les critères	155
4.4	Genres, esthétiques et réalisation	157
4.4.1	Genres, esthétiques et éléments de spatialisation	157
4.4.2	Film, fiction, animation, documentaire, sport	157
4.4.3	Musique	158
4.4.4	L'Acousmatique et l'Acousmonium par Daniel Teruggi	162
4.4.5	Radio	164
4.4.6	Jeu vidéo par Thierry Dilger	165
4.4.7	Éléments de réalisation	171
<hr/> CHAPITRE 5 – LA PRISE DE SON <hr/>		175
5.1	Rappels sur la prise de son stéréophonique	176
5.1.1	Différence d'intensité	176
5.1.2	Différence de temps	176
5.1.3	Combinaison intensité et temps	177
5.1.4	Notion d'angle de prise de son	180
5.1.5	Directivités	180
5.2.	Principe de la prise de son multicanal	181
5.3	Les systèmes principaux espacés	183
5.3.1	MMAD	183
5.3.2	OCT Surround	197
5.3.3	Arbre Decca	201
5.3.4	Arbre Fukada	204
5.3.5	INA5	204
5.4	Les systèmes principaux matricés	207
5.4.1	Double MS	207
5.4.2	Soundfield, système ambisonique d'ordre 1	211
5.4.3	Sphère Schoeps KFM 360	217

5.5	Systèmes d'ambiance	219
5.5.1	Double ORTF	219
5.5.2	ORTF Surround Schoeps	220
5.5.3	Croix IRT	222
5.5.4	Carré Hamasaki	224
5.5.5	Omni square et variantes	226
5.5.6	Holophone H2 Pro	226
5.5.7	DPA 5100	227
5.6	La haute résolution spatiale	228
5.6.1	Haute résolution spatiale et Trinnov SRP	230
5.6.2	High Order Ambisonics	235
5.7	Prise de son binaurale	237
5.8	Prise de son en tournage	240
5.8.1	Les enregistreurs	240
5.8.2	Les différents systèmes : pièges et astuces, avec la participation de Vincent Magnier	240
5.9	Applications	241
5.9.1	Film	241
5.9.2	Documentaire à l'image	242
5.9.3	Documentaire, reportage, fiction radiophonique par Guy Senaux, ingénieur du son Radio France	243
5.9.4	Autres expériences Radio	248
5.9.5	Sport	249
5.9.6	Musique	249
<hr/>		
	CHAPITRE 6 – LA POSTPRODUCTION	251
6.1	Généralités	252
6.1.1	Monitoring	252
6.1.2	Normes sur l'ordre des canaux	253
6.1.3	Montage son en multicanal	253
6.1.4	DAW	257
6.1.5	Consoles	257
6.2	La mesure	258
6.3	Manager multicanal	261
6.4	Gestion du Lfe, bass management	264
6.5	Pan-Pot, gestion du centre et des canaux arrière	267
6.5.1	Les lois de pan	267
6.5.2	Le type de pan-pot	269
6.5.3	Gestion du centre, divergence centrale	276

6.5.4	Divergence vers les autres canaux, Width ou Depth	277
6.5.5	Gestion des canaux arrière	279
6.6	Réverbération	280
6.6.1	Réverbération à convolution	281
6.6.2	Lexicon 960 L, Surround PCM96	284
6.7	Traitement Dynamique	284
6.7.1	Compresseurs multicanal	284
6.7.2	Limiteurs multicanal	288
6.8	Autres effets	289
6.8.1	Autopan	289
6.8.2	Délais	289
6.9	Downmix	289
6.10	Upmix	290
6.11	Méthode de mixage et analyses	292
6.11.1	Mixage film : les stems	292
6.11.2	Mixage Musique	293
6.11.3	Classement des sources	293
6.12	La synthèse binaurale	294
6.13	Exploitation des nouveaux formats, mixage objets	294
6.13.1	Auro-3D, 22.2 NHK	294
6.13.2	Exploitation de la WFS	297
6.13.3	WFS, Iosono Spatial Audio Workstation	299
6.13.4	Exploitation Dolby Atmos	301
6.14	Création sonore et interactivité avec la participation de Thierry Dilger	303
6.15	Applications	304
6.15.1	Film	304
6.15.2	TVHD : habillage sonore Arte HD 2008	308
6.15.3	Musique de film : Requiem pour une tueuse	308
<hr/> CHAPITRE 7 – LE CODAGE ET LA DIFFUSION <hr/>		309
7.1	Introduction	309
7.1.1	Définition	309
7.1.2	Pourquoi réduire le débit ou la taille ?	310
7.2	Principes	310
7.2.1	Codage entropique	310
7.2.2	Codage perceptif, enlever l'in audible	310
7.2.3	Les bases du codage	311
7.2.4	Bases de psychoacoustique	313
7.2.5	Modèles psychoacoustiques	315

7.3	Qualité subjective	321
7.3.1	Méthodes d'évaluation perceptuelle	322
7.3.2	Mesures perceptuelles	325
7.3.3	Considérations pratiques	325
7.3.4	Comparaison de codage	326
7.3.5	Artefacts typiques de codage	326
7.3.6	Codage en cascade	326
7.3.7	Canaux discrets et canaux matricés	327
7.4	Codage sans pertes	327
7.4.1	PCM	327
7.4.2	Lossless	327
7.4.3	DSD	328
7.5	Codages perceptifs	328
7.5.1	Famille MPEG Audio	328
7.5.2	Famille Dolby	332
7.5.3	Famille DTS	339
7.5.4	Sony ATRAC et SDDS	341
7.5.5	Auro-3D Octopus codec	341
7.5.6	Codages libres de droit	342
7.5.7	Évolution du codage	342
7.5.8	Droit et protections	342
7.5.9	Vidéo et audio	343
7.5.10	Supports physiques	345
7.5.11	Diffusion	351
—— CHAPITRE 8 – EXPLOITATION SON BROADCAST, SUPPORTS, NOUVEAUX MÉDIAS ——		359
8.1	La chaîne d'exploitation Broadcast	360
8.1.1	Synoptique du signal depuis la captation à la diffusion	360
8.1.2	Dolby Surround	361
8.1.3	Les metadata Dolby	361
8.1.4	Les metadata AAC	369
8.1.5	Metadata DTS	371
8.1.6	Intégration des encodeurs Dolby E	372
8.1.7	DM100, vérification de la bonne synchronisation des trames Dolby E et vidéo	374
8.1.8	Utilisation du DP570, metadata Dolby	374
8.1.9	Encodage/décodage Dolby Digital, Dolby Digital Plus, Dolby Pulse	380
8.1.10	Solutions Logiciels	380

8.2	Mesure du loudness	380
8.2.1	Principe, mesure Dolby	380
8.2.2	Recommandation ITU BS 1770-1	381
8.2.3	Loudness EBU R128 Europe, recommandation ITU BS 1770-3	383
8.3	Normes françaises PAD TVHD	385
8.3.1	Mesure des niveaux	386
8.3.2	Anciennes normes françaises PAD CST RT16, CST RT17 v2, CST RT19	387
8.3.3	Norme européenne actuelle EBU R128	387
8.3.4	Normes françaises PAD R128 2011 CST RT-17-TV version V3	392
8.3.5	Gestion du loudness des différents flux à la diffusion	399
8.4	Exploitation des outils de traitement broadcast 5.1	401
8.4.1	Traitement du Loudness et de la dynamique 5.1	401
8.4.2	« Upmixeurs »	402
8.4.3	Encodage Dolby D, Dolby D +, Dolby pulse et gestion des metadata	405
8.5	Exploitation des formats DTS	406
8.5.1	DTS Neural	406
8.5.2	Exploitation des formats DTS pour le support	406
8.6	Diffusion 5.1 et binaural pour les nouveaux médias	406
8.6.1	Quelques recommandations	407
8.6.2	Les codecs	407
8.6.3	Les lecteurs compatibles 5.1	410
8.6.4	Les lecteurs compatibles 5.1, 3D et binaural	411
<hr/>		
CONCLUSION		413
<hr/>		
BIBLIOGRAPHIE		415
<hr/>		
INDEX		417

À propos des auteurs

Les auteurs

Bergame Périaux

Bergame Périaux est diplômé du Conservatoire National Supérieur de Musique de Paris, de la Formation Supérieure aux Métiers du Son. Il rentre à l'INA en 2001, au département Formation Son. Il travaille en parallèle sur des productions musicales pour le CD, le DVD et la télévision. Il crée, au sein de la formation professionnelle son, la filière Multicanal dans laquelle il anime des formations pour les ingénieurs du son de Radio France, France Télévisions, France Media Monde, Arte, AMP VISUAL TV, Euromedia, Canal Plus, Eurosport, Technicolor, GTHP et de nombreux intermittents du spectacle.

En 2006, il fonde la société de production Bmusic Productions, spécialisée dans la captation, le mixage, l'édition sonore musicale et le consulting pour la gestion de projets son 3D et binaural.

Il communique sur l'actualité du son multicanal sur son blog www.lesonmulticanal.com.

Jean-Luc Ohl

Ingénieur ENSEA, diplômé en acoustique et bricoleur audio, Jean-Luc Ohl a participé aux technologies numériques liées à l'audiovisuel, dans le monde professionnel comme dans les produits grand public.

D'abord chef de produit « numérique » du groupe Thomson au lancement du Compact-Disc, il a pris une part active à l'authoring des premiers DVD produits en France. Ayant aussi assuré la direction générale de Revox-France, il œuvre maintenant pour la société 44.1, présente dans le Broadcast et le Cinéma, et a également une activité de consultant.

Un de ses passe-temps favoris est de programmer des logiciels liés à la psychoacoustique. Pour plus de détails, voir son site www.ohl.to.

Patrick Thévenot

Passionné depuis toujours par le son et diplômé de l'École Centrale de l'Électronique, du CNAM et de l'École Nationale Supérieure Louis Lumière, Patrick Thévenot a été responsable du laboratoire d'acoustique à l'INA. Il est co-fondateur de la société A2t et concepteur de la marque. Membre du comité de normalisation de l'AES, il est l'auteur de conférences et de nombreux articles dans des revues spécialisées.

Il partage actuellement son temps entre des activités d'ingénierie systémique au sein du bureau d'étude Taylor Made System tout en conservant son poste de formateur à l'INA.

Plus récemment, il est à l'origine de la marque d'enceintes Prosodia dont la finalité est d'associer neutralité et musicalité.

Spécialisé en psycho-acoustique, il évolue entre art et science grâce à son approche globale et systémique du son. Son but dans cet ouvrage est de permettre à l'ingénieur du son d'acquérir une vision claire et scientifique de l'écoute multicanal, en relation avec son vécu auditif et professionnel.

Avec la participation de...

Ont collaboré à cet ouvrage, dans l'ordre alphabétique :

Vincent Arnardi, mixeur film ; Philippe Barbeau et Martine Todisco, preneurs de son pour le documentaire animalier ; Etienne Corteel, Sonic Emotion ; Arnaud Damien, Euphonia ; Hervé Déjardin, responsable R&D sur le multicanal, Radio France ; Thierry Dilger, designer sonore ; Thierry Le Bon, mixeur film ; Bernard Lagnel, preneur de son Radio France ; Guillaume Le Dû, ingénieur du son, Radio France ; Didier Lozahic, mixeur film ; Vincent Magnier, preneur de son à l'image ; Rozenn Nicol, ingénieur, son 3D Orange Labs ; Gaël Nicolas, monteur son film ; Cyrille Richard, monteur son film ; Guy Senaux, ingénieur du son, Radio France ; Daniel Teruggi, directeur de la recherche, INA.

Remerciements

L'écriture de ces pages a été motivée par l'ambition et la passion de chaque contributeur, dont mes co-auteurs Patrick Thévenot et Jean-Luc Ohl et de nombreux collaborateurs professionnels.

Mes remerciements s'adressent à :

- Mike Williams et Guillaume Le Dû qui ont transmis généreusement leurs savoirs et leurs pratiques sur le MMAD ;
- Guy Senaux pour sa contribution sur les documentaires, reportages et fictions de Radio France en multicanal ;
- Vincent Magnier qui a collaboré sur la prise de son à l'image ;
- Les mixeurs film Vincent Arnardi, Thierry Le Bon et Didier Lozahic qui ont apporté leurs bonnes pratiques en mixage et leurs points de vue ;
- Philippe Barbeau et Martine Todisco pour ce petit voyage en plein documentaire animalier ;
- Rozenn Nicol qui a généreusement communiqué son savoir sur le binaural et sur la perception auditive ;
- Hervé Déjardin pour sa contribution sur le reportage Interception et les échanges fructueux sur l'évolution du multicanal à Radio France ;
- Bernard Lagnel pour sa générosité et le partage de ses brillantes expérimentations sur le multicanal et le binaural ;
- Thierry Dilger pour sa collaboration sur la partie création sonore, interactivité et sur l'analyse sonore des jeux vidéo ;
- Matthieu Parmentier pour la communication des avancées conséquentes sur les nouvelles technologies de la R&D à France Télévisions ;
- Jérôme Bordier pour la communication sur les avancées d'Eurosport sur la production de programmes sport en 5.1 ;
- Gaël Nicolas pour sa collaboration sur la prise de son multicanal au Soundfield ;
- Bernard Fouquet pour les discussions et le partage sur les normes de codage et de diffusion ;

- Trinnov Audio pour le partage de leurs avancées technologiques sur l'audio 3D et la haute résolution spatiale ;
- Daniel Dal Pio Luogo pour son éclairage sur l'exploitation des formats Dolby ;
- Étienne Corteel et Arnaud Damien pour leur collaboration sur l'exploitation de la WFS ;
- Cyrille Richard pour sa collaboration sur la partie montage son pour le film ;
- Florian Camerer qui a rédigé la préface ;
- l'équipe de l'INA qui a soutenu le projet et Daniel Teruggi pour sa collaboration sur le GRM, l'acousmatique et l'acousmonium ;
- les éditions Dunod, et en particulier Jean-Baptiste Gugès pour avoir soutenu le projet.

Sigles, termes techniques, conventions

Un certain nombre de sigles ou termes techniques sont utilisés par les auteurs, certains se traduisent difficilement de l'anglais vers le français et sont généralement exploités tels quels. En voici la liste et la définition.

Principaux canaux de mixage et de diffusion multicanal

- L ou FL** : Left ou Front Left, canal frontal gauche
- Lc** : Left Center, canal frontal centre gauche
- C ou FC** : Center ou Front Center, canal frontal centre
- RC** : Right Center, canal frontal centre droite
- R ou FR** : Right ou Front Right, canal frontal droite
- Ls ou SL** : Left Surround, canal surround gauche
- Rs ou SR** : Right Surround, canal surround droite
- Lfe ou LFE** : Low Frequency effect, canal d'effet basse fréquence
- Lss** : Left Side Surround, canal latéral gauche
- Rss** : Right Side Surround, canal latéral droite
- BSL** : Back Surround Left, canal arrière gauche
- BSR** : Back Surround Right, canal arrière droite
- CS** : Center Surround, canal arrière centre
- Sub, Sub Bass ou Subwoofer** : caisson de grave

Formats de mixage ou de prise de son usuels

5.1 : L, C, R, Ls, Rs, Lfe

5.0 : L, C, R, Ls, Rs

4.0 : L, R, Ls, Rs appelé format quad

3.0 : L, C, R

2.0 : L, R

Downmix, downmixer : opération de fabrication d'un format inférieur à partir d'un format supérieur. On utilise généralement ce terme pour fabriquer un format stéréo ou mono depuis un format 5.1

Upmix, upmixer : opération inverse de fabrication d'un format supérieur à partir d'un format inférieur. On fabrique généralement une version 5.1 à partir d'un format stéréo ou mono

Pan-pot, pan, panner, pan-poter : opération de spatialisation d'une source utilisant le panoramique lors du mixage

Monitoring, monitorer : écoute, contrôler à l'écoute

Manager multicanal : outil de gestion des canaux d'une piste multicanal

Stem : prémix

Channel oriented : contenu audio directement envoyé vers les enceintes (*speaker oriented*), mixage lié à un format de diffusion

Object oriented, objet : contenu audio + métadonnées de localisation, indépendant du système de diffusion

Scene oriented : contenu audio type ambisonique contenant l'information directionnelle du champ sonore

Bed : mixage traditionnel type 5.1, 7.1, 9.1 orienté channel, mixage constituant la base principale du contenu audio (« Mix core »)

Routing, router : fonction d'acheminement des différents bus ou canaux du multicanal

Hard center : centre physique porté par l'enceinte centrale seule

Sweet spot : point de convergence du système d'écoute, en multicanal ITU le centre du cercle d'écoute

Link, linké : on utilise le terme « link » pour indiquer que des pistes sont reliées entre elles

F/A : rapport frontal/arrière

CD/CR : rapport champ direct sur champ réverbéré

Son direct : ce terme a plusieurs sens suivant son contexte ; dans le contexte de l'acoustique des salles, il caractérise un son sans réverbération, il est opposé au son réverbéré et définit aussi, dans le contexte du film, le son enregistré en direct lors du tournage

Binaural : le terme désigne l'écoute spatialisée au casque bien que la définition exacte du mot binaural soit « ayant trait aux deux oreilles »

Introduction

Cet ouvrage est né d'une véritable passion pour la spatialisation du son que l'auteur principal exploite en production et enseigne depuis ses premières écoutes en 5.1. À chaque nouvelle écoute, les sensations d'espace en multicanal sont étonnantes de réalisme ou de créativité. Soit on reproduit la réalité sonore, soit on crée un univers imaginaire, les deux approches étant riches d'émotions.

L'envie a été de s'entourer de co-auteurs experts dans leurs domaines et de collaborateurs professionnels preneurs de son, mixeurs, monteurs son, réalisateurs, ingénieurs pour la recherche... Cette collaboration est bien sûr d'une grande richesse et fait de ce travail un partage et la réunion de points de vue différents et complémentaires.

Ces pages abordent principalement l'exploitation du son multicanal au format 5.1 pour différentes applications, le film, la TVHD, la radio, la musique, le jeu vidéo, la scénographie, sur des aspects techniques et artistiques. Comment l'histoire d'un tel format s'est-elle construite ? Quels en sont les intérêts, les vecteurs ? Quelles techniques de spatialisation trouve-t-on ? Quelles sont les règles qui dirigent l'écriture, la réalisation et qui constituent le langage de ces nouvelles images sonores ?

Cet ouvrage s'adresse aux professionnels en exploitation autant qu'aux étudiants en son. La lecture peut se faire à différents niveaux, le lecteur peut approfondir une technique de prise de son particulière, comme il peut souhaiter simplement comprendre les intérêts de produire un son spatialisé ou tout simplement écouter des extraits sonores en 5.1 ou en binaural.

À chaque début de chapitre sont exposés les points essentiels du contenu, ces informations servent à guider le lecteur. Les principes sont développés dans chacun des chapitres et souvent illustrés par des schémas, des photographies, des expériences, des points de vue et des interviews de professionnels. Certains contenus complémentaires ou plus spécifiques sont mis en ligne et accessibles gratuitement sur le site www.dunod.com et sont signalés tout au long de l'ouvrage. Ces extraits sont également disponibles sur le site www.lesonmulticanal.com. L'écoute d'extraits sonores est sans aucun doute un des points forts de cet ouvrage. Ce sont des sons au format 5.1 ou en binaural qui donnent des références indispensables à l'apprentissage de ce nouveau mode d'écoute. Le lecteur trouvera aussi dans la partie web une session Pro Tools test faisant référence

au chapitre sur le système d'écoute. Cet outil permet à un exploitant de diagnostiquer une écoute 5.1 ITU sans aucune mesure acoustique.

Le contenu est divisé en huit chapitres, le premier introduit les généralités et l'historique. Il définit les intérêts du multicanal, ses vecteurs, les différents marchés et présente les nombreux formats d'exploitation. Le deuxième chapitre aborde le dispositif 5.1 ITU et l'écoute cinéma, les normes, la calibration et ouvre vers d'autres systèmes plus évolués en 3D ou d'autres technologies de diffusion comme la WFS (Wave Field Synthesis). Le chapitre 3 explore les données psychoacoustiques liées au multicanal, il pose ainsi les bases théoriques nécessaires à l'analyse de l'image sonore et à la compréhension des concepts de la prise de son et de mixage. Le chapitre 4 fait l'analyse de l'image sonore avec la présentation des différents critères, il traite aussi de l'esthétique et de la réalisation des contenus spatialisés. Le chapitre 5 expose le principe de la prise de son multicanal, les différents systèmes utilisés ainsi que le cas particulier de la prise de son binaurale, il est illustré par le retour d'expériences de preneurs de son. Le chapitre 6 aborde la postproduction : les concepts de montage son pour le film, les outils de mixage 5.1 et quelques méthodes de travail sont exposés. Ce chapitre est illustré par le retour d'expérimentations de mixeurs professionnels et l'analyse d'œuvres mixées en 5.1. Un aparté est fait sur la synthèse binaurale et l'interactivité dans le son. Le chapitre 7 étudie les différentes familles de codage, Dolby, DTS, AAC, les différents supports du multicanal ainsi que les différents modes de diffusion. Enfin, le chapitre 8 aborde la chaîne de production broadcast multicanal, la gestion des metadata, la mesure du loudness et ouvre vers l'exploitation du 5.1 et du binaural pour les nouveaux médias.

Chapitre 1

Généralités

L'espace est une invitation au voyage

POINTS ESSENTIELS

- Les différentes façons d'écouter un son.
- Le multicanal : historique, intérêts, vecteurs, problématiques.
- Les différentes techniques de spatialisation : stéréophonie, ambisonie, WFS, objets, binaural.
- Les formats du multicanal.
- État des lieux, étude des marchés du multicanal, statistiques.

Dans ce premier chapitre sont exposées les généralités du multicanal. Elles permettent de situer le contexte de la spatialisation et de poser les questions essentielles sur la fabrication du son. Comment l'historique s'est-il fait depuis les premières expérimentations ? Quelles sont les différentes technologies utilisées pour spatialiser le son ? Quels sont les intérêts de produire en son multicanal ? Quels sont les marchés et les vecteurs ? Comment diffuse-t-on ce type de son ? Quelles sont les nouvelles problématiques d'exploitation ? Enfin quels sont les différents formats du multicanal ?

1.1 Écouter

Lorsque l'auditeur écoute, il s'applique à entendre un son.

Il existe aujourd'hui différentes façons d'écouter le son : on peut s'informer ou se divertir. Ces deux modes d'écoute sollicitent une attention auditive très variable, motivée par une attente chez l'auditeur et des émotions recherchées très différentes. Le multicanal peut alors se définir par la recherche d'émotions plus, fortes passant par l'auditif.

L'écoute du bulletin d'information à la radio, ou à la télévision, demande une bonne intelligibilité de la parole. Le spectateur peut être mobile, suivre les reportages qui l'intéressent le plus, ignorer les autres, dans ce cas une construction d'image sonore n'est pas recherchée : une image monophonique ou stéréophonique est le format le plus adapté. L'écoute d'une émission télévisée ou radiophonique peut demander une plus grande attention auditive. Dans ce cas le spectateur souhaite suivre l'intégralité du programme mais il peut l'écouter avec un certain recul, en faisant une autre activité. Il peut se déplacer et modifier son point d'écoute, la réception du signal ne sera pas alternée et dans ce cas la construction spatiale de l'image sonore sera peu exploitée.

L'écoute 5.1 sur home-cinéma d'un concert, d'un événement sportif, d'un documentaire ou d'un film suscite une plus grande attention auditive du spectateur. Ce dernier est généralement immobile, centré devant l'image de son écran de TV, en attente d'un spectacle pour lequel la construction de l'image sonore présente un intérêt. Le spectateur apprécie la spatialisation du son, mais celle-ci n'est pas indispensable.

Le dernier degré d'écoute implique davantage l'auditeur. C'est par exemple le cas de la diffusion en salle d'un film. Les différents canaux améliorent l'intelligibilité du son pour tous les spectateurs et renforcent particulièrement le rapport du son à l'image. D'autres contenus denses, comme la musique contemporaine, s'écoulent plus naturellement en 5.1 qu'en stéréo, certains contenus s'écrivent même directement avec l'espace, comme la musique acousmatique. Dans un jeu vidéo le joueur s'identifie au personnage grâce à la spatialisation, cette dernière décuple les sensations réalistes de situation et de déplacement. On trouve généralement dans ces contenus des séquences sonores qui enveloppent l'auditeur. Dans ce cas, la spatialisation du son fait partie de l'écriture de l'œuvre.

1.2 Historique

L'historique du multicanal nous renvoie à soixante-quinze ans d'expérimentations, depuis les expériences de Disney en 1940, jusqu'aux nouvelles technologies de spatialisation que nous connaissons aujourd'hui.

1.2.1 Cinéma

Le cinéma s'ouvre au multicanal à partir des premières expériences des ingénieurs de Disney en 1940, autour du film *Fantasia*. Le procédé utilisé est le Fantasound : le

film est sonorisé en cinq canaux à partir de trois canaux de production LCR. La diffusion était LCR derrière l'écran et LsRs en fond de salle, ce qui définit les bases de notre format 5.1. Entre 1946 et 1954, le cinéma connaît un déclin dû à l'arrivée de la télévision. Le nombre de spectateurs chute de moitié. De 1950 à 1970 s'enchaînent une multitude de procédés, entre autres le Cinérama et le Todd AO, précurseurs du format 7.1 SDDS, disposant cinq enceintes derrière l'écran, et le Cinémascope : un procédé en quatre canaux LCRS. Dans les années 1970 apparaît le procédé IMAX (Image Maximum) qui tend à développer la taille de l'image et l'angle de vision avec combinaison de projecteurs multiples pour les grandes salles. L'IMAX se décline en IMAX DOME (appelé initialement OMNIMAX), prévu pour la projection sur des écrans inclinés et en relief grâce à l'IMAX 3D. Ce dernier apporte un effet immersif. En 1974 apparaît le Subwoofer, avec notamment le procédé Sensurround, qui étend la bande passante de la diffusion dans le grave.

En 1976, ces différents procédés cèdent la place au Dolby Stereo qui, grâce à l'utilisation d'une matrice 4-2-4, permet de diffuser le mixage LCRS de l'époque à partir de deux canaux LtRt (Left total, Right total, voir chapitres 7 et 8). En 1976, *A star is Born* marque le début de la production des films en Dolby stereo. Ce sera le premier film sur pellicule 35 mm utilisant ce procédé. Par la suite, un certain nombre d'améliorations font évoluer le mode de diffusion : en 1977, le producteur de *Star Wars* souhaite renforcer la diffusion dans les graves avec Dolby pour mieux simuler les scènes de guerre dans l'espace et ajoute pour cela un canal dédié au Subwoofer appelé Lfe (Low Frequency Effect ou Baby Boom). En 1978, *Superman* est le premier film expérimentant des canaux surround stereo. Apparaît en 1983 la norme THX avec le *Retour du Jedi* : elle impose une norme pour le réglage son des salles de cinéma. Et en 1986, Dolby intègre à travers le procédé Dolby SR (Spectral Recording) son réducteur de bruit.

En 1987, une avancée importante : le dispositif 5.1 du cinéma est normé. L'année 1990 marque la sortie du centième film en Dolby Stereo SR, *Robocop 2*.

En 1992, Dolby révolutionne encore le rendu sonore au cinéma avec une version numérique du SR : le *Dolby SRD (Spectral Recording Digital)*. Le *Dolby SRD* ajoute sur la pellicule film, à côté des deux pistes analogiques *Dolby SR Lt Rt* toujours présentes pour garantir la compatibilité, un signal numérique AC-3 imprimé entre les perforations. Ainsi, on peut obtenir jusqu'à six canaux totalement discrets en configuration 3/2/1, le sixième canal correspondant au Lfe. Ce nouveau procédé numérique est inauguré la même année avec le *Retour de Batman*. En 1993, DTS et SONY se joignent à cette évolution numérique en proposant le DTS 5.1 et le SDDS 7.1 (le SDDS place cinq enceintes derrière l'écran). En 1996, environ quatre mille salles sont équipées en Dolby Digital, la même année, le Dolby Drive, système entièrement basé sur disque dur conçu pour remplacer à terme les défileurs 35 mm magnétiques perforés, est présenté à l'AES de Copenhague.

Nom	Année	Format Pellicule	Technologie	Codage Son	Format
Fantasound	1940	35 mm	3 pistes analogiques optiques		LCR
Cinérama	1952 à 1962	35 mm	7 pistes analogiques magnétiques		LLcCRcRLsRs
CinemaScope	1953 à 1967	35 mm	4 pistes analogiques magnétiques		LCRS
Todd-AO	1955 à 1992	70 mm	6 pistes analogiques magnétiques		LLcCRcRS
Dolby Stereo	1976 à 2006	35 mm	2 pistes analogiques matricées optiques	Dolby A	LCRS
Ultra Stereo	1984	35 mm	2 pistes analogiques matricées optiques		LCRS
Dolby Discrete 6 tracks	1976 à 2006	70 mm	6 pistes analogiques magnétiques	Dolby A	LLcCRcRS
Dolby « Baby Boom » 6 tracks	1977 à 2006	70 mm	6 pistes analogiques magnétiques	Dolby A	LCRLsRsLfe
Dolby « Split Surround » 6 tracks	1979 à 2006	70 mm	6 pistes analogiques magnétiques	Dolby A	LCRSLfe
Dolby Stereo SR	1986 à 2006	35 mm	2 pistes analogiques matricées optiques	Dolby SR	LCRS
Kodak CDS	1990 à 1991	35 mm 70 mm	6 canaux numériques optiques	Delta Modulation	LCRLsRsLfe
Dolby Digital	1992 à 2006	35 mm	6 canaux numériques optiques	AC3	LCRLsRsLfe
DTS	1993 à 2006	35 mm	6 canaux numériques optiques	DTS Coherent Acoustics	LCRLsRsLfe
SDDS	1993 à 2006	35 mm	6 canaux numériques optiques	ATRAC	LLcCRcRLsRsLfe

Figure 1.1 – Les différents procédés de diffusion son du cinéma sur pellicule 35 mm ou 70 mm. L'année 2006 marquant la fin du 35 mm avec l'arrivée du cinéma numérique.

En 1999, Star Wars, avec *La Menace fantôme*, fait de nouveau évoluer le format 5.1 en intégrant un canal centre arrière grâce au Dolby Digital Surround EX, procédé co-développé par Dolby et Lucas Film THX. DTS propose en 2000 l'équivalent avec son format DTS ES 7.1.

2006 marque le démarrage du cinéma numérique avec la sortie du premier film *Chicken Little* diffusé en Dolby Digital Cinema. Les salles de cinéma s'équipent progressivement en numérique. Le retour de l'intérêt pour l'image 3D sera par la suite un moteur pour ce passage au tout numérique. En 2009 a lieu la sortie de *Batman, Le Chevalier Noir* en format IMAX. De nombreux succès cinématographiques tels que trois des volets d'*Harry Potter*, *Batman Begins* ou *Superman Returns* ont été gonflés au tirage pour permettre des projections en IMAX. La France compte aujourd'hui une quinzaine de salles IMAX dont les plus récentes sont équipées IMAX Numérique 3D. Si le succès de ces nouvelles salles IMAX Numérique se confirme, trente-cinq autres sites potentiels en France sont éligibles à cette reconversion selon la société IMAX Corporation.

En 2010, Dolby présente son nouveau format Dolby Surround 7.1 dont il souhaite faire état comme étant la norme son pour l'image 3D. Cette norme comporte quatre canaux surround, dont deux sur les côtés et deux à l'arrière. *Toy Story 3* est alors le premier film diffusé en Dolby Surround 7.1.

À partir de 2010, le cinéma numérique évolue vers des formats son 3D, et l'on voit apparaître différents procédés qui équipent les salles de haut-parleurs au plafond. Ainsi voient le jour les dispositifs Imm Sound (Imm Sound équipait quelques salles Gaumont-Pathé dont la première en France à Brumath en Alsace), Auro-3D, Dolby ATMOS et la technologie WFS. Ces procédés concurrentiels multiplient le nombre de canaux et tentent de normaliser des formats 11.1, 13.1, 14.1, 23.1 et d'autres, supérieurs, compatibles 5.1 et 7.1. En juillet 2012, Dolby acquiert Imm Sound et se propulse dans le marché du son 3D pour le cinéma. En France, la même année, les deux premières salles équipées Dolby Atmos sont inaugurées à Paris et à Bordeaux. En 2013, un format d'échange libre de droit pour le mixage objet, le MDA (multi-dimensionnel audio) est supporté par DTS. En 2015, il existe plusieurs complexes équipés Dolby Atmos : environ dix-sept salles de cinéma, deux salles de vision Disney/Warner, cinq auditoriums (Dubbing Brothers, Cinéphase, Ink Production, Creative Sound, les Auditoriums de Saint-Ouen), Titra films et le labo DCP (Eclair KDM).

1.2.2 Home-cinéma, DVD, TVHD, Radio, jeu vidéo, Internet

Le film s'installe chez le particulier grâce au home-cinéma dès que sa diffusion devient possible, notamment avec l'arrivée de nouveaux supports et avec le développement de solutions de matrices Dolby Surround.

La fin des années soixante et des années soixante-dix sont marquées par les expérimentations de la quadriphonie, notamment avec l'apparition de nouveaux supports matricés : le QS quadrasonic sound, le SQ stereo quadrasonic, le Matrix H (développé par la BBC pour être diffusé sur la FM) et des supports discrets tels que le CD4 (JVC, RCA), le UD4/UMX (Denon), le Q4 et le Q8, ces derniers étant des formats sur bande 1/4 pouce et cartouche, tous les autres étant des formats sur disque vinyle. Cette évolution technologique des supports est soutenue par les maisons de disques, on voit apparaître notamment de nouvelles productions musicales, comme *Dark Side of the Moon* des Pink Floyd en 1973.

HISTORIQUE CINÉMA

1931	Blumlein, 2 canaux L/R stéréo		analogiques et Dolby Stereo Digital, proposée par Cinemeccanica
1933	Bell Labs, 3 canaux L/C/R stéréo		
1940	<i>Fantasia</i> , procédé Fantasound 3 canaux diffusés en LCR/LsRs	1993	<i>Jurassic Park</i> , DTS 5.1 <i>Last Action Hero</i> , SDDS 7.1
1952	<i>This is Cinerama</i> , procédé Cinérama 7 canaux LlcCRcR/LsRs	1995	Présentation au Show East d'Atlantic City du processeur de cinéma numérique Dolby CP500 capable de lire à la fois les pistes analogiques optiques et la piste Dolby D
1953	<i>The Robe</i> , procédé Cinemascope 4 canaux LCRS, 20th Century Fox	1996	Environ 4 000 salles sont équipées dans le monde pour lire le Dolby Digital Le Dolby Drive, système entièrement basé sur disque dur, conçu pour remplacer à terme les défailleurs 35 mm magnétiques perforés, est présenté à l'AES de Copenhague
1955	<i>Oklahoma !</i> , procédé Todd AO 6 pistes LlcCRcR/S	1998	15 000 salles de cinéma équipées en Dolby Digital dans le monde, 14 000 en DTS
1970	<i>Tiger Child</i> , le premier film IMAX, est projeté durant l'exposition universelle de 1970, à Osaka	1999	<i>La Menace fantôme</i> , Dolby Digital Surround EX, format 6.1 co-développé par Dolby et Lucas Film THX 20 000 salles équipées en Dolby Digital
1974	<i>Earthquake</i> procédé Sensurround, apparition des Subwoofers	2000	DTS ES 7.1 Plus de 25 000 salles équipées Dolby Digital
1976	<i>A star is born</i> (version Barbara Streisand, Kris Kristofferson), premier film 35 mm avec des effets surround et première utilisation de la matrice 4 : 2 : 4 du Dolby Stereo	2006	Cinéma Numérique, <i>Chicken Little</i> Dolby Digital Cinema
1977	<i>Star Wars</i> , première utilisation du canal Lfe	2009	<i>Avatar</i> , regain d'intérêt pour le film en 3D
1977	<i>Close Encounters of the Third Kind</i> , Canal Lfe dédié	2010	<i>Toy Story 3</i> , Dolby 3D 7.1
1978	<i>Superman</i> , premier film magnétique 70 mm expérimentant des canaux arrière stéréo	2011	<i>Mission impossible 4</i> , sortie au format IMAX
1979	<i>Apocalypse Now</i> premier film en Dolby Stereo six pistes magnétiques 70 mm à être distribué commercialement avec un son surround stereo <i>Don Giovanni</i> , premier mixage film Dolby Stereo en France	2012	<i>Red Tails</i> , Lucas Film, sortie du premier film Auro-3D 11.1 en salle de cinéma
1983	<i>Return of the Jedi</i> , Norme THX	2012	<i>The Impossible</i> , sortie au format Imm Sound 14.1
1986	Le format Dolby Stereo SR (spectral recording) 35 mm optique est présenté à l'Academy of Motion Picture Arts and Sciences	2012	Dolby ATMOS, 128 canaux, 64 haut-parleurs, sortie de <i>Rebelle</i> , Disney/Pixar, premier film en Dolby Atmos
1987	Premiers films en Dolby Stereo SR <i>Innerspace</i> , <i>Robocop</i> . Le format 5.1 est normé pour le film	2012	Dolby acquiert Imm Sound
1990	Sortie du centième film en Dolby Stereo SR <i>Robocop 2</i>	2012	Inauguration des 2 premières salles de cinéma équipées Dolby Atmos en France, à Paris et à Bordeaux
1991	Premières démonstrations du format Dolby Stereo Digital (Dolby SRD) intégrant l'AC3 en complément des 2 pistes LtRt	2013	Développement du format d'échange MDA par DTS pour le mixage objets
1992	<i>Batman Returns</i> , Dolby Digital 5.1 (Dolby SRD)	2015	Installations Dolby Atmos en France : 17 salles de cinéma, 2 salles de vision, 5 auditoriums, un labo DCP. 170 films produits en Dolby Atmos de 2012 à 2015.
1993	Première tête de lecture optique 35 mm capable de lire à la fois les bandes son		

Chacun des supports quad est soutenu par un groupe de maisons de disques. Le QS est supporté par ABC, Advent, Bluesway, Candide, Command, Decca, Impulse, Longines, MCA, Ovation, Pye, Turnabout et Vox. Le SQ est supporté par Angel, CTI, Columbia, EMI, Epic, Eurodisc, Harvest, HMV, Seraphim, Supraphon et Vanguard. Et le CD4 est supporté par Arista, Atlantic, Capricorn, Elektra, Fantasy, JVC, Nonesuch, A & M, Reprise & Warner Bros Records. Les radios s'impliquent de leur côté dans les recherches et dans les productions : entre autres Radio France, la BBC et des radios américaines, qui diffusent des émissions ou des musiques en format quad.

L'échec de la quadriphonie tient principalement à un trop grand nombre de supports quad concurrentiels et vient aussi de la réticence du public à ajouter deux enceintes supplémentaires dans son salon. D'autre part, l'esthétique musicale définie par le format quad n'était peut-être pas suffisamment mûre dans l'esprit des producteurs. L'étude psychoacoustique du format était imparfaite, l'absence de compatibilité stéréo et la localisation imprécise des images sonores limitaient le procédé. L'absence de compatibilité avec le 5.1 du cinéma a sans aucun doute aussi joué un rôle.

Le home-cinéma prend son essor en 1987, dès l'intégration du Dolby Surround dans les amplificateurs AV (audiovisuels) sous forme de décodeurs Pro Logic. La cassette VHS, le vidéo disc ou le laser disc, comportent la version 5.1 des films en Dolby Surround. Ils sont largement dépassés par le marché du DVD vidéo qui apparaît en 1996. À la télévision, les films du soir sont généralement diffusés en Dolby Surround, compatible avec les premiers amplificateurs AV. Les technologies discrètes Dolby Digital et DTS sont intégrées progressivement à tous les produits audiovisuels : le film, la musique (quelques productions en CD et DVD), le jeu vidéo et bientôt la télévision (Dolby Digital).

En 2005 le marché de la HD révolutionne le secteur audiovisuel, notamment celui de la télévision. La TVHD est lancée, aidée par la vente conséquente d'écrans plats. Certaines chaînes HD diffusent leurs programmes en Dolby Digital 5.1, comme le sport ou des films. Dolby propose aux chaînes de télévision son procédé de codage Dolby E permettant de transporter huit canaux sur une simple interface AES3, ce qui permet de véhiculer un son 5.1 et sa version stéréo simultanément sur une infrastructure stéréo. Ce marché est relayé par le support Blu-ray, le remplaçant du DVD, capable d'intégrer une image HD et un son multicanal non compressé compatible 7.1. En 2009, une version audio du Blu-ray appelée Pure Audio Blu-ray voit le jour. Malheureusement, peu de productions existent encore aujourd'hui. En 2010, le multimédia suit cette évolution avec la première démonstration de diffusion HD en Dolby Digital plus 5.1 depuis un mobile, utilisant une interface mini HDMI. On trouve aussi cette interface HDMI sur les ordinateurs et sur les tablettes numériques. L'utilisateur peut donc télécharger un film en HD son Dolby Digital 5.1 et le diffuser sur son ampli home-cinéma.

En février 2013, Radio France inaugure son nouveau site de diffusion NouvOson sur lequel l'auditeur peut écouter des programmes en 5.1 et en binaural grâce à un player compatible. Début 2015 Orange Labs sort la nouvelle version de son application Orange Radio compatible binaural, celle-ci permet la diffusion binaurale de contenus 5.1 à partir de flux de webradios tels que NouvOson ou RFI Labo.

HISTORIQUE HOME-CINÉMA, DVD, TVHD RADIO, JEU VIDÉO

1964	La NHK commence le développement de la TVHD		
1971	Quadriphonie, supports matrixH (BBC), SQ stereo quadraphonic (CBS), QS quadraphonic sound (SANSUI), CD4 (JVC et RCA)		Les systèmes Virtual Dolby Surround et Virtual Dolby Digital, permettant un effet de son surround à partir de seulement 2 haut-parleurs (transaural) sont présentés au CES de Las Vegas
1982	Dolby introduit le procédé Dolby Surround La NHK commence les productions TVHD		Le Dolby Digital est ratifié comme étant le format audio obligatoire pour les DVD Pal/Sécam, de même que pour les disques NTSC, ce qui permet une sortie mondiale d'un titre, avec la même bande-son
1984	Procédé Dolby Soundlink adaptive delta modulation plus connu sous l'appellation AC-1, algorithme de compression numérique pour les transmissions TV en satellite et par câble		<i>Le Singe Soleil</i> , premier documentaire radio enregistré en multicanal, diffusé sur France Inter
1985	La chaîne de télévision musicale VH1 commence à émettre aux USA en utilisant le Dolby AC1 pour sa distribution par satellite Premier vidéo-disc et première cassette distribués avec un son Dolby Surround	1998	Premier jeu vidéo en Dolby Digital Dolby Headphone développé par Lake DSP (Sydney, Australie) DAB
1987	Lancement du Dolby Pro Logic	1999	Dolby E ; SACD Singapore Airlines propose à ses passagers le système Dolby Headphone
1987	CBS diffuse le 21 ^e Super Bowl en Dolby Surround		Un flux Dolby Digital 5.1 est envoyé par Internet de Montréal à la convention AES de New York
1988	TDF1&2, transmission TV satellite puis câble D2MacPaquet compatible TVHD, transmission 8 canaux audio numériques	2000	Dolby Pro Logic 2 ; DVD Audio Norme D2Mac abandonnée au profit du DVB MPEG2
1989	Les codecs numériques AC-2 Dolby 500 sont présentés à l'AES de New York La NHK émet en test 1 heure par semaine des programmes TVHD (High Vision)		NHK première chaîne HD 24h/24 par satellite numérique
1991	La NHK émet en test 8 heures par jour des programmes TVHD	2005	Codec sans perte Dolby TrueHD pour les supports vidéo HD T-DMB Première démonstration par la NHK de l'Ultra High Definition TV
1994	Le premier service de diffusion directe par satellite avec encodage AC-3 (le DMX for Business) est lancé Fox Network diffuse toute la saison de football américain (NFL) en Dolby Surround Norme 5.1 ITU-R BS 775	2006	Blu-ray
1995	Premiers produits grand public intégrant la technologie AC3 dont le Laser disc, Toshiba et Time Warner font la démonstration d'un prototype de DVD avec de l'audio en Numérique AC-3 La BBC retransmet la première émission radiophonique en Dolby Surround (<i>Bomber</i> , BBC Radio 4)	2007	Le marché du DVD américain compte 126 millions de lecteurs commercialisés avec 75 000 titres disponibles DAB+
1996	Dolby signe des accords avec Microsoft, d'une part, pour développer l'intégration dans les PC de techniques AC-3 et Surround Pro Logic, et avec Digidesign, d'autre part, pour marier plus efficacement leurs technologies La version française de <i>Judge Dredd</i> (1995) est le premier LaserDisc NTSC européen à comporter une bande-son en Dolby Digital. En 1996, on compte déjà plus de 100 titres Laser Disc disponibles aux USA avec un son surround 5.1 Dolby Digital Le Dolby Net, une version à faible débit du Dolby Digital, est présenté pour les applications de streaming Internet à bande passante réduite <i>Le Cataclysme Sonore</i> , première émission radiophonique en France mixée en multicanal, diffusée sur France Inter en Dolby Surround	2008	Lancement en France de la TNT HD
1997	DVD vidéo La première automobile équipée d'un système audio Surround Dolby Pro Logic est proposée par Volvo au Detroit Motor Show Le premier décodeur Dolby Digital software exploitant la technologie MMX™ est annoncé par Intel	2009	Pure Audio Blu-ray ; DTS Neural Surround
		2010	Première présentation publique de l'intégration du son 5.1 dans un téléphone portable utilisant la technologie Dolby Mobile
		2012	Le Super Bowl est produit en DTS Neural, transporté en Dolby E aux chaînes TVHD puis diffusé en Dolby Digital
		2012	Jeux Olympiques de Londres : la NHK, la BBC et OBS captent en Ultra High Definition une partie des jeux, son 22.2, avec retransmission en direct par satellite dans les salles de cinéma, à Londres, Glasgow, Bradford, Washington, Tokyo et Osaka
		2013	Lancement du site internet NouvOson Radio France comportant des programmes en 5.1 en binaural
		2014	Développement du standard audio 3D MPEG-H 3D par Fraunhofer
		2014	Norme ITU-R BS 2051 Advanced sound system for programme production
		2015	Orange Labs sort l'application Orange Radio compatible binaural
		2015	Sortie d'une quinzaine de films sur Blu-ray en Dolby Atmos
		2015	The <i>Expendables 2</i> , premier film sur Blu-ray en DTS Neo X

1.2.3 Écriture, édition et diffusion musicales

Dès le xvi^e siècle, les compositeurs ont manifesté un intérêt pour l'espace et le mouvement dans l'écriture musicale. Ils disposent de façon circulaire les chœurs et certains cuivres en coulisses. Les musiciens sont dans le public et se déplacent. L'époque classique et romantique tend à fixer l'orchestre devant le public, tandis que le $xvii^e$ siècle et la musique contemporaine utilisent beaucoup la notion d'espace. La densité de l'écriture musicale contemporaine, avec l'utilisation de plusieurs orchestres et un dispositif d'enceintes, tend à spatialiser les musiciens et les sources autour de l'auditeur.

Les premiers compositeurs qui se sont intéressés à l'espace après Gabrieli, Berlioz, Mahler ou Debussy, sont Edgar Varèse, Iannis Xenakis, Karlheinz Stockhausen et Pierre Boulez. Varèse déclare en visionnaire : « la musique de demain sera spatiale » et « les sons donneront l'impression de décrire des trajectoires dans l'espace, de se situer dans un univers sonore en relief ». Iannis Xenakis et Karlheinz Stockhausen sont les compositeurs qui ont le plus réfléchi sur la spatialisation de la musique. Leurs réflexions portent aussi bien sur la répartition du public ou de l'orchestre au sein de la salle de concert que sur l'écriture même et débouchent sur la notion de relief, de mouvements et d'ubiquités sonores.

À partir de la seconde moitié du xx^e siècle, le haut-parleur est intégré dans la composition musicale comme un instrument à part entière : les œuvres sont alors plus facilement écrites pour l'espace.

Le Groupe de Recherches Musicales de l'INA (Institut National de l'Audiovisuel), le GRM, a été créé en 1958 par Pierre Schaeffer, douze ans après son invention de la *musique concrète*, pour explorer les possibilités de la création musicale associant des outils technologiques. Dès ses débuts, et notamment lors de la création de l'œuvre emblématique qu'il a composé avec Pierre Henry, la *Symphonie pour un homme seul* en 1951, l'organisation du son dans l'espace a été au centre de la restitution des musiques en concert. Après la première expérience avec quatre haut-parleurs contrôlés depuis la scène, le nombre de haut-parleurs a été incrémenté progressivement pour aboutir en 1974 à la mise en œuvre de l'Acousmonium, conçu par François Bayle pour pouvoir distribuer la musique dans l'espace d'une salle de concert en fonction de ses caractéristiques timbrales ou fréquentielles.

L'Acousmonium a été conçu pour permettre à des interprètes de « jouer » avec le son stéréophonique dans une salle de concert à partir d'une console de contrôle placée dans son centre. Suite à l'évolution des formats multicanal, le dispositif s'est complexifié pour comprendre aujourd'hui environ quatre-vingts haut-parleurs disposés de manière circulaire, environnant le public et de façon frontale. Environ vingt concerts par an sont réalisés sur le dispositif dans des salles de volume différent ; plus de huit cents concerts ont été donnés depuis le début du cycle de concerts « Multiphonies » en 1978. Des outils complémentaires, les GRM Tools¹, permettent de contrôler et d'organiser les sons dans l'espace lors du processus de composition des œuvres.

1. <http://www.inagrm.com/grmtools>

L'Ircam, équipé depuis peu en WFS circulaire, fait aussi partie des lieux de diffusion spatialisée pour les œuvres de musique contemporaine. Dans *Répons*, Pierre Boulez utilise la technologie de pointe de l'Ircam, il allie un dispositif électroacoustique : vingt-quatre musiciens, six solistes ainsi que six haut-parleurs permettant d'amplifier, de spatialiser les sons et de modifier ceux des solistes. Le public est placé au centre.

À la fin du xx^e siècle, la période quadriphonique est marquée par la sortie de *Dark Side of the Moon* des Pink Floyd en 1973. En 1967, Pink Floyd donne le premier concert multicanal, en format quadriphonique, à la salle Queen Elizabeth de Londres. Par la suite, s'enchaînent des expérimentations musicales en Dolby Surround, des albums 5.1 édités en CD et DVD DTS dont notamment *Hell Freezes Over* des Eagles, une des premières exploitations commerciales en CD DTS. Puis, l'évolution des supports, avec l'apparition du DVD audio et du SACD, donne un véritable élan à la production musicale 5.1 : des artistes s'y consacrent tels que Björk (albums SACD, DVD hybride et live DVD / Blu-ray), Peter Gabriel (albums SACD et live DVD / Blu-ray), les Beatles (album *Love* produit en DVD audio), Diana Krall (SACD, DVD audio), etc. Cet élan, qui motive quelques labels à investir dans le SACD 5.1, est éphémère. La crise du CD freine alors les éditions en 5.1, notamment pour la production de musiques actuelles. Aujourd'hui, le monde de la musique classique continue à produire en 5.1 avec certains artistes renommés comme Jordi Savall : les enregistrements sont édités en SACD, DVD hybride ou Pure Audio Blu-ray par des labels spécialisés comme Alia Vox, Aix records ou 2L (Lindberg Lyd). Depuis 2011, la plateforme d'achat en ligne Qobuz propose des éditions musicales classique et jazz en 5.1. L'année 2012 marque une nouvelle étape avec la sortie du premier Blu-ray 9.1 Auro-3D intégrant un son 3D, il s'agit de TrondheimSolistene – Souvenir, édité par 2L. Auro Technologies commercialise les premiers décodeurs Auro-3D sur les amplificateurs AV fin 2013.

En 2012 a lieu l'inauguration de l'installation WFS de l'espace de projection de l'Ircam. La couronne est équipée de deux cent soixante-quatre haut-parleurs et complétée par un dôme de soixante-quinze haut-parleurs pour une diffusion tridimensionnelle en mode ambisonique.

1.3 Intérêts

1.3.1 Une écoute naturelle : réalisme et immersion

L'historique nous a révélé qu'il existait de nombreux dispositifs de spatialisation avec des champs d'applications diverses, mais c'est bien le monde du cinéma qui a réussi à faire adopter plus généralement la spatialisation du son. Les autres applications du multicanal sont très peu connues et gagnent à l'être, notamment grâce à la simplification de la diffusion.

Souvent, on compare le son multicanal au son stéréo : la stéréo c'est regarder à travers une fenêtre et le multicanal c'est être immergé dans un champ sonore.

HISTORIQUE, ÉCRITURE, ÉDITION ET DIFFUSION MUSICALES

- | | | | |
|------|--|------|---|
| 1570 | <i>Spem in alium</i> , Thomas Tallis, motet à 40 voix, 8 chœurs à 5 voix (soprano, alto, ténor, baryton, basse) circulaires placés dans une salle de château octogonale | 1989 | Premier enregistrement de musique en Dolby Surround, The Home Video Album, distribué en CD et en cassette audio par RCA Victor |
| 1585 | <i>Sacrae Symphoniae</i> , Giovanni Gabrieli, double chœur, musique polyphonique spatialisée, exécutée dans la basilique Saint-Marc à Venise qui comprend alors 2 tribunes opposées recevant chacune 1 chœur | | <i>Hay que caminar</i> , Luigi Nono, œuvre pour 2 violons, musiciens en mouvement
<i>Le noir de l'étoile</i> , Gérard Grisey, œuvre pour 6 stands de percussions, public entouré par les musiciens |
| 1830 | <i>Symphonie Fantastique</i> , Hector Berlioz, les cuivres sont disposés derrière les spectateurs, une instruction précise : un hautbois est placé hors scène, impliquant une certaine distance | 1993 | Premier enregistrement rock mixé en Dolby Surround pour le titre <i>Pictures at an Exhibition</i> , dernière plage de l'album <i>The Return of the Manticore</i> , d'Emerson, Lake & Palmer |
| 1837 | <i>Requiem</i> , Hector Berlioz, 4 orchestres disposés Nord, Est, Sud et Ouest | 1995 | Édition de l'album <i>Hell Freezes Over</i> , The Eagles, en CD DTS |
| 1894 | <i>Symphonie n° 2</i> , Gustav Mahler, sur la partition est mentionné : « les trompettes sont placées hors scène, gauche droite, le plus loin possible » | 1997 | Le DVD <i>Spectacular</i> est le premier titre DVD audio, distribué par Delos International, avec un son en Dolby Digital 5.1 |
| 1908 | <i>The Unanswered Question (A Cosmic Landscape)</i> , Charles Ives, distribution spatiale des instruments | 2001 | <i>Vespertine</i> , premier album de Björk écrit pour le 5.1, édition SACD |
| 1923 | <i>Hyperprism</i> , Edgar Varèse, 16 percussions et 9 instruments à vent, première tentative chez Varèse de spatialiser la musique | 2006 | <i>Love</i> , nouvel album de musiques compilées et remixées des Beatles, DVD audio en 5.1, produit par George et Giles Martin en collaboration avec le Cirque du Soleil. Le spectacle est sonorisé par 6 341 haut-parleurs à la salle The Mirage, à Las Vegas. |
| 1958 | <i>Gruppen</i> , Karlheinz Stockhausen, 109 musiciens divisés en 3 groupes quasi identiques
<i>Poème électronique</i> , Edgard Varèse, disposition de plus de 400 haut-parleurs à l'intérieur du pavillon Philips conçu par Le Corbusier et réalisé par Iannis Xenakis pour l'Exposition universelle de Bruxelles de 1958 | 2006 | Sortie de la discographie <i>Surrounded</i> de Björk en 5.1, coffret des albums, titres studio et vidéo clips en stéréo et 5.1 |
| 1966 | <i>Terretêktorh</i> , Iannis Xenakis, 88 musiciens dispersés dans le public | 2009 | <i>Divertimenti</i> , orchestre de chambre <i>TrondheimSolistene Lindberg Lyd</i> , édition SACD/Pure Audio Blu-ray, 2L
<i>Piano Concerto</i> , Edvard Grieg, un des premiers Blu-ray 7.1, édité par le label 2L |
| 1966 | <i>Terretêktorh</i> , Iannis Xenakis, 88 musiciens dispersés dans le public | 2010 | <i>This is it</i> , documentaire musical sur la préparation de la tournée 2009 de Michael Jackson, édition DVD/Blu-ray |
| 1973 | <i>Dark Side of The Moon</i> , Pink Floyd, format quadriphonique
Création du premier orchestre de haut-parleurs au GMEB à Bourges | 2012 | <i>TrondheimSolistene – Souvenir</i> , premier Blu-ray 9.1 Auro-3D, 5.1, 7.1, stéréo, édité par 2L |
| 1974 | Constitution de l'Acousonium du GRM par François Bayle (environ 80 projecteurs sonores, répartis sur 16 à 24 canaux) | | Inauguration de l'installation circulaire WFS de l'espace de projection de l'Ircam, 264 haut-parleurs régulièrement répartis sur une couronne, complétés par un dôme de 75 haut-parleurs, diffusion tridimensionnelle en mode ambisonique |
| 1981 | <i>Répons</i> , Pierre Boulez, ensemble instrumental (cordes, bois, cuivres), instruments solistes (deux pianos, harpe, vibraphone, xylophone, glockenspiel, cymbalum) et système électroacoustique pour transformer et spatialiser le son des solistes, composé de 6 enceintes | 2015 | Sortie du titre <i>Stonemilker</i> tiré de l'album <i>Vulnicura</i> de Björk, en version 3D binaurale. |

Dans la réalité, la fenêtre n'existe pas, on entend toujours un son spatialisé : une ambiance en extérieur ou une acoustique de salle ou de pièce. C'est pourquoi, restituer une ambiance en multicanal a toujours son effet persuasif lors de la première écoute et la comparaison avec la stéréo révèle un véritable appétit pour l'espace. La spatialisation permet aussi la fidèle répartition des sources dans l'espace sans repliement, alors que la stéréo impose le frontal, avec superposition du champ arrière à la scène frontale. Donc le premier intérêt du multicanal est d'enrichir la représentation mentale que l'on se fait pendant l'écoute, avec une sensation d'espace plus forte. L'auditeur est immergé dans le champ sonore, comme s'il se retrouvait sur le lieu du tournage, au concert, au théâtre, dans la scène du film, sur le stade ou dans la peau du personnage du jeu vidéo. Et de ce fait, il devient acteur.

1.3.2 *Un espace : de nouvelles possibilités d'écriture*

L'espace en multicanal ouvre de nouvelles perspectives dans l'écriture de l'œuvre. Le réalisateur peut introduire plusieurs points de vue, avec plusieurs actions pouvant se dérouler simultanément à des endroits différents. C'est aussi l'opportunité d'introduire le mouvement et de le développer, de faire interagir plusieurs directions tout autour de l'auditeur, de renforcer l'interaction avec ce dernier par des effets d'internalisation et des contenus basse fréquence.

Ce nouveau champ est à la fois une grande liberté mais il est soumis à des règles, d'autant plus contraignantes dans les rapports avec l'image. La spatialisation du son trouve son intérêt pour des programmes audiovisuels dont la réalisation sonore est pensée en multicanal dès le début. Avant de commencer à produire, il faut écrire pour l'espace.

1.3.3 *Définition d'image, angle de vision, distance et dispositif sonore*

Le schéma de la **figure 1.2** a été réalisé d'après des publications de la NHK : il représente l'évolution du format de l'image et sa configuration audio. Il part de la télévision numérique SD (son stéréo) définie par une image de 720 par 576 pixels, vue à une distance de quatre à huit fois la hauteur de l'écran et sous un angle de vision de 15°, jusqu'au format Super Hi-Vision développé par la NHK, d'une définition d'image de 8 k, imposant un angle de vision de 100° et accompagné du format son 22.2.

La TVHD se situe entre les deux configurations. L'image de définition 2K est vue sous un angle de 30°, à une distance de 3H (H étant la hauteur de l'image) avec un dispositif son normé 5.1. Le cinéma numérique projette des images de définition 4K, avec un angle de vision de 55° et à une distance de 1,5 fois la hauteur de l'écran. Son dispositif audio est le 5.1 mais on trouve d'autres formats supérieurs.

L'angle de vision augmente, il suit le format de l'image et plonge le spectateur dans un champ visuel de plus en plus réaliste et immersif tout comme le dispositif sonore, contraint alors de suivre l'image afin de garder une cohérence. Le nombre de canaux du dispositif audio croît avec l'angle de vision du dispositif image.

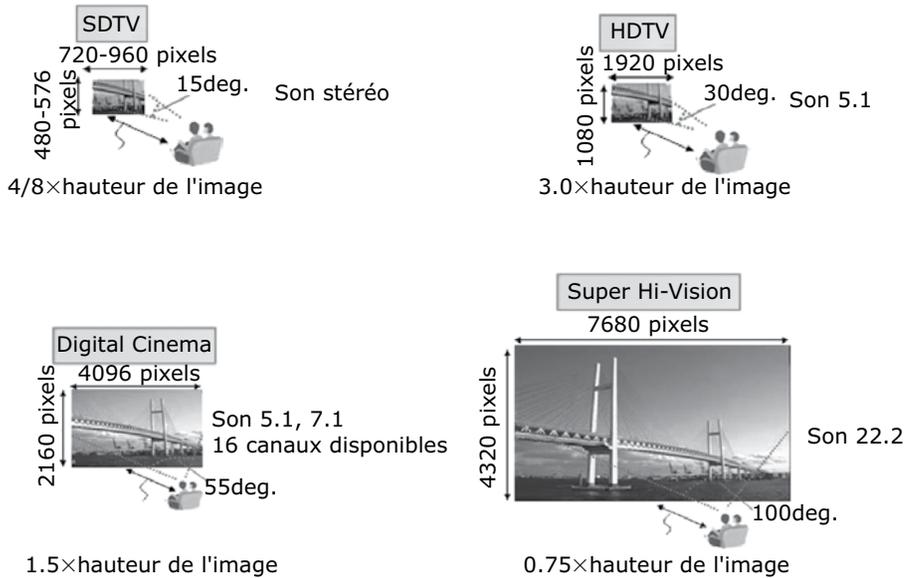


Figure 1.2 – Évolution de la relation définition d'image, angle de vision, distance de vision et dispositif de diffusion sonore, © B. Périaux d'après une publication NHK.

1.3.4 Meilleure intelligibilité, extension de la bande passante, interaction avec le corps

Le format 5.1 apporte une meilleure intelligibilité du signal grâce à une répartition des sources sur six canaux. La voix, par exemple, peut être positionnée dans le centre avec très peu d'effet de masque : les ambiances, les effets, les bruitages et la musique occupent toutes les directions. Le subwoofer étend la bande passante vers le bas. Le grave est alors entendu et ressenti aussi par le corps de l'auditeur.

1.4 Vecteurs

Il existe aujourd'hui de nombreux vecteurs du son multicanal, le principal et le plus ancien étant le cinéma.

1.4.1 Le cinéma : passage au numérique, installations son 3D Barco Auro-3D et Dolby Atmos

La plupart des films sortent en salle au format 5.1 en Dolby ou DTS et 7.1 pour le format Sony SDDS. Le passage au cinéma numérique relance le marché par les nouvelles technologies et ravive l'intérêt pour la spatialisation. Seize canaux audio ont été pensés pour la norme DCP (Digital Cinema Package) qui équipe les salles numériques. La diffusion du son se fait alors directement en fichiers PCM.

Le marché de l'image 3D relance aussi le développement du multicanal en imposant son nouveau format 7.1 déployant quatre canaux sur la couronne de haut-parleurs surround. Les autres technologies comme la WFS, l'Auro-3D, le Dolby Atmos surpassent les normes et multiplient les canaux. La salle de cinéma de demain sera peut-être équipée de vingt, trente ou quarante canaux, Dolby Atmos limitant le rendu à soixante-quatre haut-parleurs.

1.4.2 *La TVHD, l'UHDTV et le Super Hi-Vision*

La TVHD est lancée entre 2000 et 2005, elle utilise la technologie Dolby. Certaines chaînes HD diffusent en Dolby Digital 5.1 des programmes de sport, des films, des documentaires ou des concerts. Le sport a réellement été un programme clef pour le multicanal depuis les jeux olympiques d'Athènes en 2004, diffusés en 5.1. La HD et le 5.1 ont relancé le marché de la télévision avec la vente d'écrans plats HD et de home-cinéma. Le film peut maintenant être diffusé en qualité cinéma sur une chaîne HD. Les écrans de télévision de plus en plus fins ne permettent plus d'intégrer un son de bonne qualité, si bien qu'il devient nécessaire d'investir dans un home-cinéma et utiliser l'interface HDMI, pratique pour véhiculer le son. Aujourd'hui en expérimentation en France, la TV 3D sollicite la création de chaînes 3D telles qu'Orange 3D ou Canal Plus 3D (Arte se lance aussi dans la production de programmes 3D), bien que l'intérêt du particulier pour la stéréoscopie chez lui ne soit pas encore concret. La raison est certainement la fatigue éprouvée par le spectateur et son manque de confort car la notion de profondeur dans l'image est sans aucun doute acceptée par tous comme une amélioration du rendu visuel. L'UHDTV (Ultra High Definition TV) utilise un format 4K pour l'image et des études sur son format audio 9.1 ou 10.1 sont en cours. L'UHDTV 8K ou Super Hi-Vision s'accompagne du 22.2. La NHK prévoit des essais de diffusion audio 3D 22.2 pour les jeux olympiques de 2016 afin de valider une diffusion satellite pour 2020. L'ATSC 3.0 étudie également les possibilités de diffusion d'un son 3D.

1.4.3 *La radio numérique*

La radio numérique permet de diffuser le son en 5.1 à travers le DAB, le DAB+ et le DMB. En Europe, le DAB+ est le standard le plus probable. En France le DAB+ semble en bonne voie bien que le DMB reste une technologie possible. Certains pays européens comme la Suisse, le Danemark, la Suède, l'Angleterre ou la Norvège diffusent en DAB+. Certains constructeurs de voitures équipent leurs véhicules de récepteurs DAB.

1.4.4 *Les supports Blu-ray, DVD, SACD*

Les supports DVD et Blu-ray représentent un autre marché pour le multicanal, ils proposent généralement un son en 5.1. Le film a lancé le marché du DVD suivi par le live. Les supports son uniquement en 5.1 pour la musique sont très peu connus du public et seuls quelques éditeurs de musique classique et de jazz sortent aujourd'hui des SACD, DVD ou Pure Audio Blu-ray en 5.1. Une version 4K du Blu-ray, l'ultra HD

Blu-ray, est prévue pour la fin 2015 et différents codecs audio 3D comme le Dolby Atmos, l'Auro-3D, le DTS-UHD ou le MPEG-H 3D pourraient être intégrés.

La voiture est aussi un marché en développement pour ces supports : des installations en 5.1 sont proposées avec les formats Dolby, DTS ou Auro-3D. En 2015, les premiers modèles haut de gamme sont équipés en son 3D, c'est le cas de Audi pour son Q7, BMW X5, Mercedes ML Class ou Volvo XC90.

1.4.5 *Le jeu vidéo*

Le jeu vidéo profite de l'évolution technologique des supports. Les rendus sonores des moteurs de programmation des jeux sont réalisés en 5.1. L'interactivité du jeu vidéo, notamment les jeux de rôles, rend la spatialisation du son évidente. Les PS3, PS4 et la Xbox sont compatibles avec le format 5.1, généralement en Dolby Digital (la Wii est stéréo et compatible Dolby Surround). On trouve le plus souvent une version Dolby Digital et DTS sur les jeux PS4.

1.4.6 *La salle de spectacle, le théâtre, la retransmission d'opéras au cinéma*

La sonorisation utilise de plus en plus fréquemment un canal de diffusion central et, dans certains spectacles, plus rares, des canaux arrière. Des artistes comme Yann Tiersen, Matthieu Chedid, Jean-Michel Jarre, ont déjà expérimenté la sonorisation de leurs concerts en multicanal. La retransmission d'opéras en direct dans les salles de cinéma devient de plus en plus fréquente, même si le rendu sonore n'est pas idéal. Néanmoins, de nouveaux systèmes de diffusion son 3D comme l'Auro-3D ou le Dolby Atmos apparaissent et améliorent considérablement la qualité de restitution. Par ailleurs, de nombreuses expérimentations en WFS ont marqué l'année 2011, notamment en France à l'Ircam, au théâtre du Châtelet, à l'Institut du Monde Arabe et à France Télévisions. La WFS permettra sans aucun doute d'améliorer la qualité du son dans ce type de lieux et peut être aussi dans les salles de cinéma. Ces technologies de spatialisation permettront de diffuser des programmes et des spectacles plus diversifiés, avec une meilleure compatibilité sur le contenu sonorisé (film, musique, théâtre).

1.4.7 *Internet*

Internet est aujourd'hui un lieu de convergence des technologies multimédias : ainsi se côtoient les chaînes de télévision, les radios, les plateformes de diffusion YouTube et dailymotion. De nouvelles formes d'écriture émergent telles que le web documentaire, le podcast, avec de nouveaux types de diffuseurs webradios et webTV. La plupart des chaînes de télévision proposent aujourd'hui la rediffusion de leurs programmes ainsi que le flux en direct sur leur site internet. Les radios proposent le même type de service grâce à des lecteurs multimédias sur leurs sites dont la base de données est alimentée par le flux en direct et les podcasts. Internet permet aussi plus facilement l'interactivité : le spectateur choisit alors son programme à travers sa navigation. Les vidéos HD, notamment les bandes annonces des films, ont lancé la diffusion Internet en 5.1 sous forme de vidéos QuickTime HD disponibles en téléchargement ou en streaming. Aujourd'hui il

existe même des webradios qui diffusent en 5.1 : NouvOson, Concertzender, Antenne Radio, Rock Antenne, les radios suédoises et norvégiennes mettent à disposition leurs programmes sous forme de fichiers Dolby Digital ou DTS téléchargeables. Il faut savoir que les applications comme iTunes, QuickTime, VLC player ou Windows Media Player sont compatibles 5.1. Un simple lien vers ces applications permet de lire un fichier 5.1. Depuis peu, le langage HTML5 est directement compatible avec les fichiers 5.1. Il suffit d'envoyer le flux constitué des six canaux vers un ampli home-cinéma via une interface HDMI ou une carte son pour profiter du son 5.1. Aussi, depuis que les écrans de télévision se connectent au réseau Internet, la diffusion de contenus HD avec un son 5.1 par l'interface HDMI devient directement accessible.

1.4.8 *Le téléphone mobile, la tablette numérique*

Enfin, le téléphone mobile et la tablette numérique peuvent diffuser un son 5.1 par le biais de l'interface HDMI. Ce sont des lecteurs mobiles qui contiennent nos bibliothèques de photos, de sons, de vidéos, de films. Le particulier peut diffuser par exemple un film en HD avec un son AAC ou Dolby Digital grâce au procédé Dolby mobile depuis son téléphone portable, sur iOS ou Android, et le diffuser directement en 5.1 via l'interface HDMI. Progressivement, des procédés de son binaural sont proposés dans ces lecteurs multimédias. On peut imaginer que, dans une dizaine d'années, l'écoute au casque sera sans doute spatialisée.

1.5 **Problématiques**

L'apprentissage du multicanal met en évidence plusieurs problématiques : la maîtrise du système d'écoute, la définition d'une image sonore en multicanal, des concepts nouveaux de mixage et de prise de son, la gestion de metadata pendant le mixage telle que le downmix stereo, la mesure du loudness (aujourd'hui présente aussi pour la stéréo), les profils de compression Dolby, et la diffusion.

1.5.1 *Le système d'écoute*

La maîtrise du système de monitoring est primordiale car l'exploitant doit pouvoir le diagnostiquer assez rapidement avant de démarrer son travail.

1.5.2 *L'image sonore multicanal*

L'image sonore multicanal est une nouvelle écoute. Les programmes en 5.1 ont déjà trouvé leurs points de repères : le film a développé son langage, ainsi que le sport, la musique classique, la fiction radio, la fiction TV, le documentaire ou la musique acoustique. Tous les types de production qui se prêtent à la spatialisation ont aujourd'hui leur esthétique. Mais il existe bien des contraintes et des subtilités dans l'apprentissage de cette nouvelle image sonore qui seront abordées aux chapitres 3 et 4.

1.5.3 *La prise de son spatialisée*

La captation multicanal utilise des systèmes composés de quatre ou cinq microphones. Ces derniers permettent de prendre une photo très réaliste de l'environnement : ils découpent l'espace en différents secteurs et utilisent, pour certains, des procédés de matricage souvent complexes. Leur déploiement dans les productions est souvent difficile, ce qui nécessite une méthode de travail particulière en fonction des différentes applications.

1.5.4 *Des concepts de mixage nouveaux*

De nouveaux concepts de mixage s'imposent. Le mixeur utilise des outils de spatialisation, des outils de management de pistes ou d'objets multicanal, il gère des réverbérations multicanal, des algorithmes multicanal de traitement dynamique et des procédés de downmix et d'upmix. Par ailleurs, le mixage objets apparaît à travers les procédés de diffusion son 3D pour la salle de cinéma ou la WFS. C'est encore une nouvelle approche de mixage.

1.5.5 *La diffusion, les metadata, le mixage Dolby*

La diffusion de metadata type Dolby pour les chaînes TVHD demande au mixeur un travail supplémentaire. Ce dernier doit les renseigner pendant le mixage : il mesure alors le loudness du programme, simule le downmix stereo et optimise le traitement dynamique par le biais des profils de compression DRC (Dynamic Range Control). Ces metadata accompagnent le transport du signal dans le Dolby E jusqu'à la diffusion Dolby Digital. On utilise aussi le DTS ou le Dolby pour les supports DVD/Blu-ray et d'autres codecs types AAC pour la diffusion radio DAB+ et le streaming sur Internet. Il existe aujourd'hui un grand nombre de procédés qui permettent de diffuser un son multicanal adapté à la TV, la radio, au support, à Internet ou à la salle de cinéma. On assiste donc à une réelle évolution des métiers, associée à de nouvelles compétences.

1.6 **Les différentes techniques de spatialisation**

Il existe différentes techniques de spatialisation qui ont toutes leurs avantages et leurs inconvénients. On peut les classer en quatre familles : la Stéréophonie ou Multichannel Stereo, les techniques ambisoniques, la WFS (Wave Field Synthesis) et les techniques binaurales.

1.6.1 *Stéréophonie ou Multichannel Stereo*

On définit par Stéréophonie ou Multichannel Stereo (« channel oriented » ou « direct speaker »), les techniques qui reposent sur une décomposition de l'espace en canaux discrets. Chaque canal est diffusé directement sur un haut-parleur (certains formats diffusent le même canal vers une couronne d'enceintes, pour la salle de cinéma, par

exemple). On trouve alors généralement une correspondance exacte entre le nombre de canaux du mixage et le nombre de haut-parleurs de diffusion. Dans ce groupe on classe tous les formats discrets diffusés en stéréo, 4.1, 5.1, 6.1, 7.1, 9.1, 10.2, 11.1, 13.1, 22.2 etc.

1.6.2 L'Ambisonie (« scene oriented »)

Les techniques ambisoniques (« Ambisonics ») reposent sur une décomposition mathématique de l'espace en harmoniques sphériques. Ces harmoniques vont de l'ordre 1 à des ordres plus élevés comme l'ordre 3, 4 ou 5. Cette technique encode le son en composantes spatiales suivant l'axe x, y et z du repère cartésien, appelées format B, indépendantes du dispositif de restitution. Les signaux du format ambisonique n'alimentent pas directement les enceintes, mais transportent l'information directionnelle d'un champ sonore, c'est pourquoi on utilise le terme « scene oriented ». Le format ambisonique d'ordre 1, appelé format B Soundfield, contient les composantes bidirectionnelles du repère cartésien de l'espace X, Y, Z et l'information omnidirectionnelle W. De ce format codé, on peut extraire un format 5.1, 6.1, 7.1, etc. Les ordres supérieurs (High Order Ambisonics) apporteront une meilleure résolution spatiale dans la restitution.

1.6.3 La WFS

Les techniques WFS (Wave Field Synthesis) reposent sur la décomposition du champ sonore en une multitude de sources ponctuelles, elle est basée sur le principe physique d'Huygens. Les haut-parleurs multiples, agissant comme des sources sonores ponctuelles issues de microphones placés en ligne, reproduiraient alors une copie conforme du son d'origine dans la zone d'écoute. La restitution du champ sonore est faite par synthèse du front d'onde.

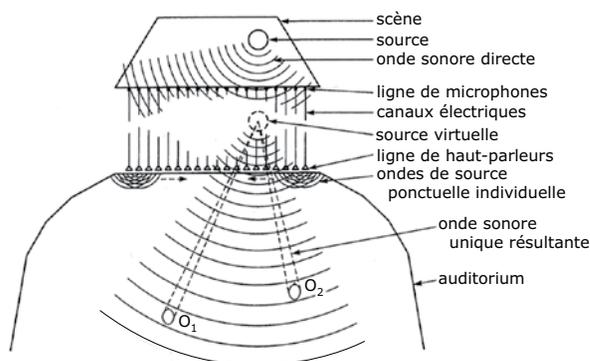


Figure 1.3 – Principe de la WFS, © Snow.

1.6.4 Le binaural

Les techniques binaurales permettent de reproduire un espace au casque stéréophonique. Il existe des techniques de prise de son binaurales utilisant des têtes artificielles, des mannequins ou le preneur de son lui-même, comme il existe des techniques de synthèse binaurale reposant sur les fonctions de transfert HRTF (Head-Related Transfer Functions) et sur la simulation d'un champ acoustique. Ces techniques sont mises à contribution pour les casques multicanal ou pour la diffusion multicanal compatible binaurale.

1.6.5 Les objets (« object oriented »)

Tout mixage est généralement porté par un format multicanal « stéréophonique » que l'on définit par un nombre de canaux de diffusion (par exemple six pour le 5.1). Ce type de spatialisation n'offre pas de compatibilité simple entre formats, car la manipulation de la scène sonore qu'il définit est limitée. On trouve cependant des solutions de downmix et d'upmix acceptables. Aujourd'hui, ce type de multicanal offre un bien grand nombre de formats. Se pose alors la question de compatibilité d'un programme spatialisé avec les différents dispositifs de diffusion. Dans l'approche objet, chaque source est définie par une piste et son information spatiale peut alors être positionnée dans l'espace dans n'importe quel format multicanal de diffusion. Ce procédé simplifie la représentation de l'espace mais impose au processeur de diffusion d'intégrer un module de spatialisation avec simulation de l'effet de salle ou bien de coder la réverbération par objets.

L'approche objet est utilisée dans la WFS et dans les nouveaux procédés de diffusion cinéma tels que le Dolby ATMOS ou le MDA. On la trouve aussi dans le codage spatial MPEG4, MPEG7, MPEG-H 3D.

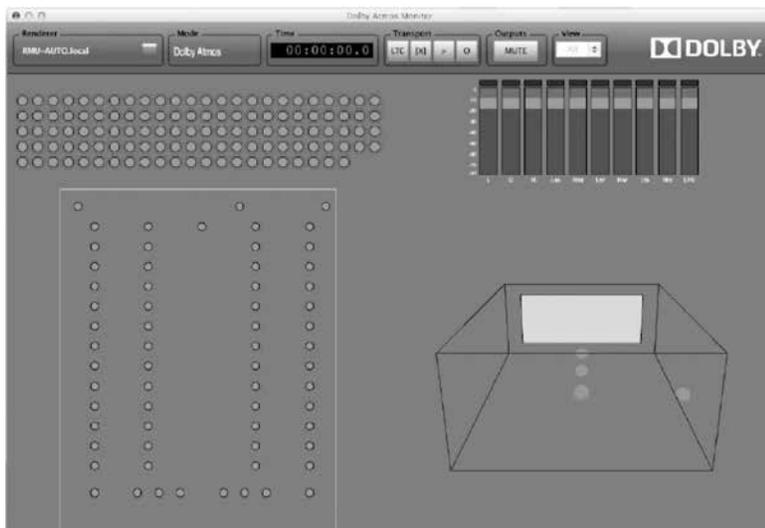


Figure 1.4 – Application Dolby Atmos Monitor, © Dolby. En couleur sur dunod.com

Le tableau suivant présente une synthèse des techniques de spatialisation.

Techniques de spatialisation	Les signaux enregistrés	Manipulation de la scène encodée	Dispositif de restitution
Stéréophonie, « channel oriented » (2.0, 5.1, 7.1, 10.2, Auro-3D 13.1, 22.2, etc.)	généralement 1 signal par enceinte	limitée, scène sonore figée	fixe, lié au format
Ambisonie, « scene oriented » (Soundfield, HOA) ordre m, composantes	3D : $(m+1)^2$ 2D : $2m+1$	globale (rotation, distorsion de perspective)	adaptable, indépendant du format
WFS	objets + metadata de localisation	chaque source peut être traitée de façon indépendante	synthèse du front d'onde dispositif adaptable, indépendant du format (objets)
Objets, « object oriented » (Dolby Atmos, MDA, MPEG-H 3D, MPEG 4, MPEG 7...)	objets + metadata de localisation	chaque source peut être traitée de façon indépendante	adaptable, indépendant du format
Binaural	2 signaux L, R	limitée, scène sonore figée	casque stéréo

1.7 Formats multicanal

Qu'appelle-t-on format multicanal ? Ici, il s'agit de définir le type d'installation sonore spatialisée, la configuration étant souvent liée à un procédé de codage. Ainsi, les codages 5.1 Dolby Digital ou DTS utilisent les configurations normées 5.1 ITU-R BS775-1 ou 5.1 cinéma définies au chapitre 2. Les industriels développent généralement leurs propres formats bien avant la normalisation : c'est le cas aujourd'hui, notamment avec le 7.1 qui ne correspond pas à une norme. C'est pourquoi on trouve de nombreuses dispositions d'enceintes. Deux types d'installations se distinguent : l'écoute cinéma et l'écoute broadcast/home-cinéma. Le tableau de la **figure 1.5** donne une vision de l'ensemble des configurations, de la quadriphonie à la WFS.

Nom du Format		Canaux de Diffusion <i>plan horizontal Fr/ Side/Rear Sub plan supérieur Fr/Side/Rear/Top plan inférieur Fr Sub</i>	Canaux de Mixage	Codage de diffusion
Quadriphonie 4.0	4.0 L R à ± 45° Ls Rs à ± 135°	2/2	LRLsRs	
5.1 ITU-R BS 775-1 & 2	Dolby Surround LtRt	3/2 + Sub	LCRS	Dolby Surround LtRt
	5.1 discret C à 0° L R à ± 30° Ls Rs à ± 110°	3/2 + Sub	LCRLsRsLfe	Dolby Digital, DTS, MPEG
5.1 option 4 enceintes arrière ITU-R BS 775-1 & 2	5.1 C à 0° L R à ± 30° Ls1/Ls2 Rs1/Rs2 entre 60° et 150°	3/4 + Sub	LCRLsRsLfe	
5.1 amélioré Günther Theile	5.1 Ls1/Ls2 Rs1/Rs2 entre 60° et 150°	3/4 + Sub	LCRLsRsLfe	
5.1 Cinéma	5.1 Ls et Rs sur couronne arrière	3/2 + Sub	LCRLsRsLfe	Dolby Digital, DTS
6.1 Cinéma	6.1 Cs centre arrière	3/2/1 + Sub	LCRLsRsCsLfe	Dolby Digital EX, DTS ES
6.1/7.1 IMAX Cinéma	6.1 / 7.1 Ls et Rs sur couronne arrière FH hauteur frontal	3/2/1 + Sub 1	LCRLsRsCsLfe FH	
7.1 SDDS Cinéma	7.1 5 canaux en frontal	5/2 + Sub	LLcCRcRLsRsLfe	SDDS
7.1 Cinéma 3D	7.1 2 canaux latéraux Ls Rs 2 canaux arrière BsL BsR	3/2/2 + Sub	LCRLsRsBsLsRLfe	Dolby Surround 7.1
7.1 Blu-ray Dolby 7.1 Dolby IIx	7.1 latéraux Ls Rs à ± 90° Lb et Rb à ± 150°	3/2/2 + Sub	LCRLsRsLbRbLfe	Dolby Digital TrueHD Dolby Surround IIx

Figure 1.5 – Les différents formats en multicanal (Suite).

Nom du Format		Canaux de Diffusion <i>plan horizontal Fr/ Side/Rear Sub plan supérieur Fr/Side/Rear/Top plan inférieur Fr Sub</i>	Canaux de Mixage	Codage de diffusion
7.1 Blu-ray DTS	7.1 Wide Lw Rw à ± 60°	5/2 + Sub	LCRLwRwLsRsLfe	DTS Master Audio
	7.1 Side Lss Rss à ± 90°	3/2/2 + Sub	LCRLssRssLsrRsrLfe	DTS Master Audio
	7.1 Front Height LH RH à ± 30°	3/2 + Sub 2	LCRLsRsLfe LHRH	DTS Master Audio
	7.1 Height Side LHS RHS à ± 90°	3/2 + Sub 2	LCRLsRsLfe LHSRHS	DTS Master Audio
	7.1 110/150 Side à ± 110° Ls Rs à ± 150°	3/2/2 + Sub	LCRLssRssLsrRsrLfe	DTS Master Audio
7.1, 9.1, 11.1 selon Mike Williams	5.1 ou 7.1 75/120 Hauteur 4 canaux à +45° d'élévation 0°, -90°, +90°, 180°	3/2/2 + Sub 1/2/1	LCRLmRmLs RsLfe Hc, Hl, Hr, Hb	
8.1 Musique acousmatique format octophonique	8.1 8 secteurs égaux de 45° LR à ± 45° Lss Rss à ± 90° Lsr Rsr à ± 135° CS à 180°	3/2/3 + Sub	LCRLssRssLsr RsrCsLfe	
9.1 Dolby Surround IIz	9.1 Side à ± 90° Front Height à ± 30° arrières à ± 150°	3/2/2 + Sub 2	LCRLsRsLfeLrsRrs LvHrVh	Dolby Surround IIz 9.1
Auro-3D Home-Cinéma/ Studio	9.1 Front Height à ± 30° Rear Height à ± 110° Tilt de 30°	3/2 + Sub 2/2	LCRLsRsLfe FlhFrhRlhRrh	Auro-3D Octopus
	10.1 Front Height à ± 30° Rear Height à ± 110° Tilt de 30° Top TH	3/2 + Sub 2/2/1	LCRLsRsLfe FlhFrhRlhRrhTH	Auro-3D Octopus

Figure 1.5 – Les différents formats en multicanal (Suite).

Nom du Format		Canaux de Diffusion <i>plan horizontal Fr/ Side/Rear Sub plan supérieur Fr/Side/Rear/Top plan inférieur Fr Sub</i>	Canaux de Mixage	Codage de diffusion
Auro-3D Cinéma	11.1 hauteur centre frontal HC Top TH	3/2 + Sub 3/2/1	LCRLsRsLfe FlhHCFrhRlhRrh TH	Auro-3D Octopus
	12.1 arrière centre CS	3/2/1 + Sub 3/2/1	LCRLsRsCSLfe FlhHCFrhRlhRrh TH	Auro-3D Octopus
	13.1 hauteur arrière centre HRC	3/2/1 + Sub 3/2/1/1	LCRLsRsCS Lfe FlhHCFrhRlhRrh THHRC	Auro-3D Octopus
10.2 T. Holman	10.2 LH et RH à 45° horizontal tilt de 45°	5/2/1 + 2 Sub 2	LCRLwRwLsRs CSLLfeRLfe LhRh	
11.1 DTS Neo X	11.1 Lw Rw à ± 60° Lss Rss à ± 90° Ls Rs à ± 110° Lsr Rsr à ± 150° Lh Rh à ± 30°	5/2/2 + Sub 2	LCRLwRwLss Rss(ouLsRs)Lsr Rsr LhRh	DTS Neo X
22.2 NHK	22.2 3 couches horizontale supérieure	5/2/3	FLFLcFCFRcFR SiLSIRBLBCBR	
		3/2/3/1	TpFLTpFCTpFR TpSiLTpSiRTpBL TpBCTpBRTpC	
	inférieure	3 + 2 Sub	BtFLBtFCBtFR Lfe1Lfe2	
Imm Sound (Dolby) Cinéma	14.1 ou 23.1 plafond supérieure horizontale latéraux et arrière sur couronne inférieure	3/6/2 + 2 Sub 3/5 3		Mixage objets

Figure 1.5 – Les différents formats en multicanal (Suite).

Nom du Format		Canaux de Diffusion <i>plan horizontal</i> Fr/ Side/Rear Sub <i>plan supérieur</i> Fr/Side/Rear/Top <i>plan inférieur</i> Fr Sub	Canaux de Mixage	Codage de diffusion
Dolby ATMOS Cinema	64 HP Max 2 couronnes supérieures horizontale	2 rangées dans la longueur 5 frontal/côtés variable/arrières variable 1 Sub avant centre et 2 Sub arrière	Bed 7.1 : LRCLfeLssRssLsrRsr Bed OH : LTsRTs Objets 1-118	Mixage objets
Dolby ATMOS Audi montage, Local renderer	9.1		LRCLfeLssRss LsrRsrLtsRts	Mixage objets
	11.1 (1 side, 1 back, 2 top per side)		LRCLfeLssRss LsrRsrLts1Lts2 Rts1Rts2	
	11.1 (2 side, 1 back, 1 top per side)		LRCLfeLss1Lss2 Rss1Rss2LsrRsrLts Rts	
	13.1 (2 side, 1 back, 2 top per side)		LRCLfeLss1Lss2 Rss1Rss2LsrRsrLts1 Lts2Rts1Rts2	
	13.1 (2 side, 2 back, 1 top per side)		LRCLfeLss1Lss2 Rss1Rss2Lsr1Lsr2 Rsr1Rsr2LtsRts	
	15.1 (2 side, 2 back, 2 top per side)		LRCLfeLss1Lss2 Rss1Rss2Lsr1Lsr2 Rsr1Rsr2Lts1Lts2 Rts1Rts2	
	15.1 (3 side, 1 back, 2 top per side)		LRCLfeLss1Lss2 Lss3Rss1Rss2Rss3 LsrRsrLts1Lts2Rts1 Rts2	

Figure 1.5 – Les différents formats en multicanal (Suite).

Nom du Format		Canaux de Diffusion <i>plan horizontal Fr/Side/Rear Sub</i> <i>plan supérieur Fr/Side/Rear/Top</i> <i>plan inférieur Fr Sub</i>	Canaux de Mixage	Codage de diffusion
Dolby ATMOS Home	34 HP Max	24 couche medium 10 plafond		Mixage objets
	5.1.2			
	5.1.4			
	7.1.2			
	9.1.2			
ITU-R BS 2051 Advanced Sound Systems	49.2 3 couches	16 Couche inférieure 16 Couche medium 16 Couche supérieure + 1 top 2 Sub gauche et droite		
WFS Cinéma, Salle de spectacle	couronnes d'enceintes nombre très variable de HP entre 25 et 200 Couches medium et supérieure possibles			Mixage objets

Figure 1.5 – Les différents formats en multicanal.

1.7.1 Quadriphonie

Le format quadriphonique définit une configuration de quatre enceintes placées sur un carré. Les enceintes frontales sont disposées à $\pm 45^\circ$ et celles de l'arrière à $\pm 135^\circ$.

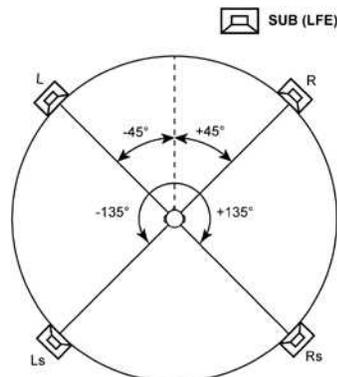


Figure 1.6 – Le format de la Quadriphonie.

1.7.2 5.1 ITU

La norme 5.1 Broadcast ITU-R BS 775-1 est préconisée pour la plupart des contenus 5.1 autres que ceux diffusés en salle de cinéma. Elle est utilisée par les principaux codages, Dolby Digital et DTS.

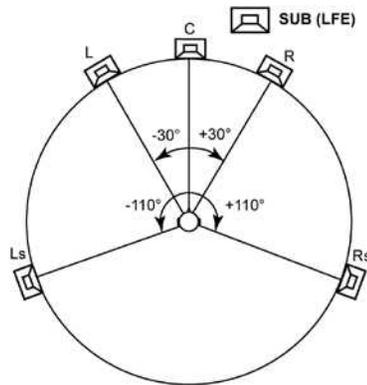


Figure 1.7 – La norme 5.1 Broadcast ITU-R BS 775-1.

1.7.3 LCRS Dolby Surround (TV, home-cinéma)

Le Dolby Surround est un codage à l'origine pour le format de mixage quatre canaux LCRS. Il utilise la norme 5.1 ITU-R BS 775-1 et est destiné à la télévision et au home-cinéma. Le canal de mixage S alimente les deux voies arrière Ls et Rs.

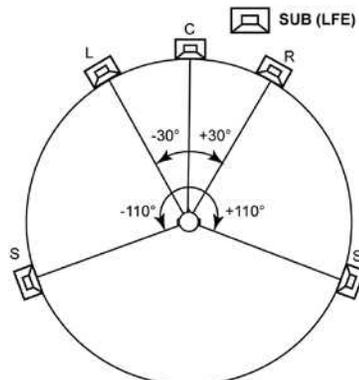


Figure 1.8 – Format Dolby Surround.

1.7.4 5.1 ITU 4 enceintes arrière

Il existe dans les normes ITU-R BS 775-1 et 2 une configuration optionnelle avec quatre enceintes à l'arrière. Les enceintes arrière, Ls1, Ls2, Rs1, Rs2, sont positionnées entre 60° et 150° , de part et d'autre du centre. Ces enceintes couvrent donc sur l'arrière deux zones de 90° . La norme mentionne les possibilités de réglage de décorrélation des enceintes ainsi que le délai entre elles, recommandés pour les grandes installations. En revanche, aucun matriçage entre les canaux Ls1/Ls2 et Rs1/Rs2 n'est proposé.

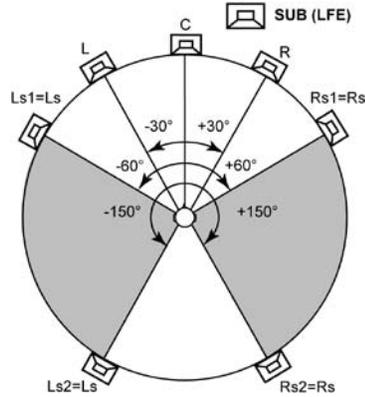


Figure 1.9 – Format 5.1 avec 4 enceintes à l'arrière, en option dans la norme ITU-R BS 775-1 & 2.

1.7.5 5.1 amélioré Günther Theile

Le 5.1 amélioré, préconisé par Günther Theile (IRT), reprend la configuration optionnelle : quatre enceintes arrière ITU-R BS 775-1, et utilise un matriçage entre les canaux L/Ls et R/Rs pour alimenter respectivement l'enceinte Ls1 et l'enceinte Rs1. Ce matriçage décorrèle les canaux arrière, il apporte ainsi un meilleur rendu spatial et augmente la zone d'écoute. Les quatre enceintes arrière pourraient être placées entre $\pm 60^\circ$ et 150° .

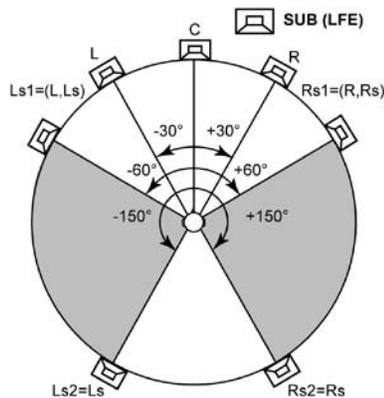


Figure 1.10 – Format 5.1 amélioré avec quatre enceintes à l'arrière préconisé par Günther Theile.

1.7.6 5.1 cinéma

La configuration 5.1 type cinéma (ISO 2969/1987(E) / SMPTE ST202-2010) dispose les trois enceintes frontales identiques, alignées derrière l'écran et une couronne d'enceintes pour le surround, dont le nombre diffère en fonction de la taille de la salle ou de l'auditorium cinéma.

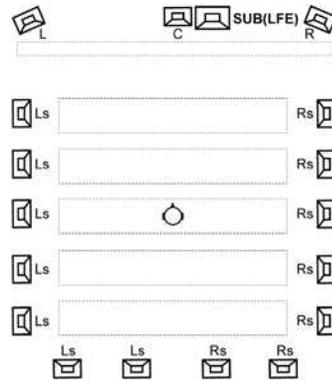


Figure 1.11 – Format 5.1 pour la salle de cinéma.

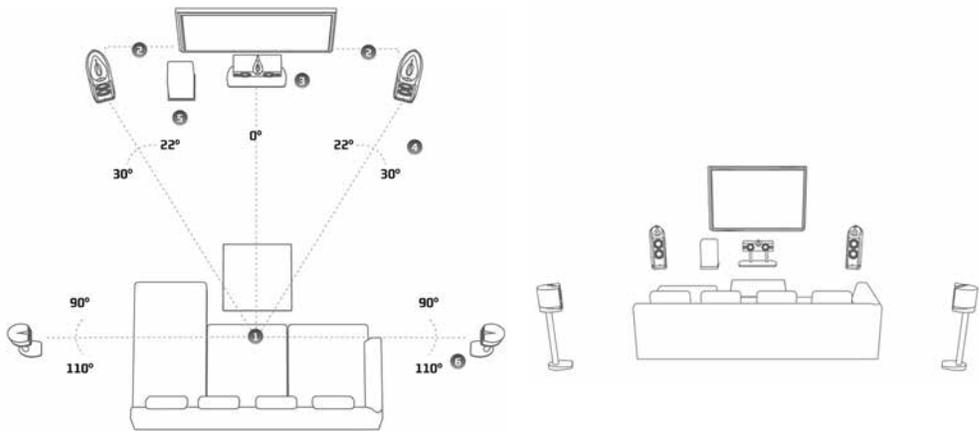


Figure 1.12 – Recommandation Dolby home-cinéma 5.1, © Dolby.

- (1) position d'écoute ; (2) enceintes frontales gauche et droite ; (3) enceinte centrale ;
 (4) angle des enceintes frontales gauche et droite ; (5) subwoofer ;
 (6) enceintes arrière gauche et droite.

1.7.7 5.1 cinéma Dolby Stereo

La diffusion LtRt en salle de cinéma utilise la même disposition d'enceintes dans la salle que la configuration 5.1 cinéma, les enceintes Ls et Rs étant alimentées par le canal surround S.

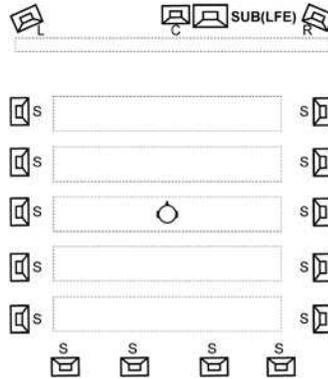


Figure 1.13 – Format 5.1 Dolby Stereo pour la salle de cinéma.

1.7.8 6.1 cinéma

La configuration cinéma 6.1 ajoute un canal central arrière supplémentaire, appelé CS (Center Surround). Cette configuration est utilisée par les codages Dolby Digital EX et DTS ES.

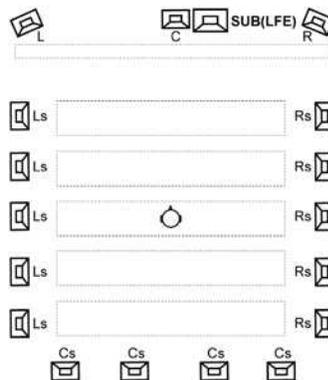


Figure 1.14 – Format 6.1 cinéma.

1.7.9 6.1, 7.1 IMAX

Le format IMAX est un format pour les grandes salles de cinéma, il utilise la configuration cinéma 5.1, à laquelle on ajoute un canal de hauteur H, intégré au plafond généralement dans le frontal et un canal arrière pour le 7.1 IMAX. La Géode est équipée en OMNIMAX avec un format 12.4 élaboré par la société Cabasse, il rajoute des enceintes au-dessus et en dessous des spectateurs.

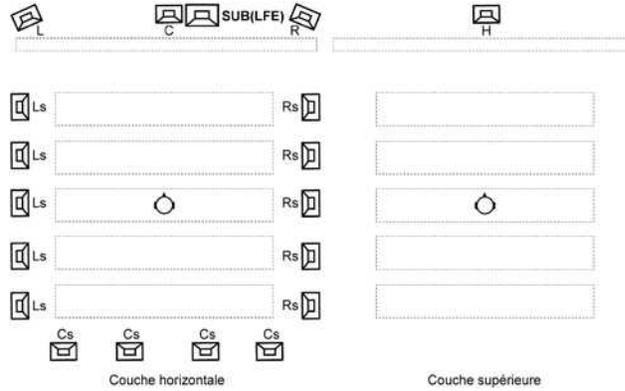


Figure 1.15 – Format 7.1 IMAX cinéma.

1.7.10 7.1 cinéma SDDS

Le SDDS est un format cinéma 7.1 qui ajoute deux enceintes supplémentaires frontales à $\pm 15^\circ$ de part et d'autre du centre. On obtient ainsi cinq enceintes identiques, alignées derrière l'écran. La couronne d'enceintes pour le surround ne change pas.

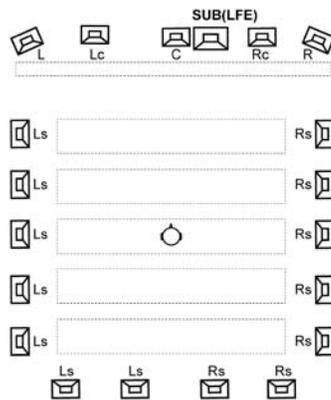


Figure 1.16 – Format 7.1 SDDS cinéma.

1.7.11 7.1 cinéma image 3D : DTS, Dolby Surround 7.1

Dolby et DTS préconisent une configuration 7.1 pour le cinéma avec image 3D. Elle utilise des canaux supplémentaires appelés Side (Side Left et Side Right) qui alimentent les enceintes surround à 90° . Les canaux surround BsL et BsR alimentent la partie arrière de la couronne surround.

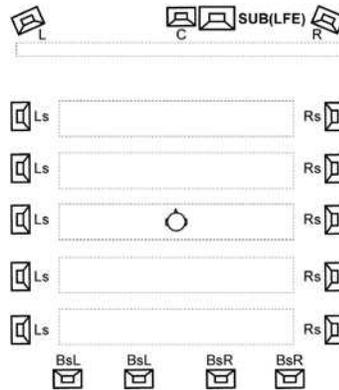


Figure 1.17 – Format 7.1 cinéma 3D.

1.7.12 7.1 Blu-ray Dolby

Dolby préconise une disposition 7.1 pour le Blu-ray qui reprend la norme ITU-R BS 775-1 en la modifiant pour disposer à l'arrière deux enceintes Side (Side Left et Side Right) Ls et Rs à $\pm 90^\circ$, et deux enceintes arrières BsL et BsR à $\pm 150^\circ$.

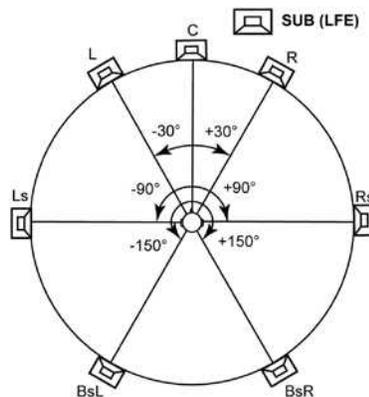


Figure 1.18 – Format 7.1 Blu-ray Dolby.

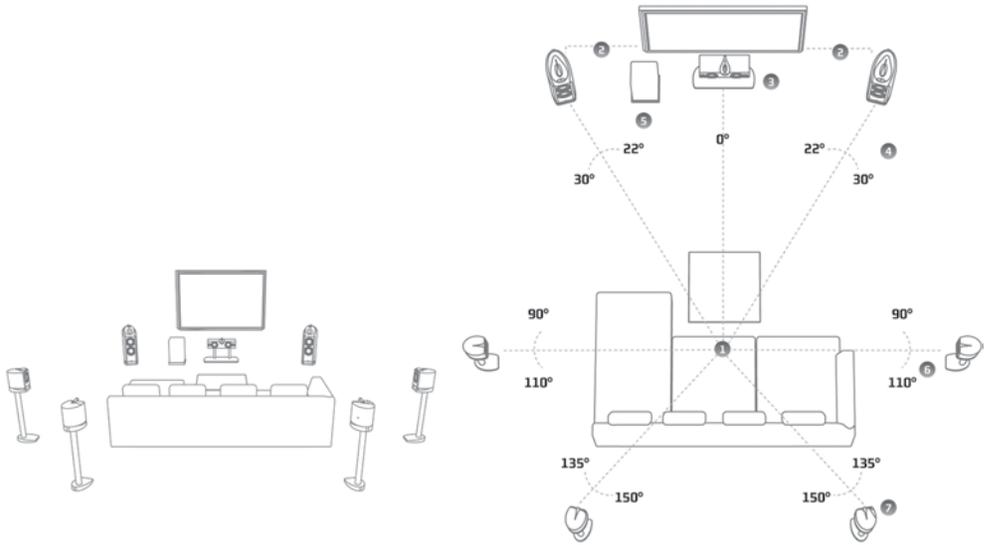


Figure 1.19 – Recommandation Dolby home-cinéma 7.1, © Dolby.

(1) position d'écoute ; (2) enceintes frontales gauche et droite ; (3) enceinte centrale ; (4) angle des enceintes frontales gauche et droite ; (5) subwoofer ; (6) enceintes arrière gauche et droite ; (7) enceintes arrière LB et RB.

1.7.13 7.1 Blu-ray DTS

DTS communique sur sept dispositions possibles pour le 7.1 du Blu-ray, dont trois principales : « Wide », « Side » et « Front Height » représentées sur le schéma 1.20.

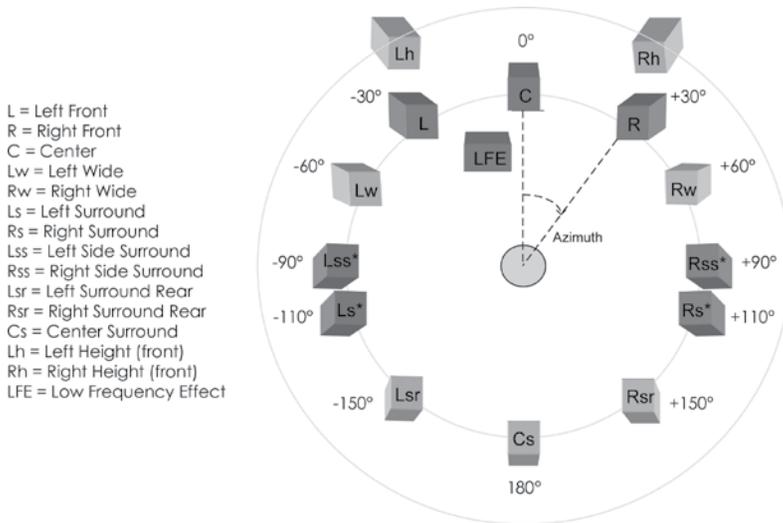


Figure 1.20 – Format 7.1 Blu-ray DTS, configurations principales, © B. Périaux d'après DTS

◆ **Wide**

La configuration Wide utilise la disposition ITU 5.1 et rajoute deux enceintes frontales larges, LW et RW, disposées à $\pm 60^\circ$.

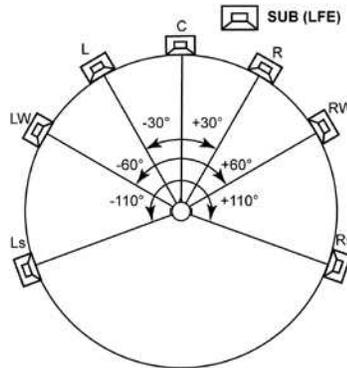


Figure 1.21 – Format 7.1 Blu-ray DTS, configuration Wide.

◆ **Side**

La configuration Side dispose deux enceintes Side, SL et SR, à $\pm 90^\circ$, accompagnées des enceintes surround Ls et Rs à $\pm 150^\circ$.

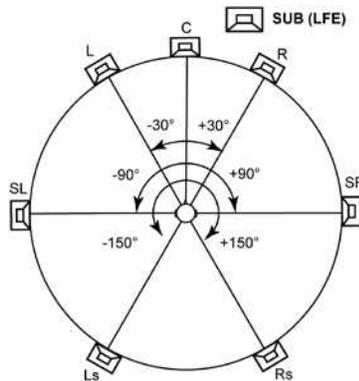


Figure 1.22 – Format 7.1 Blu-ray DTS, configuration Side.

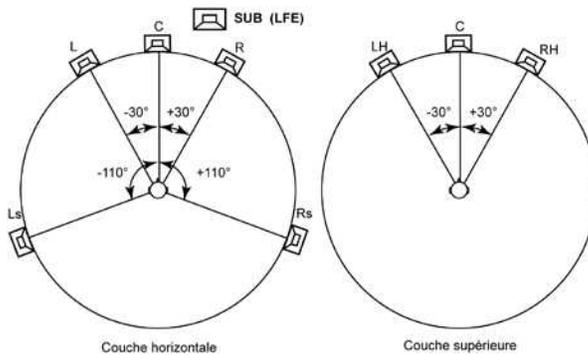


Figure 1.23 – Format 7.1 Blu-ray DTS, configuration Height.

◆ Front Height

La configuration Front Height reprend la norme 5.1 ITU à laquelle on ajoute deux enceintes de hauteur, LH et RH, disposées à $\pm 30^\circ$.

Il existe quatre autres configurations DTS 7.1 :

- la configuration disposant deux enceintes de hauteur sur le latéral, appelées Left Height Side et Right Height Side, à $\pm 90^\circ$, et complétées par la configuration ITU pour la couche horizontale ;
- la configuration 110°/150° pour les quatre canaux arrière ;
- la configuration ITU avec un Centre arrière CS et une enceinte de hauteur juste au-dessus de la tête appelée COH (Center Over Head) ;
- la configuration ITU avec une enceinte de hauteur dans le frontal, appelée CH (Center Height) et un canal central Surround CS.

1.7.14 7.1, 9.1, 11.1 selon recommandation de Mike Williams

Dans la configuration 7.1 préconisée par Mike Williams, les quatre enceintes surround sont disposées à $\pm 75^\circ$ pour les Side (Lm et Rm) et à $\pm 120^\circ$ pour les canaux Ls

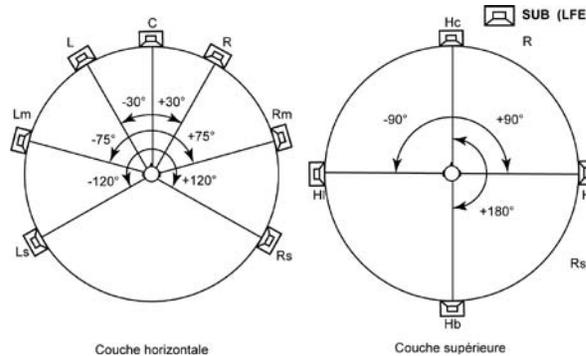


Figure 1.24 et 1.25 – Formats 7.1 75/120, 9.1, 11.1 préconisés par Mike Williams, photo d'une disposition 9.1 à l'INA, © B. Périaux.

et Rs. C'est une configuration qui donne une bonne homogénéité de l'image sonore, et qui évite ainsi la direction 90° des latéraux limitant la construction des images fantômes. Elle peut être complétée par quatre enceintes de hauteur d'angle d'élévation $+45^\circ$, Hc à 0° (Height Center), Hb à $+180^\circ$ (Height Back), Hl à -90° (Height Left) et Hr à $+90^\circ$ (Height Right). Cette disposition permet une restitution d'élévation sur les triangles isocèles formés par une enceinte de hauteur (le sommet du triangle) et les deux enceintes du bas correspondantes (la base du triangle). Cette disposition encore en expérimentation donne des résultats de localisation en hauteur performants.

1.7.15 8.1 Musique acousmatique octophonique

Les créations sonores de type musique acousmatique sont souvent spatialisées au format 8.1, constitué de huit enceintes réparties de façon homogène sur le cercle, tous les 45° . Les enceintes L, R sont à $\pm 45^\circ$, les latéraux à $\pm 90^\circ$, les enceintes Ls et Rs à $\pm 135^\circ$ et le centre arrière CS à 180° . C'est une disposition très symétrique mais sans compatibilité stéréo.

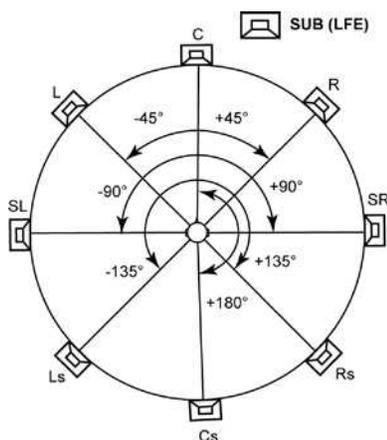


Figure 1.26 – Format 8.1 pour la musique acousmatique, appelé format octophonique.

1.7.16 7.1 Dolby Surround IIX, 9.1 Dolby Surround IIZ

Le Dolby Surround IIX utilise la configuration Blu-ray Dolby 7.1, c'est un matricage sur deux canaux, de type LtRt. Le Dolby Surround IIZ apporte la notion de hauteur au format 5.1 :

- en ajoutant deux enceintes Left Height (LH) et Right Height (RH), disposées au-dessus des enceintes Left et Right, ce qui constitue alors un format 7.1 ;
- il peut aussi intégrer les deux canaux latéraux placés à $\pm 90^\circ$, on parle alors de format Dolby Surround IIZ 9.1.

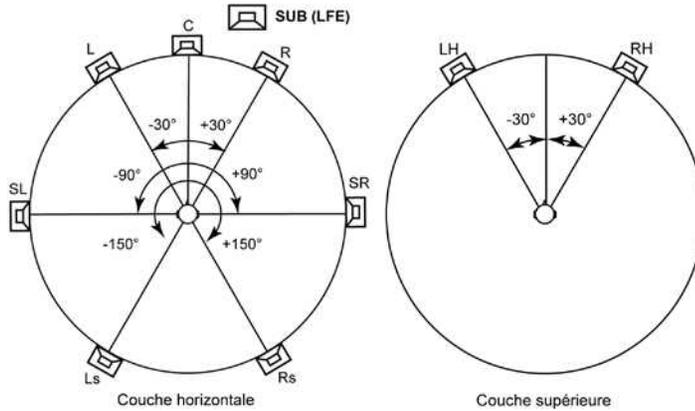


Figure 1.27 - Format Dolby Surround IIz 9.1.

1.7.17 Auro-3D 9.1 ou 10.1 home-cinéma/studio

La technologie Auro-3D a été développée par Wilfried Van Baelen (des Studios Galaxy). Elle intègre la notion de hauteur. La version home-cinéma offre deux possibilités, un format 9.1 construit à partir de la norme 5.1 ITU à laquelle on ajoute :

- quatre enceintes de hauteur, Left Height, Right Height, Rear Left Height Surround et Rear Right Height Surround, au-dessus des enceintes L, R, Ls, Rs (les enceintes de hauteur sont placées avec un angle d'élévation de 30° par rapport au plan horizontal) ;
- il existe une option pour le Top Height (TH), placé au plafond et centré au-dessus de l'auditeur, elle complète alors l'installation en 10.1.

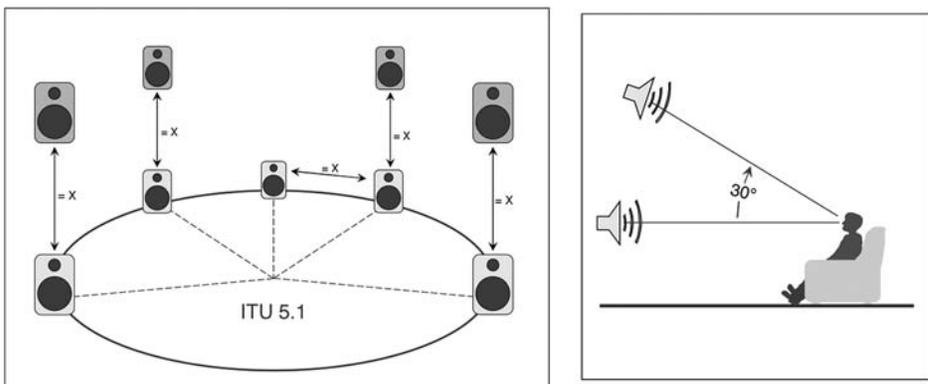


Figure 1.28 - Format Auro-3D 9.1, configuration, © Auro-3D.

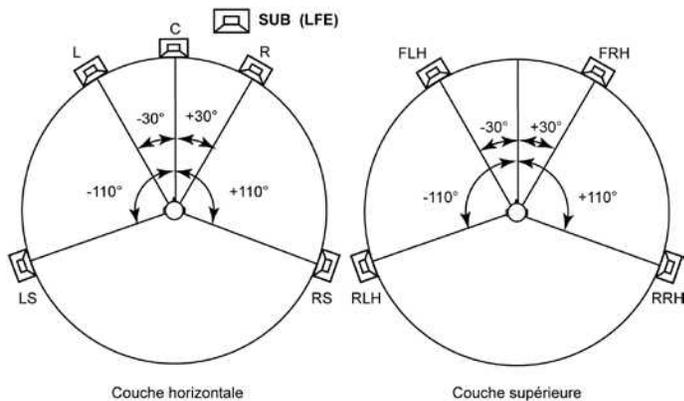


Figure 1.29 – Format Auro-3D 9.1 home-cinéma/studio.

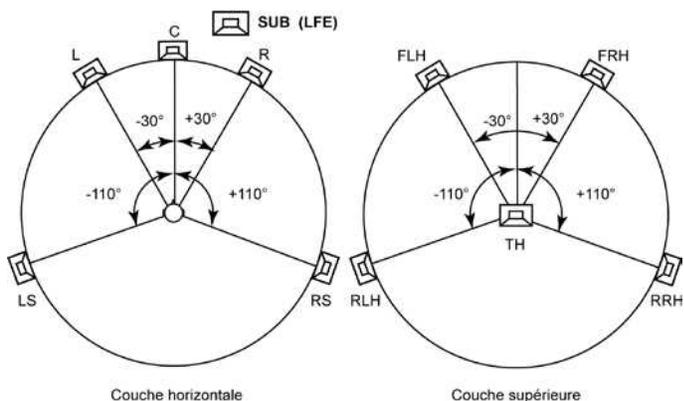


Figure 1.30 – Format Auro-3D 10.1 home-cinéma/studio.

1.7.18 Auro-3D 11.1, 12.1, 13.1 cinéma

Trois versions sont proposées pour la salle de cinéma :

- l'Auro-3D 11.1 : il ajoute l'enceinte Height Center (HC) placée en hauteur, centrée, dans le frontal ;
- l'Auro-3D 12.1 : celui-ci rajoute l'enceinte centrale arrière (Rear Center) ;
- l'Auro-3D 13.1 : complète l'installation avec une enceinte centrale arrière, en hauteur (Height Rear Center).

Il existe aujourd'hui quelques salles équipées. Le Dôme UGC La Défense est la première en France : elle est équipée en 20.1.

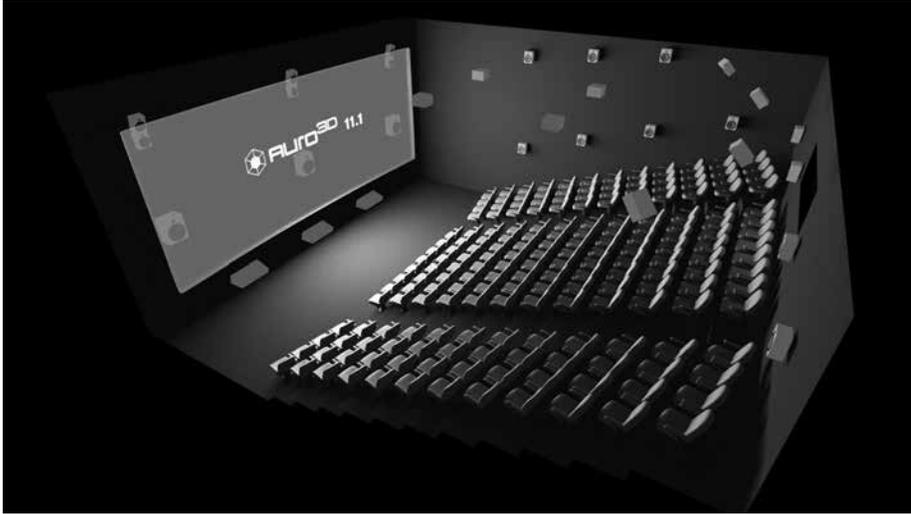


Figure 1.31 – Exemple d’une salle équipée Auro-3D en 11.1, © Auro-3D. En couleur sur dunod.com

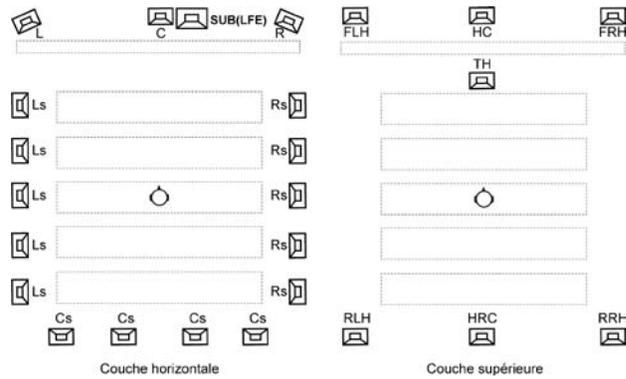


Figure 1.32 – Format Auro-3D 13.1 cinéma.

1.7.19 10.2 Tomlinson Holman

Tomlinson Holman (TMH Labs), créateur de THX, propose un format 10.2 qui repose sur la norme 5.1 ITU à laquelle on ajoute :

- deux enceintes Wide Left et Wide Right, à $\pm 55^\circ$;
- une enceinte centrale arrière CS (Center Surround) ;
- deux canaux Lfe : Left Lfe et Right Lfe, positionnés à $\pm 90^\circ$;
- deux enceintes de hauteur : Left Height et Right Height, placées à $\pm 45^\circ$ avec 45° d’élévation.

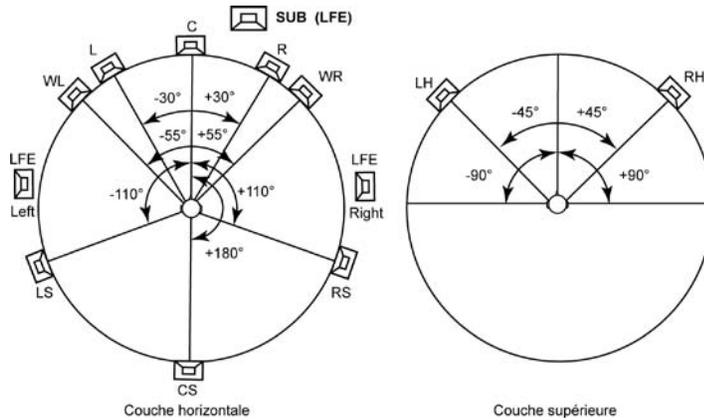


Figure 1.33 – Format 10.2 Tomlinson Holman.

1.7.20 11.1 DTS Neo X

DTS propose depuis 2011 le format DTS Neo X 11.1. Il repose sur la configuration 5.1 ITU à laquelle on ajoute :

- deux enceintes Wide Left et Wide Right à $\pm 60^\circ$ ou deux enceintes latérales Left Side Surround et Right Side Surround ;
- un canal centre arrière (CS) ;
- deux enceintes de hauteur Left Height et Right Height, placées à $\pm 30^\circ$.

Voir la **figure 1.20**.

1.7.21 22.2 NHK (9/10/3.2)

Le format 22.2 de la NHK a été développé pour accompagner le format Image 8K Super Hi-Vision, dont les programmes sont diffusés pour une salle de cinéma. Ce format est composé de trois couches :

- une première couche horizontale constituée de dix enceintes, cinq frontales derrière l'écran, deux sur le latéral à $\pm 90^\circ$ et 3 à l'arrière ;
- une couche supérieure composée d'une couronne de neuf : enceintes, trois en frontal, deux en latéral, trois à l'arrière et une enceinte centrée, au plafond ;
- une troisième couche inférieure composée de trois enceintes derrière l'écran, complétée par deux canaux Lfe gauche et droite.

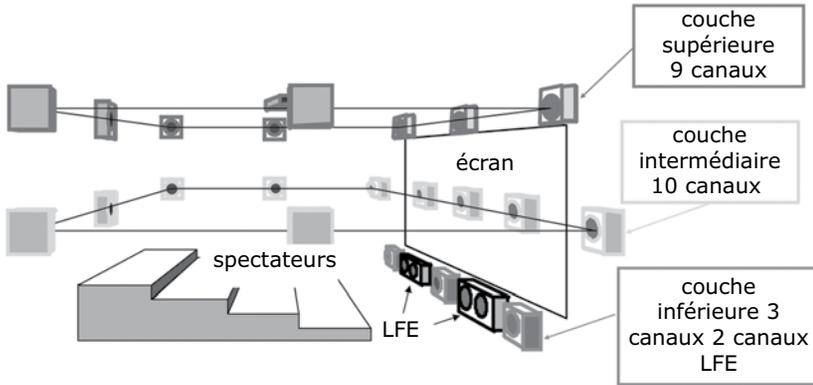


Figure 1.34 – Format 22.2 NHK, © NHK.

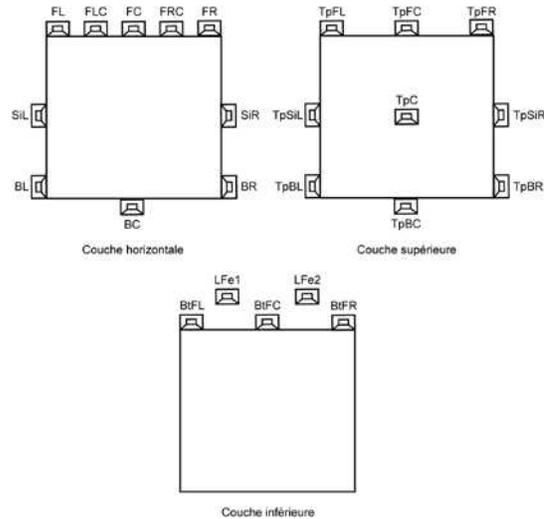


Figure 1.35 – Format 22.2 NHK.

1.7.22 Imm Sound 23.1 ou 14.1 (Dolby)

Le procédé de diffusion Imm Sound a équipé les grandes salles de cinéma, il permet la restitution sur trois couronnes d'enceintes en 23.1 ou 14.1.

- la couche supérieure est constituée de trois enceintes en frontal et de cinq enceintes au plafond ;
- la couche horizontale est définie par trois enceintes en frontal (L, C, R), six canaux répartis sur dix enceintes pour les latéraux et deux canaux arrière répartis sur six enceintes ;

- la couche inférieure comporte trois enceintes en frontal complétées par deux Subwoofer alimentés par un seul canal Lfe.
- Il existe aujourd'hui une trentaine de salles équipées en Imm Sound, dont une en France, à Brumath (en Alsace). Ces salles devraient vraisemblablement être transformées en Dolby Atmos.

1.7.23 Norme ITU-R BS 2051

La norme ITU-R BS 2051 a été pensée pour un système sonore avancé avec ou sans image (*Advanced sound system for programme production*) au-delà de la norme ITU-R BS 775. Elle comporte trois niveaux d'enceintes, seize canaux pour la couche medium, seize pour la couche du bas, dix-sept pour la couche du haut ainsi que deux canaux LFE.

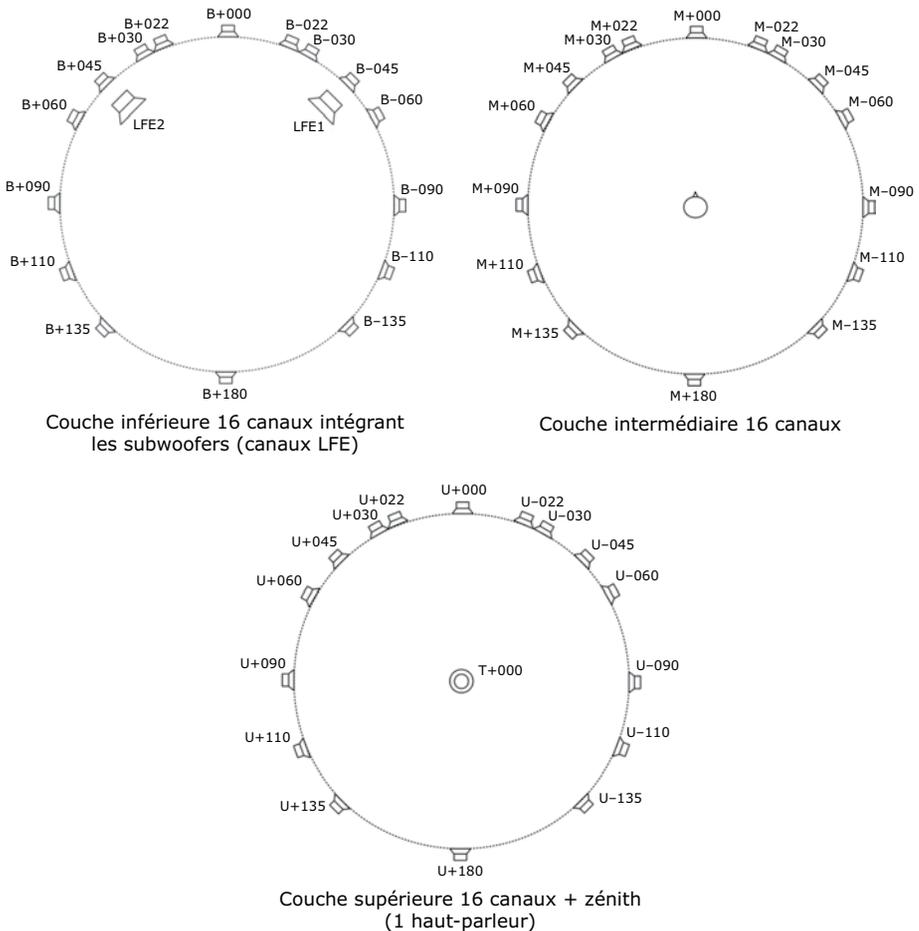
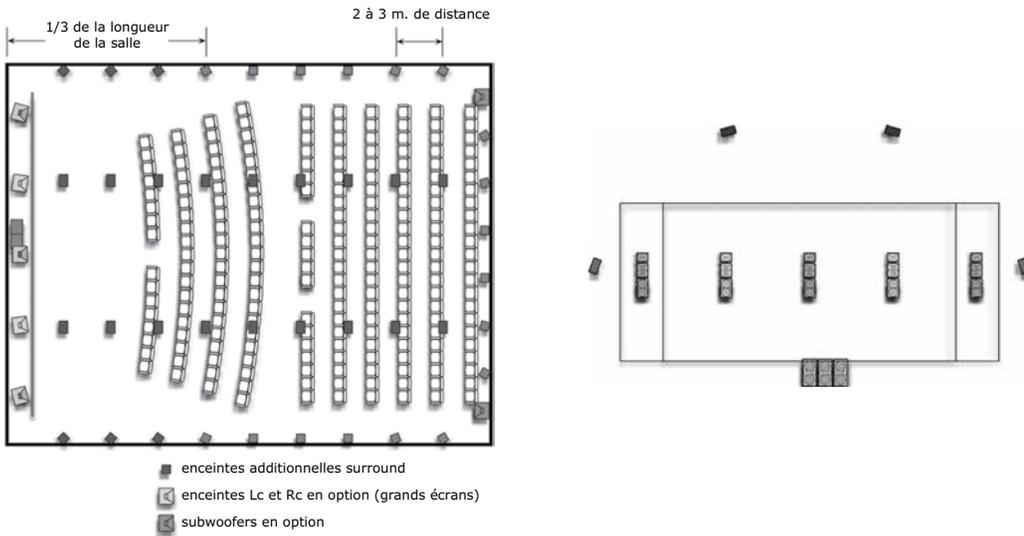


Figure 1.36 – Norme ITU-R BS 2051, © ITU.

1.7.24 Dolby Atmos

Dolby Atmos est un procédé pour la diffusion en salle de cinéma qui permet de gérer jusqu'à cent vingt-huit canaux et soixante-quatre haut-parleurs. Le procédé permet la restitution de la hauteur mais s'adapte aux différents formats du multicanal existants comme le 5.1 ou le 7.1. Il intègre :

- trois haut-parleurs frontaux derrière l'écran (plus deux Lc Rc de part et d'autre du centre en option pour les écrans de plus de 12 mètres) ;
- des haut-parleurs Side Surround, Rear Surround et Top Surround ;
- deux subwoofers qui viennent compléter ceux placés au centre et en bas de l'écran, ils sont placés à l'arrière de la salle, à gauche et à droite.



Figures 1.37 et 1.38 – Configurations Dolby Atmos, © Dolby. En couleur sur dunod.com

Il existe aussi les configurations Dolby Atmos Home, représentées sur la **figure 1.39**.



Figure 1.39 – Configuration Dolby Atmos Home 7.1.4, © Dolby.

1.7.25 WFS

Le système de diffusion WFS est constitué d'une couronne d'enceintes placées tout autour de l'auditeur, au minimum une trentaine, généralement une centaine, parfois disposées sur deux couronnes.

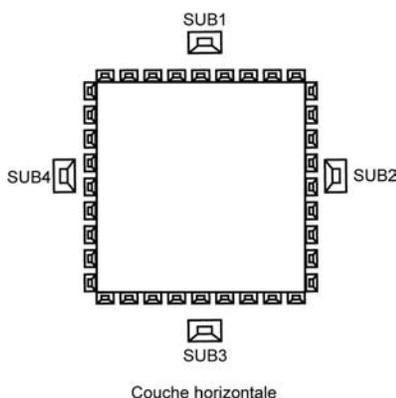


Figure 1.40 – Dispositif multicanal WFS.

1.8 État des lieux des marchés du multicanal

1.8.1 Statistiques

◆ Évolution du home-cinéma et de la HD

Il est difficile d'obtenir des statistiques sur le multicanal et sur le nombre de foyers équipés dans le monde. Nous avons plus de visibilité sur le nombre de lecteurs Blu-ray/DVD ou sur le nombre d'écrans Full HD vendus, qui représentent un certain potentiel en taux d'équipement. Le taux d'équipement en home-cinéma 5.1 des foyers français, au deuxième trimestre 2009, est de 10 à 15 %, celui du lecteur DVD est proche de 90 %, et celui du téléviseur compatible HD est autour de 40 %. Aujourd'hui, un tiers des foyers français est équipé d'un lecteur Blu-ray contre plus de la moitié en Angleterre et aux États-Unis.

Les opérateurs télécoms ont un rôle à jouer sur l'évolution technologique des équipements avec notamment le développement des box compatibles TVHD équipées d'une interface HDMI.

◆ Évolution du multimédia

L'interface HDMI qui permet la connexion à l'ampli home-cinéma rend le multicanal accessible. Elle équipe l'ordinateur, notamment toute la gamme Apple, via le port thunderbolt ou directement HDMI, et, progressivement, la gamme des PC Portables et des PC de bureau. On la trouve sur certaines tablettes numériques et sur certains téléphones mobiles et dans les consoles de jeu. Par ailleurs, les sorties numériques pour le son codé transportées via l'interface SPDIF (les sorties coaxiales ou mini-jack), la sortie

optique Toslink ou bien les sorties analogiques individuelles sur connectique RCA, sont aussi des solutions qui équipent la grande majorité des PC, en particulier pour les jeux.

1.8.2 Premier bilan de la TVHD

En janvier 2009, l'EBU publie un premier bilan de la TVHD dans lequel est exposée une analyse du marché européen des écrans plats HD, des contenus HD des chaînes, et diffuse une étude sur l'évolution du déploiement de la télévision dans le futur. Le document rapporte que l'année 2008 a été importante pour la TVHD, du fait de l'apparition d'un certain nombre de facteurs qui ont contribué à l'essor de cette technologie. La TVHD progresse en effet sur tous les fronts, de la production à la distribution, en passant par les appareils grand public, en particulier au Royaume-Uni et en France.

On répertorie en 2014 une quarantaine de chaînes HD européennes diffusant un son 5.1 et privilégiant les genres tels que le sport, le film, la fiction et le documentaire. Le film est mixé en 5.1 pour la salle de cinéma, une version 5.1 compatible LtRt pour la télévision est généralement disponible (voir chapitres 6, 7 et 8). Les grands événements sportifs sont captés en 5.1, notamment les coupes du monde de football, de rugby, les Jeux Olympiques (depuis ceux d'Athènes en 2004), le tennis. Depuis quelques années, France Télévisions produit certains matchs de Roland Garros en 5.1 : en 2011 trois courts de tennis étaient équipés de systèmes de prise de son multicanal. À cette liste s'ajoutent l'US Open, tous les matchs de football ligue 2 diffusés le lundi soir, la formule 1, le handball, la natation. Le match de football diffusé sur Canal + chaque dimanche soir est aussi produit en 5.1. Les chaînes de sport telles qu'Eurosport ou Canal+ diffusent en 5.1 leurs génériques et leurs bandes annonces.

France Télévisions a démarré la diffusion en Dolby Digital avec France 2 HD depuis quelques années. Le départ antenne est alimenté par un processeur d'upmix de diffusion Isostem, la diffusion est depuis passée en Dolby Digital +. Dès la mise en route du nouveau centre de diffusion prévue pour fin 2015, les quatre chaînes nationales France 4, France 2, France 5 puis France 3 devraient démarrer leurs services 5.1.

La diffusion 5.1 des programmes sur Arte HD devrait progressivement démarrer en 2015 essentiellement pour les fictions, les documentaires et les opéras. La régie de diffusion est équipée d'un processeur Aeromax qui permet également « d'upmixer » un programme stéréo. On trouve aussi quelques chaînes de documentaires (National Geographic HD, Animal Planet HD, Discovery HD, BBC HD) et quelques chaînes musicales (MTV Live HD, NRJ12 HD, Deluxe Music, BravaHDTV). D'autres chaînes européennes diffusent en Dolby Digital 5.1 : BBC HD 1, 2 et 3, SKY 3D, BR Nord HD, BR Sud HD, ORF2 HD, SF1 HD, NRK HD 1, 2, et 3, SVT 1 et 2... La chaîne BBC 3 a expérimenté début 2014 une diffusion 4.0 sur leur site internet en HTML5 MPEG dash.

La France compte début 2014 une dizaine de chaînes HD françaises diffusant un son en 5.1 généralement en Dolby Digital à 384 kbps ou Dolby Digital + à 256 kbps : France 2 HD, M6HD, Canal+HD, Arte HD, W9 HD, 6 Ter HD, Eurosport HD (diffusion 5.1 pour l'étranger), NRJ12 HD, Ciné plus premier HD, TCM HD, Disney Channel

HD. Cette liste est susceptible d'être modifiée. À cela s'ajoute la VOD HD, les box Free et Bouygues Telecom supportent le DTS Neural. Il existe un site Digitalbitrate qui analyse les différents flux des chaînes HD (www.digitalbitrate.com) dont la **figure 1.4** présente un exemple de mesures.

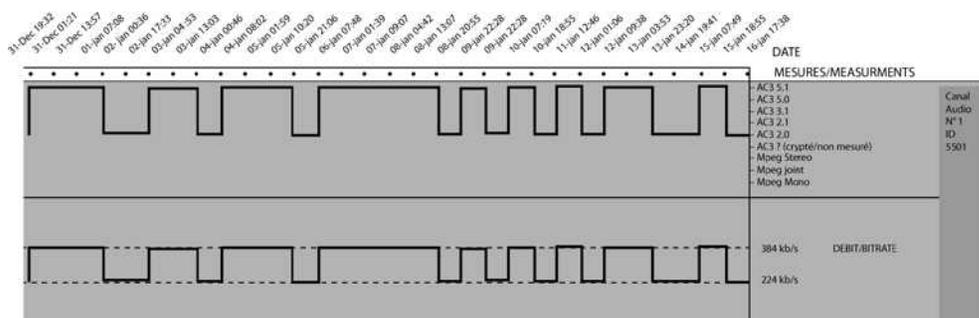


Figure 1.41 – Exemple de mesure audio de la BBC HD sat Astra, © Digitalbitrate.
En couleur sur dunod.com

La diffusion peut alterner entre programmes stéréo et 5.1 ou bien être en 5.1 en continu, le flux est alors généralement issu d'un processeur d'upmix.

1.8.3 Perspectives sur la production radiophonique

En Europe, la diffusion d'émissions radiophoniques en 5.1 a été initiée par les radios norvégienne, suédoise, allemande et autrichienne. La radio norvégienne NRK diffuse sur sa webradio des programmes 5.1 disponibles en streaming. La radio suédoise SR possède une page Internet webradio, sur laquelle elle met à disposition des programmes en 5.1, DTS ou Dolby Digital, que l'utilisateur peut télécharger et lire sur un DVD. Certaines radios allemandes et autrichiennes diffusent en 5.1 leurs programmes.

En France, avant l'arrivée de la TVHD, Radio France a produit un certain nombre de documentaires et de fictions radiophoniques. Ces premières expérimentations datent des années 1990, la radio produit aussi le son d'opéras filmés, édités en DVD et Blu-ray. Depuis 2010, un retour à la production multicanal se fait par le biais de magazines sur France Inter comme *Interception*, dont quatre éditions ont été produites en multicanal en 2011. D'autres expérimentations sont menées sur des festivals de musique comme La Folle Journée de Nantes, les Vieilles Charrues, le festival d'Orange, l'orchestre national de jazz de Montpellier ou Jazz à Vienne. Des projets d'expérimentations autour d'artistes avec des recherches de réalisation vidéo comme « Andrew Bird », « Erikm » ont lieu. Ce développement au sein du département des nouveaux médias (DNM) a abouti au lancement de la webradio NouvOson qui diffuse des programmes en 5.1 et en binaural. Ce lancement est rejoint par le Mouv qui se restructure début 2014 avec la même dynamique : celle de produire des artistes en multicanal. Ces flux sont disponibles en binaural pour le mobile, la tablette numérique et en 5.1 pour une diffusion sur home-cinéma par l'interface HDMI.

1.8.4 *Événements, séminaires, groupes de recherche, formation*

Le son multicanal est un sujet aujourd'hui très exposé dans les séminaires sur le son : citons à titre d'exemples les conventions AES (*Audio Engineering Society*), les séminaires de l'International VDT Symposium (*Verband Deutscher Tonmeister*) et bien d'autres rencontres organisées par des institutions du broadcast, de la recherche, des associations, des industriels, etc.

En France le FISM (Forum International du Son Multicanal), les journées AFSI, France Télévisions, l'Ircam, la Semaine du Son proposent des rencontres sur différents aspects du multicanal.

Dans le domaine de la recherche, en France, des groupes de travail se sont constitués afin de promouvoir le binaural et le son 3D. BiLi (Binaural Listening) est un consortium français pour le développement de technologies de production et de diffusion de nouveaux médias en binaural. Edison a pour but de promouvoir la création de contenus audio 3D et de développer les technologies associées pour la production et la diffusion.

Enfin, le multicanal est maintenant abordé dans les différents cursus son de formations initiale et professionnelle.

Chapitre 2

Le système d'écoute

*L'écoute monitoring est à l'ingénieur du son
ce que la lumière du soleil est au peintre*

Patrick Thévenot

POINTS ESSENTIELS

- Critères d'écoute pour l'analyse d'un système de monitoring multicanal.
- Bass management.
- Norme ITU.
- Norme Cinéma.
- Cabines de mixage multicanal.
- Méthodes de mesure.
- L'influence du local d'écoute.
- Corrections acoustiques et électro-acoustiques, THX.
- Test entièrement auditif de votre écoute 5.1.
- Le WFS.
- Le Transaural.

2.1 Caractérisation auditive d'une écoute

2.1.1 Pourquoi apprendre à écouter ?

L'écoute d'un son, d'un objet sonore, l'appréciation d'une scène sonore sont étroitement liés à notre culture auditive, mais aussi à notre humeur du moment. En bref, nous restons constamment dans le domaine affectif (« j'aime, je n'aime pas ») lors de la caractérisation d'une écoute. Nous avons donc besoin d'apprendre à analyser objectivement ce que nous entendons afin d'enregistrer, mixer les sons, mais aussi afin d'évaluer notre système d'écoute. Il s'agit de l'écoute professionnelle à laquelle tout ingénieur du son s'est exercé, bien souvent seul et sans méthodologie, d'où une efficacité très moyenne et surtout sans transmission possible de cette méthode à ses confrères. C'est un peu le même problème pour deux personnes qui ne parlent pas la même langue. Il est donc nécessaire, avant tout test auditif devant être comparé à la mesure, d'acquérir une méthodologie d'écoute, objective et professionnelle, fiable.

Ces critères, un peu différents de ceux utilisés pour la prise de son et le mixage, sont adaptés à la caractérisation d'un système d'écoute (enceintes + salle) en 2.0, 5.1 & 7.1, ceci en corrélation avec des mesures que nous détaillerons plus tard. En stéréophonie, notre expérience auditive peut relativement « corriger » les défauts en s'adaptant au système d'écoute utilisé. Par contre, dès que le nombre de canaux augmente, notre audition « intelligente » devient impuissante devant le nombre de paramètres à gérer : par conséquent il est impératif d'acquérir une technique d'écoute objective et systématique.

2.1.2 Critères objectifs nécessaires à la caractérisation d'une écoute

Nous déterminerons des **critères de construction**, des **critères d'équilibre**, des **critères de définition**, des **critères de stabilité** et des **critères de réverbération**. Ces critères découlent de la convergence d'expérimentations rigoureuses menées au Conservatoire National Supérieur de Musique de Paris (CNSM) et à Radio France par Guy Laporte, au Conservatoire National des Arts et Métiers (CNAM) par Jacques Jouhaneau et à l'Institut National de l'Audiovisuel (INA) par Sébastien Noly et moi-même.

Cette méthode quantifiée d'analyse sonore permet de s'affranchir en grande partie de ses goûts et de sa culture d'écoute grâce à une prise en conscience de celle-ci. Il n'est évidemment pas question de ne plus prendre de plaisir à écouter, mais de pouvoir rapidement distinguer l'aspect esthétique et affectif de l'écoute, de son côté neutre et objectif, ceci dans des phases différentes et séparées.

Ces familles de critères se subdivisent en sous-critères selon le plan suivant :

◆ Critères de construction

Complément Web 2.1

- Le type et le degré d'immersion ;
- La linéarité d'espace ;
- La profondeur des plans ;
- Le relief.

◆ Critères d'équilibre

Complément Web 2.2

- L'équilibre spectral ;
- L'équilibre dynamique.

◆ Critères de définition

Complément Web 2.3

- La séparation des timbres ;
- La transparence ;
- L'homogénéité spatiale de la définition.

◆ Critères de stabilité

Complément Web 2.4

- La stabilité de l'image globale ;
- La stabilité des sujets ;
- L'homogénéité spatiale de la stabilité.

◆ Critères de réverbération

Complément Web 2.5

- La quantité ;
- Le type ;
- L'homogénéité.

2.1.3 Apprentissage d'une méthodologie d'écoute

Complément Web 2.6

L'apprentissage de cette méthode passe avant tout par l'affect et pas uniquement par l'intellect. Elle doit donc obligatoirement s'appuyer sur un support audiovisuel. C'est pourquoi nous renvoyons le lecteur aux compléments numériques afin de l'expérimenter.

Cette méthode consiste à identifier les sensations sonores correspondant à chaque critère défini précédemment en faisant correspondre les illustrations visuelles des pages précédentes avec les extraits sonores proposés dans le support fourni. **La quantification des impressions sonores** (évaluation numérique) est très importante et représente le meilleur moyen de rester dans le domaine objectif (en utilisant le cerveau gauche, centre cartésien, et non le droit, centre émotionnel). Cependant, l'analyse doit se faire sur un temps très court (une minute au maximum), la première impression étant toujours la bonne. En effet, le cerveau gauche possède la faculté, fâcheuse dans ce cas, d'enregistrer la première séquence sonore et de réécouter ensuite en boucle l'enregistrement de celle-ci et non le nouveau message sonore rentrant... Les exploitants connaissent bien ce phénomène lors de l'écoute répétée d'une phrase sonore pour le choix d'un effet, lors du mixage, par exemple.

L'évaluation objective et le diagnostic d'un système de restitution multicanal sera largement facilité par l'utilisation de cette méthode.

Complément Web 2.7

Des fiches d'évaluation téléchargeables sur le site WEB permettent de se familiariser et de s'entraîner à l'analyse d'écoute en fonction de ses besoins et de son niveau d'expérience.

2.2 Finalité d'une écoute monitoring en multicanal

2.2.1 *Domaine artistique, domaine objectif*

Nous retrouvons des systèmes d'écoute en beaucoup d'endroits de la chaîne audiovisuelle. Ces écoutes ont des finalités très différentes suivant qu'elles se trouvent dans le domaine artistique ou le domaine objectif. Nous devons donc faire la distinction entre ces différents types et ces deux domaines.

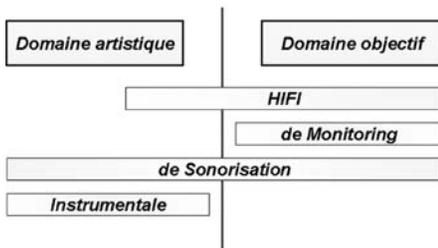


Figure 2.1 – Les différents types de systèmes d'écoute.

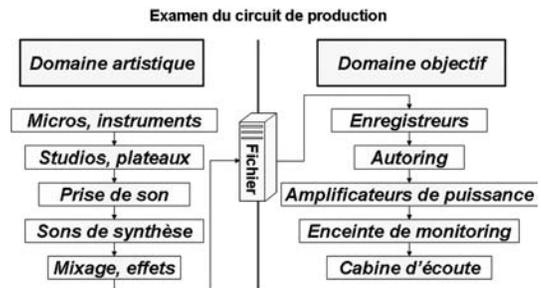


Figure 2.2 – Phases objectives et artistiques d'un circuit de production.

Nous devons également examiner le circuit de production/diffusion en fonction de ces deux domaines.

2.2.2 *Problématique spécifique au multicanal*

En diffusion stéréophonique à deux canaux, les oreilles très aiguës de l'ingénieur du son lui permettent de s'adapter et de compenser en grande partie les défauts d'une écoute, ce qui explique que l'on trouve d'excellents enregistrements, y compris très anciens, alors que les systèmes d'écoute, comparés à leurs homologues actuels, étaient d'une neutralité contestable, voire très « colorés ».

L'introduction du multicanal, avec l'augmentation du nombre de canaux (5.1 et 7.1) ne permet plus à une oreille, même très entraînée, d'interpréter et de compenser les défauts d'un système d'écoute multidirectionnel. En effet, les paramètres deviennent trop nombreux et trop complexes pour être extrapolés et analysés afin d'assurer la transportabilité de l'enregistrement ou du mixage jusqu'à l'auditeur final.

Il devient donc indispensable de mettre en œuvre une méthodologie de réglage ainsi que des corrections acoustiques et électroniques permettant de maîtriser enfin ce que l'on entend. Cette possibilité, tout simplement impensable il y a vingt ans, est maintenant réalisable, grâce à l'introduction de mesures informatisées basées sur la réponse impulsionnelle (une meilleure connaissance de notre perception ainsi que des outils de corrections numériques très puissants).

2.3 Les différents types de mesures

Complément Web 2.8

Nous distinguerons chronologiquement plusieurs méthodes. Toutes sont détaillées et illustrées sur le site Web.

2.3.1 *Fréquence glissante*

La plus ancienne. Utilisable quasi exclusivement en chambre sourde. Champ d'application limité. Peu révélatrice auditivement.

2.3.2 *Bruit rose*

Toujours utilisé depuis les années soixante-dix. Rapidité d'interprétation. Bonne approche auditive de l'équilibre tonal.

2.3.3 *MLS, CHIRP, TDS*

Mesures modernes basées sur le calcul FFT. Permet d'accéder au domaine temporel et fréquentiel. Difficilement interprétable par le néophyte. Bien corrélé à l'impression auditive sous réserve de rester dans le domaine linéaire (pas de distorsion). Limites : pas de prise en compte du retard dynamique contribuant au respect de l'émotion sonore.

2.3.4 Retard dynamique

Mesure visuelle récente. Permet l'évaluation du respect du facteur émotionnel du son enregistré. Complémentaire à la mesure MLS.

2.3.5 Synthèse des différentes mesures

En résumé, et pour synthèse, la **figure 2.3** permet de comparer les différents signaux de mesures avec leurs possibilités d'analyse et de calcul... et leurs limites. Toutes ces mesures sont décrites en détail sur le site Web.

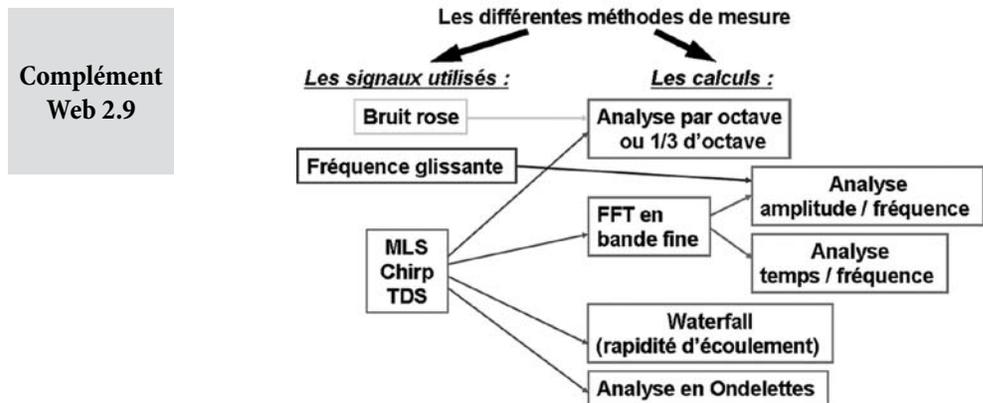


Figure 2.3 – Les différentes méthodes de mesure.

2.4 L'enceinte acoustique

2.4.1 Différents types

Pour une utilisation monitoring nous distinguerons les enceintes de proximité, les enceintes de mixage principales et les écouteuses dites « clients ». Ces dernières peuvent être très « typées » et différentes en fonction de l'utilisation finale et de la culture d'écoute.

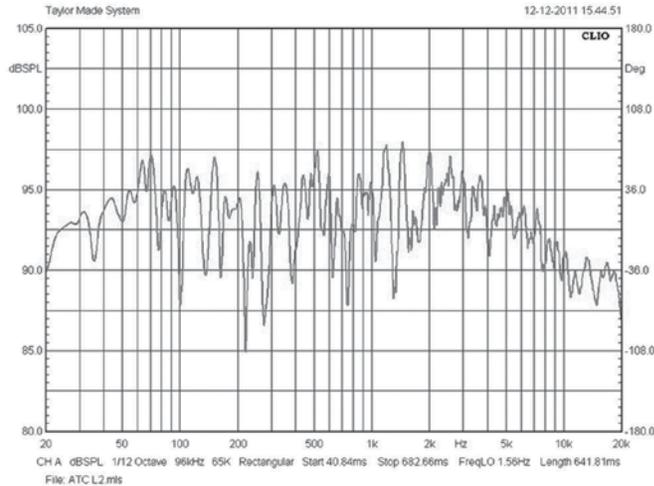
Nous examinerons, sans être exhaustif, trois cultures d'écoute : le **télévision** et le **multimédia**, le **CD et DVD** et le **cinéma**. Les différences seront souvent dans la taille, le niveau, la dynamique, la directivité et la neutralité.

Dans tous les cas, en multicanal, les enceintes doivent être identiques (à l'exception du caisson de grave), ce qui n'était pas le cas en Dolby Surround où les enceintes arrières étaient plus petites que les enceintes frontales.

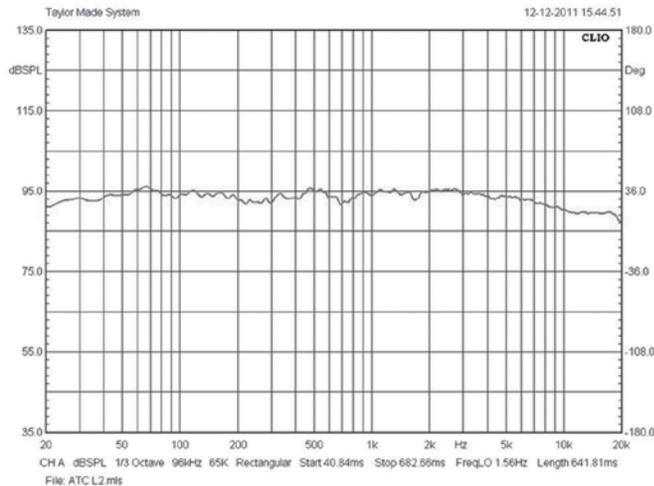
2.4.2 Mesures et interprétations auditives

◆ Courbe de réponse en fonction de la fréquence

La courbe de réponse « niveau/fréquence » reste la première mesure, significative de l'équilibre spectral perçu. Cette mesure est traditionnellement faite dans l'axe de l'enceinte (à 0°). Il est important d'identifier clairement la résolution en dB (0.5 à 2 dB/division) ainsi que le lissage (1/12 d'octave à 1/3 d'octave).



(a)



(b)

Figure 2.4 – Courbe de réponse en niveau en fonction de la fréquence.
Même mesure avec des échelles et des résolutions différentes.

◆ Courbes de directivité

La directivité d'une enceinte est un facteur primordial lors de son exploitation en mixage en fonction de l'acoustique du lieu. Il existe plusieurs manières de l'illustrer. Nous ne retiendrons que la plus courante, sous forme des courbes de réponse effectuées à 0°, 30°, 60° et 90°. Elles nous donneront de précieuses informations sur la manière dont l'enceinte rayonne dans le local et donc, sur son positionnement optimal, mais aussi sur la réponse en puissance acoustique totale qui est la sommation en dB des courbes de directivité.

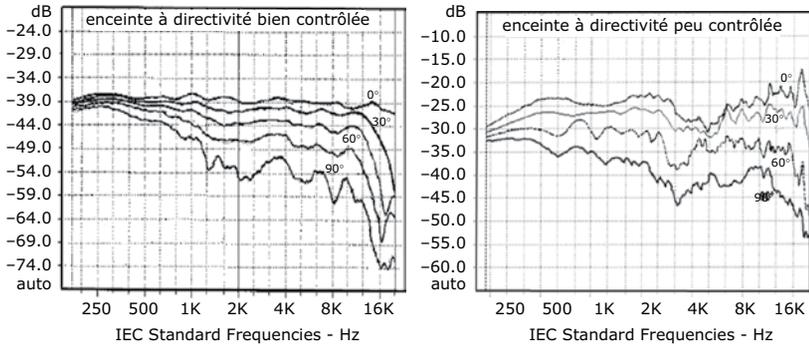


Figure 2.5 – Courbes de directivité en fonction de la fréquence.

Nous pouvons distinguer trois familles de directivité :

- Les enceintes omnidirectionnelles ;
- Les enceintes à directivité croissante ;
- Les enceintes à directivité constante.

Le facteur de directivité Q est le rapport de l'énergie émise dans l'axe de l'enceinte sur l'énergie totale fournie par celle-ci. Q est un coefficient sans unité qui varie de 1 à l'infini.

En sonorisation, les constructeurs donnent souvent l'angle d'ouverture de l'enceinte. Il s'agit de l'angle correspondant à une chute de 3 dB ou 6 dB du niveau par rapport à l'axe. Cette indication n'est significative que pour les enceintes à directivité constante et n'a aucun sens dans les autres cas, étant donné que cet angle varie à chaque fréquence !

Complément Web 2.10

L'utilisation d'enceintes dont la directivité est croissante et bien contrôlée peut avoir certains avantages en monitoring comme la garantie de la transportabilité du mixage chez l'auditeur et l'agrandissement de la zone d'écoute stable (sweet spot). Cette étude est développée sur le site Web.

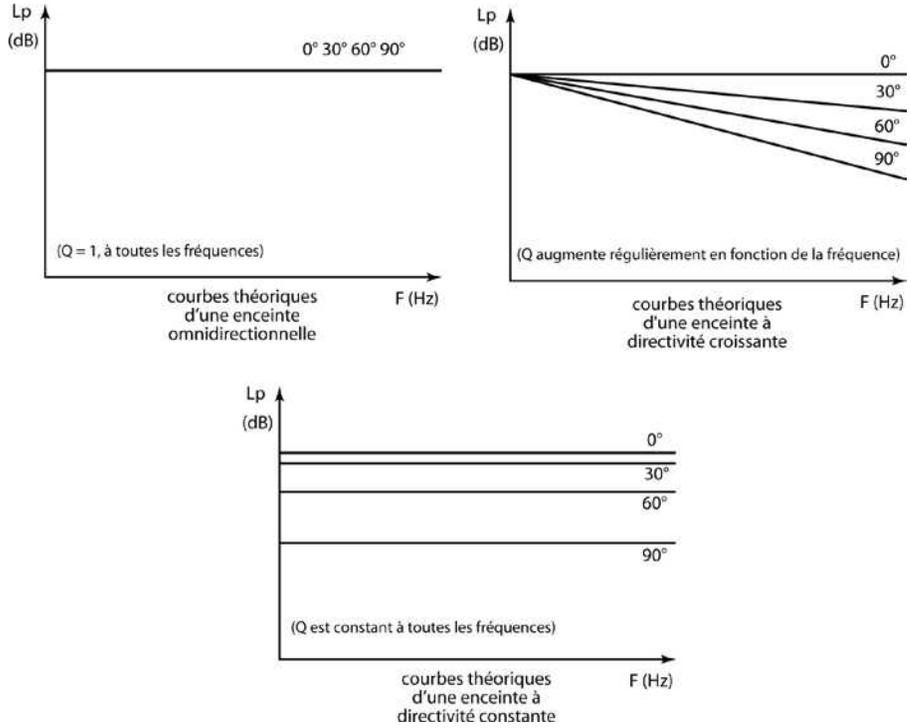


Figure 2.6 - Les différents types de directivités.

◆ **Courbe de retard en fonction de la fréquence**

Le retard relatif (en régime harmonique) entre les fréquences est lié à la courbe de phase. Appelé « temps de propagation de groupe », celui-ci est très important et conditionne la reconstruction en profondeur et relief des plans sonores. L'allure de cette courbe par rapport à la fréquence modifie également largement l'équilibre spectral et dynamique de l'enregistrement.

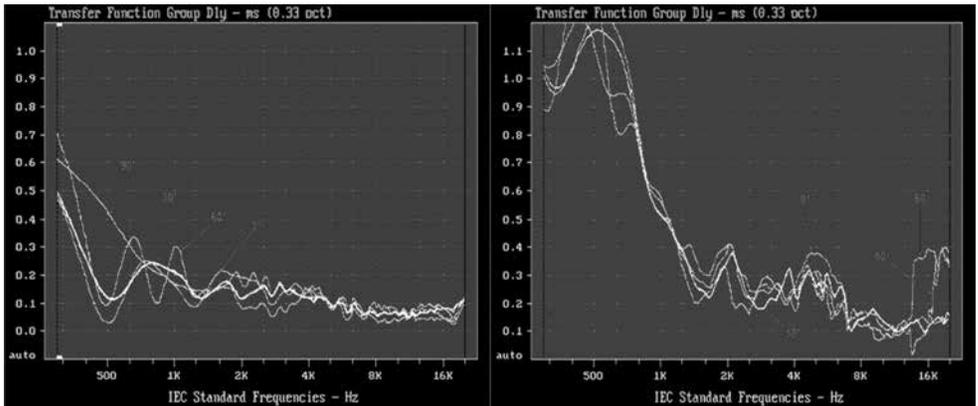


Figure 2.7 - Courbes de temps de propagation de groupe. En couleur sur dunod.com

◆ Influence du retard temporel sur le niveau subjectif de grave

En dessous de 200 Hz, si le grave est retardé progressivement en fonction de la fréquence, notre impression de niveau augmente proportionnellement. Ceci est observé sur certaines enceintes qui génèrent un grave très typé, qui peut plaire, mais n'est pas transposable à d'autres systèmes.

Complément Web 2.11

Ce phénomène est étudié et illustré auditivement en complément numérique.

◆ Rapidité de l'écoulement de l'énergie (Waterfall)

Le « waterfall » (chute d'eau) illustre la rapidité avec laquelle l'énergie se dissipe en fonction du temps à chaque fréquence. Cette information, très importante (bien que peu médiatisée !) permet de mesurer la capacité de séparation en timbre et en transparence de l'enceinte, mais aussi de l'équilibre dynamique.

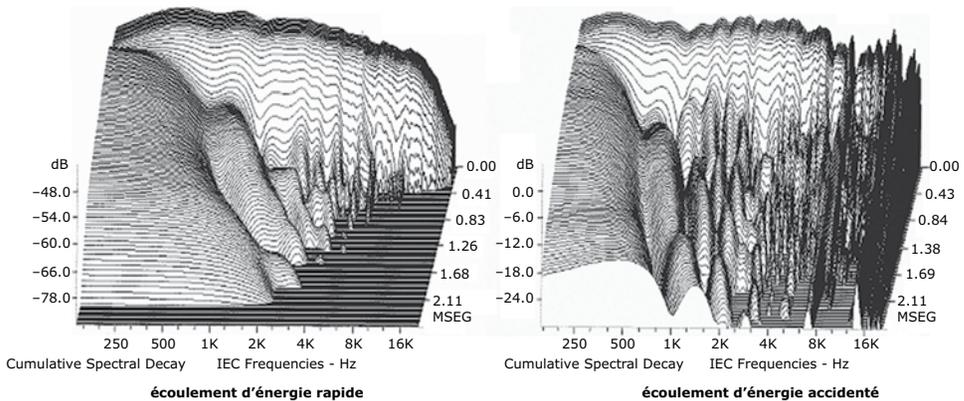


Figure 2.8 – Mesures du waterfall.

Attention, une mauvaise rapidité de dissipation dans l'aigu peut engendrer un embelissement de la musicalité et une fausse transparence de l'enregistrement, pouvant compenser un équilibre dynamique pauvre (par exemple pour des sources compressées) ou bien une prise de son de proximité un peu « dure ».

◆ Analyse par ondelettes

Enfin, l'analyse par ondelettes (Wavelet) montre en une seule image la manière dont l'énergie sonore s'écoule en fonction du temps et de la fréquence, avec une vision claire du « traînage » éventuel à certaines fréquences, même pour des niveaux résiduels très faibles (pourtant très bien perçus par notre audition).

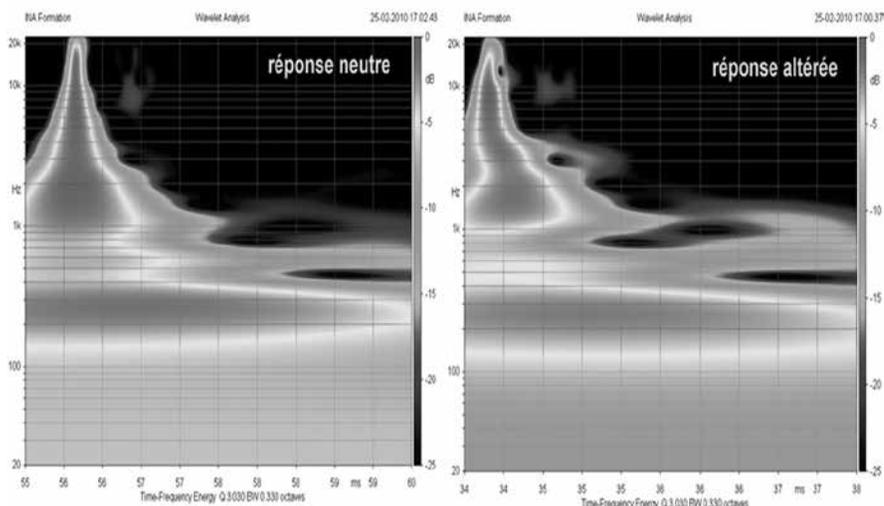


Figure 2.9 – Analyses en ondelettes. En couleur sur dunod.com

Cette mesure nous informe sur certaines colorations de timbre perçues à l'oreille mais qui ne sont révélées, ni par la courbe de réponse, ni par la courbe de retard de groupe, mais aussi sur la stabilité de l'image sonore.

◆ Rendement et sensibilité

Le rendement d'une enceinte acoustique est primordial afin que celle-ci puisse restituer la dynamique attendue en fonction du type d'utilisation. On confond souvent le rendement et la sensibilité. Bien qu'elles soient liées, ces deux mesures ne sont pas équivalentes.

Le rendement R_d représente le rapport entre la puissance acoustique restituée et la puissance acoustique électrique fournie par l'amplificateur. Il se mesure en %.

$$R_d = (W_a/P_e) \times 100$$

Avec :

W_a : puissance acoustique en watt acoustique.

P_e : puissance électrique en watt électrique.

Le rendement représente donc toute l'énergie dissipée par l'enceinte tout autour d'elle.

La sensibilité est une mesure de niveau de pression acoustique (dB) effectuée en champ libre dans l'axe d'écoute (0°) à 1 mètre de distance pour une puissance appliquée de 1 W, soit 2.83 V efficaces sur une enceinte de 8Ω d'impédance nominale (mesure normalisée suivant IEC 60268-5 3^{ed}).

$$S_{dB} = L_p \text{ à } 0^\circ / 2.83 \text{ V}$$

avec L_p : niveau de pression à 1 m en dB.

La sensibilité mesurée dans l'axe de l'enceinte est donc liée au rendement. Ainsi, pour un même rendement, une enceinte très directive aura une meilleure sensibilité qu'une enceinte plus omnidirectionnelle. En champ libre, la différence de sensibilité sera parfaitement perçue tandis qu'en milieu réverbérant, l'impression de niveau sonore sera très semblable et significative du même rendement.

Le rendement et la sensibilité n'ont de sens que pour les enceintes passives. Pour les enceintes amplifiées et (ou) actives, les constructeurs donnent le niveau de pression maximal ($L_p \text{ max}$) à 1 m en champ libre, mais cette valeur peut être supérieure suivant les caractéristiques de l'environnement acoustique.

2.4.3 Conditions de fidélité et choix pour le multicanal

Toutes les courbes doivent être les plus régulières possibles et sans accident, le but d'une écoute monitoring n'étant pas d'embellir le son, mais de juger rapidement de la qualité d'un enregistrement ou un mixage sans se tromper et avec un minimum de fatigue auditive.

◆ L'écoute de proximité

Enceinte compacte par nécessité (en général deux voies), celle-ci doit posséder une courbe de réponse régulièrement descendante au-dessus de 2000 Hz, une courbe de TG (Temps de propagation de Groupe) la plus droite possible, un Waterfall à décroissance rapide au-dessus de 500 Hz et une dynamique raisonnable ($L_p \text{ max} > 90 \text{ dB}$ à la distance d'écoute). La directivité peut être assez large mais contrôlée (sans accidents). Elle est utilisée principalement en enregistrement et mixage musical TV, CD ou DVD. Les cinq enceintes doivent être identiques et un caisson de grave peut être adjoint en renfort de grave (avec réglage correct du bass management), ou uniquement pour le LFE.

$L_p \text{ max}$: niveau sonore maximal que peut fournir en permanence chaque enceinte acoustique à la position de mixage. On utilise un bruit rose en pleine bande (20 Hz à 20 000 Hz) et la mesure est effectuée par un sonomètre réglé avec une grande constante de temps (position slow) et en mesure linéaire ou quasi linéaire (pondération C).

Le **bass management** consiste à envoyer dans le caisson de grave (« Sub Bass ») les fréquences graves des cinq canaux afin de protéger les enceintes. La fréquence de transition se situe entre 50 Hz et 120 Hz avec un filtrage électronique à 24 dB/octave au minimum (ordre 4). Le choix des fréquences de coupure ainsi que des pentes et structures de filtre devra être soigneusement ajusté *in situ*. En effet, l'environnement acoustique joue un rôle très important dans le résultat final. Le processeur utilisé doit être capable de gérer non seulement le filtrage, mais aussi les retards relatifs entre les enceintes et le Sub ainsi qu'une égalisation de type paramétrique. Le réglage doit être réalisé à la mesure par un spécialiste et non « à l'oreille » comme on le rencontre souvent ! L'enjeu d'un réglage réussi est une parfaite complémentarité entre le son des enceintes et celui du Sub : on ne doit pas localiser celui-ci afin d'obtenir une immersion correcte dans la scène sonore. Si on entend le caisson de grave, le système est mal réglé... Ces remarques sont également valables pour l'écoute principale.

◆ L'écoute principale

Pour une utilisation en TV, CD ou DVD, l'enceinte sera de taille plus volumineuse (avec deux ou trois voies) et doit également posséder une courbe de réponse régulièrement descendante au-delà de 7 kHz, une courbe de TG la plus droite possible, un Waterfall à décroissance rapide au-delà de 500 Hz et une dynamique plus élevée ($L_p \text{ max} > 100 \text{ dB}$ à la distance d'écoute).

La directivité doit être plus marquée, mais toujours bien contrôlée (sans accidents). L'addition d'un caisson de grave est impérative, au moins pour la voie LFE, et doit être bien réglée, surtout en cas de bass management. Les cinq enceintes doivent également être identiques.

Pour une utilisation en mixage « cinéma », trois voies sont nécessaires et la courbe de réponse doit descendre régulièrement au-dessus de 2 kHz, avec une courbe de TG s'atténuant rapidement au-dessus de 1 kHz, un Waterfall à décroissance rapide au-dessus de 500 Hz et une dynamique encore plus élevée ($L_p \text{ max} > 105 \text{ dB}$ à la distance d'écoute). La directivité doit être très importante mais toujours bien contrôlée (sans accidents). Les enceintes surround peuvent être plus petites mais en plus grand nombre que les trois enceintes frontales. Le caisson de grave, uniquement utilisé en LFE sera suffisamment dimensionné afin de délivrer le niveau nécessaire (115 dB max) entre 20 Hz et 120 Hz, à la position de mixage.

L'avance temporelle au-dessus de 1 kHz est liée aux mesures statistiques effectuées sur un grand nombre de systèmes de diffusion en salle de cinéma. Cette caractéristique commune est liée à l'effet dispersif temporel dû à la plupart des pavillons utilisés. Les valeurs mesurées peuvent aller jusqu'à 3 ms dans certains cas. Cette avance temporelle peut toutefois être corrigée grâce aux processeurs actuels ou à l'évolution des technologies de conception des enceintes. Par conséquent, Cette courbe doit être ajustée en

fonction du circuit de production considéré. L'incidence sur l'audition, et donc sur le mixage, sera un effet de « son projeté » qui a pour conséquence une modification de l'équilibre spectral ainsi que des plans sonores (en particulier sur les voix).

Il est à noter que cette avance temporelle est dynamique et varie en fonction du niveau instantané : nous sommes donc en présence d'un phénomène non linéaire, ce qui représente la limite actuelle des possibilités de correction par processeurs.

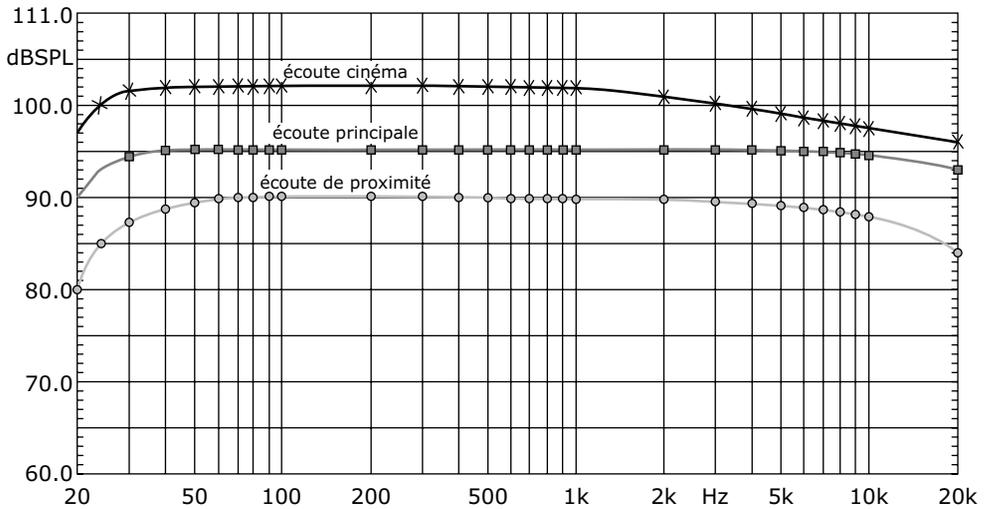


Figure 2.10 – Courbes de réponse type en amplitude/fréquence.

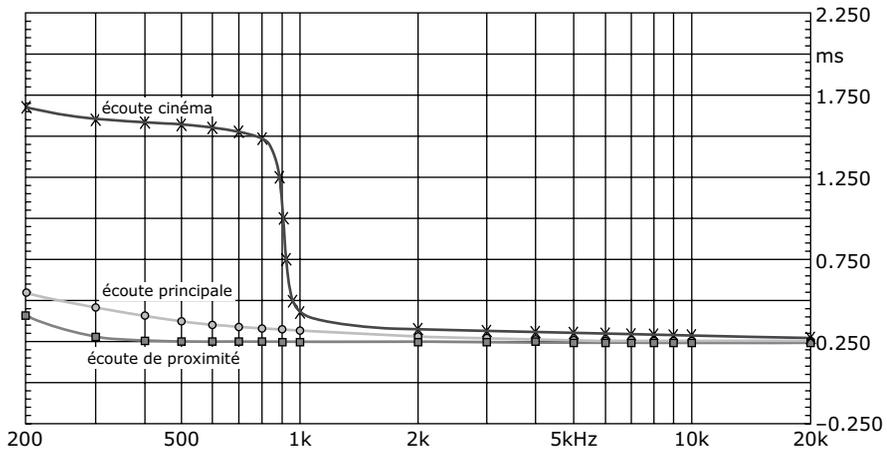


Figure 2.11 – Courbes de réponse type en retard de groupe/fréquence.

◆ L'écoute « client »

Celle-ci devrait être installée dans une salle acoustiquement plus vivante (semi-réverbérante). Elle est principalement utilisée pour une production CD ou DVD et peut être plus typée suivant l'utilisateur final. Pour la musique classique, le jazz, la musique ethnique, l'enceinte doit être neutre (courbe de réponse régulièrement descendante au-dessus de 10 000 Hz, courbe de TG la plus droite possible, Water-fall pouvant être moins rapide au-dessus de 5 000 Hz, dynamique assez élevée avec $L_p \text{ max} > 95 \text{ dB}$ à la distance d'écoute).

La directivité peut être large, mais toujours bien contrôlée (sans accidents). L'addition d'un caisson de grave est impérative au moins pour la voie LFE et toujours bien réglé surtout en cas de bass management. Pour la musique variété ou moderne, l'enceinte est identique mais peut avoir une courbe de TG descendante au-dessus de 2 kHz avec une variation de l'ordre de 0.3 à 0,5 ms.

Pour une production TV, celle-ci peut être calquée sur une écoute Home-cinéma associée à un écran de téléviseur grand public, avec un système large bande (une voie) associé à un petit caisson de grave avec bass management. La dynamique est évidemment moins grande avec un niveau maximum situé vers 85 dB max. Dans tous les cas, les cinq enceintes doivent être identiques.

2.5 L'influence du local

2.5.1 Réponse dans le grave, couplage physique

Dans les fréquences graves (en dessous de 200 Hz), la courbe de réponse de l'enceinte est complètement tributaire du local d'écoute. L'enceinte transmet plus ou moins d'énergie en fonction de son type, de la dimension du local et de son positionnement dans ce local. Il s'agit là d'un phénomène d'adaptation d'impédance acoustique entre le générateur (enceinte) et le récepteur (local), aggravé par la présence d'ondes stationnaires dans la cabine.

La courbe de réponse de l'enceinte devient très irrégulière (même si celle-ci est droite en champ libre !).

Par conséquent, seul un réglage par égalisation électronique et optionnellement, un traitement acoustique additionnel, permet d'obtenir une courbe de réponse plate et régulière au point d'écoute dans cette plage de fréquence. Si on désire en plus augmenter la zone d'écoute (sweet spot), on devra atténuer les ondes stationnaires du local par un traitement acoustique adapté. Des solutions plus ou moins complexes existent suivant le budget et le volume disponible.

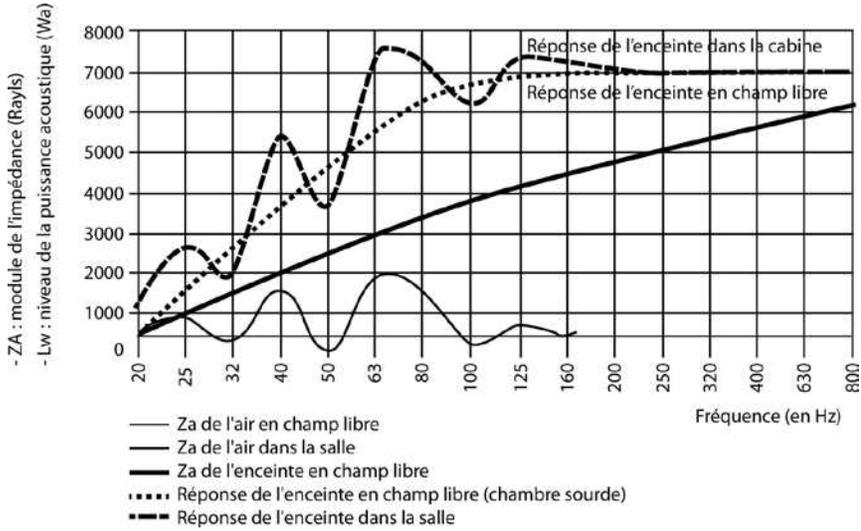


Figure 2.12 – Adaptation d'impédance acoustique.

En multicanal, le positionnement du caisson de grave (SUB) doit être réalisé en tenant compte de ce couplage afin d'éviter les accidents trop importants, en sachant, qu'un excès de couplage (par exemple : « bosse » de +10 dB) peut toujours être corrigé par égalisation, mais un manque de couplage (par exemple : « trou » de -10 dB) demande une attention particulière afin d'assurer la linéarité de la correction. Quelques fois, si ce « trou » est issu d'une onde stationnaire principale (mode propre) du local, la seule solution est de recourir à une correction acoustique permettant d'atténuer ce mode, à moins de disposer d'un correcteur électronique contrôlant avec précision l'amplitude et la phase séparément (filtrage FIR) et faire, en quelque sorte, de l'anti-signal.

En cas de **bass management**, la sommation correcte en amplitude et en phase des « satellites » avec le SUB doit être effectuée avec précision. Il faut tenir compte que le couplage à la salle de ceux-ci n'est pas du tout identique autour de la fréquence de raccordement. Par conséquent, le recours à une correction électronique complexe est incontournable afin d'harmoniser cette sommation ; chaque cas étant un cas particulier, le bass management s'élabore sur mesure et *in situ*. Contrairement aux idées reçues, quand le bass management est bien fait, la transition entre le SUB et les enceintes ne s'entend pas !

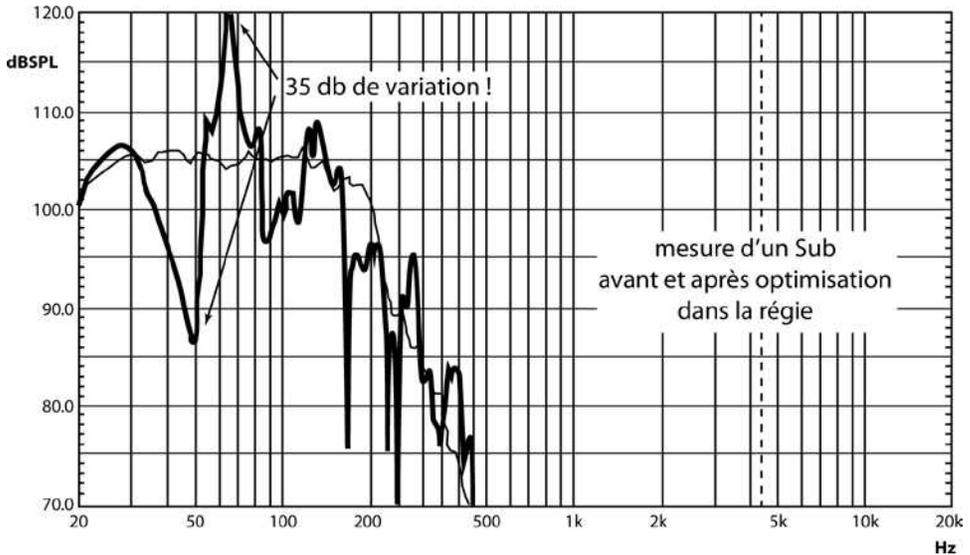


Figure 2.13 – Exemple d'égalisation d'un Sub à la position d'écoute.

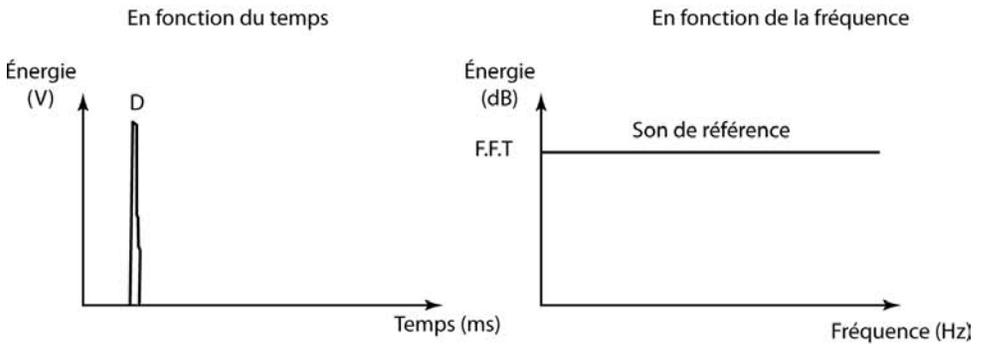
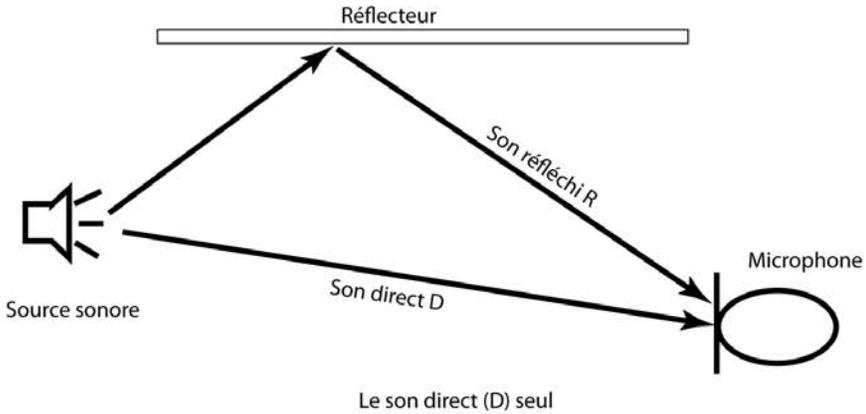
Dans tous les cas, il est important de bien dimensionner le SUB (HP + amplificateur) afin qu'il puisse accepter des corrections de niveau important sans distorsion. Rappelons qu'une correction de +10 dB équivaut à multiplier par 10 la puissance électrique délivrée !

2.5.2 Influence des premières réflexions, couplage psycho-acoustique

Au-dessus de 200 Hz, la dimension de la salle devient très grande par rapport à la longueur d'onde, par conséquent, l'enceinte se comporte comme en rayonnement libre et la salle n'a plus d'influence physique sur celle-ci.

En revanche, les premières réflexions (avant 10 ms) modifient notre perception de la source (son direct) : l'enceinte acoustique ne change pas, mais notre perception modifie le timbre, la localisation, la stabilité et la transparence de la source sonore initiale ; il s'agit de **l'effet de fusion**.

Ceci est dû au pouvoir séparateur de notre système auditif (constante de temps de l'oreille) qui varie, en fonction du retard relatif entre les premières réflexions et le son direct, et en fonction de la fréquence. Dans l'extrême aigu, nous sommes capables de séparer facilement celles-ci à toutes les fréquences (pas d'effet de fusion). Quand la fréquence diminue, les premières réflexions sont progressivement amalgamées au son direct (effet de fusion partiel) et nous percevons un détimbrage variable en fonction des accidents aux fréquences considérées ; il s'agit du **couplage psychoacoustique**. Dans le



Le son direct avec la première réflexion (D+R)

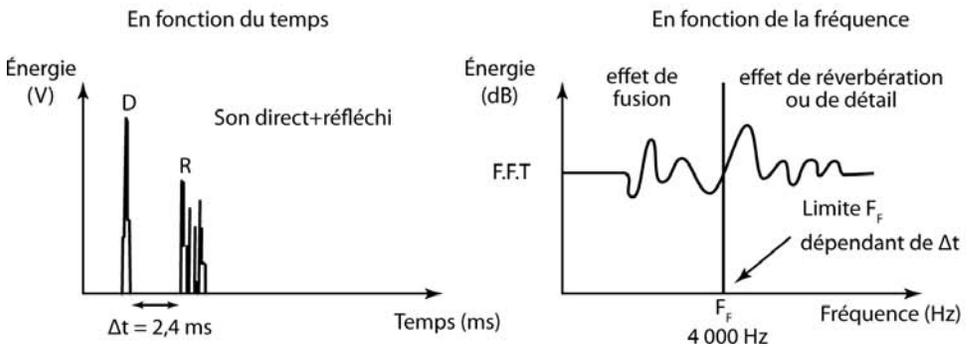


Figure 2.14 – Principe de l'effet de fusion.

grave, nous amalgamons toutes ces réflexions à la source initiale (effet de fusion total). Ces accidents, parfaitement mesurables par méthode MLS, peuvent être corrigés ou compensés (suivant la valeur de ces retards relatifs) grâce à une méthode d'égalisation dite **physiotemporelle**TM et suivant une courbe qui tient compte du temps et de la fréquence.

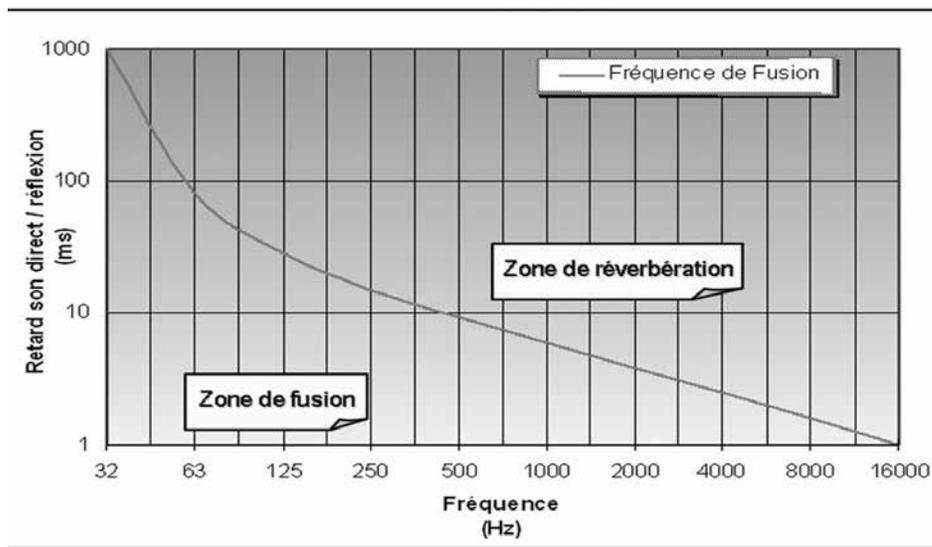


Figure 2.15 – Courbe d'effet de fusion/réverbération.

Cette courbe a été obtenue à la suite d'études effectuées en chambre anéchoïque à l'INA sur plus de cent cinquante sujets où ont été corrélées la mesure objective (par réponse impulsionnelle) et la perception auditive utilisant les critères d'analyses précédemment exposés.

Complément Web 2.12

Une illustration de cet effet est en écoute sur le site Web.

Cette égalisation particulière est très différente de celle obtenue en bruit rose, qui ne tient absolument pas compte de l'effet de fusion.

Elle nécessite l'utilisation de correcteurs paramétriques très précis en amplitude et en phase afin de compenser (ou corriger) ces premières réflexions. On entend souvent dire qu'il faut se méfier des égalisations modifiant la phase du signal, mais ce raisonnement fait abstraction des premières réflexions qui, par nature, déforment la phase du son direct perçu. Au contraire, il est nécessaire d'agir sur la phase en même temps que sur l'amplitude afin de supprimer ce défaut. Cette correction nécessite d'être très précis, d'où le choix impératif d'égaliseurs numériques à haute résolution.

Une expérience très simple à réaliser en cabine de mixage consiste à déplacer régulièrement une source sonore (voix, par exemple) de gauche à droite, en passant par l'enceinte centrale (pan pot). Sans **égalisation physiotemporelle**TM, le déplacement est peu précis et pas linéaire. Avec une égalisation physiotemporelleTM, la source glisse harmonieusement de gauche à droite, sans matérialisation auditive des enceintes.

2.5.3 Influence du champ réverbéré tardif, équilibre tonal et transmissibilité des mixages

Les réflexions tardives (après 10 ms), en dehors de l'effet de fusion, ne modifient plus le timbre ni l'identité du son direct. Il faut distinguer les réflexions arrivant entre 10 ms et 20 ms et le champ réverbéré (en général plus diffus) qui arrive après 20 ms.

L'énergie arrivant entre 10 ms et 20 ms vient modifier la précision en localisation des sources, sans altérer leur timbre. En mixage multicanal, on travaille souvent cette zone temporelle afin de créer une impression d'immersion sonore pour l'auditeur. Ce phénomène est utilisé différemment en sonorisations multipoints afin de modifier virtuellement la position d'une source et est plus connu sous le terme d'« effet Haas ».

L'énergie après 20 ms affecte la balance tonale (équilibre tonal global) et la profondeur du mixage. Il s'agit de l'effet de réverbération bien connu. En effet, cette énergie tardive diminue la transparence des sources sonores dans l'aigu en masquant les attaques. On a donc l'impression d'avoir moins d'aigus. Cet équilibre tonal subjectif peut être évalué en calculant le rapport en fonction de la fréquence, de l'énergie avant 20 ms sur l'énergie après 20 ms. Cette mesure peut être matérialisée par l'indice de clarté C20 (en dB), dont la formule est :

$$C20 = 10 \log \left(\int_0^{20 \text{ ms}} p(t) \cdot dt / \int_{20 \text{ ms}}^{\infty} p(t) \cdot dt \right)$$

On obtient donc une courbe significative de cet équilibre. Si elle est montante dans l'aigu, ce qui est le cas de beaucoup de cabine de mixage trop amortie, celui-ci sera surévalué en niveau et donc sous modulé dans le mixage. Chez l'utilisateur final (production CD, DVD ou TV), la courbe sera plus descendante dans l'aigu, donc avec moins d'attaques et le mixage paraîtra sous modulé dans cette zone spectrale.

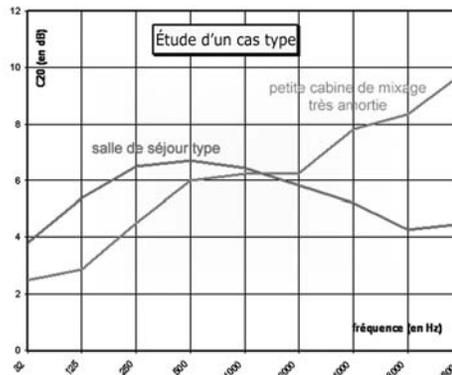


Figure 2.16 – Cas d'étude de la courbe C20/f.

Ce qui explique que la courbe de réponse en fréquence des enceintes monitoring soit descendante dans l'aigu afin de compenser ce phénomène et d'assurer la transportabilité du mixage chez le client final. Une autre solution élégante (applicable en stéréophonie à deux canaux, mais aussi en 5.1 sur les enceintes gauches et droites) consiste à utiliser des enceintes à directivité régulièrement croissante dans l'aigu (courbes de réponse décroissante en dehors de l'axe) et de faire converger les enceintes devant le point d'écoute. Ainsi, on arrive à corriger ce défaut (il s'agit d'une correction dans ce cas et non d'une compensation !) avec en plus un élargissement significatif de la zone d'écoute stable (sweet spot) grâce à un rattrapage acoustique de l'éloignement relatif par une augmentation inverse du niveau relatif dans le médium aigu (zone fréquentielle permettant la localisation). Évidemment, l'enceinte centrale reste dirigée sur l'auditeur, mais avec une égalisation différente des enceintes gauches et droites. À l'expérience, l'enceinte centrale fonctionne plus en source réelle que les autres, cela n'étant pas gênant. Les enceintes surround gauches et droites peuvent également converger derrière le point d'écoute pour les mêmes raisons.

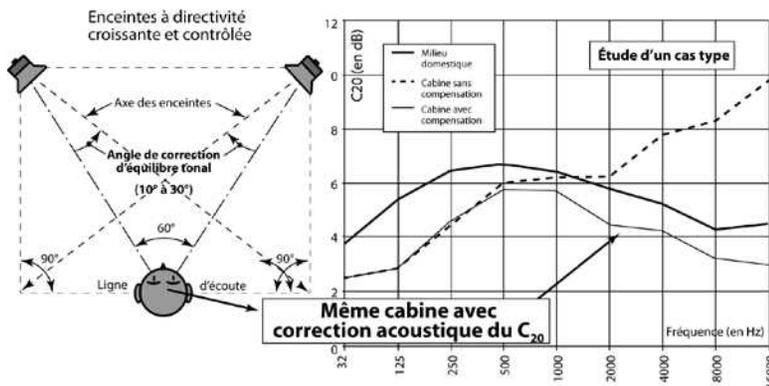


Figure 2.17 - Correction du C20 par la directivité des enceintes.

2.5.4 Le local et l'enceinte, synthèse globale

Cette synthèse globale montre l'influence du local sur le mixage et permet de juger de la complexité, mais aussi la logique de ses effets.

L'application des mêmes phénomènes de réverbération en sonorisation, montre l'interaction du local avec l'enceinte pour ce type d'utilisation.

Dans tous les cas et au-delà de 50 ms, l'énergie devient complètement nuisible et détruit l'intelligibilité ; fort heureusement, en cabine de mixage, il n'y a plus beaucoup d'énergie dans le médium aigu au-delà de ce délai.

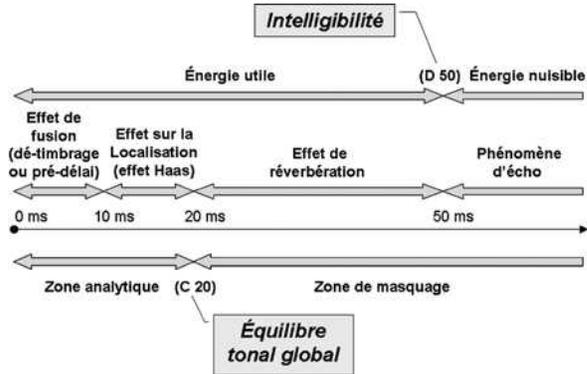


Figure 2.18 – L'influence du local d'écoute : synthèse.

2.6 Les différents types de cabines de mixage

2.6.1 Type « Tom Hidley »

Le célèbre concept Tom Hidley date des années soixante-dix. Il repose sur le principe d'une acoustique très « vivante » à l'avant de la cabine, avec les enceintes encastrées et une zone arrière très absorbante afin de réduire l'effet de masque et la perte d'intelligibilité. Quand le concept est bien réalisé, le RT60 est assez court avec un C20 relativement faible, garantissant une bonne transportabilité du mixage, mais avec un manque de transparence, voire une fausse transparence flatteuse (embellissement des attaques) correspondant à un son adapté à la mode et aux besoins de cette période.



Figure 2.19 – Studio Paramount Recoding Group.

Cette cabine est, par contre en générale agréable et peu fatigante.

Les enceintes avant devraient être très directives (ce qui n'est pas toujours le cas !) afin d'atténuer les premières réflexions, sources de dénaturation du son direct comme nous l'avons vu précédemment.

Conçu pour la stéréophonie, ce type de cabine est bien sûr complètement inadapté en multicanal. Notons que Tom Hidley a depuis, adapté et modifié son savoir-faire afin d'être parfaitement compatible au multicanal !

2.6.2 LEDE

Le concept **LEDE** (Live End, Dead End, qu'on peut traduire par *zone vivante et zone morte*), introduit en 1980 par Don Davis est encore utilisé à l'heure actuelle. Celui-ci a été également conçu pour la stéréophonie. Utilisable en Dolby Surround (années quatre-vingt-dix), ce concept n'est pas du tout adapté au multicanal actuel (5.1).

Son principe repose sur une acoustique très absorbante à l'avant (avec des enceintes encastrées, ou pas) et un volume arrière très diffusant afin de ne pas trop diminuer le RT60. En théorie, l'échogramme ne doit pas montrer d'énergie entre le son direct (0 ms) et 15 ms, puis une réverbération diffuse au-delà afin d'entendre avec précision les premières réflexions de l'enregistrement, sans les dénaturer par l'effet de fusion vu précédemment.

L'échogramme représente la décroissance de l'énergie en dB en fonction du temps. Ce calcul est effectué à partir de la réponse impulsionnelle au point d'écoute.

Malheureusement, la console de mixage, les baies périphériques et les patchs viennent perturber la zone morte de la partie avant et il n'est pas rare de voir apparaître de très fortes réflexions émergentes dans la zone entre 0 et 15 ms, provoquant des colorations par effet de fusion et distorsion d'espace en largeur comme en profondeur. De plus, le C20 en général élevé et ascendant dans l'aigu (s'il n'y a pas assez d'amortissement dans le grave) rend le mixage très difficilement transposable chez le client, si bien que le mixage réel se fait sur des écoutes de proximité et que les écoutes principales sont souvent utilisées comme écoutes « client ».

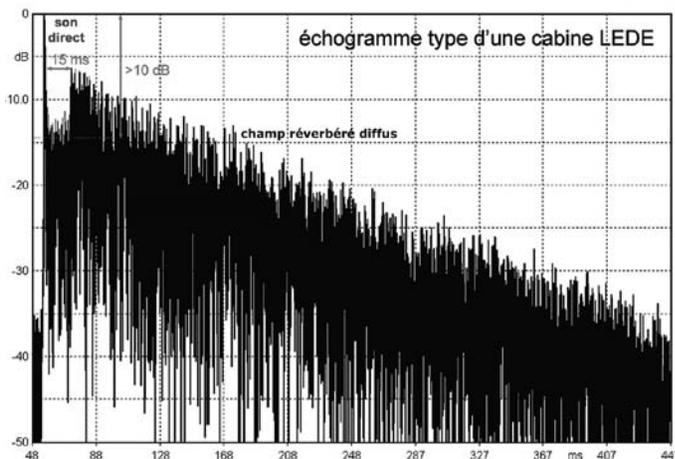


Figure 2.20 – Mesure de l'échogramme d'une cabine LEDE.

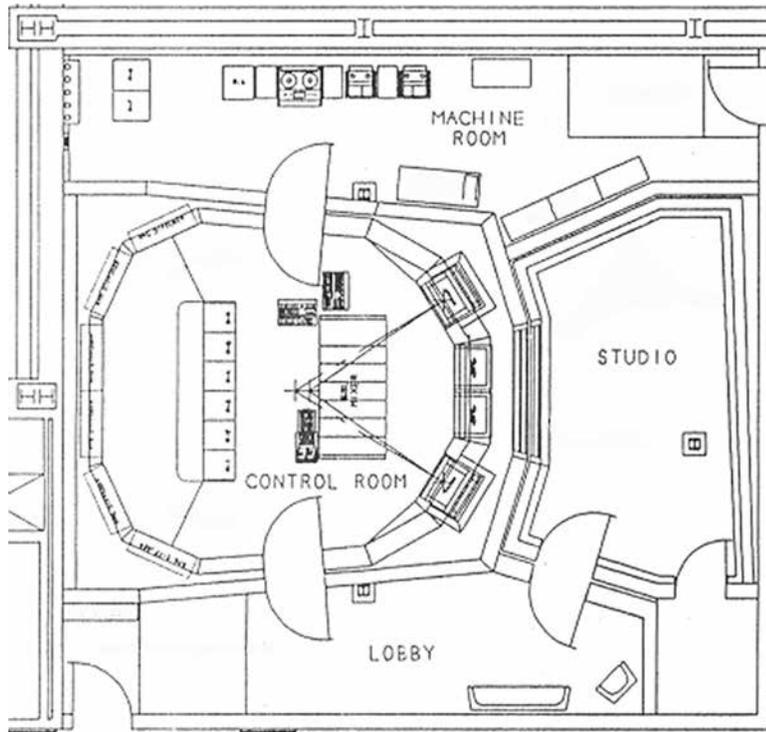


Figure 2.21 – Exemple de cabine LEDE.

De plus, l'écoute prolongée peut provoquer une fatigue auditive assez importante, si la rupture avant/arrière est associée à un RT60 global trop court.

2.6.3 Type « à amortissement réparti »

Le principe de cette cabine, exposé par R. Walker (BBC) dès les années quatre-vingt-dix est parfaitement pensé pour le multicanal actuel.

Basé sur les préconisations ITU (voir chapitre ultérieur sur les normes 5.1) et inspiré de l'idée initiatrice de la cabine LEDE, à savoir : éviter l'énergie dans les quinze premières millisecondes, R. Walker préconise une solution très différente qui consiste à maîtriser la direction des premières réflexions par la géométrie interne du local en alternance avec des zones absorbantes judicieusement choisies.

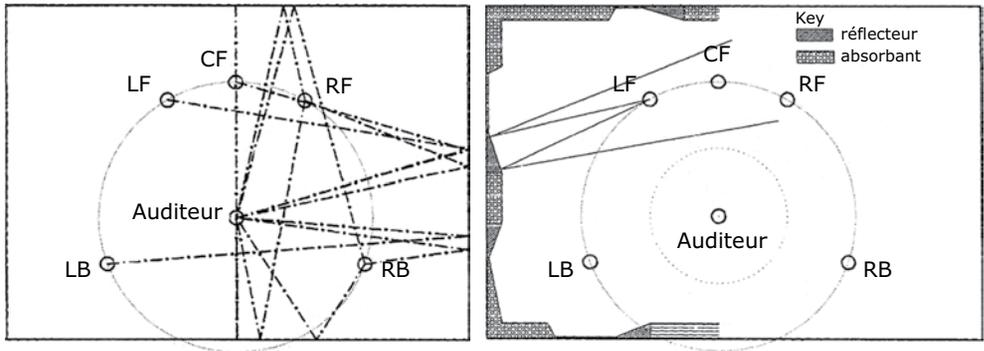


Figure 2.22 – Principe de la cabine à amortissement réparti (d'après les travaux de R. Walker).

Nous obtenons alors un C20 moyennement important du fait du peu d'énergie précoce reçue au point d'écoute, comparée à l'énergie tardive plus grande, mais avec un RT60 plus élevé et plus proche de celui d'un salon domestique grâce à la répartition homogène des réverbérations avant/arrière après 15 ms.

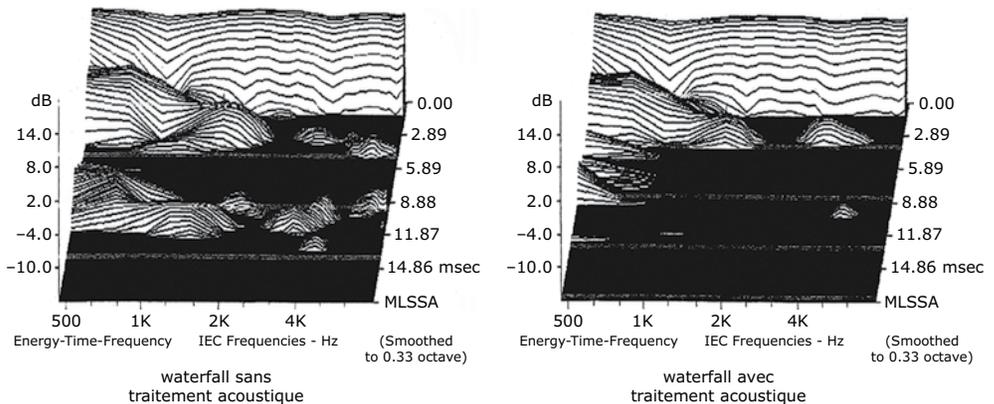


Figure 2.23 – Waterfall mesuré d'une cabine à amortissement réparti.

Cette cabine, non polarisée est donc particulièrement bien adaptée au multicanal en 5.1 (ou plus) à condition de bien maîtriser, au point d'écoute le positionnement des absorbatants et réflecteurs en fonction de la position des enceintes.

Les enceintes peuvent être apparentes, sur pied, suspendues ou encastrées, ce qui ne change rien au principe, le choix étant fait suivant le type d'enceintes choisies et le volume de salle dont on dispose.

La transmissibilité du mixage est meilleure mais le manque de précision des attaques peut être compensé par une courbe de réponse plus plate des enceintes dans l'aigu ou bien par l'utilisation d'enceintes plus directives (mais bien contrôlées) augmentant le C20. Par contre, la fatigue auditive sera faible dans la durée et le confort acoustique très appréciable.

2.6.4 Une nouvelle approche : la cabine à « diffusion répartie »

Inspirée des travaux de R. Walker (BBC Research) et ensuite développé par la société française d'ingénierie Taylor Made System (TMS) spécialisée dans le *sound systemic engineering*, un nouveau concept est proposé, basé sur la distribution harmonieuse des premières réflexions permettant d'obtenir une précision et une résolution de mixage (identique à une cabine très mate) en même temps qu'une grande transportabilité et une absence de fatigue auditive.

En effet, on considère toujours l'acoustique d'une cabine dans sa globalité bien qu'il n'y ait pas qu'une seule acoustique dans un lieu, mais une infinité d'acoustiques suivant le chemin acoustique considéré. Ainsi, il est intéressant de considérer et d'optimiser séparément le chemin acoustique enceinte/auditeur déterminant pour le mixage et le chemin acoustique auditeur/auditeur correspondant à l'ambiance de travail dans le lieu.

Le premier chemin doit être associé à un C20 élevé et aussi plat que possible en fréquence et avec un TR60 assez faible. Par contre, la répartition des premières réflexions doit être la plus régulière possible en fonction du temps avec aucune émergence notoire afin de ne pas « attirer » l'attention de l'ingénieur du son et colorer ainsi le son direct. Ce champ réfléchi précoce, suffisamment complexe et décorrélié du son direct, sera donc assimilable à un champ aléatoire ne se mélangeant pas auditivement aux premières réflexions contenues dans l'enregistrement. Il ne doit y avoir aucune rupture dans la répartition temporelle des premières réflexions sur les vingt premières millisecondes ainsi qu'un écart minimum de 10 dB entre le son direct et les premières réflexions. Cela suppose d'utiliser des enceintes à directivité croissante ou constante parfaitement bien contrôlées, ainsi qu'un encastrement sans rupture d'impédance acoustique avec la paroi les intégrant.

Le second chemin auditeur/auditeur peut avantageusement avoir un C20 plus faible et un RT60 plus élevé, correspondant à une acoustique plus conviviale à l'image d'un salon domestique, peu fatigant.

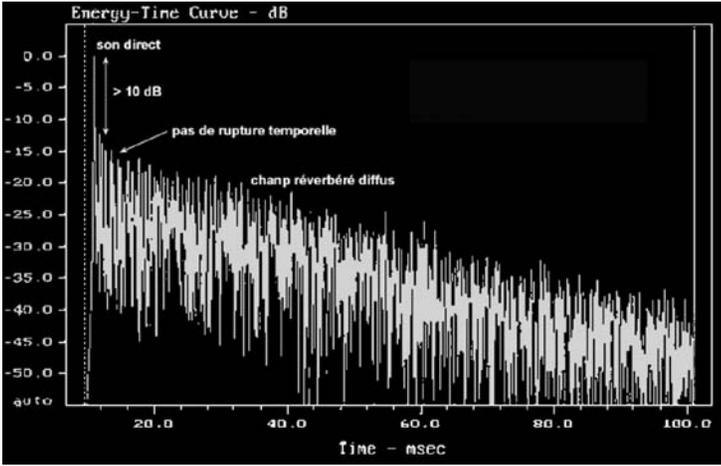


Figure 2.24 – Mesure de l'échogramme en mode « mixage » dans une cabine à diffusion répartie.

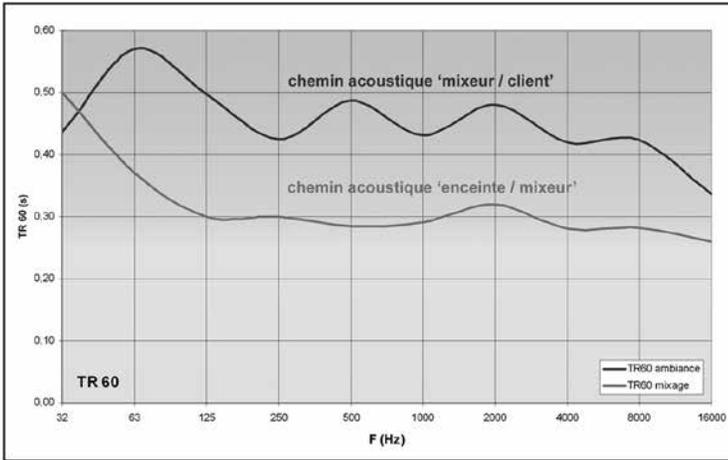


Figure 2.25 – Mesure du TR60 d'une cabine à diffusion répartie.

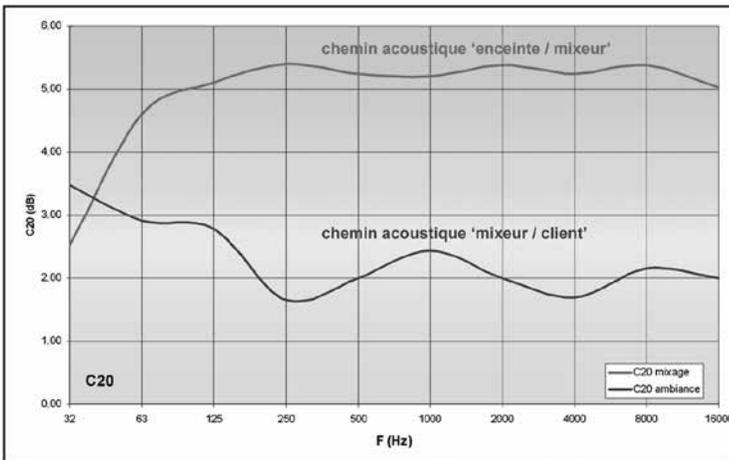


Figure 2.26 – Mesure du C20 d'une cabine à diffusion répartie.

La satisfaction de ces critères conduit à une forme de cabine assez complexe, comportant des diffuseurs 3D judicieusement placés, associée aux enceintes choisies précédemment et orientée de manière particulière (convergentes devant le point de mixage), mais respectant les normes ITU concernant les angles « enceintes/mixeur ». Nous avons vu que la qualité de l'encastrement joue un rôle primordial dans l'obtention d'une distribution régulière et sans accident des premières réflexions. Si les enceintes ne sont pas encastrées, leurs courbes de réponse et leur directivité doivent être irréfutables ainsi que leur positionnement et orientation. Dans les deux cas, la convergence des enceintes devant le point d'écoute permet de régler le C20 à la valeur désirée en même temps que d'augmenter la zone d'écoute (sweet spot).

Seule, la paroi arrière comporte une zone absorbante qui permet l'extinction de l'énergie tardive et donc de maîtriser le RT60.

Dans le grave, l'utilisation d'éléments absorbants (basstrap, panneaux résonnant, etc.) est souhaitable afin de réguler le C20 et le RT60, comme dans toute cabine.

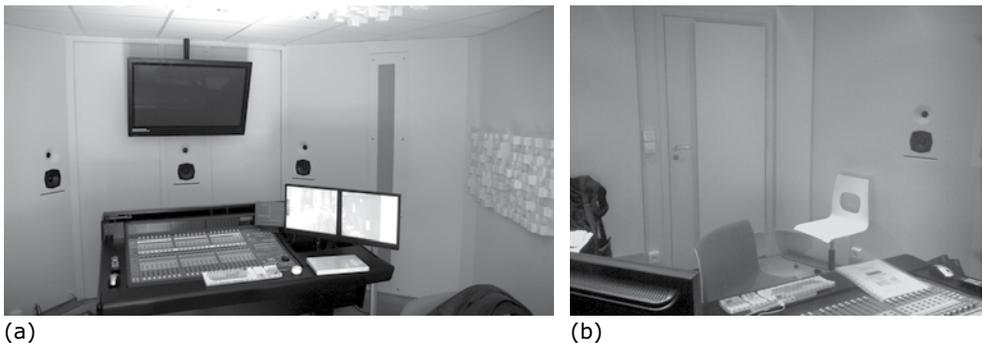


Figure 2.27 – Réalisation de cabines à diffusion répartie (avant et arrière de la cabine).

Bien entendu, cette cabine non polarisée est particulièrement bien adaptée au mixage multicanal en 5.1 ou 7.1.

2.6.5 Avantages et inconvénients en multicanal

Nous distinguons deux types de cabines : les cabines à acoustique polarisée et celles à acoustique répartie.

Les cabines à **acoustique polarisée** de type Tom Hidley et LEDE sont conçues pour la stéréophonie en 2.0. Elles n'ont pas de sens en multicanal 5.1 et le seul moyen de les utiliser malgré tout en 5.1 est de travailler en proximité, au détriment de la zone d'écoute alors très réduite, voire ponctuelle.

Les cabines à **acoustique répartie**, non polarisées de type « à amortissement réparti » ou « à diffusion répartie » sont pensées pour le multicanal 5.1 ou 7.1 tout en étant parfaitement compatibles en 2.0 et représentent donc la meilleure solution pour ce format. Le choix entre ces deux concepts se fera en fonction du besoin en dynamique

et en niveau sonore maximal, la cabine « à diffusion répartie », plus conservatrice de l'énergie, permet l'obtention d'un niveau de travail et d'une dynamique supérieure pour une même puissance d'enceintes, mais est plus difficile à mettre en œuvre et demande un ajustement acoustique et un réglage électronique précis, ainsi qu'un choix judicieux d'enceintes.

2.7 L'influence de l'amplificateur

2.7.1 *Les différentes technologies*

Nous distinguerons les amplificateurs analogiques et les amplificateurs dits « numériques ».

Complément Web 2.13

Ces technologies sont détaillées sur le site Web.

2.7.2 *Limites des mesures classiques, interactions avec les hauts parleurs et incidence sur l'écoute en 5.1*

La mesure d'un amplificateur n'est pas récente. Elle est traditionnellement réalisée en suivant les normes internationales IEC 60268-3 Ed3 avec des signaux sinusoïdaux, en fréquence glissante, celui-ci étant chargé par une résistance pure. On en déduit la bande passante à -3 dB, la courbe de réponse, le temps de montée, la courbe de phase, la distorsion, etc. Le problème est, comme pour les enceintes, que ces mesures sont très peu corrélées aux impressions d'écoute du fait de l'utilisation de tels signaux n'intégrant pas le régime dynamique couramment rencontré en utilisation réelle.

D'autre part, un haut-parleur se comporte très différemment d'une résistance pure et son effet microphonique (force contre électromotrice), renvoie en permanence dans l'amplificateur des courants générés par son propre mouvement. J'ai mesuré personnellement des courants crêtes dépassant trente ampères sur quelques millisecondes que l'amplificateur devait envoyer au haut-parleur afin de maîtriser le déplacement mécanique de la bobine mobile ! Si l'amplificateur ne peut pas générer cette énergie au bon instant, la dynamique reproduite est altérée, modifiant l'équilibre tonal et dynamique, mais aussi la définition et timbre et la transparence.

On perçoit donc mieux l'importance et le rôle subtil que joue l'amplificateur sur la réponse en transitoire d'un son complexe. Les mesures MLS, générant un signal composé d'un pseudo-bruit blanc (ou rose) dont le facteur de crête est de l'ordre de 3 dB, ne peut donc pas complètement mesurer cet effet parfaitement perceptible à l'oreille, très sensible aux variations rapides de niveau sonore. Il faut donc recourir à des mesures moins directes et des savoir-faire que nous ne détaillerons pas ici.

En revanche, on comprend tout l'intérêt d'avoir des câbles les plus courts possibles entre l'amplificateur et le haut-parleur, et donc l'intérêt des enceintes amplifiées. En multicanal, les enceintes amplifiées sont un plus, mais nécessitent un minimum de ventilation et bien sûr, une alimentation secteur.

Cependant, dans le cas d'amplificateurs « numériques », ces interactions amplificateur/haut-parleur sont beaucoup moins sensibles aux courants microphoniques générés par les haut-parleurs et ce type de technologie permet de placer un amplificateur multicanal plus loin des enceintes, dans la baie technique, par exemple, pour autant que le câble soit de bonne section.

Différentes configurations sont possibles, un choix souvent retenu est un filtrage passif pour le médium aigu et actif pour le grave. Un filtre passif (au-dessus de 500 Hz) peut donner d'excellents résultats à condition qu'il soit bien pensé.

Dans tous les cas, les amplificateurs doivent être identiques sur les cinq voies exceptées pour l'amplificateur du SUB qui doit être plus puissant, mais de même technologie, sous peine de modifier considérablement l'équilibre dynamique et tonal de construction avant/arrière ainsi que le type d'enveloppement.

2.8 Normalisation d'une écoute multicanal

2.8.1 Normes ITU-R BS.775.1 et 2

◆ Configuration géométrique

Il est très important de respecter les angles de $\pm 30^\circ$ pour FR et FL (Front Left, Front Right), ce qui permet une parfaite compatibilité avec la stéréophonie en 2.0. L'angle vertical doit être de 0° (à hauteur d'oreilles).

FC se trouve évidemment au centre (à 0°) horizontal et à 0° vertical.

SL et SR (Surround Left, Surround Right) sont angulés entre $\pm 100^\circ$ et $\pm 120^\circ$ horizontal et entre 0° et $+15^\circ$ vertical.

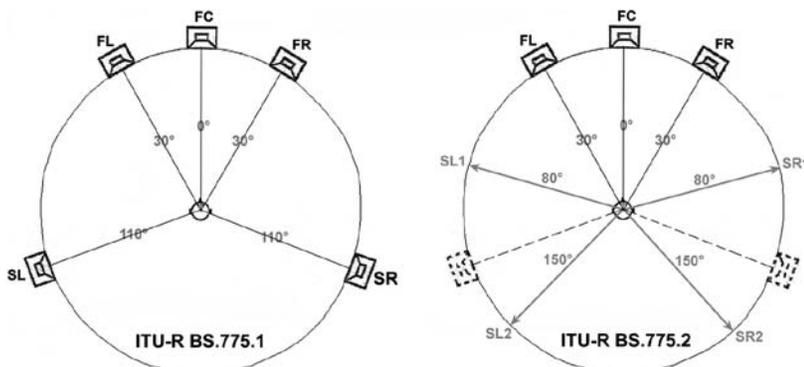


Figure 2.28 – Implantation géométrique du système d'écoute en 5.1.

Les enceintes surround (SL et SR) peuvent être dédoublées avec RS1/LS1 et RS2/LS2 situées à $\pm 80^\circ$ et $\pm 150^\circ$ horizontal. L'angulation verticale ne change pas.

En format 7.1, il suffit d'affecter les canaux SL, SR, BSL et BSR aux enceintes surround déjà prévues dans la configuration 5.1 avec enceintes surround dédoublées. Cette configuration est facilement compatible avec le format 5.1 sans changer la position des enceintes.

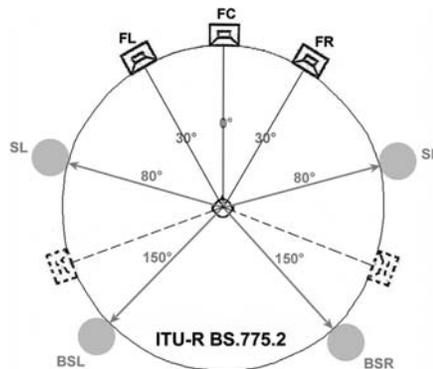


Figure 2.29 – Implantation géométrique du système d'écoute en 7.1.

D'autres préconisations peuvent être possibles, avec par exemple, des angles de 90° et 120° pour les enceintes surround. Comme nous l'avons vu, les sept enceintes doivent être identiques.

◆ Bass Management

Le bass-management (évoqué précédemment) est indispensable étant donné que la très grande majorité des installations grand public en est pourvue. En effet, si on dispose d'enceintes acoustiques ayant une bonne capacité à reproduire le spectre en pleine bande avec un niveau suffisant, il peut être tentant de mixer sans utiliser ce procédé. Cependant, il est important de comprendre qu'une sommation électrique (génération de filtres en peigne, détimbrage) n'est pas équivalente auditivement à une sommation acoustique (génération d'un effet d'espace et d'enveloppement). Par conséquent, si l'utilisateur final possède un système avec bass management, il est fortement conseillé de mixer dans cette configuration et le risque est grand de ne pas retrouver un mixage effectué sans bass-management sur une installation l'utilisant !

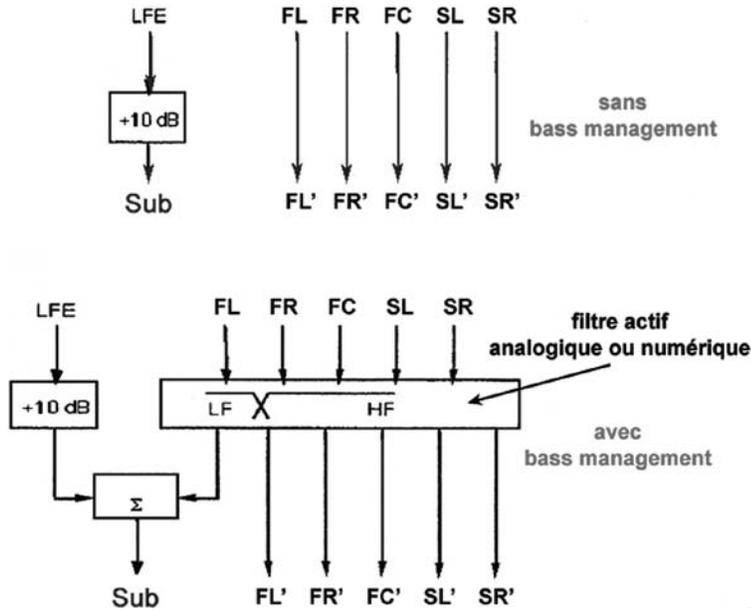


Figure 2.30 – Principe du bass-management.

La norme ITU prévoit d'aiguiller les fréquences basses (< 80 Hz en codage DTS et < 120 Hz en codage Dolby) dans le Sub Bass et, ainsi, de le mélanger au canal LFE (Low Frequency Effect). La pente des filtres électroniques utilisés doit être supérieure ou égale à 24 dB/octave. Une fois de plus, ce réglage doit être pratiqué avec soin par un opérateur compétent !

Le réglage doit être réalisé avec précision afin que la transition soit harmonieuse et naturelle.

L'augmentation de 10 dB du canal LFE sur le monitoring est obligatoire en général et provient historiquement de la norme imposée par Dolby. Cependant, ce réglage à +10 dB n'est pas utilisé pour les mixages musicaux purs (nous préciserons pourquoi par la suite).

◆ Niveaux d'alignement pour production TV, DVD et Blu-ray

Pour une production DVD ou Blu-ray, le niveau d'alignement, mesuré au sonomètre au point d'écoute doit être de 79 dB_C par canal et correspondre à un niveau électrique de sortie de console de -18 dB_{FS}. Le signal de calibrage est un bruit rose large bande.

Si les voies surround sont dédoublées, on veillera à ce que l'addition soit réalisée avec soin telle que [SR1 + SR2 = 79 dB_C] et [SL1 + SL2 = 79 dB_C].

Le niveau d'alignement de la voie LFE est de + 10 dB par rapport aux autres canaux (ce qui permet de sous moduler le signal LFE d'autant en mixage). On parle de gain de bande de fréquence (« In-band-gain »).

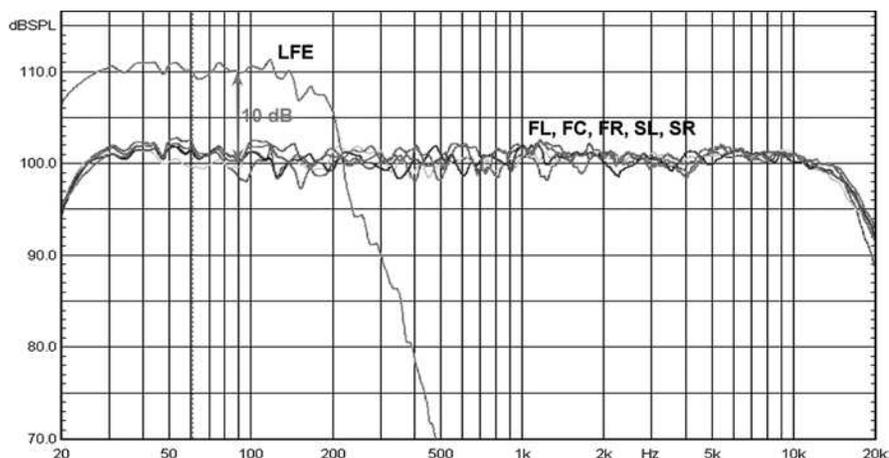


Figure 2.31 – Mesure de l'alignement d'une écoute 5.1 en production TV, DVD et Blu-ray.

ATTENTION !

- Si on dispose d'un **analyseur de spectre**, on mesurera directement les **+ 10 dB** sur l'écran de l'analyseur.
- Si on ne possède qu'un **sonomètre** et compte tenu des lois de sommations acoustiques en fonction des largeurs de bande, **le niveau d'alignement du LFE sera de 83dB_C**, c'est-à-dire le **niveau de référence augmenté de +4 dB**.
- L'augmentation de 10 dB s'effectue **uniquement sur la voie LFE** et non sur le caisson de grave (Sub Bass) en cas de bass-management.

◆ Niveaux d'alignement pour production Pure Audio Blu-ray, DVD audio et SACD

Pour une production Pure Audio Blu-ray ou SACD, les niveaux d'alignement seront les mêmes qu'en DVD et Blu-ray (79 dB_C par canal/-18 dB_{FS}/bruit rose) excepté pour le réglage de la voie LFE qui doit se faire au même niveau relatif (0 dB), mesuré à l'analyseur de spectre.

Si on dispose uniquement d'un sonomètre, le niveau d'alignement du LFE sera de 73 dB_C afin de tenir compte de l'étroitesse de la bande de fréquence de ce canal.

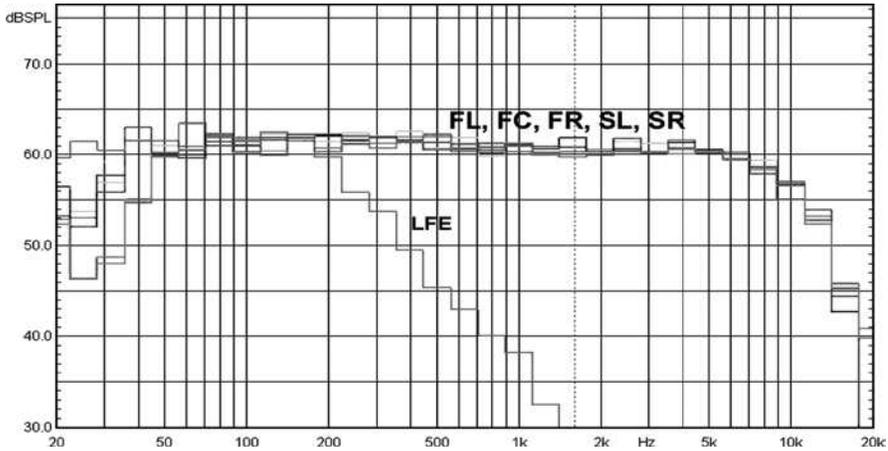


Figure 2.32 – Mesure de l'alignement d'une écoute 5.1 en production musicale pure (1/3 d'octave au point d'écoute).

Le bass-management s'effectue de la même façon et avec les mêmes précautions.

◆ Les préconisations acoustiques et électriques

Complément Web 2.14

L'ITU préconise des configurations géométriques, des réglages de niveaux et des performances acoustiques pour les cabines qui sont décrites et commentées sur le site Web.

2.8.2 Normes Cinéma (ISO 2969/1987(E)/SMPTE ST202-2010)

◆ Configuration géométrique

La configuration est très différente de la norme ITU réservée au Broadcast et aux applications grand public et n'est pas compatible avec le format 2.0.

L'angle de FR et FL est de $\pm 20^\circ$, tandis que le nombre d'enceintes surround, placées sur les parois latérales peut varier entre quatre et douze suivant la dimension de l'auditorium de mixage ou de la salle de projection.

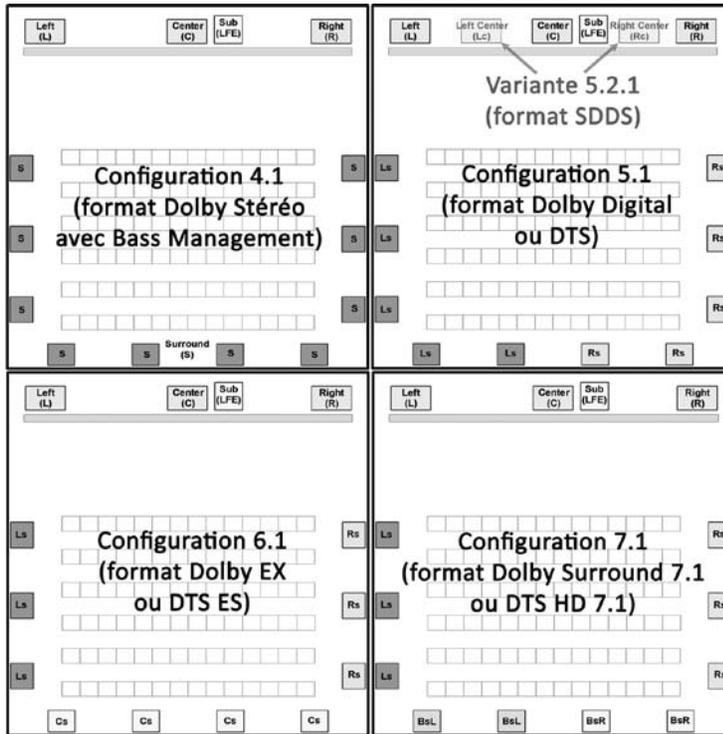


Figure 2.33 – Géométries du système de diffusion LtRt, 5.1 Dolby Digital / DTS, 6.1 Dolby Digital EX/DTS ES, 7.1 SDDS, 7.1 Dolby Surround, © CST.

◆ Niveau d'alignement

Les niveaux d'alignement et de travail sont également très différents, comparés aux normes ITU.

Pour les voies avant (FR, FC et FL), on veut obtenir 85 dB_C par canal au point d'écoute (derrière la console en mixage ou au $2/3$ de la salle en projection) pour un niveau de sortie de console de -20 dB_{FS} avec un bruit rose large bande.

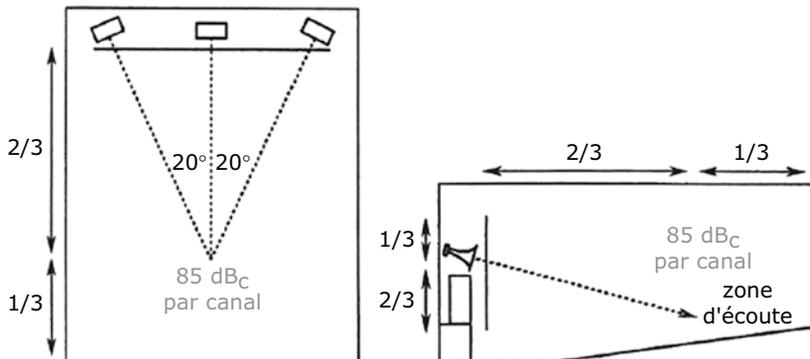


Figure 2.34 – Alignement des enceintes frontales.

Pour les enceintes surround, c'est la sommation de toutes les enceintes qui doit générer 85 dB_C dans les mêmes conditions (à -20 dB_{FS} et au point d'écoute).

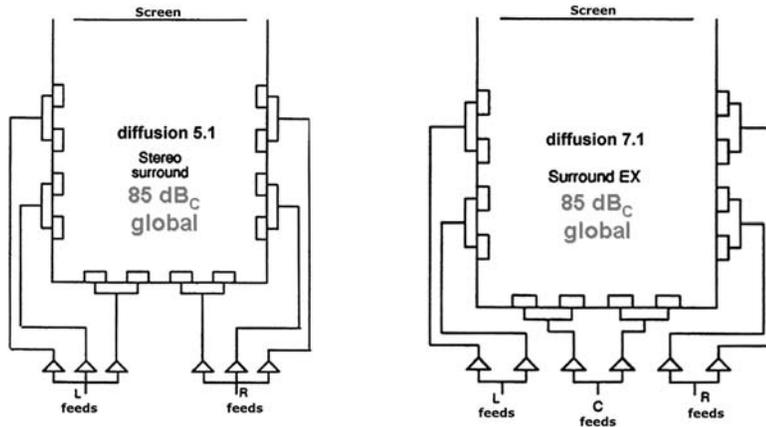


Figure 2.35 – Alignement des enceintes surround.

Comme pour la norme ITU (DVD, TV, Blu-ray), le niveau d'alignement de la voie LFE doit être à $+10 \text{ dB}$ par rapport aux enceintes principales, mesuré par un **analyseur de spectre 1/3 d'octave**. Si on utilise un sonomètre, le niveau de calibrage sera de 89 dB_C , pour tenir compte des largeurs de bande.

ALIGNEMENT TEMPOREL DES ENCEINTES FRONTALES

Il est à noter qu'aucune spécification ne prévoit de normaliser l'alignement temporel des enceintes frontales :

Celles-ci peuvent être alignées physiquement. La voie centrale est alors en avance temporelle par rapport aux deux autres.

Celles-ci sont alignées temporellement grâce à l'utilisation des retards inclus dans les processeurs actuels et le signal de chaque enceinte frontale arrive en même temps au point de réglage dans la salle.

Des expérimentations auditives ont montré que les mixages différaient notablement suivant ces deux écoles, notamment pour le placement spatial des voix.

◆ Courbes ISO X et SMPTE

La courbe de pondération dite ISO X utilisée pour égaliser au bruit rose l'ensemble enceintes + salle existe depuis les années soixante environ. Celle-ci a été mise au point empiriquement et auditivement en tenant compte de la technologie de l'époque. Les pavillons d'aigu très long étant très dispersifs temporellement (avance de l'ordre de 3 ms au-dessus de $1\,000 \text{ Hz}$) ; les écrans de cinéma étaient peu transparents acoustiquement (une atténuation de 10 dB dans l'aigu n'étant pas rare) ; l'acoustique des salles

était très absorbante et croissante dans l'aigu ; les mesures au bruit rose, discutées précédemment donnaient une image faussée de l'équilibre tonal perçu ; enfin la distance d'écoute était très grande. C'est ainsi que cette courbe ISO X a été élaborée !

Parfaitement d'actualité, l'ISO X est encore utilisée, bien que la technologie ait beaucoup changé (pavillons moins dispersifs et mieux contrôlés en directivité, écrans plus transparents acoustiquement, possibilité de moduler la réponse impulsionnelle du système, etc.). Il est donc important que la norme préconise de l'utiliser, avec des modifications en fonction de la dimension de la salle, ce qui donne trois courbes ISO X différentes.

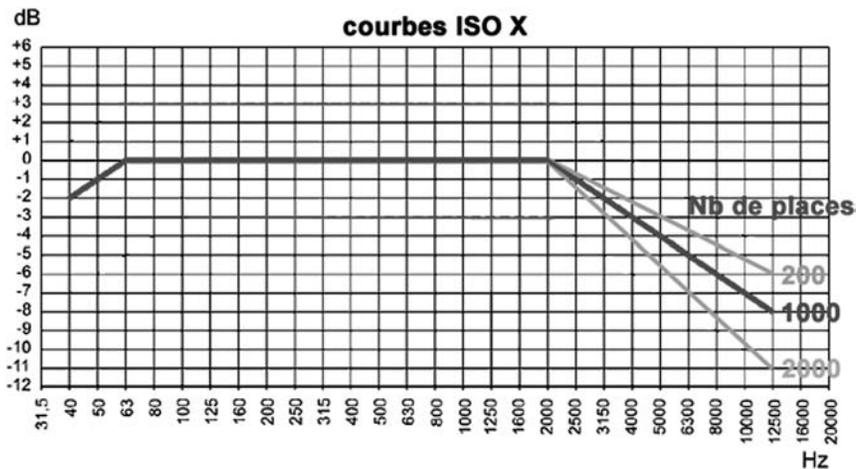


Figure 2.36 – Courbes ISO 2969 X.

Plus récemment, et en fonction de progrès des transducteurs utilisés en salle, la courbe SMPTE 202 m introduit une variante en fonction de la fréquence et permet de préciser l'allure de la courbe dans le médium aigu.

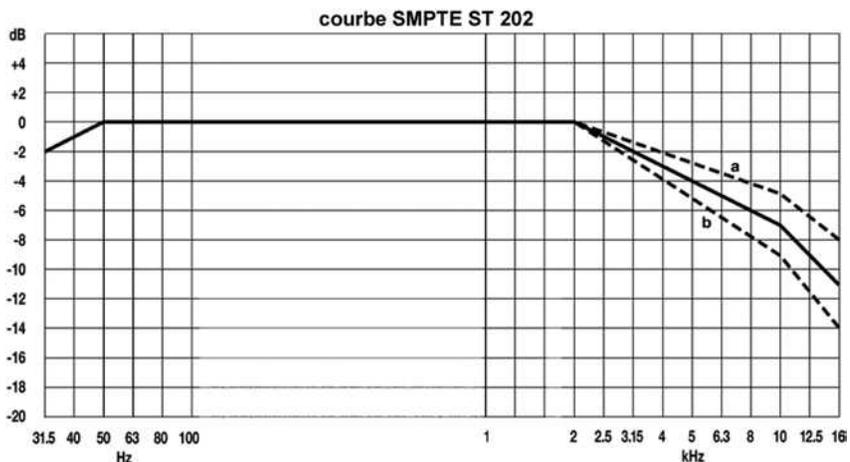


Figure 2.37 – Courbes SMPTE 202 (2010).

◆ Les préconisations acoustiques et électriques

Complément Web 2.15

Les configurations géométriques, les niveaux d'alignements et les caractéristiques acoustiques des auditoriums de mixage sont décrits et commentés sur le site Web, ainsi que les recommandations THX.

2.8.3 L'écoute au casque multicanal

Jusqu'à présent, l'écoute au casque permettait d'appréhender les critères d'équilibre, de définition et de réverbération. Cependant, l'estimation de la construction et de la stabilité de l'image sonore en stéréophonie 2.0, 5.1 ou 7.1 étaient impossibles avec un casque binaural classique, performant ou non.

L'écoute au casque se différencie de l'écoute stéréophonique sur enceintes par son impossibilité à reproduire une localisation identique en largeur et en profondeur. En effet, en perception binaurale naturelle, chaque oreille reçoit un signal de la source sonore.

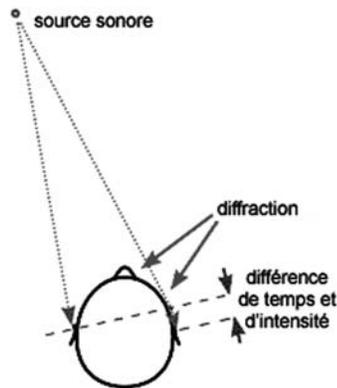


Figure 2.38 – Écoute binaurale.

Ces deux signaux sont caractérisés par une différence de niveau, une différence de temps mais aussi des déformations en amplitude et en phase très complexes dans la bande médium-aigu, étudiées et connues sous le nom de fonctions HRTF (*Head Related Transfer Function*).

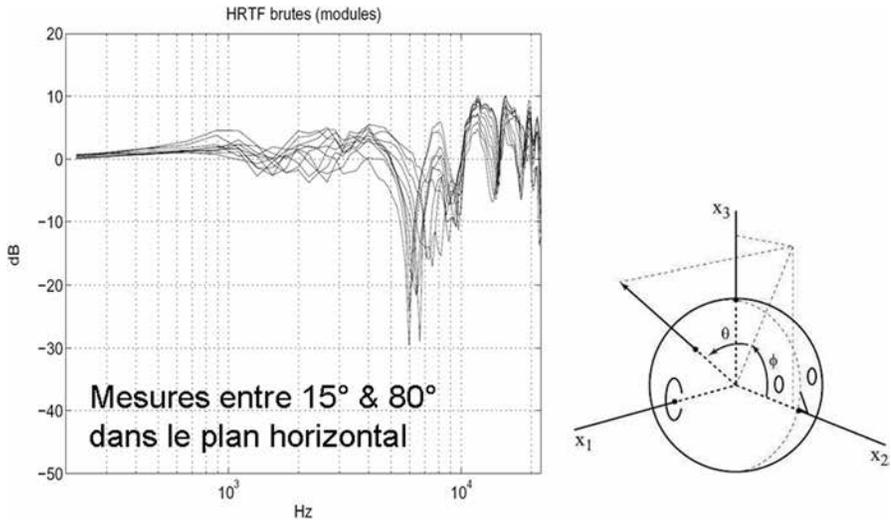


Figure 2.39 – Exemple de fonctions HRTF.

Ces déformations sont dues à la forme particulière de notre oreille, tête et torse et donnent une signature sonore différente pour chaque angle solide (horizontalement et verticalement) ; celles-ci sont identifiées et comparées à une base de données existante dans le cerveau, ce qui nous permet de localiser un son dans les trois dimensions. Lors de l'écoute sur deux enceintes situées à $\pm 30^\circ$ horizontal, le cerveau peut utiliser cette faculté et localiser une ou des sources virtuelles à l'intérieur de cet angle, par un artefact qu'on appelle la stéréophonie (du latin *stereo* qui signifie *relief*).

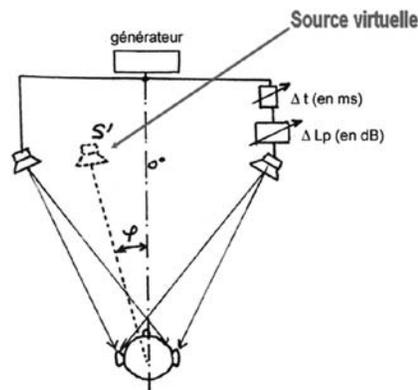


Figure 2.40 – Perception stéréophonique.

Un casque n'envoie qu'un seul canal à chaque oreille avec un angle fixe de 90° par rapport à l'axe de notre tête et se trouve donc incapable de restituer cet espace, d'où l'impossibilité de l'utiliser pour caractériser en construction une prise de son.

Il est évident, qu'en multicanal à cinq ou sept canaux, le problème est encore plus insoluble. De multiples essais ont été tentés pour reproduire cet espace horizontal que

propose ce format en fabriquant des casques à haut-parleurs multiples situés à l'avant et à l'arrière de la tête, sans véritable succès, par manque de précision et détimbrage.

La seule voie intéressante était de « reconstituer » les fonctions HRTF correspondant aux angles de restitution réels des enceintes, ce qui est à présent possible grâce aux processeurs de plus en plus puissants.

Trois problèmes restaient alors à résoudre :

- les fonctions HRTF ne sont pas universelles et varient d'un individu à l'autre ;
- lors de l'écoute naturelle, nous bougeons sans arrêt notre tête, souvent inconsciemment, afin de faire rapidement varier ces fonctions et donc localiser plus facilement la source ;
- la réponse en amplitude et en phase d'un casque dépend principalement du couplage physique avec notre oreille du grave à l'aigu, le volume de la cavité oreille/casque étant très faible.

Très récemment, ces problèmes ont été résolus et l'on trouve sur le marché des casques d'une part, munis d'un calibrage individuel permettant mesure et mise en mémoire de nos propres fonctions HRTF et, d'autre part, pourvus d'un capteur d'angle situé au sommet de la tête permettant le calcul en temps réel des variations HRTF liés à nos mouvements. Enfin, une égalisation fine permet de compenser les variations de niveau et de phase dues au couplage physique avec notre oreille.

En plus, le système proposé permet de mesurer et de mettre en mémoire les caractéristiques d'une écoute multicanal en 5.1 ou 7.1 réelle *in situ* et de reproduire celle-ci au casque, avec ses défauts éventuels et son environnement !

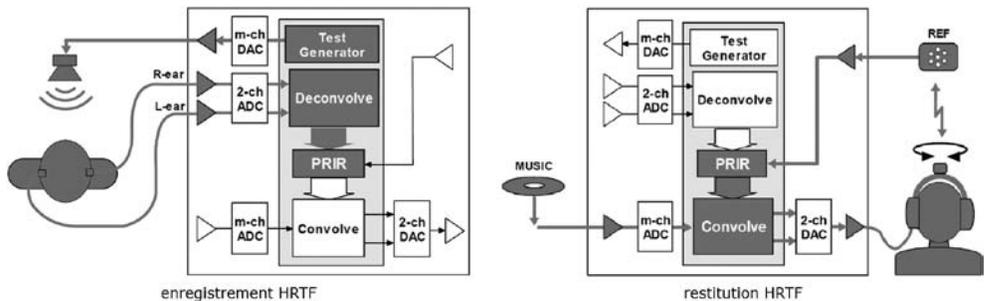


Figure 2.41 – Principe du casque Smyth SVS.

Cette technologie est évidemment basée sur des calculs utilisant la réponse impulsionnelle précédemment décrite.

Malgré tout, l'écoute au casque ne remplacera jamais une écoute sur enceinte, plus confortable et moins contraignante, mais peut être considérée comme complémentaire suivant le contexte et les besoins en production.

Radio France propose également des émissions en codage binaural avec un choix de fonctions HRTF adaptées aux différentes typologies d'oreilles.

2.9 L'écoute multicanal et l'image

2.9.1 Influence sur notre perception

A priori, on constate que la présence de l'image modifie notablement notre perception du son.

Ce chapitre est développé en détails sur le site Web.

2.9.2 Rapports dimensionnels entre les imageries sonores et visuelles

Le format de l'image ainsi que sa qualité imposent une distance de vision optimale en relation avec ceux-ci. Les préconisations ci-après seront valides pour une image de type HD.

Par rapport à cette distance optimale, si la distance d'écoute est trop courte, les distorsions angulaires (indice SDS) peuvent dépasser les 12 % pour des sujets sonores spatialisés virtuellement et le rapport entre différents sujets sonores sera modifié (relief et linéarité d'espace). À l'inverse, si cette distance est trop grande, la précision de localisation sera altérée.

Il existe donc des préconisations précises (étudiées par G. Steinke : *Paper International Tonmeister Symposium*, novembre 2005) permettant d'optimiser ces combinaisons en mixage TV et en mixage Cinéma et représentées dans les **figures 2.42, 2.43 et 2.44**.

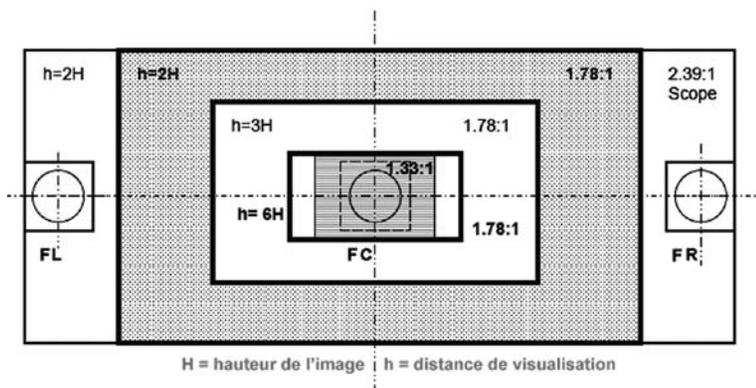


Figure 2.42 – Rapport hauteur d'image et distance de visualisation.

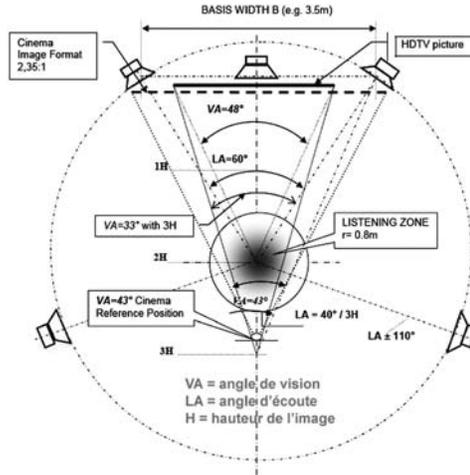


Figure 2.43 – Configuration pour une distance d'écoute $h = 2H$.

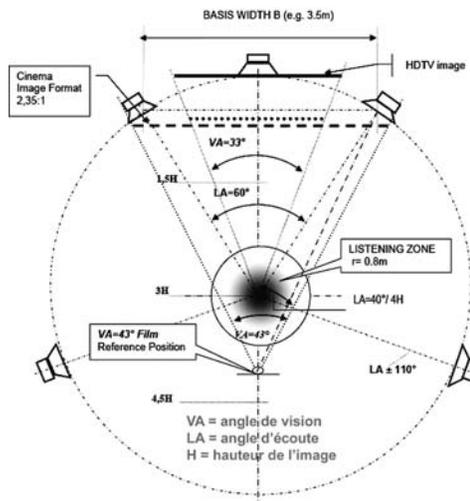


Figure 2.44 – Configuration pour une distance d'écoute $h = 3H$.

Le fait d'augmenter le nombre de canaux arrière (7.1) ou simplement le nombre d'enceintes surround en ajoutant de la décorrélation permet en pratique une augmentation de la zone d'écoute très appréciable en mixage à l'image.

2.10 L'écoute domestique en multicanal

2.10.1 De la HI-FI au Home-cinéma

Avec la démocratisation de la stéréophonie, les années soixante-dix ont vu naître commercialement la Haute-Fidélité. Les débuts de cette technologie se bornaient principalement à recréer des sources réelles localisées dans les enceintes. Assez rapidement,

les mélomanes et amateurs éclairés ont souhaité une reproduction plus naturelle, avec l'envie d'assister à un concert, poussant en avant les technologies de plus en plus sophistiquées et faisant appel à des prises de son dites **acoustiques naturelles**. Les sources réelles ont fait place à une imagerie sonore reproduisant des sources virtuelles beaucoup plus riches en émotions. L'espace stéréophonique s'étalait non seulement en largeur, mais aussi en profondeur. L'équilibre spectral et dynamique est devenu de plus en plus grand.

Poussé par le label THX, la montée du home-cinéma fut alors inévitable dans les salons. Les audiophiles hifistes cèdent peu à peu la place aux cinéphiles. Mais ces deux mondes ont eu bien du mal à s'unifier. En effet, les qualités recherchées semblaient assez incompatibles, les audiophiles recherchant avant tout la qualité et les nuances musicales alors que les cinéphiles étaient sensibles au niveau et à la dynamique. Les matériels dédiés à ces deux segments commerciaux n'avaient rien en commun et semblaient même incompatibles.

LE LABEL THX VERSION GRAND PUBLIC

Notons plusieurs évolutions et extensions de ce label dédié au home-cinéma grand public :

- D'une part, le THX Ultra (pour grandes salles) et le THX Select (version économique pour petites salles) en 5.1 et garantissant une grande puissance sur tous les canaux avec des niveaux de 105 dB sans distorsion.
- D'autre part, le THX Ultra 2 le THX Select 2, plus élitistes et préconisant une restitution en 7.1 canaux.

Une autre particularité est de proposer des enceintes surround dipolaires permettant d'obtenir une immersion accrue dans le champ acoustique et l'emploi de diffuseurs 2D et 3D sur les parois latérales de la salle.

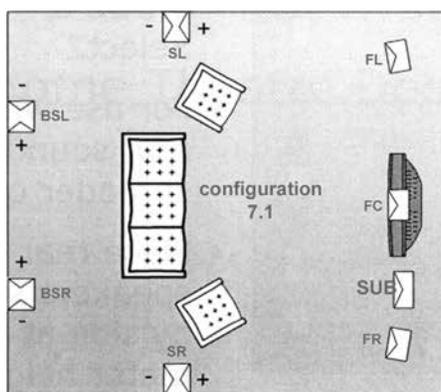


Figure 2.45 – Configuration d'une installation home-cinéma THX.

Les enceintes Back Surround (BSL et BSR) sont alimentées par le même canal de mixage, mais le processeur THX inclus dans le préamplificateur home-cinéma assure une décorrélation temporelle de celui-ci afin de créer un champ arrière plus diffus et non focalisant, améliorant ainsi l'immersion.

Ce système selon THX est très intéressant en home-cinéma, mais il n'est pas adapté pour la reproduction audiophile en HI-FI multicanal (SACD et Blu-ray music).

Actuellement, grâce aux progrès des mesures électroacoustiques, les systèmes home-cinéma et les systèmes HI-FI se sont réconciliés et la norme ITU rend compatible les deux types utilisation avec un même matériel permettant des réglages et optimisations adaptées (grâce au recours à des processeurs numériques de plus en plus évolués et pertinents). La télévision HD facilite sans doute le développement et la convergence de ce marché qui sera relayé dans les années à venir par les systèmes multimédias intégrant aussi les jeux vidéo et Internet. Les supports Blu-ray permettent également de satisfaire les mélomanes les plus exigeants ainsi que les cinéphiles avertis. Les liaisons (fibre optique) et serveurs à grand débit vont également amplifier cette mutation.

Actuellement une installation audiovisuelle domestique de ce type se configure selon le schéma suivant :

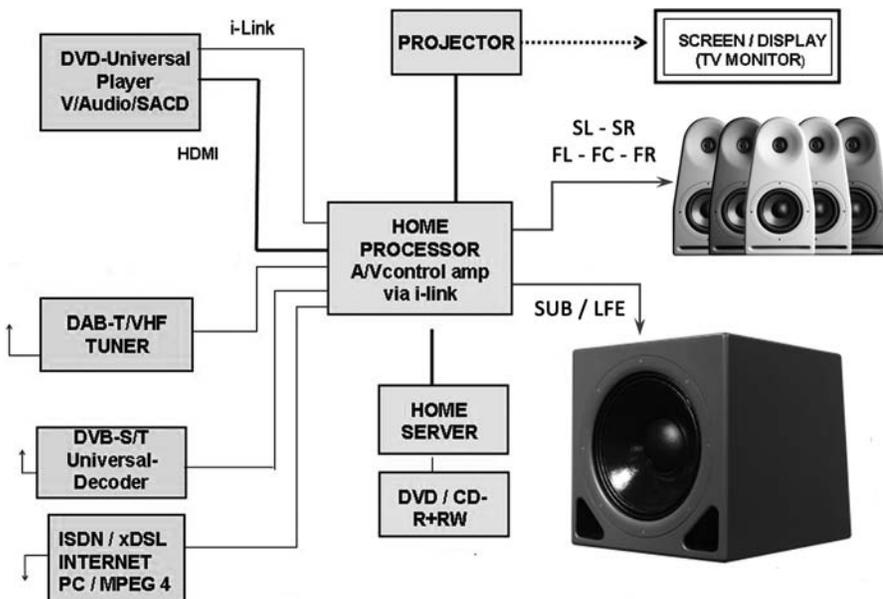


Figure 2.46 – Synoptique d'une installation domestique moderne.

Du point de vue de l'équilibre dynamique et de la transparence, les supports et réseaux numériques rejoignent et dépassent les performances objectives et subjectives des

meilleurs systèmes analogiques grâce aux formats à hautes résolutions du type PCM 24/96 et aux algorithmes de compression de dynamique de plus en plus performants.

Cependant, il est à craindre que l'évolution de ces technologies génère deux marchés différents et divergents : le multimédia de masse avec des sons très compressés privilégiant la quantité à la qualité et l'audiovisuel hi-tech de haut de gamme visant un public socialement plus érudit, à l'image d'une société élitiste à deux vitesses...

Il est important de comprendre que le fait d'habituer les jeunes générations à écouter des sons compressés avec une fausse dynamique (MP3, Smartphone, ordinateur, etc.) modifie l'apprentissage du langage en l'appauvrissant et par voie de conséquence risque de modifier leur manière de penser et le développement de leur esprit critique. Des études très sérieuses mais peu connues ont été menées à ce sujet.

ÉCOUTE HI-FI, HOME-CINÉMA ET MULTIMÉDIA

- Le home-cinéma, apparu dans les années quatre-vingt-dix a créé une diversion par rapport au monde de la HI-FI. Les motivations étant différentes, les systèmes d'écoute sont également très divergents, voire incompatibles dans certains cas.
- Le niveau sonore et la dynamique sont primordiaux chez les cinéphiles alors que le timbre, la transparence et le relief sonore sont essentiels pour les hifistes !
- Actuellement, l'introduction de processeurs de plus en plus sophistiqués ainsi que le développement des mesures par réponse impulsionnelle tend à favoriser la convergence et la compatibilité des deux mondes. Les systèmes haut de gamme actuels permettent à la fois une écoute fine et dynamique. Même les écrans actuels (plasma, LED, etc.) incorporent des systèmes audio de qualité bien supérieure à leurs homologues de la fin du XXe siècle.
- On assiste également à l'entrée du monde du jeu vidéo et de l'informatique de réseau dans l'environnement home-cinéma et HI-FI, sous la forme de serveurs très sophistiqués. Ceux-ci représentent sans doute l'avenir de ce marché.

2.11 Calibrage d'une écoute et optimisation par corrections électroniques

2.11.1 Correction ou compensation ?

Avant d'aborder le calibrage d'une écoute tel que nous l'avons vu précédemment, il est nécessaire de s'assurer de la neutralité de chaque enceinte associée à son environnement. Après avoir optimisé au maximum l'acoustique de la régie pour obtenir un écho-gramme régulier avec le moins d'accidents possibles, ceci en fonction des contraintes physiques (et budgétaires) du local, nous sommes amenés à envisager des solutions

électroniques afin de « maximaliser » la qualité de notre écoute par rapport à nos besoins et suivant le type de production.

Il y a deux solutions possibles, souvent confondues, et qui ne sont pas équivalentes en niveaux de performances :

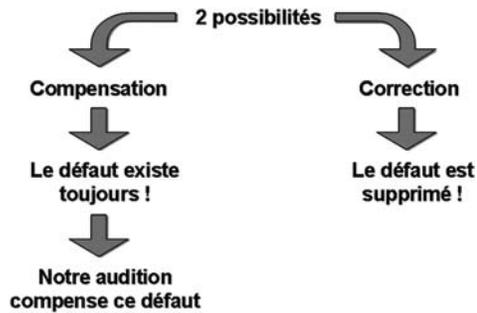


Figure 2.47 – Correction ou compensation ?

- les **corrections** corrigent les défauts en les supprimant réellement.
- les **compensations**, compensent les défauts vis-à-vis de notre audition, souvent par l'addition d'autres défauts psycho-acoustiquement opposés, mais ne les suppriment pas physiquement.

L'exemple de la **figure 2.48** montre la compensation de l'équilibre spectral perçu d'une enceinte par une analyse et égalisation en 1/3 d'octave. Si nous effectuons après correction, une analyse bande fine en mesure MLS, nous voyons que les défauts de couplage physique au local (dans le grave) et les défauts propres à l'enceinte associée aux premières réflexions (dans le médium-aigu), existent toujours et auront une incidence sur la transparence et la localisation des sources. L'équilibre tonal perçu est compensé mais les défauts d'origines restent non corrigés.

2.11.2 Les différents moyens électroniques, filtres analogiques, filtres numériques IIR, filtres numériques FIR, les optimiseurs de champ acoustique

Dans les différentes solutions de correction et/ou de compensation, nous distinguerons les filtres analogiques et les filtres numériques. Les filtres analogiques, un peu obsolètes à l'heure actuelle, ne sont plus beaucoup utilisés, au bénéfice des filtres numériques.

Dans les filtres numériques, il nous faut distinguer les filtres IIR (Infinite Impulse Response) des filtres FIR (Finite Impulse Response).

Les **filtres IIR** sont des filtres dont l'amplitude et la phase sont liées (une correction amplitude/fréquence génère obligatoirement un déphasage temporel associé) permettent de faire des compensations et/ou des corrections limitées, suivant les processeurs et la méthode utilisée.

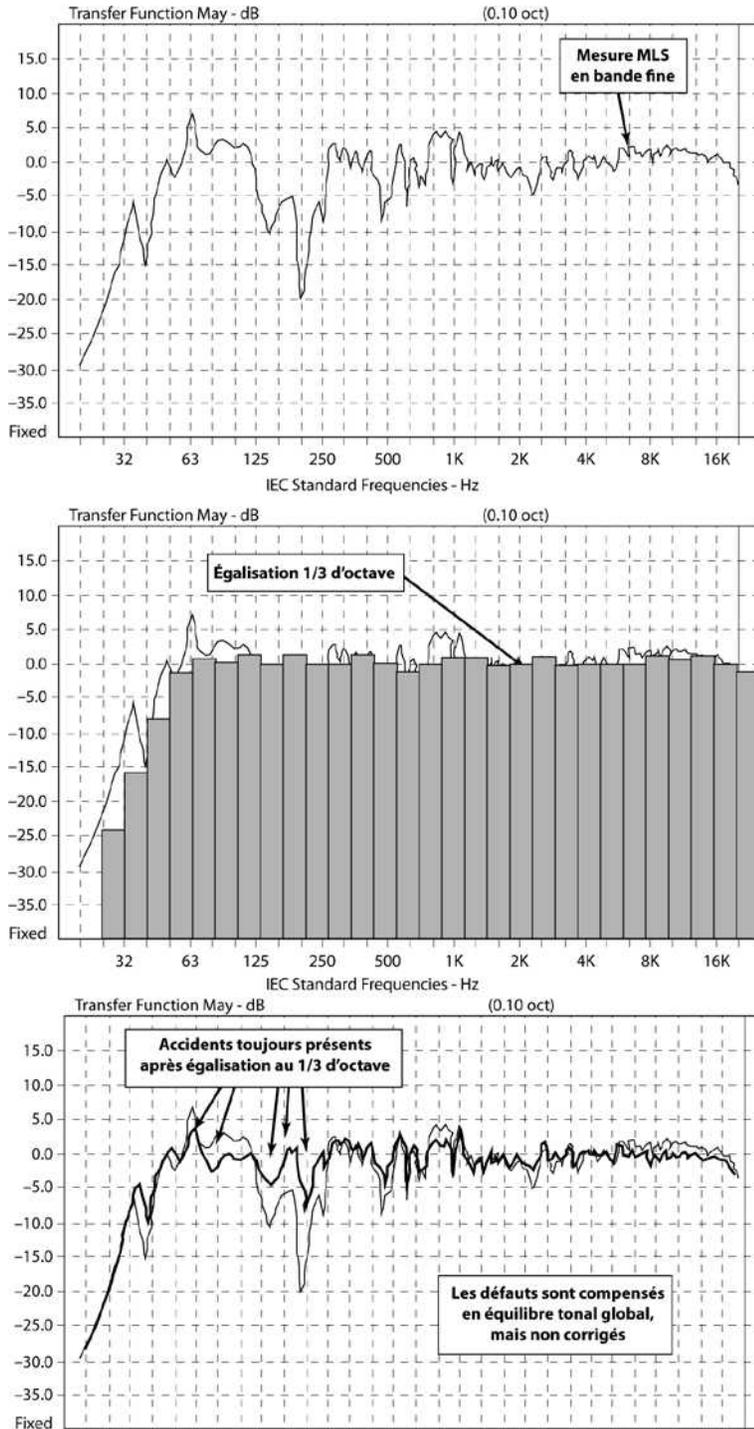


Figure 2.48 – Exemple d'égalisation au 1/3 d'octave : compensation.

Les **filtres FIR** sont des filtres dont l'amplitude et la phase sont indépendantes (possibilité de correction en amplitude/fréquence sans déphasage temporel) permettent de faire des corrections beaucoup plus puissantes et précises et donc plus pertinentes.

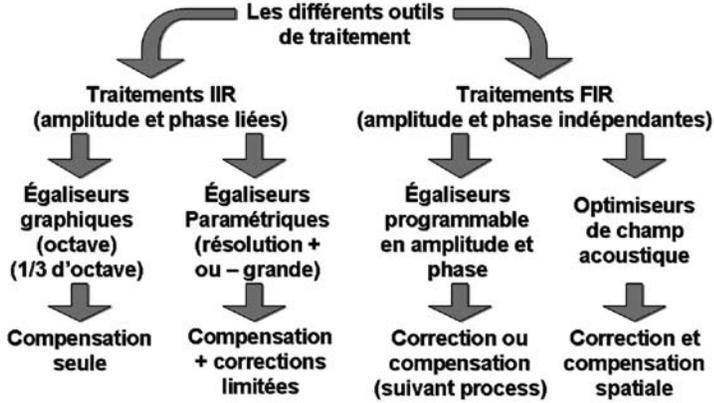


Figure 2.49 – Corrections et compensations par moyens électroniques.

Ces filtres sophistiqués peuvent être associés à un simple microphone de mesure omnidirectionnel, à une **sonde à directivité croissante**, appelée **physio-angulaire™** et développée par la société **Taylor Made System**, qui permet de rejeter les réflexions arrières non amalgamées au son direct par notre audition.

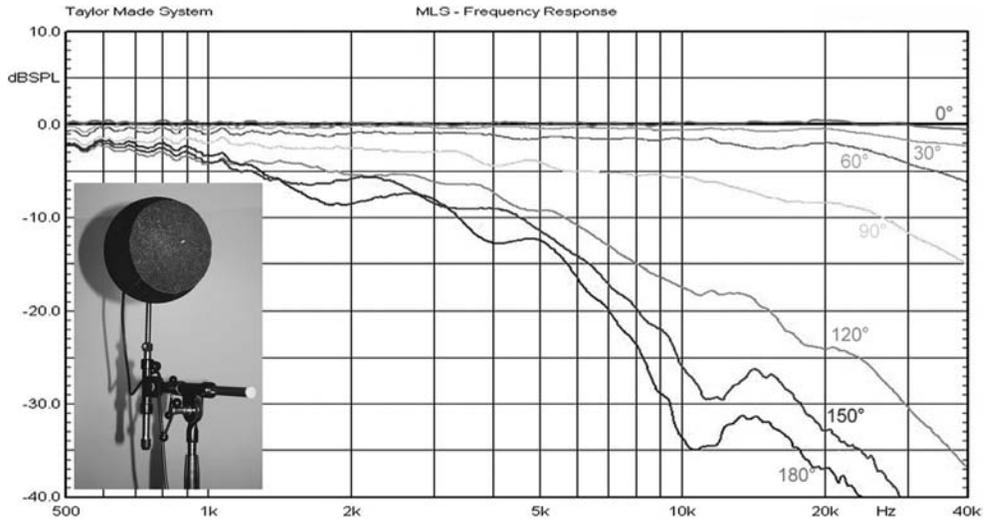


Figure 2.50 – Sonde à directivité croissante.

On peut aussi utiliser une **sonde tridimensionnelle**, développée par la société **Trinnov Audio** qui est assimilable à une sonde intensimétrique. Dans ce dernier cas, nous pouvons parler de correction de champ acoustique ou de correction et compensation

spatiale. La différence entre la pression acoustique et le champ acoustique réside dans le pouvoir séparateur de ce dernier suivant la provenance et la direction de la source incluant ses premières réflexions.

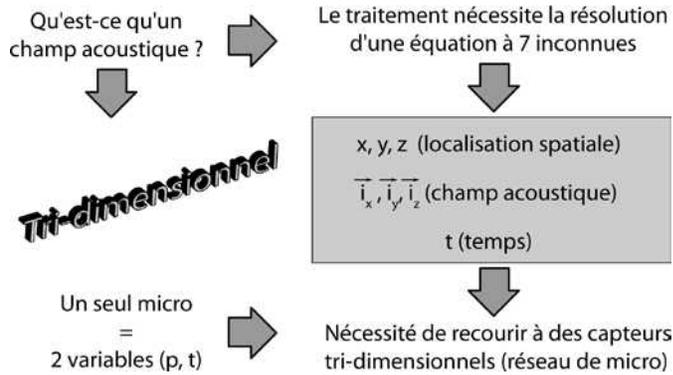


Figure 2.51 – Mesure et calcul d'un champ acoustique.

INTENSITÉ ACOUSTIQUE I_a

Contrairement à la pression acoustique qui est mathématiquement un scalaire, l'intensité acoustique est un vecteur et donc inclut des informations de direction en plus de celle de niveau :

$$I_a \text{ (en w/m}^2\text{)} = W_a / S$$

où I_a est un flux d'énergie dans une direction donnée (W_a étant une puissance acoustique et S la surface élémentaire considérée).

Ceci permet une correction sélective suivant la direction et même de synthétiser un champ acoustique réel provenant d'une source virtuelle (qui n'existe pas physiquement !).

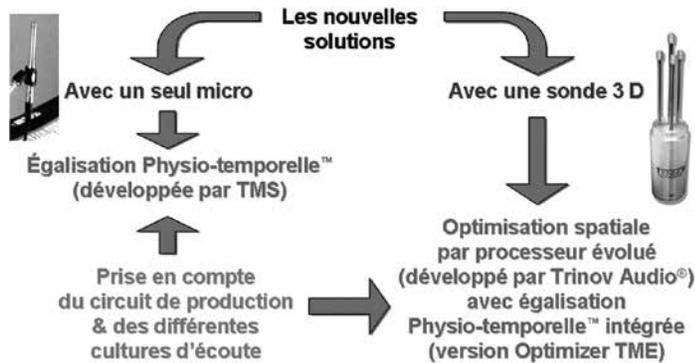


Figure 2.52 – Optimisation avancée d'un champ acoustique.

À ce stade, il faut noter la différence avec le principe de la stéréophonie, où l'image de la source sonore virtuelle naît d'un artefact de notre audition (ni la source, ni le champ acoustique n'existent !).

Dans le cas de la synthèse d'un champ acoustique, la source n'existe pas, mais le champ acoustique est bien réel !

Par exemple, une enceinte acoustique mal placée au regard de la norme ITU (pour des raisons d'ergonomie ou d'encombrement), peut être repositionnée correctement et apparaître très plausiblement au bon endroit. On parle alors de repositionnement ou « remapping » de la source sonore à la position d'écoute ; la seule restriction étant que ce repositionnement n'est optimal qu'au point d'écoute (sweet spot) avec néanmoins un élargissement de cette zone plus ou moins grand en fonction du contrôle de la directivité des enceintes et de l'environnement propre de celles-ci.

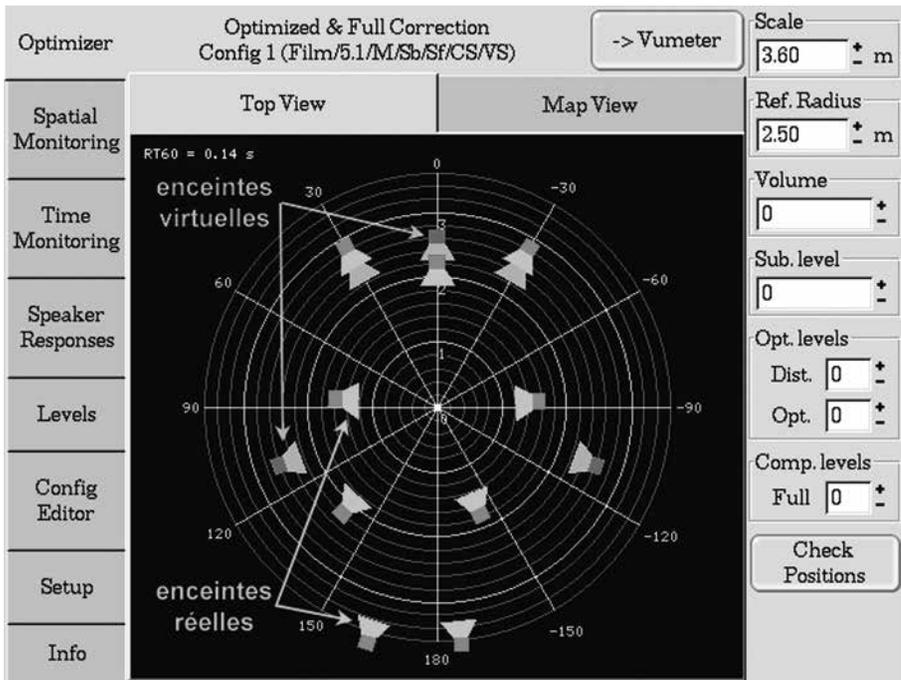


Figure 2.53 – Exemple de remapping par l'Optimizer de Trinnov Audio.

Cette application est particulièrement intéressante dans les cars régis ou les cabines exigües, où le placement des enceintes est très restrictif.

OUTILS DE TRAITEMENT ET DE CORRECTION

D'autres outils de traitements numériques existent. Sans être exhaustif, on peut citer :

- Processeur London de chez BSS (technologie IIR);
- Processeur DME de chez Yamaha (technologie IIR);
- Processeur Lake de chez Lab Groupen (technologie FIR).

Tous ces processeurs évoluent rapidement en fonctionnalités et puissance de calcul, ce qui et va dans le sens des besoins actuels induits par l'évolution et la pertinence des systèmes de mesures vus précédemment.

2.11.3 *Prise en compte des circuits de production et des cultures d'écoute*

Les compensations et corrections en vue de rendre neutre le système d'écoute (enceintes + salle) ne sont pas suffisantes en enregistrement et mixage. Une deuxième phase de traitement est nécessaire afin de garantir la compatibilité avec l'utilisation finale, compatibilité qui dépendra du type du circuit de production.

On peut donc distinguer différents circuits de production :

- La production musicale sur support CD, DVD ou Blu-ray;
- La production TV;
- La production cinématographique;
- Les productions spécifiques (muséographie, parcs d'attraction, planétarium, etc.);
- Les productions médias globales.

Grâce aux mesures par réponses impulsionnelles et aux processeurs FIR, il devient possible de mesurer et de reproduire une courbe cible correspondant aux circuits de production type, ceci d'une manière statistique. D'autres ajustements, plus subtils doivent tenir compte des différences d'effets de masque dus aux acoustiques différentes entre le lieu de mixage et le lieu de diffusion. Ces ajustements peuvent se faire par modification du traitement acoustique et/ou par introduction d'un « anti-signal » généré par les enceintes via un processeur FIR évolué. Dans tous les cas, ce travail ne peut se faire qu'avec la complicité et la confiance de l'ingénieur du son et avec le recul suffisant. On pourrait comparer cette coopération avec celle du pilote et du préparateur dans le domaine de la Formule 1 en automobile.

2.11.4 *Diagnostic et test d'un système d'écoute 5.1 par méthode entièrement auditive à l'intention des ingénieurs du son*

Il existe de nombreux CD et DVD de réglage destinés aux amateurs HI-FI, aux cinéphiles et aussi aux ingénieurs du son. Généralement, ces productions sont, ou ludiques, ou complexes, ou les deux et traitent souvent en plus du réglage de l'image. Dans tous

les cas observés dans les applications professionnelles, il est nécessaire de disposer d'un sonomètre étalonné pour faire les réglages afin de régler le niveau d'alignement absolu.

La méthode, proposée sur le site Web, n'utilise que notre audition et fait appel à ses caractéristiques physiologiques. En effet, notre sensation de niveau sonore varie avec la fréquence du son. Cette variation est donc incluse et compensée dans le contenu audio de la session de test. Le niveau de référence est important puisqu'il détermine la courbe d'isotonie utilisable pour les compensations auditives. Un instrument de musique coutumier des exploitants son a été utilisé pour ce faire. Le seul instrument nécessaire est donc l'oreille exercée de l'ingénieur du son.

Par ailleurs, celui-ci ne disposant en général que de peu de temps, la méthode privilégie ici l'efficacité et la rapidité d'analyse plutôt que l'aspect ludique, la finalité de ce test étant de diagnostiquer rapidement le système d'écoute et de se rassurer quant au niveau de confiance du mixage réalisé.

◆ **Diagnostic d'une écoute 5.1 en session Protocols**

Complément Web 2.16

Le test auditif est proposé intégralement sur le site Web avec possibilité de téléchargement de la session.

Les objectifs sont :

- diagnostiquer rapidement la qualité d'une écoute 5.1 sans recourir à la mesure ;
- Calibrer l'écoute non seulement en niveau, mais aussi en polarité et en alignement temporel ;
- Égaliser sommairement chaque voie en tenant compte de son environnement (effet de fusion) et de l'acoustique tardive de la salle (équilibre spectral lié au C_{20}).

2.12 Perspectives et évolutions

2.12.1 L'augmentation du nombre de canaux (7.1 à 22.2 ou 23.1)

Partant du principe de base du multicanal 5.1 et la capacité des processeurs sans cesse en augmentation, une extension du nombre de voies est tentante afin d'augmenter l'immersion sonore de l'auditeur et accéder ainsi à des effets spéciaux de plus en plus sophistiqués.

Ces différents formats ayant déjà été présentés précédemment, on se limitera à les citer sans être exhaustif le format 7.1, le format 10.2, le format 22.2, etc.

Chacun présente des avantages et évolutions par rapport au 5.1, mais reste basé sur le principe de la stéréophonie et de la création de sources acoustiques virtuelles

rendues de plus en plus stables et précises grâce au rapprochement des enceintes entre elles.

En même temps, la complexité et les contraintes physiques des installations limitent le développement commercial et la vulgarisation de ces formats, dont le champ d'application privilégié demeure les diffusions publiques comme les cinémas, parcs d'attractions, salles de spectacle, etc.

Pour les salles de cinéma, on voit apparaître des voix célestes très intéressantes pour créer des effets sonores et des ambiances de zénith. C'est le cas pour certains planétariums, parc d'attractions et salles IMAX.

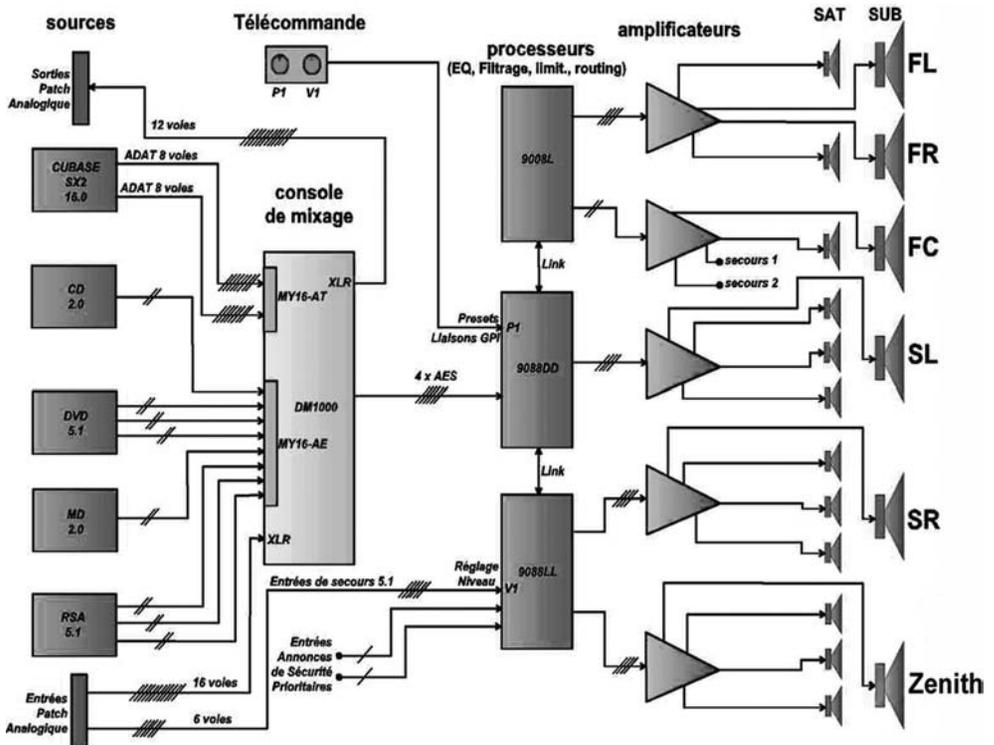


Figure 2.54 - Synoptique d'installation en 6.2 avec voie zénith au Planétarium de la Cité des Sciences à Paris (étude Taylor Made System).

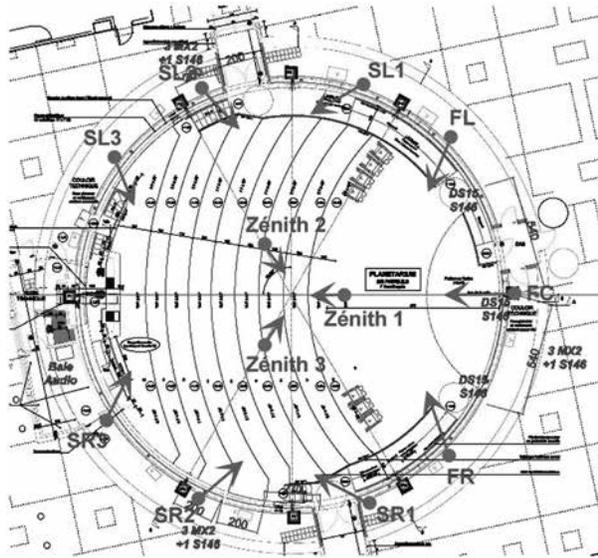


Figure 2.55 – Schéma d'implantation du système son.

◆ **Imm Sound en 23.1**

La technologie 14.1 à 23.1 récemment introduite par **Imm Sound** présente l'intérêt d'une grande immersion (éventuellement très réaliste) dans le champ réverbéré recréé. Pour ce faire, elle s'appuie uniquement sur des sources de diffusion en hauteur dans les trois dimensions (au zénith et derrière l'écran), mais aussi sur une extension du nombre de voies latérales et arrières, le tout associé à un processeur permettant de gérer les effets et le placement des sources. Ce système de diffusion est compatible avec les formats 5.1 et 7.1, via la gestion du processeur.

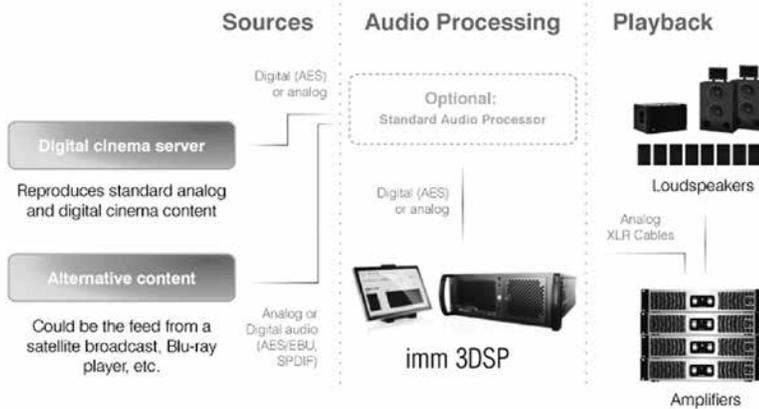


Figure 2.56 – Principe du système.

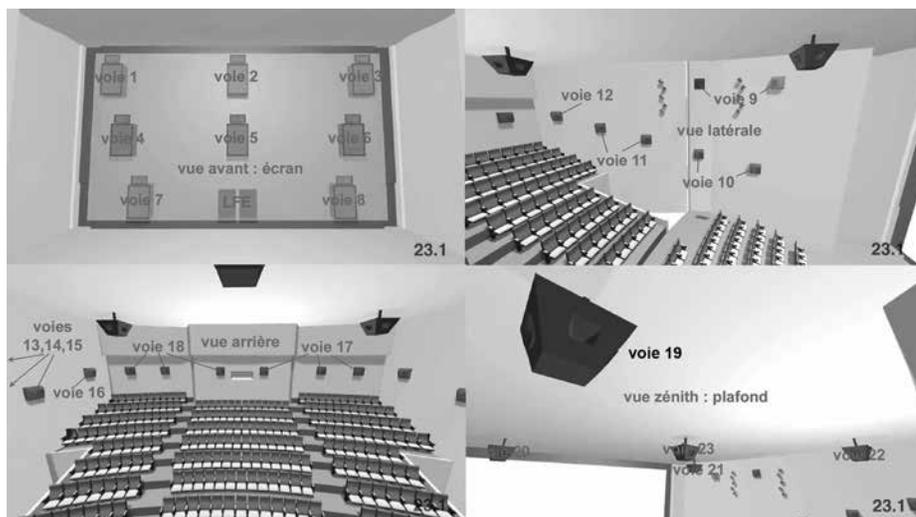


Figure 2.57 – Implantation des (23 +1) enceintes.



Figure 2.58 – Exemple de réalisation 23.1 à Düsseldorf.

Nous imaginons facilement que le nombre de voies ne peut pas augmenter indéfiniment si on souhaite des formats reproductibles d'un lieu à l'autre et restant viables financièrement. D'autres solutions radicalement différentes voient le jour actuellement. Des solutions qui ne sont plus basées sur l'imagerie sonore virtuelle.

◆ Dolby ATMOS

Cette évolution met en œuvre de multiples enceintes supplémentaires permettant la reproduction de sources sonores dans le plan de la hauteur. Elle vise donc à une immersion en 3-D dans le champ acoustique. Le positionnement dans le plan vertical est

établi avec rigueur afin de permettre une reproductibilité des effets sonores quelle que soit la salle. Par ailleurs, des voies latérales avant comblent l'espace habituellement vide entre les enceintes de façade et les enceintes surround classiques. Des caissons à graves (LFE) additionnels peuvent également être positionnés à l'arrière de la salle.

**Complément
Web 2.17**

Cette technologie récente est détaillée sur le site Web ainsi que le mixage objet associé.

◆ **NHK, TV Ultra Haute Définition en 22.2**

La chaîne de télévision japonaise NHK, sous l'impulsion de Kimio Hamasaki, responsable du service recherche de la chaîne propose un format son multicanal 22.2 associé à une image de très haute qualité. Le système est composé de trois couches d'enceintes :

- une couche supérieure avec neuf canaux ;
- une couche médiane avec dix canaux ;
- une couche inférieure avec trois canaux et deux voies LFE.

Comme pour le système Imm Sound, le but recherché est une meilleure immersion dans le champ réverbéré, plus de possibilités artistiques de spatialisation et aussi un sweet spot (zone d'écoute correcte) bien supérieure à un système 5.1.

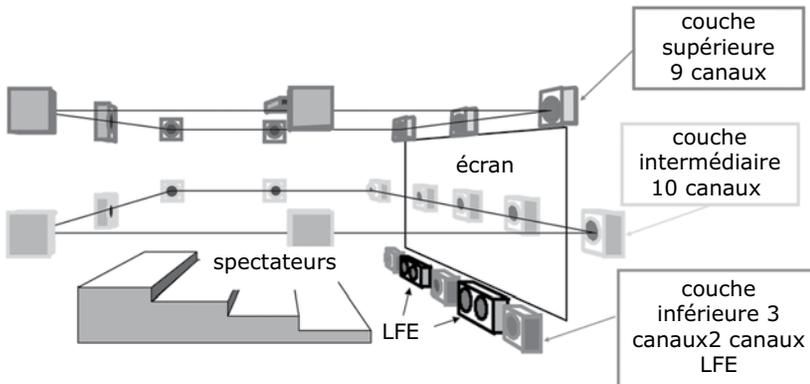


Figure 2.59 – Principe du format 22.2.

Le système est destiné à des diffusions de grandes dimensions, les expériences menées utilisaient un écran de projection de 10 m de base avec une résolution permettant un angle de vision de 100°, mais celui-ci peu être décliné pour la télévision Ultra Haute Définition pour usage domestique.

	HDTV	UHDTV
Nombre pixels	1080 × 1980	4320 × 7680
Angle de vue	30 ° à l'horizontal	Plus de 100 ° à l'horizontal
Résolution	Équivalente à un film en 35 mm	Plus de 2 fois un film en 70 mm

Figure 2.60 – Comparaison d'une image HDTV et UHDTV.

La normalisation du standard 22.2 prévoit un alignement temporel de toutes les enceintes par rapport à un point d'écoute central, ce que schématise la **figure 2.61**. Le niveau d'alignement en bruit rose est de 85 dB_C par canal pour -18 dBFS mesuré au sonomètre avec +10 dB pour chaque canal LFE, écart bande par bande mesuré à l'analyseur 1/3 d'octave.

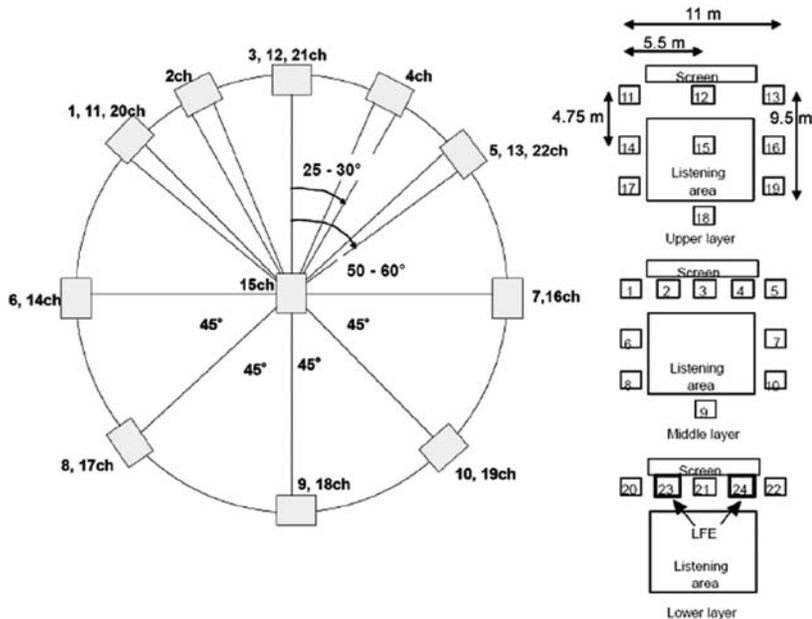


Figure 2.61 – Normalisation du format 22.2.

Ce format de diffusion est compatible avec un support 5.1 et même 2.0 et les enregistrements en 22.2 sont également transférables (down mixage) dans les formats 5.1 et 2.0.

2.12.2 L'après multicanal ? Le WFS et le Transaural et la 3D

◆ Le principe WFS

Le principe du WFS (Wave Field Synthesis) ou synthèse de front d'onde, n'est pas récent et résulte de recherches menées en laboratoire comme à l'IRCAM ou à l'université de Delft au Pays-Bas dont les premières expérimentations remontent aux années

quatre-vingt. Ces techniques ont connu de très importants développements au début des années deux mille, dans le cadre du projet européen Carrouso. Sur ces bases les sociétés Sonic Emotion en Suisse et Euphonia en France ont sorti ces études du laboratoire en rendant le système plus économique.

Le principe de base, assez simple est radicalement différent de celui de la stéréophonie qui vise à la création d'images sonores et de champs acoustiques virtuels. Il s'agit ici de recréer des champs acoustiques réels à partir de sources sonores qui peuvent être virtuelles selon le **principe d'Huygens**, le tout étant géré par un ordinateur contrôlant chaque haut-parleur séparément, avec autant de voies d'amplification. À ce niveau, on retrouve l'approche du remaping de Trinnov Audio qui tend à optimiser un champ acoustique en un point précis d'écoute, alors qu'ici, il est question de synthétiser un front d'onde réel perçu naturellement sur une large zone d'écoute, comme dans la réalité d'une source réelle.

PRINCIPE D'HUYGENS

Huygens en 1678 a énoncé ce principe très intuitif en observant le comportement d'ondes à la surface de l'eau : les vibrations qui se propagent à l'extérieur d'une surface fermée S_0 contenant la source sont identiques à celles qu'on obtiendrait en supprimant cette source et en la remplaçant par des sources convenablement réparties sur la surface S_0 .

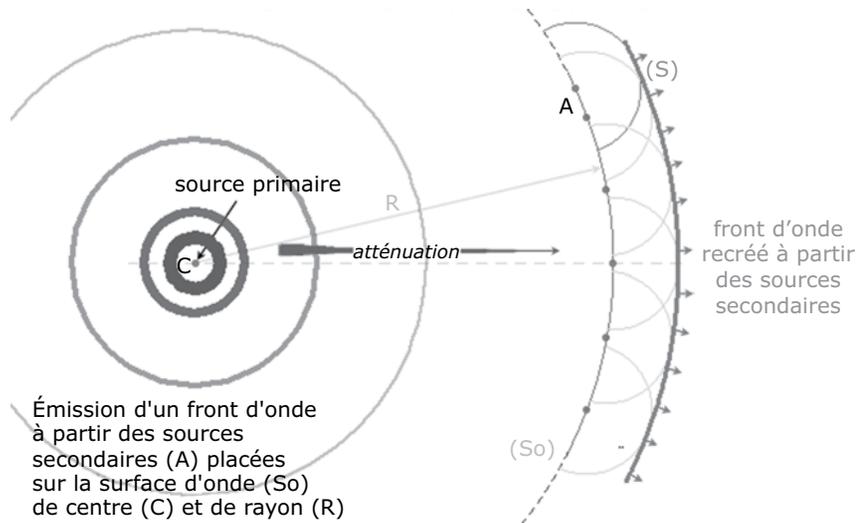


Figure 2.62 – Principe d'Huygens.

La conséquence de ce qui précède est que l'on sort du paradigme qui prévaut le plus souvent en sonorisation et qui consiste à disposer d'une source par canal de restitution (une enceinte pour le canal gauche, une pour le canal droit, etc.). Dans le cas de la WFS, on s'appuie sur un réseau de transducteurs régulièrement espacés, puis on gère le nombre de sources sonores virtuelles correspondant à autant de positions de sources spatialement localisées que désiré. Le champ sonore se déploie alors sur toute la zone de public en fonction des positions des sources virtuelles comme s'il s'agissait de sources réelles. L'auditeur n'entend plus la sonorisation, mais une scène sonore beaucoup plus naturelle !

Complément Web 2.18

Le développement de cette nouvelle technologie est largement exposé sur le site Web.

Les applications sont très nombreuses, en théâtre, concert live, parcs d'attraction, mais aussi cinéma 3D et ce système représente sans doute une des solutions d'avenir du multicanal !

◆ La technologie Transaurale

Il s'agit ici de recréer exactement le stimulus d'origine perçu par les oreilles lors d'une écoute naturelle. La technologie Transaural permet la localisation des sources sonores grâce à l'utilisation par notre cerveau des fonctions HRTF vues par ailleurs. La solution la plus simple est de reproduire au casque un enregistrement effectué à l'aide d'une tête artificielle de type binaurale avec, si possible, un tracking des mouvements de la tête afin de compenser l'effet de sa rotation.

Une autre solution est possible avec deux haut-parleurs, il s'agit des techniques transaurales dont le principe repose sur une annulation des chemins croisés (crosstalk ou cancelling) : il faut donc traiter le signal électrique de façon à ce que l'oreille gauche ne reçoive que le signal du canal gauche et inversement (comme dans un casque, mais avec un angle d'attaque par rapport à l'oreille beaucoup plus réduit). Ces solutions ont d'abord été proposées par Schroeder et Attal, puis développées par Cooper et Bauck.

L'écoute stéréophonique, grâce à l'interférence entre les canaux gauche et droite, crée une imagerie virtuelle comme le montre le schéma de la **figure 2.63**.

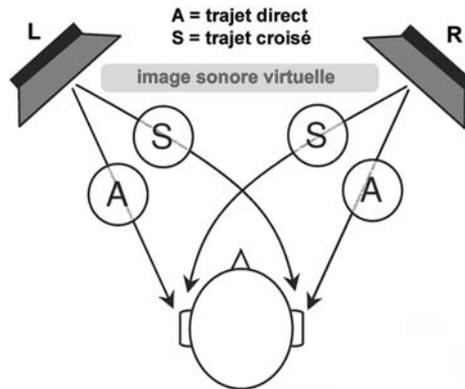


Figure 2.63 – Restitution stéréophonique standard.

L'écoute transaurale tente de supprimer ces interférences (chemins croisés) en utilisant les fonctions HRTF qui correspondent aux angles d'incidence de chaque chemin par rapport à chaque oreille et en les convoluant (traitement mathématique utilisant le principe de convolution mathématique) de manière à ce qu'il ne reste à chaque oreille que le signal original de l'enregistrement binaural.

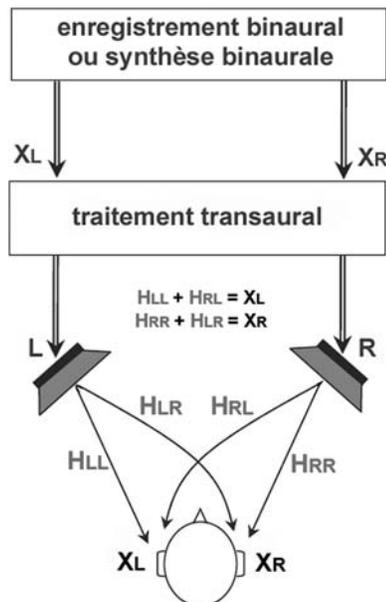


Figure 2.64 – Restitution transaurale.

Le traitement des signaux binauraux, issus par exemple d'une tête artificielle permet une restitution 3-D par la technique de l'annulation de ces chemins croisés. Le réalisme de reproduction par cette technique peut être tout à fait étonnant : on peut parfaitement entendre des sources placées dans tout l'espace y compris à l'arrière. Mais ce résultat est

conditionné à l'annulation exacte des signaux croisés et donc à la précision de la position de la tête et en particulier à une position axiale. Cette approche est tout à fait adaptée à une écoute type « jeu vidéo » pour laquelle la place de l'auditeur devant un écran est figée. Pour éviter cette trop grande sensibilité aux déplacements, il a été proposé par l'ISVR, Université de Southampton, une solution un peu différente, la technologie **Stéréo Dipôle** : en approchant les deux enceintes de reproduction jusqu'à réduire l'angle à environ 10° , on peut diminuer l'influence des déplacements et des colorations. Mais ceci a une contrepartie : l'annulation des trajets croisés nécessite des puissances importantes dans le grave. Afin de limiter cet inconvénient, différentes propositions ont été envisagées :

- changer l'écartement des haut-parleurs en fonction des fréquences en écartant plus les HP de grave par rapport aux aigus ;
- limiter le traitement des signaux croisés dans une zone de fréquence entre 500 Hz et 1 500 Hz, cette zone contenant les principales informations vis-à-vis de la localisation.

Yamaha a proposé un circuit DSP de traitement Stereo Dipole qui a été utilisé, entre autres, par CreativeLabs, dans le système Playworks PS2000. D'autres traitements sont implémentés dans des cartes audio pour ordinateurs. Ces technologies ont sans doute aussi un avenir pour les écrans de télévision.

La mesure et le calcul des fonctions transaurales peuvent être faits à l'aide d'une tête artificielle ou par une tête naturelle en chambre anéchoïque suivant le principe ci-après.

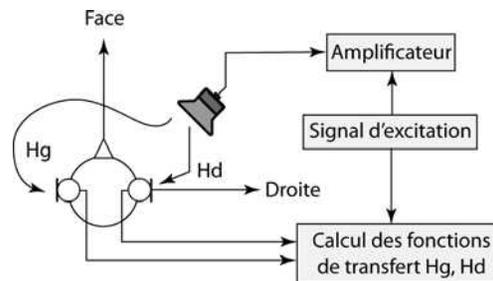


Figure 2.65 – Enregistrement des fonctions transaurales.

L'utilisation des techniques transaurales peut également être appliquée à l'enregistrement et au mixage en complément de systèmes multicanal en 5.1, en particulier afin de stabiliser la localisation des sources latérales surround et ainsi agrandir sensiblement la zone d'écoute pour l'auditeur. Des études pratiques ont été réalisées entre autres, au Conservatoire National Supérieur de Musique de Paris (CNSMDP).

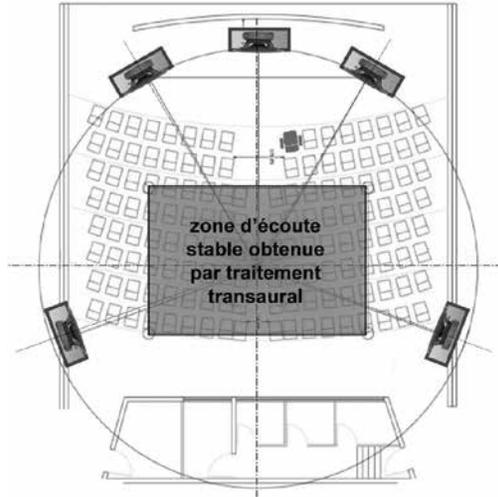


Figure 2.66 – Exemple d'expérimentation effectuée au CNSMDP par traitement transaural sur une écoute 5.1.

◆ La restitution en 3D

On introduit ici la dimension d'informations en hauteur déjà initiée par le système Dolby Atmos, mais avec une approche d'imagerie sonore virtuelle étendant le principe de la stéréophonie dans le sens vertical.

Nous pouvons citer sans être exhaustif, le système Auro-3D, le système HOA, le WFS 3D, les recommandations ITU-R BS-2051, etc.

Complément Web 2.19

Ces nouvelles technologies prometteuses sont détaillées et décrites sur le site Web.

Chapitre 3

Les facteurs perceptifs

*En stéréo on regarde une scène à travers une fenêtre,
en multicanal on se situe à l'intérieur de la scène*

POINTS ESSENTIELS

- Le standard stéréo : l'image fantôme et la diaphonie acoustique, les limites du standard.
- Le 5.1 et ses contraintes.
- Localisation dans l'espace.
- Principe de fonctionnement, fonctions HRTF, précision de localisation, angle minimum audible, effet Haas et la multiplication des sources, perception de la distance, perception des sources latérales, zone d'écoute, influence de la vision sur la localisation, internalisation et externalisation, synthèse binaurale.
- Impression spatiale, acoustique des salles, taille de l'environnement, spaciousness (largeur, enveloppement), réverbérance.
- Attention auditive, rapports son à l'image, dislocalisation, distraction.
- Ébauche de solutions pour l'analyse de l'image sonore multicanal.

La perception en multicanal est liée à la culture d'écoute de l'auditeur. S'il n'est pas accoutumé au format, ce dernier aura besoin d'un temps d'adaptation pour se

défaire des automatismes de la stéréo. Certaines esthétiques, dites **frontales** (développées dans le chapitre 4) s'écoutent plutôt facilement car elles font référence à des **images réelles** que le cerveau traite quotidiennement dans la perception naturelle de l'espace. Cette approche consiste à reconstruire l'image sonore perçue dans une situation où l'action se déroule dans l'espace frontal, de façon scénique : le sport, le documentaire, la fiction TV, le concert en sont des exemples concrets. Dans ce contexte, l'image sonore spatialisée donne un sens à tout spectacle ou événement dans lequel, un effet de salle ou d'ambiance extérieure est présent. Nous verrons par la suite que le cinéma est un cas particulier où la narration issue de l'imaginaire du spectateur s'oppose très souvent à la perception réaliste auditive, il fait appel à l'imaginaire du spectateur.

Lorsque l'exploration de l'espace se fait autour de l'auditeur, intégrant le champ réverbéré et le champ direct, l'esthétique est **équirépartie**, elle définit très souvent des **images sonores virtuelles**, peut-être plus difficiles à apprivoiser pour certains. L'écriture du scénario est nouvelle et remet en question la réalisation : c'est le cas par exemple de la fiction sonore en radio, la création sonore, la musique acousmatique, les musiques actuelles, une scène de film qui immerge le spectateur dans l'action.

Le dispositif 5.1 est certainement plus complexe que le standard stéréophonique (voir chap 2) mais il offre plus de possibilités. Il permet de placer le spectateur dans un champ sonore. On parle alors d'effet d'**immersion** : les sources sonores sont localisées sur le cercle défini par le dispositif d'écoute 5.1 et/ou au-delà, on parle alors d'**enveloppement**, mais certaines sont perçues à l'intérieur. Lorsque les sources sont localisées à l'intérieur du cercle, proche de la tête de l'auditeur, on parle alors d'**internalisation**. Une des difficultés du 5.1 consiste à raccorder le champ frontal avec le champ arrière afin d'obtenir une homogénéité. Dans ce sens, le 7.1, utilisant quatre enceintes pour la restitution latérale et arrière, est un standard plus homogène et simplifie la démarche.

La psycho acoustique définit un certain nombre de facteurs perceptifs répertoriés dans ce chapitre qui aideront le lecteur à une meilleure compréhension des règles du 5.1 : définition du standard stéréo et du fonctionnement de l'image fantôme, l'apport du 5.1, les contraintes du dispositif, la perception sonore spatialisée (le principe de localisation, l'impression spatiale), le principe de l'attention auditive. L'analyse de ces facteurs perceptifs nous permettra de définir l'image sonore au chapitre 4.

3.1 Le standard stéréo

Définissons tout d'abord le standard stéréophonique et rappelons le principe de fonctionnement de l'image fantôme.

Une image fantôme se crée lorsque les deux champs sonores issus des deux enceintes gauche et droite sont quasi identiques, et lorsque l'acoustique de la pièce est adaptée et que l'auditeur est centré sur son « sweet spot » (point de convergence du système d'écoute). Alors, l'auditeur peut localiser l'image sonore par sommation des deux champs sonores. L'image fantôme est une image virtuelle (qui s'oppose aux sources réelles que sont les haut-parleurs stéréo), car elle peut générer une infinité de sources que l'auditeur localisera sur un angle de 60° , s'il est positionné au « sweet spot ».

Cette image fantôme stéréo a deux limites.

La première limite vient de l'effet de précedence (défini par la suite) qui contraint l'auditeur à se placer au « sweet spot ». Il est bien connu que l'image sonore est sensible à la position de l'auditeur, si ce dernier se décentre légèrement et s'oriente vers sa gauche ou sa droite, sa localisation se porte rapidement vers l'enceinte correspondante, et pour un déplacement important la localisation par sommation ne fonctionne plus, l'auditeur perçoit alors un champ sonore droite ou gauche. Le déplacement de l'auditeur vers l'avant ou l'arrière modifie aussi la perception.

Le standard stéréo trouve aussi ses limites dans l'effet de diaphonie acoustique inhérent au procédé. Lors de l'écoute stéréo, quatre champs sonores sont à considérer. L'enceinte gauche produit du son en direction de l'oreille gauche et l'oreille droite, elle diffuse ainsi deux champs sonores. L'enceinte droite produit aussi, de façon symétrique, les deux champs sonores opposés. Le champ sonore issu du haut-parleur gauche, qui se déploie vers l'oreille droite, est considéré comme de la diaphonie acoustique. Une source réelle émettrait un seul champ sonore direct à chaque oreille tandis qu'une image fantôme en produit deux. Le son issu du haut-parleur gauche arrive à l'oreille droite avec une différence de marche d'environ $200 \mu\text{s}$, en comparaison avec l'oreille gauche et subit la diffraction due à la tête. Il en résulte, par sommation des champs sonores, un filtrage en peigne qui donne une courbe de réponse en fréquence accidentée, présentant une forte atténuation vers 2 kHz, dans la zone de présence, et des réjections vers l'aigu. L'image fantôme sonne donc plus distante, moins précise, tandis que le centre réel porté par l'enceinte centrale, appelé « hard center » ou centre physique, sonne plein, riche.

IMAGE FANTÔME STÉRÉO

- L'auditeur est contraint de se positionner au « sweet spot » pour bien localiser.
- Atténuation du spectre vers 2 kHz, conséquence de la diaphonie acoustique.
- Les sources centrées sont virtuelles, les sources latéralisées sur les enceintes sont réelles et sonnent pleines.

3.2 Apport du 5.1

Le dispositif 5.1 est un développement de la stéréo en deux dimensions, sur le plan horizontal. Il faut faire appel à des formats supérieurs pour apporter l'information d'élevation, tels que le format Dolby Atmos, l'IMAX (utilisé à la Géode), le 10.2 proposé par Tomlinson Holman, les formats Auro-3D, ou le 22.2 de la NHK.

Le dispositif 5.1 élargit tout d'abord le point d'écoute, notamment grâce à l'enceinte centrale qui permet à l'auditeur de se décentrer tout en gardant une image bien localisée. L'effet de décalage de localisation existe toujours avec le centre mais il est nettement moins important que pour la stéréo.

L'enceinte centrale introduit la notion de source réelle centrée, en comparaison avec l'image fantôme virtuelle, portée par les enceintes gauche et droite. Le centre réel n'est pas sujet à la diaphonie acoustique, il sonne plein, non accidenté contrairement au centre virtuel. L'enceinte centrale, combinée aux enceintes gauche et droite, permet d'obtenir une scène frontale plus riche, plus large, offrant une bonne stabilité des sources, une localisation plus précise avec moins de distorsion angulaire, un meilleur timbre et une plus grande clarté pour les sources centrées. Cependant, l'utilisation des trois canaux frontaux peut s'avérer imprécise, si l'on ne procède à aucun réglage du système d'écoute. L'optimisation du dispositif permet d'harmoniser les enceintes, et, dans ce cas, la combinaison d'un centre fantôme et d'un centre physique est plus homogène. Sans optimisation, l'image est dispersée, on entend souvent les enceintes séparément. Les sources mono divergées dans le frontal, réparties entre le centre, la gauche et la droite, sont démultipliées en trois sources au lieu de converger vers une seule source centrale.

L'image sonore obtenue par le dispositif 5.1 reproduit sur 360° les réflexions latérales précoces et le champ réverbéré, lesquels se déploient tout autour de l'auditeur donnant ainsi une meilleure lecture de l'architecture de l'événement sonore. La perception est naturelle, sans repliement d'espace, avec une sensation d'enveloppement et d'immersion dans le champ sonore. La représentation mentale est plus forte.

Enfin, les canaux arrière trouvent un meilleur équilibre grâce à la présence de l'enceinte centrale.

APPORT DU 5.1

- Deux types de centres : centre fantôme issu des canaux L et R ; hard center, ou centre physique, provenant de l'enceinte centrale seule ; combinaison de sources centrées réelles et virtuelles.
- Avec l'utilisation du centre, les sources frontales sont plus stables, meilleure restitution de leurs timbres, meilleur rapport son à l'image.
- Enveloppement et immersion dans le champ sonore.
- Internalisation possible.

Dans cet apport du 5.1, deux nouveaux paramètres sont à considérer, il s'agit de l'**impression spatiale**, qui définit l'espace ressenti par l'auditeur et le **principe de l'attention auditive**. À l'écoute d'un programme multicanal, l'auditeur est sollicité par les différents messages afférents tout autour de lui, ces informations concourent à l'effet d'enveloppement. La **localisation**, à travers le dispositif 5.1, est définie en deux dimensions, dans le plan horizontal, intégrant toutes les directions externes, au-delà du cercle d'écoute défini par les enceintes, ainsi que la direction interne entrant vers le point d'écoute (voir **figures 3.26** et **3.27**). Nous verrons dans la partie attention auditive que le spectateur peut rapidement être gêné auditivement par des sons venant des côtés et de l'arrière.

3.3 Le dispositif 5.1 et ses contraintes

En théorie, le dispositif 5.1 combine dix images fantômes issues des paires d'enceintes, lesquelles introduisent une certaine corrélation dans l'image sonore. Cette corrélation est nécessaire pour créer une image fantôme. Cependant, si la corrélation est trop forte, nous obtenons alors un certain flou dans la localisation et un étalement des sources, les performances spatiales sont alors limitées. À l'inverse, si la séparation des canaux est trop forte, l'image fantôme ne fonctionne plus. On constate des trous dans l'environnement sonore, une baisse du niveau sonore des images fantômes, et les images s'agglutinent sur les enceintes. Cette diaphonie acoustique dépend des technologies de spatialisation, à la prise de son (voir haute résolution spatiale), au mixage et à la diffusion (voir WFS). Il existe donc, en fonction du système de diffusion, une séparation optimale des canaux. La séparation obtenue par le dispositif 5.1 n'est certainement pas optimale pour la restitution du champ sonore.

L'utilisation du système 5.1 à l'origine n'est pas prévue pour une localisation précise sur 360°. Le 5.1 vient culturellement du cinéma, son centre est utilisé pour les dialogues du film, ses canaux frontaux restituent une localisation précise pour les bruitages, les effets et ses canaux surround apportent les ambiances, la réverbération, les effets ponctuels, la musique en coordination avec les canaux frontaux. Malheureusement, pour reproduire un espace sonore, nous ne pouvons disposer d'une infinité de points de diffusion. Nous devons limiter la diffusion à un certain nombre d'enceintes compatibles avec l'économie de la production audiovisuelle et recréer artificiellement des sources virtuelles, travaillant en images fantômes. C'est pourquoi le terme de stéréophonie augmentée est peut-être plus approprié.

La **figure 3.1** présente une étude de la BBC sur l'analyse de l'image quadriphonique.

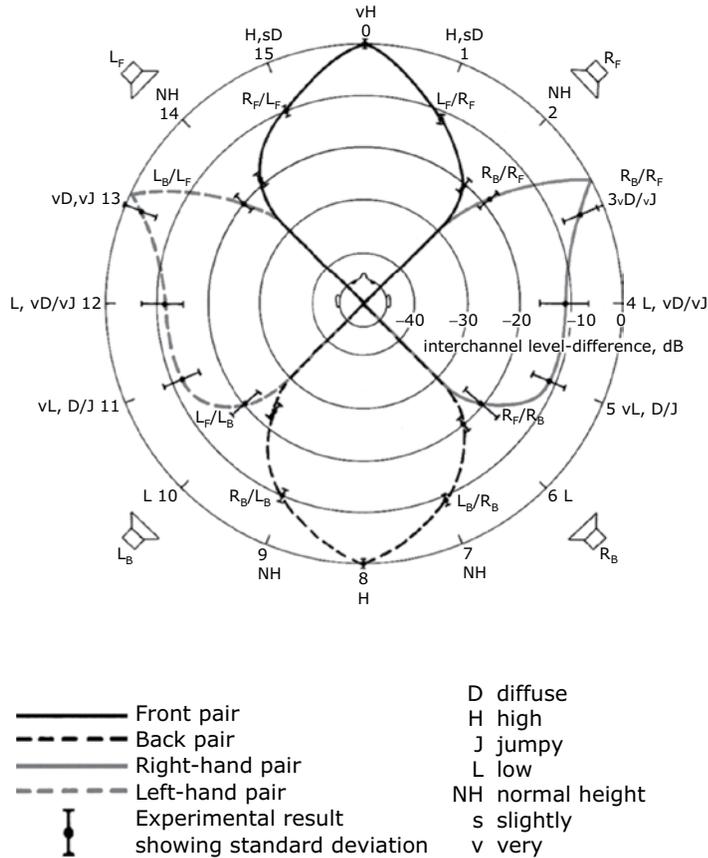


Figure 3.1 – Étude de la perception du format quadriphonique, © P. Ratliffe, BBC R & D.

Ce schéma est une étude faite en 1975 par la BBC sur la perception des images fantômes en quadriphonie en fonction de la différence de niveau entre paires d'enceintes adjacentes. Elle révèle la fragilité des images latérales (images diffuses, instables). Elle indique aussi la dissymétrie avant arrière du format quadriphonique (localisation à 65° de part et d'autre du centre pour un son également réparti en niveau entre l'avant et l'arrière) et le passage rapide de l'avant vers l'arrière des sources. Ceci est dû à une trop faible différence interaurale sur les paires d'enceintes latérales situées du même côté de la tête.

En 5.1, les angles entre les enceintes arrières (140°) et entre les enceintes latérales (80°), sont trop importants pour la création d'images fantômes stables, il existe une distorsion angulaire conséquente et des phénomènes de « bascules » entre les enceintes ont lieu. De ce fait, la localisation des sources latérales est extrêmement imprécise, indépendamment du système auditif.

Le 7.1 améliore la perception latérale et offre une meilleure cohérence frontale/arrière.

3.4 Localisation

3.4.1 Principe de fonctionnement

PARAMÈTRES DE LOCALISATION D'UNE SOURCE

- ITD, IPD différences interaurales de temps et de phase, basses fréquences.
- ILD différences interaurales de niveau, hautes fréquences.
- Indices spectraux (colorations spectrales des HRTF) : précision de la localisation dans le plan horizontal.
- Changement des indices spectraux, de l'ITD et de l'ILD par rotation de la tête ou mouvement de la source (indices dynamiques) : permet la distinction avant/arrière des sons.
- Localisation en élévation : indices spectraux correspondant aux fonctions HRTF monaurales induites par les effets de filtre en peigne des lobes et du pavillon de l'oreille externe.

La localisation des sons repose principalement sur la théorie Duplex définie par l'existence de critères interauraux, le système auditif étant équipé de deux oreilles.

Pour localiser un son dans l'espace, l'oreille utilise plusieurs paramètres : les critères interauraux de temps ITD (Interaural Time Difference), de phase IPD (Interaural Phase Difference) dans les basses fréquences, d'intensité ILD (Interaural Level Difference) dans les hautes fréquences supérieures à 200 Hz mais aussi les indices spectraux, définis comme étant les colorations spectrales des fonctions de transfert HRTF (Head Related Transfer Function). Ces colorations spectrales varient largement d'un individu à l'autre.

On définit l'oreille **ipsilatérale** comme étant l'oreille éclairée par l'onde sonore, et l'oreille **controlatérale**, celle dans l'ombre.

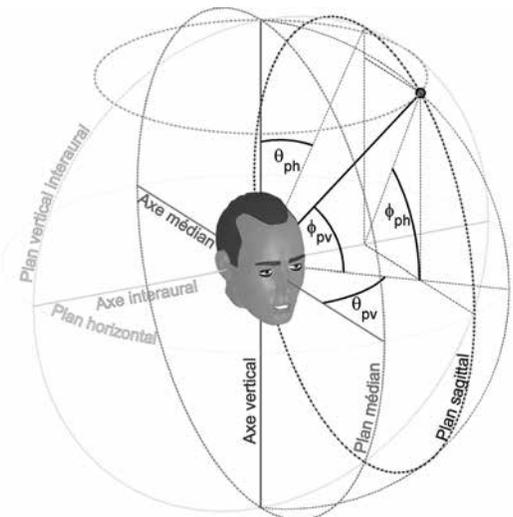
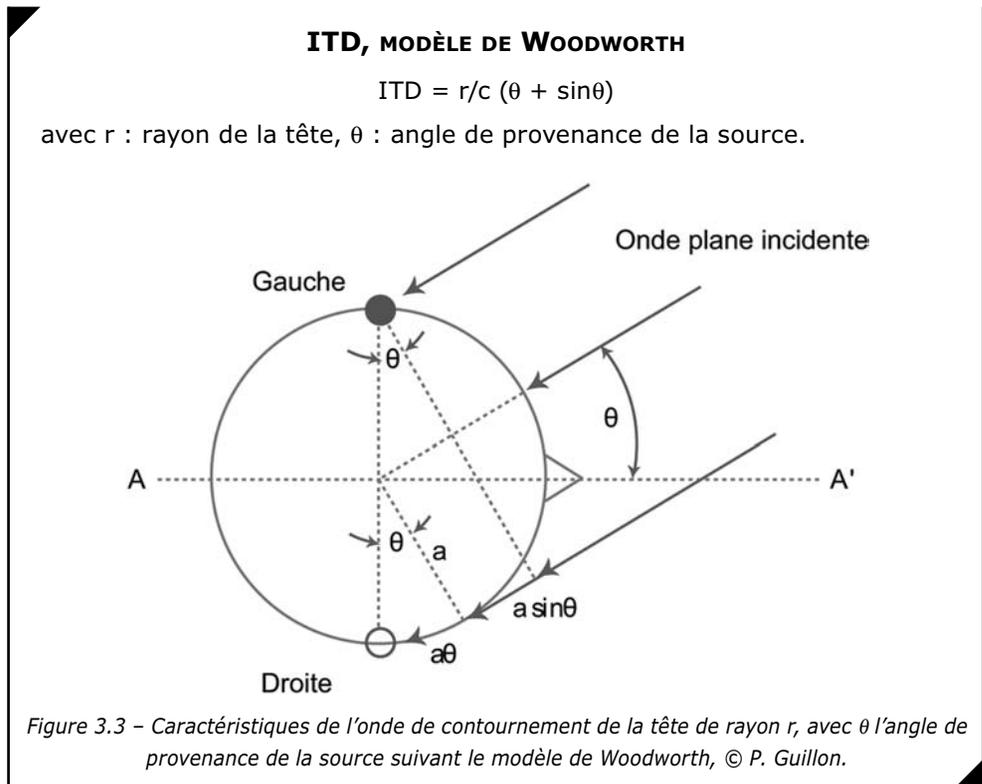


Figure 3.2 – Paramètres de positionnement des sources dans le repère de la tête de l'auditeur, © P. Guillon.

L'ITD traduit, par une valeur de temps, la différence de marche que subit l'onde sonore pour atteindre l'oreille opposée, contralatérale. Le modèle de Woodworth donne une approximation en ondes planes de l'ITD (voir **figure 3.3**), assimilant la tête humaine à une sphère :



L'ITD peut être ou non significative par rapport aux longueurs d'ondes des signaux sonores. Cela se traduit par deux types d'ITD :

- les différences temporelles de la structure fine, dites aussi IPD (différences de phase), valables pour les fréquences basses, inférieures à 1 500 Hz ;
- les différences de temps d'enveloppe, valables pour les fréquences aiguës, à partir de 200 Hz.

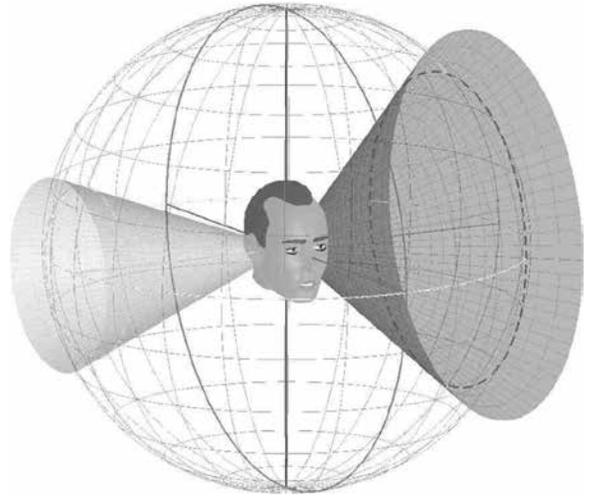
L'échelle de grandeur de l'ITD est 0-800 μs , la plus petite différence perçue étant de l'ordre de 20 μs . La distance moyenne entre nos deux oreilles est entre 14 et 15 cm, elle influe sur les valeurs maximales d'ITD.

L'ILD traduit par une différence de niveau, la diffraction et l'absorption que subit cette même onde. Aux basses fréquences, l'ILD est très faible, car la tête diffracte peu l'onde incidente.

L'ITD et l'ILD sont des critères dont la valeur est constante pour des positions de la source et de la tête fixes, ils permettent de localiser le son dans ce que l'on appelle « le cône de confusion », dont l'axe de symétrie est l'axe interaural.

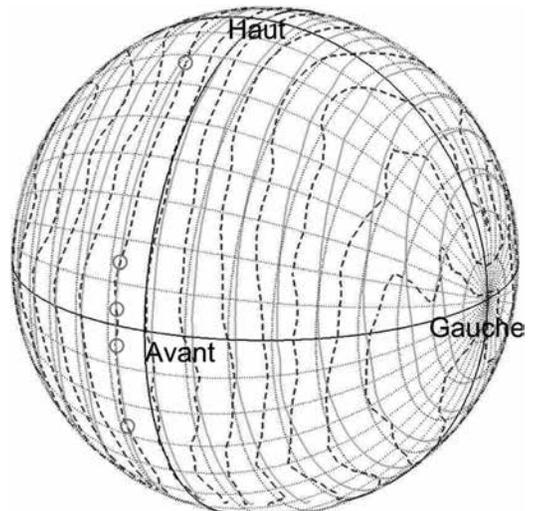
Figure 3.4 – Le cône de confusion issu des paramètres ITD et ILD seuls, mis en évidence par Woodworth, © P. Guillon.

Les hyperboloïdes correspondent chacune aux positions de sources qui génèrent un ITD constant selon le modèle de Woodworth. En gris foncé, une valeur d'ITD, en gris clair, une autre valeur.



En effet, pour un même angle d'incidence, la différence de marche est exactement la même que le son vienne de l'avant ou de l'arrière. La mesure des ITD et ILD à valeurs constantes sur une tête humaine donne des cercles concentriques, centrés sur l'axe interaural. Il existe alors une infinité de positions qui engendrent un même couple de valeurs d'ITD et d'ILD.

Figure 3.5 – Courbes iso-ITD et iso-ILD, les lignes iso-ITD (en traits pleins) sont espacées de $100 \mu\text{s}$, les lignes iso-ILD (en pointillés) sont espacées de 4 dB, les données sont issues de mesures sur une tête réelle, © P. Guillon.



L'utilisation exclusive de ces deux indices ne permet pas de localiser convenablement les sons en hauteur, la localisation des sons en élévation semble être contrôlée par leur contenu fréquentiel, comme des expériences psychoacoustiques le suggèrent. D'autre part, il semblerait que la localisation latérale est génétique, tandis que l'élévation est un mécanisme qui s'acquiert par la vision (calibration par la vision).

Les indices spectraux permettent alors la localisation en élévation et lèvent l'indétermination avant/arrière. Les **indices de localisation dynamiques** ITD et ILD obtenus par des micromouvements de la tête vont compléter l'information. Par exemple, pour une source située devant l'auditeur, dans le plan horizontal, si ce dernier tourne la tête vers la gauche, la source sera localisée sur la droite et si ce dernier tourne la tête vers la droite, la source sera localisée sur la gauche. Pour une source positionnée à l'arrière, l'effet sera inverse. L'auditeur peut ainsi discriminer l'avant de l'arrière.

Les **figures 3.6 et 3.7** représentent les variations ILD et ITD en fonction de l'angle azimut de la source.

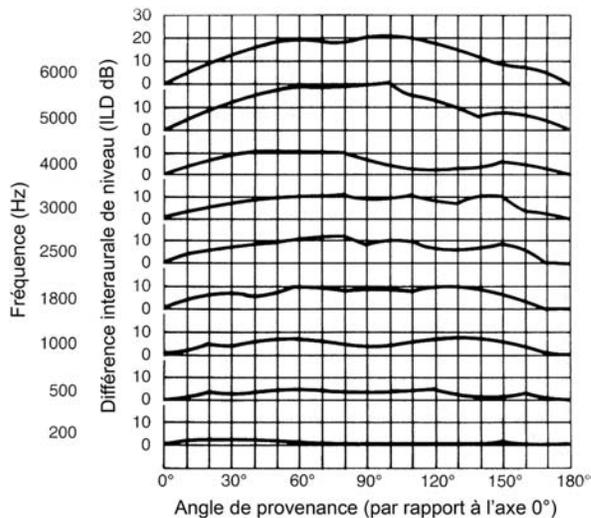


Figure 3.6 – Variations ILD pour des signaux sinusoïdaux en fonction de l'angle de la source, à différentes fréquences. Les positions 60° et 130° donnent les valeurs maximales pour un signal de fréquence 1 kHz © B.C.J. Moore.

3.4.2 Fonctions de transfert HRTF

Les fonctions de transfert relatives à la tête décrivent la propagation acoustique entre la source sonore et les oreilles de l'auditeur. Elles sont la conséquence de la diffraction due à la tête, des réflexions sur les épaules, sur le torse, et des réflexions et des diffractions issues des pavillons de l'oreille. Elles se traduisent par des profils spectraux spécifiques en fonction de l'angle de provenance de la source.

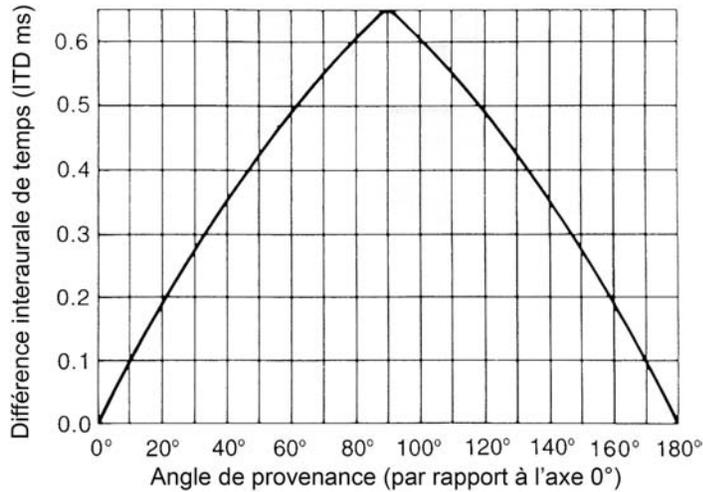


Figure 3.7 – Variations ITD en fonction de l'angle, la position à 90° donne la valeur maximale 0,65 ms, © B.C.J. Moore.

Ces phénomènes sont considérés comme l'empreinte acoustique de l'auditeur, laquelle diffère d'un individu à l'autre. Le pavillon de l'oreille fonctionne comme une cavité résonnante, il ajoute des colorations spectrales variables en fonction de la direction de la source, appelées **indices spectraux**. Il semble être l'élément clef dans la caractérisation de la fonction HRTF. Chaque direction du son a une courbe de transfert spécifique et l'oreille est capable d'extraire cette caractéristique fréquentielle pour retrouver la provenance angulaire de la source. La fonction de transfert HRTF représente la morphologie de la tête, elle est représentée par une courbe de réponse en fréquence (HRTF) ou par une réponse impulsionnelle (HRIR, Head Related Impulse Response), son équivalent en temps. C'est une fonction à trois variables, la fréquence ou le temps, la localisation de la source sonore en coordonnées sphériques par exemple (azimut, angle de radiation et d'élévation) et la morphologie de l'individu.

Les résonances induites par la forme du pavillon, qui est propre à chacun, n'affectent pas les basses fréquences en dessous de 3,5 kHz-4 kHz puisque sa taille est trop petite en comparaison avec la forme d'onde. Pour les hautes fréquences, l'effet du pavillon est décrit par des pics et des trous dont les fréquences dépendent de la direction du son.

On définit par indices spectraux les attributs fréquentiels des courbes de transferts HRTF qui présentent une certaine saillance pouvant être détectée, analysée et interprétée par l'oreille

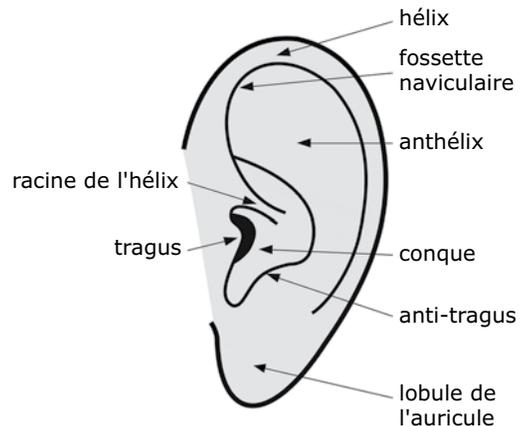


Figure 3.8 – Morphologie du pavillon de l'oreille, © R. Nicol, Orange Labs.

afin de localiser le son. Les indices spectraux, qui contribuent le plus à la localisation, sont ceux issus des effets des cavités des pavillons de l'oreille, plus précisément la conque. Ceux, issus de la tête et des épaules, sont d'un intérêt moindre et correspondent à des fréquences en dessous de 4 kHz. Les fréquences qui jouent un rôle important dans les indices spectraux des pavillons sont contenues dans la bande 4 kHz-16 kHz.

Pour localiser le son, le cerveau utilise potentiellement deux méthodes. La première consiste à utiliser l'indice spectral porté par le spectre global de la source, en comparant le spectre perçu et les spectres des HRTF enregistrés dans la mémoire. La seconde fait référence aux irrégularités créées par les pics et les creux dans le spectre perçu.

Figure 3.9 – Pics (P1, P2, P3) et creux (C1, C2, C3) observés sur les fonctions HRTF, le niveau est représenté par les couleurs, représentation du plan horizontal, © P. Guillon.
En couleur sur dunod.com

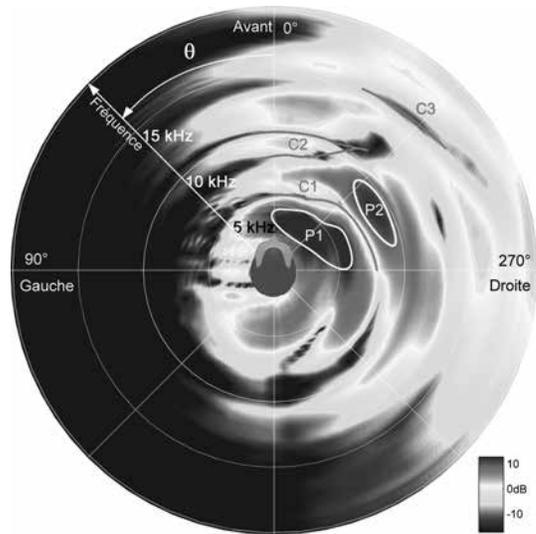
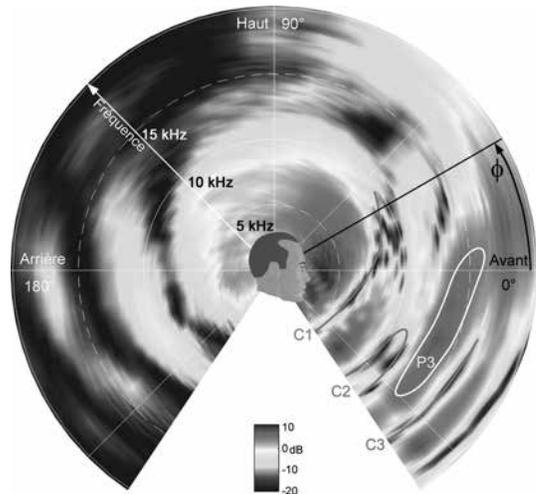
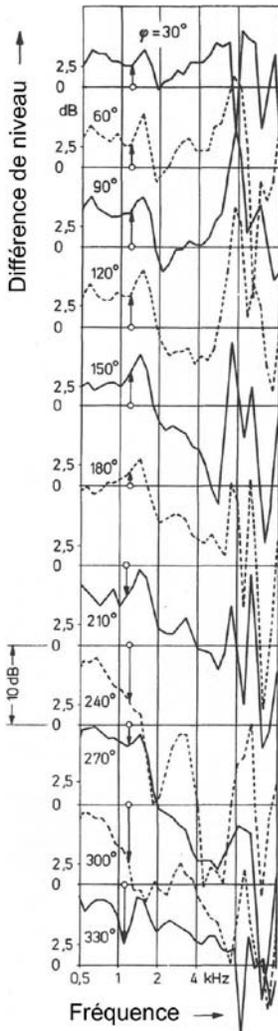
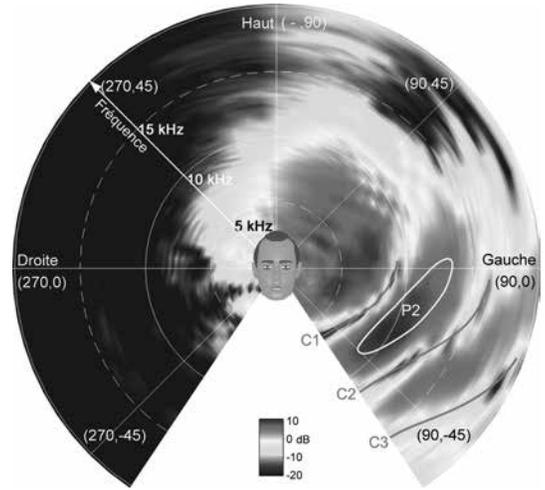


Figure 3.10 – Représentation du plan médian, la localisation des sons en hauteur semble être contrôlée par leur contenu en fréquence, © P. Guillon.
En couleur sur dunod.com



Dans le traitement des indices spectraux, l'oreille utilise concomitamment là aussi deux méthodes : elle traite de façon commune les indices recueillis par l'oreille gauche et l'oreille droite pour les sons centrés vers l'avant ou l'arrière, cela s'appelle

Figure 3.11 – Représentation dans le plan vertical interaural, © P. Guillon. En couleur sur dunod.com



le traitement binaural, et elle les traite également de façon indépendante quand la source est latéralisée, ce qui correspond à exploiter les HRTF monaurales ipsilatérales (voir la représentation des courbes HRTF interaurales et monaurales des **figures 3.12** et **3.13**). Ce filtrage, dépendant de la direction de la source, produit par le pavillon de l'oreille, permet la localisation des sons dans la direction verticale et le discernement frontal/arrière des sources.

L'histogramme de la **figure 3.14** représente une évaluation de la qualité de reproduction du son 3D faite à partir d'une étude sur dix-sept personnes, professionnels du son mais non spécialistes 3D. Cette étude menée par A. Silzle, publiée lors du séminaire VDT ICSA 2011, compare des restitutions de contenus 22.2 en Auro-3D 9.1, ITU BS 775 5.1 et stéréo, comparés en référence au 22.2. Sont représentés aussi les rendus sonores par downmix du 22.2 nommés Downmix 9.1 ou Downmix 5.1. Le rendu sonore Multichannel 9.1 est issu du 22.2 en désactivant les canaux du 22.2 non utiles, le Multichannel 5.1 est obtenu de la même façon.

Figure 3.12 – Fonctions de transfert interaurales, © J. Blauert.

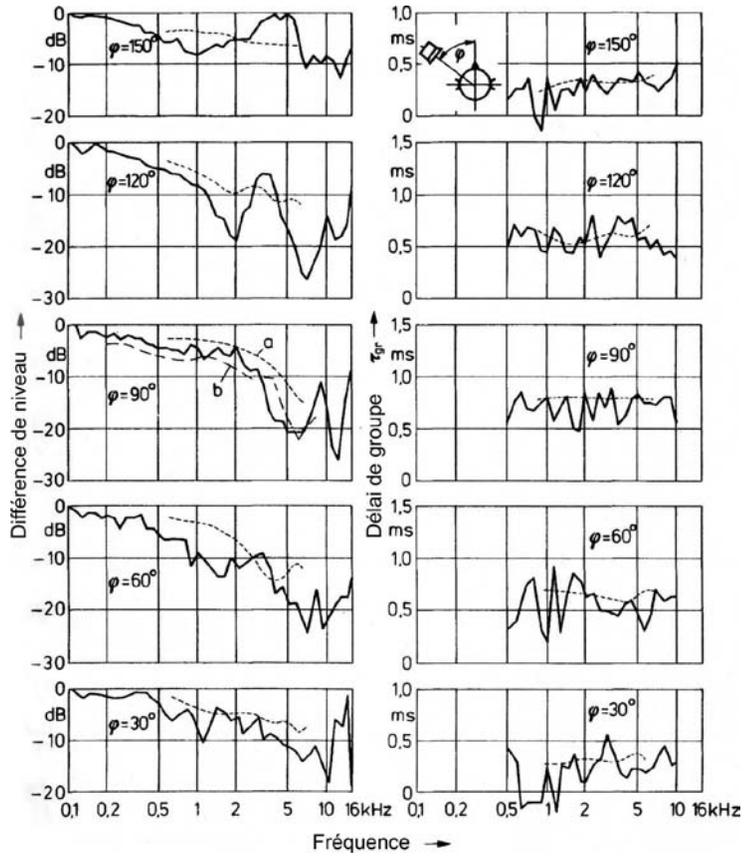


Figure 3.13 – Fonctions de transfert monaurales, © J. Blauert.

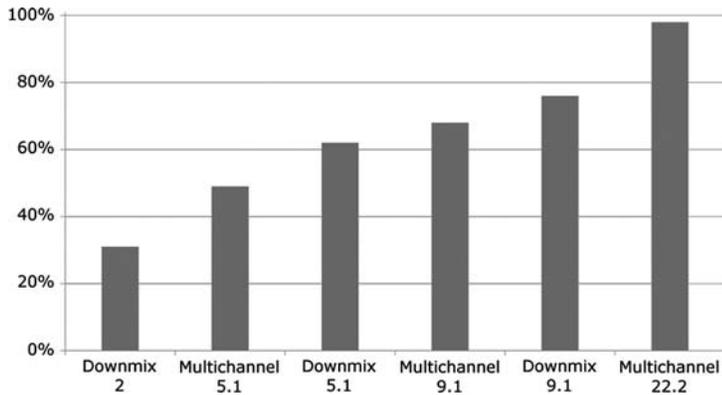


Figure 3.14 – Évaluation de la qualité de reproduction du son 3D, © B. Périaux d'après A. Silzle, S. George, E.A.P. Habets, T. Bachmann.

En **figure 3.15** une étude récente sur la perception de hauteur, menée par J. Barbour, publiée par G. Theile et H. Wittek pour une diffusion sur un format Auro-3D.

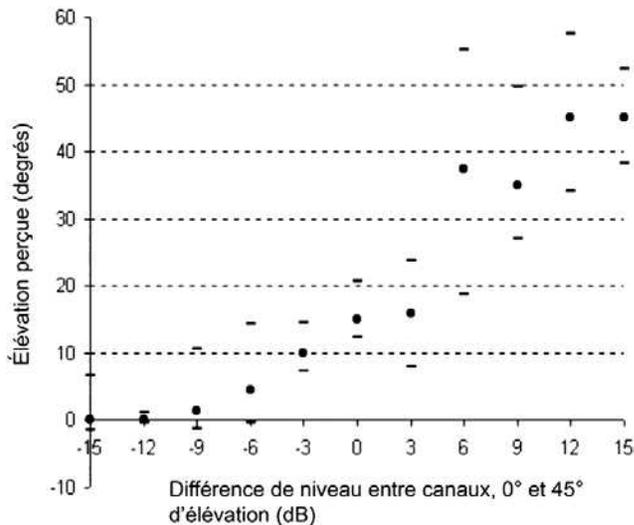


Figure 3.15 – Étude du panning d'intensité pour l'élévation, enceinte horizontale à 0°, enceinte de hauteur à +45°, © J. Barbour.

Cette étude représente la moyenne et l'écart type des positions de sources perçues dans le plan vertical, sur un format Auro-3D 9.1, suivant une différence d'amplitude entre les enceintes horizontales et les enceintes de hauteur. Dans le format Auro-3D, les enceintes de hauteur sont positionnées à un angle d'élévation de 30°. Les résultats traduisent un tassement des sources entre 0° et 15°, domaine où le pan-pot d'intensité en élévation fonctionne. Au-delà de 15°, les valeurs sont beaucoup plus dispersées. Le pan-pot d'intensité entre les enceintes horizontales et les enceintes verticales peut être utilisé, mais la perception des sources reste très instable, et les sources subissent un filtrage.

En conclusion sur la perception d'élévation, les enceintes de hauteur améliorent nettement la perception naturelle d'espace, la diffusion en hauteur de l'effet de salle est convaincante. En revanche, la localisation du son direct par image fantôme entre les enceintes horizontales et les enceintes de hauteur ayant le même azimut semble très imprécise et instable (par exemple l'enceinte frontale gauche horizontale et l'enceinte gauche en hauteur du format Auro-3D 9.1). Selon des expérimentations récentes de Mike Williams, la localisation entre une enceinte centrale placée à +45° de hauteur et les enceintes horizontales gauches et droites donnerait de meilleurs résultats, ce qui met en évidence l'intérêt de disposer les enceintes en triangle entre la couche horizontale et la couche hauteur (autre exemple dans la zone latérale, enceinte avant droite horizontale à +30°, enceinte hauteur droite à +90° d'azimut et à +45° d'élévation et enceinte arrière droite horizontale à +110°). D'autres expérimentations sont nécessaires afin de définir un format de diffusion 3D de type 9.1.

Dans la phase du mixage multicanal, chaque angle défini par la norme 5.1 fait appel à une fonction de transfert HRTF spécifique, lorsque l'on sollicite le placement des sources au-delà de 50°, l'équilibre spectral est modifié au-dessus de 4 kHz. Tomlinson Holman évoque l'expérimentation suivante, représentant la différence de spectre entre un son placé à 30° dans le frontal et à 120°, à l'arrière. La différence de spectre apparaît pour la zone de fréquence allant de 1,5 kHz jusqu'à 7 kHz, dans laquelle la source perd de son niveau, jusqu'à une dizaine de dB. Il en résulte une perte de définition, de timbre, de clarté.

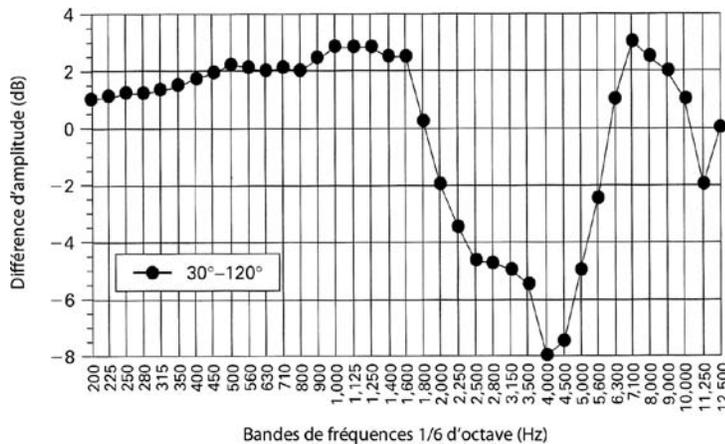


Figure 3.16 – Différences d'intensité avant/arrière (30°/120°) en fonction de la fréquence pour une configuration 5.1 avec un positionnement des enceintes arrière à 120°, © T. Holman.

Ce phénomène pose le problème d'homogénéité spatiale pour un son localisé entre l'avant et l'arrière. En effet, toute source placée dans les zones latérales utilise l'image fantôme L/LS ou R/RS, et la localisation par sommation des deux sources issues des deux enceintes ne fonctionne que très rarement car les deux champs sonores sont différents. En effet, l'auditeur perçoit plutôt deux sources qu'une seule, une source naturelle dans le frontal et une source détimbrée à l'arrière. Ceci représente une des faiblesses du format 5.1 : sa localisation latérale est limitée. Pour pallier cette faiblesse, il existe certains outils de synthèse HRTF permettant d'améliorer la localisation des sources latérales en rééquilibrant les spectres, mais cela suppose que l'auditeur ne tourne pas la tête.

3.4.3 Précision de localisation

Afin de caractériser les incertitudes de localisation, dans les expérimentations suivantes, on immobilise la tête de l'auditeur. Pour une source à 0° par exemple, on mesure l'incertitude de localisation frontale, la tête de l'auditeur faisant face à cette direction. Les mesures pour les provenances frontales, arrière, latérales et verticales ont été menées par J. Blauert. La localisation en azimuth dans le frontal et à l'arrière est précise, mais moins bonne sur les côtés ; la localisation en élévation est aussi moins bonne.

ZONES D'INCERTITUDE DE LOCALISATION

Localisation en azimut

- Dans l'axe frontal : 7 °.
- À l'arrière : 11 °.
- Sur le latéral : 20 °.

Localisation en élévation, plan médian

- 20 à 40 °.

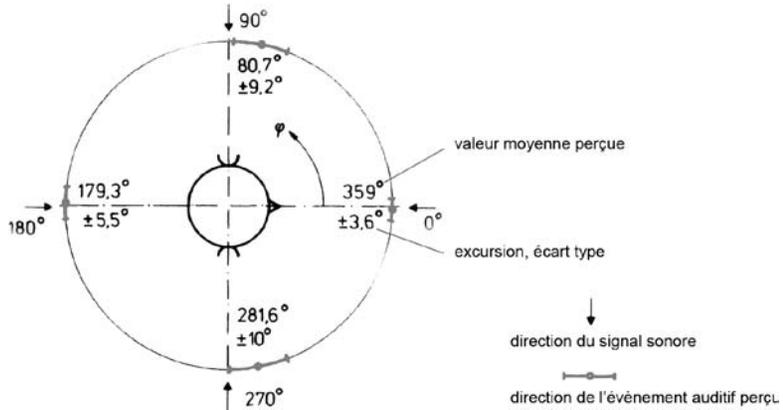


Figure 3.17 – Représentation des incertitudes de localisation en azimut des sources avec la moyenne, données recueillies auprès de 600 à 900 sujets, dont la tête était immobilisée, les stimuli étant des pulsations de bruit blanc d'une durée de 100 ms, de niveau 70 phones, © J. Blauert.

Pour un individu sourd d'une oreille, les valeurs d'incertitudes augmentent considérablement.

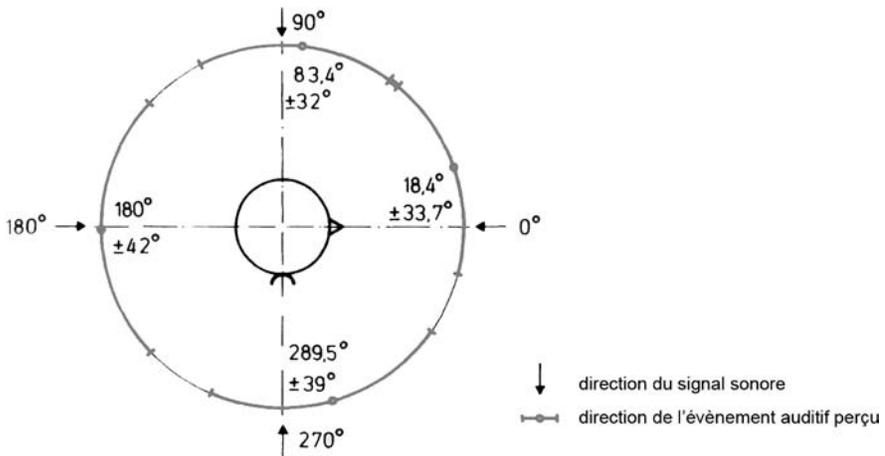


Figure 3.18 – Représentation des incertitudes de localisation en azimut des sources avec la moyenne, données recueillies auprès de 600 à 900 sujets, sourds de l'oreille gauche et dont la tête était immobilisée, les stimuli étant des pulsations de bruit blanc d'une durée de 100 ms, de niveau 70 phones, © J. Blauert.

Le schéma de la **figure 3.20** représente l'étude du filtrage dans la perception d'élévation et de localisation frontale/arrière des sources.

En synthèse sur la précision de localisation, l'oreille localise de façon très précise dans le latéral, mais le degré de précision est bien moindre pour l'élévation.

Figure 3.19 – Représentation des incertitudes de localisation en élévation des sources, données recueillies auprès de 7 sujets, dont la tête était immobilisée, les stimuli étant de la parole continue par une personne familière, d'un niveau de 65 phons © J. Blauert

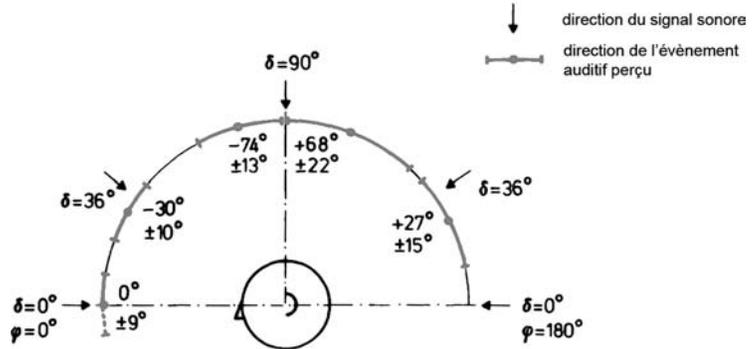
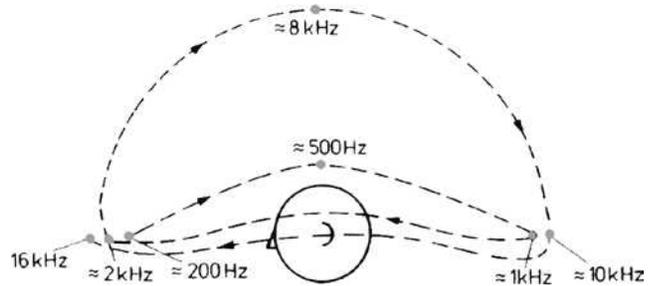


Figure 3.20 – Étude de l'effet du contenu fréquentiel sur la perception de l'élévation, frontale et arrière, les fréquences autour de 8 kHz semblent être localisées au-dessus de la tête, les fréquences 10 kHz et 1 kHz semblent provenir de l'arrière, les fréquences 200 Hz, 2 kHz et 16 kHz semblent provenir de l'avant © J. Blauert.



3.4.4 Angle minimum audible

Pour caractériser l'angle minimum audible pour des sons séquentiels, dans les expériences suivantes, l'auditeur est maintenant libre de tourner la tête afin de préciser la localisation de la source, faisant ainsi appel aux indices spectraux dynamiques. L'angle minimum que l'oreille peut discerner varie en fonction de la direction de la source sonore, il est de 1° pour une source présentée à 0° dans le plan horizontal ; dans le plan vertical, l'angle passe à 3° . Cet angle est faible dans le frontal mais augmente progressivement dans les secteurs latéraux et à l'arrière. L'étude de la **figure 3.21** représente les valeurs dans le plan horizontal.

Figure 3.21 – Variations de l'angle minimum audible en fonction de l'angle de provenance de la source pour des fréquences de 1 kHz et 500 Hz, l'auditeur étant libre de tourner la tête. On retrouve la valeur 1° dans l'axe frontal, jusqu'à 60° de provenance les valeurs restent inférieures à 3° et à partir de 60° de provenance, les valeurs croissent fortement.

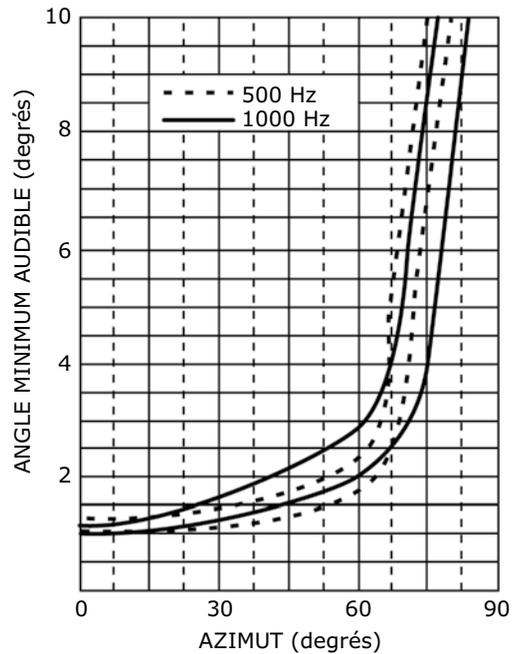
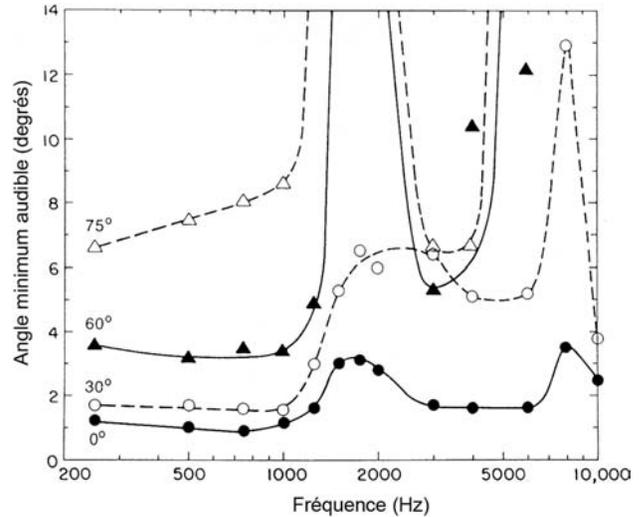


Figure 3.22 – Étude de la variation de l'angle minimum audible en fonction de la fréquence pour des valeurs de provenance de 0°, 30°, 60°, 75°.
© B.C.J. Moore.



3.4.5 Perception de la distance

La perception de la distance fonctionne à partir de plusieurs indices, notamment pour les sons familiers, le niveau perçu est un critère important. Cet indice s'avère être encore plus efficace lorsque plusieurs sources sonores sont présentes, de telle sorte que l'oreille peut comparer les différents niveaux perçus. C'est un indice efficace dans le jugement des distances absolues.

L'indice spectral est aussi utilisé pour des distances modérées, mettant en évidence les propriétés d'absorption de l'air. Les hautes fréquences sont plus rapidement atténuées que les basses fréquences, en fonction aussi de la présence de vent et du taux d'humidité de l'air. Cet indice est dépendant de la familiarité avec les sources, il est efficace dans le jugement des distances relatives mais pas dans l'évaluation des distances absolues.

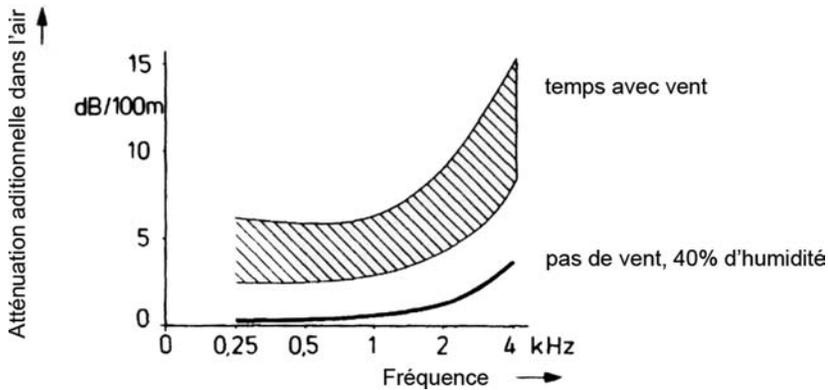


Figure 3.23 – Étude de l'atténuation en niveau de la propagation aérienne en fonction de la distance et de la présence de vent, suivant la fréquence. Les fréquences aiguës sont les premières atténuées (à partir de 1 kHz) et la présence de vent agit considérablement sur l'atténuation globale du niveau, © J. Blauert.

Pour les sons proches de la tête, ce sont les valeurs d'ILD plus importantes, qui permettent d'évaluer leur distance (effet de champ proche).

Dans une salle, l'évaluation de la distance en absolu et en relatif se fait grâce au rapport CD/CR et au temps initial, délai entre le son direct et les premières réflexions. Le spectre du son réfléchi, altéré par les premières réflexions, peut aussi servir d'indice de jugement de la distance.

Pour conclure, les informations obtenues des différents indices sont combinées pour évaluer la distance, les deux principaux critères restant le niveau perçu et le rapport CD/CR. Ces deux indices donnent des résultats dépendants du type de source, et de sa position angulaire (azimut).

Le jugement de la distance est assez imprécis, des erreurs d'environ 20 % sont assez communes pour des sons non familiers. Les valeurs perçues pour les sons proches ont tendances à être surévaluées, tandis que pour les sons lointains, les valeurs ont tendance à être sous-évaluées.

La variation de distance d'une source mobile donne des écarts de niveaux perçus qui diminuent avec la distance. On peut en conclure que la dynamique constitue très certainement un critère d'écoute intéressant dans l'évaluation de la distance.

PARAMÈTRES POUR L'ÉVALUATION DE LA DISTANCE APPARENTE

- Atténuation du niveau.
- Rapport CD/CR.
- Atténuation des fréquences aiguës.
- Dynamique des sons.
- Variation spectrale.
- Variations d'ILD plus importantes pour les sons proches.

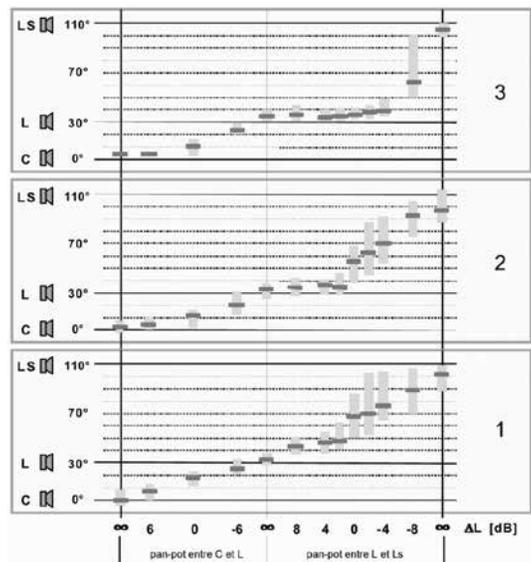
3.4.6 Perception des sources latérales et zone d'écoute

L'étude de la **figure 3.24** représente la localisation des sources sur le demi-cercle gauche de l'écoute 5.1, défini par le centre, la gauche et l'arrière gauche. On étudie les valeurs angulaires des sources perçues, en faisant varier la différence d'intensité entre les différentes enceintes, ce pour trois positions de l'auditeur :

- 1 : positionnement au sweet spot ;
- 2 : positionnement reculé et décalé vers la gauche ;
- 3 : positionnement devant le sweet spot.

Les valeurs représentées sont la moyenne en noir et l'écart type en gris.

Figure 3.24 – Étude de la localisation sur un dispositif 5.1 ITU, © G. Theile.



Cette étude révèle la fragilité de l'écoute 5.1 par rapport au placement de l'auditeur. Un mauvais placement gauche-droite a un effet limité sur la perception de l'image sonore, tandis qu'un déplacement dans la profondeur induit un décalage important des sources

vers l'avant ou l'arrière. L'étude révèle aussi, par la représentation des excursions des valeurs de localisation, une localisation des sources latérales imprécise, entre 50° et 110°.

À partir de cette étude, on peut conclure que le déplacement dans la profondeur est plus critique que le déplacement latéral, et l'image frontale est plus robuste que l'image latérale. L'image sonore 5.1 reste fragile par rapport au déplacement de l'auditeur, il faut faire appel à un dispositif optimisé (voir chapitre 2) ou aux technologies WFS pour stabiliser la construction de l'image et élargir la zone d'écoute.

Suite à ces expérimentations, la zone d'écoute est un critère important, il est indispensable de s'y déplacer pour évaluer la robustesse de l'image sonore, ainsi que les variations spectrales possibles en rapport avec la diaphonie acoustique des canaux du multicanal (l'exemple du double MS présente ces variations spectrales lorsqu'on se déplace juste autour du sweet spot).

3.4.7 Effet Haas et multiplication des sources

L'utilisation de plusieurs sources n'est pas sans difficulté. S'il est relativement aisé de se placer à distance égale des deux haut-parleurs d'un système stéréophonique, cela devient plus compliqué lorsque l'on multiplie le nombre de sources. Un seul point d'écoute au centre du cercle est alors défini pour le format 5.1.

La localisation d'une source sonore repose sur la loi du premier front d'onde : la source perçue en premier est identifiée comme l'unique provenance. Cela explique pourquoi nous pouvons localiser une source sonore dans un environnement réverbérant comme lors d'un concert par exemple.

- Pour $0 \text{ ms} < \Delta t < 1,1 \text{ ms}$, latéralisation par sommation ;
- Pour $1 \text{ ms} < \Delta t < 25\text{-}30 \text{ ms}$, loi du premier front d'onde ou effet de précedence, source latéralisée, zone du filtrage en peigne ;
- À partir de 40-50 ms, phénomène d'écho.

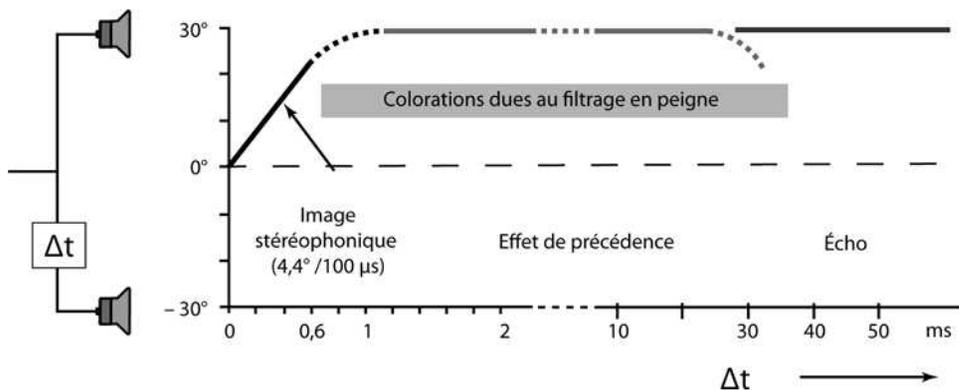


Figure 3.25 – Analyse des 50 premières ms du son, © G. Theile.

Comme la **figure 3.25** l'indique, l'effet Haas a lieu à partir d'une différence de temps de 1 ms jusqu'à environ 30 ms. Il n'est pas négligeable dans le cas d'une diffusion multi-point. Cet effet intervient avec une différence de temps d'arrivée supérieure à 1 ms, ce qui correspond à une différence de marche de seulement 34 cm. Entre 1 ms et 10 ms a lieu l'effet de fusion qui caractérise la modification du timbre perçu de la source et de sa localisation en fonction de la fréquence.

La position de l'auditeur par rapport aux différentes enceintes du dispositif 5.1 est donc sensible, la localisation fragile de l'image sonore est d'autant plus critique que les enceintes ne sont pas toujours temporellement alignées. Avec un système d'écoute optimisé, la localisation des sources est plus précise.

3.4.8 *Influence de la vision sur la localisation*

La perception sonore est constamment confrontée aux informations visuelles recueillies par le cerveau. En multicanal, on peut facilement imaginer que la spatialisation sur 360°, multipliant les informations à traiter va d'autant plus complexifier le processus.

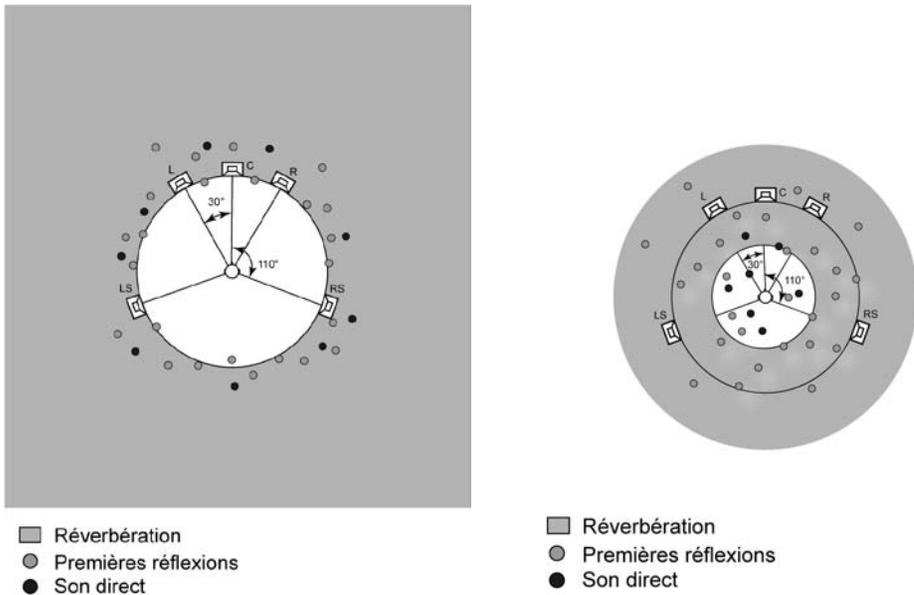
La vision peut jouer un rôle important en précisant le positionnement de la tête dans l'espace. Selon B. C. J. Moore, nous sommes capables d'identifier une source sonore comme immobile, même si l'on tourne la tête et que le son qui arrive aux oreilles en est changé. En quelque sorte, l'information de position de la tête et l'information de localisation perçue de la source sonore sont toutes deux combinées dans l'interprétation du cerveau pour en valider la cohérence. Il ne faut pas négliger l'influence de l'orientation visuelle perçue dans l'interprétation des indices sonores spatiaux. C'est-à-dire que la plus forte représentation spatiale implique l'intégration des informations provenant des différents sens.

3.4.9 *Internalisation et externalisation*

Le phénomène d'internalisation a lieu lorsque la localisation d'une source est intracrânienne. Ce phénomène est présent pour toute écoute au casque stéréo traditionnelle. La stéréo s'étend sur une ligne allant de l'oreille gauche à l'oreille droite, dans la tête de l'auditeur. Selon J. Blauert, il a lieu aussi dans un contexte de reproduction d'une même source ou bien de deux sources ressemblantes sur des haut-parleurs qui se font face, c'est le cas pour les diagonales L/Rs et R/Ls. Cette perception interne résulterait selon D. Griesinger du comportement de l'ITD. Si ce dernier varie très peu lorsque l'auditeur tourne la tête, il en résulte par association pour le cerveau un seul flux auditif. La perception n'est plus externe, la localisation est intracrânienne ou proche de la tête. Le rôle du pavillon de l'oreille, outre le fait de localiser les sources, comme précédemment, est de créer la perception externe des sons. La présence de réverbération y contribue.

Pour conclure, nous obtenons une localisation intracrânienne, ou proche de la tête, pour une source (ou deux sources très ressemblantes) répartie sur deux enceintes opposées.

Ceci détermine un nouveau paramètre pour la caractérisation de l'image sonore multicanal, que l'on peut nommer **localisation interne** ou **internalisation**.



Figures 3.26 et 3.27 – De gauche à droite, représentation d'images sonores avec externalisation et internalisation (ou localisation interne des sources), © B. Périaux.

Une corrélation forte des cinq canaux du multicanal internalise donc l'image sonore : c'est le cas pour un mixage qui ne serait pas assez décorrélé, ou pour un système de prise de son multicanal trop coïncident. D'autre part, il existe des lieux où l'internalisation est tout à fait possible et naturelle, par exemple un intérieur voiture, ou bien un ascenseur.

3.4.10 Synthèse binaurale

Il existe aujourd'hui diverses solutions de reproduction de l'espace pour le casque par synthèse binaurale. Toutes appliquent le principe de correction HRTF aux canaux entrant, à partir d'informations générales ou moyennées de formes et de dimensions de pavillons de l'oreille, de têtes et de torsos.

Sans procédé de virtualisation, un son monophonique écouté au casque positionne la source au centre de la tête, tandis qu'un signal stéréophonique positionne les sons sur une ligne de gauche à droite en passant par le centre de la tête, mais jamais en dehors comme dans la perception naturelle. Avec une source multicanal, les procédés de virtualisation binaurale ouvrent effectivement la perception et l'élargissent même au-delà du casque. Le rendu sonore diffère bien évidemment en fonction du type de casque utilisé (fermé, ouvert, intra-auriculaire).

La localisation des canaux arrière fonctionne généralement bien. En revanche, la localisation frontale est plus limitée, avec notamment un canal central difficile à percevoir devant soi. Dans tous les cas, aucun de ces procédés ne peut prétendre aujourd'hui

reproduire parfaitement la localisation obtenue à partir de cinq ou davantage d'enceintes réparties autour de l'auditeur. La personnalisation des HRTF et l'utilisation de « head tracking » améliorent la perception.

Les HRIR sont les filtres binauraux obtenus par transformée inverse de Fourier des HRTF. Il est également possible d'intégrer aux HRTF les effets liés à la propagation acoustique dans une salle, ce sont les BRIR (Binaural Room Impulse Response).

Le casque multicanal fonctionne soit à partir de réponses impulsionnelles (BRIR) intégrant l'empreinte du système d'écoute, l'acoustique, et les HRTF de l'auditeur, soit à partir de bibliothèques HRTF (mesurées en chambre sourde à une distance d'environ 1,5 m) généralement complétées par un module de simulation de pièce.

Une base de données HRTF de mille directions au moins semble suffisante pour un espace virtuel 3D convainquant. Il existe différentes bases de données, dont les caractéristiques pour les plus connues sont les suivantes (données 2013) :

- le CIPIC, Université de Davis, 45 individus, 1 250 directions ;
- l'IRCAM, base Listen, 51 individus, 187 directions, l'IRCAM procède fin 2013 à la mesure d'une nouvelle base de données HRTF dans le cadre de BiLi, faite sur 1 680 points (pas d'échantillonnage de 6°, latéral 0° à 360°, élévation -51° à +86°) avec 54 individus (42 hommes, 12 femmes) et 3 têtes artificielles dont la Neumann KU100 et la B & K 4100D ;
- le TNO FTR & D, Human Factors Research Institute, Soesterberg, 965 directions, 8 individus et une tête artificielle Head Acoustics HMSIII (base privée d'Orange) ;
- l'ISVR Institute of Sound and Vibration Research, Université de Southampton ;
- E. Grassi, Université du Maryland, 7 individus, 1 093 directions ;
- Pr. Itakura, Université de Nagoya, 96 individus, 72 directions horizontales.

Il existe depuis peu un format standard de fichier pour les HRTF nommé SOFA qui permet d'importer ses propres fonctions de transfert dans un procédé de synthèse binaurale.

Généralement le casque multicanal qui fonctionne à partir de HRTF personnalisées, issues de mesures de réponses impulsionnelles (BRIR), permet de reproduire les directions du multicanal disponibles en entrée (huit canaux généralement), l'auditeur perçoit alors les directions du système d'écoute 5.1 ou 7.1, mesurées lors de la réponse impulsionnelle.

Parmi les différentes solutions proposées, quelques procédés professionnels sérieux ont vu le jour : le casque développé conjointement dans les années 1990 par l'IRT et Studer (il n'y a pas eu de véritable développement commercial), le Realiser A8 développé par Smyth Research en 2009, le casque 3D Sound One de la société 3D Sounds Labs et le DMS BP84 resté à l'état de prototype.

Le Realiser A8 permet à l'auditeur de mesurer ses propres HRTF avec des microphones que l'auditeur place dans le creux de ses oreilles et applique le filtrage obtenu aux différentes voies du multicanal. La mesure des HRTF est faite pour trois orientations de la tête, à -30°, 0° et +30°, le processeur calcule les angles intermédiaires par interpolation. La seconde étape de calibration consiste à mesurer les HPTF (Head-Phone

Transfert Function), la réponse du casque afin d'annuler la fonction de transfert entre le casque et les oreilles de l'auditeur dans le procédé de virtualisation.

Cette étape est délicate car les HPTF dépendent de l'individu (sa morphologie) et du positionnement du casque sur les oreilles. C'est pourquoi une mesure individuelle est nécessaire. D'autre part la sensibilité du positionnement du casque est quasi-insoluble, mais moins critique dans la mesure où les variations spectrales obtenues sont moindres que celles des indices spectraux utiles pour la localisation.

Il est recommandé d'utiliser des casques de type FEC (Free Air Coupling Equivalent) afin de reproduire l'impédance équivalente aux conditions de champ libre, le casque ouvert donne de bons résultats. Le critère utilisé est le PDR (Pressure Division Ratio), ce dernier doit être le plus proche de 1.

Le Realiser A8 fait appel à trois composants critiques : la personnalisation (mesure des HRTF), le suivi des mouvements de tête horizontaux sur 60° (« head tracking ») et la capture des propriétés de n'importe quel système sonore dans son environnement d'origine. La personnalisation est réalisée par la mesure de nos propres fonctions HRTF et le suivi des mouvements de tête se fait par recalcul des paramètres de ces fonctions toutes les 5 ms.

Cette technique vise aussi à simuler une écoute multicanal en reproduisant son identité et l'acoustique du local dans lequel on fait la mesure (mesure du temps de réverbération, réponse en fréquence des enceintes...). Nous pouvons donc imaginer transporter dans le processeur notre écoute multicanal favorite, bien optimisée, n'importe où.

Le prototype BP84 DMS utilise une librairie de HRTF, et oriente l'utilisateur vers les fonctions HRTF les plus appropriées. Pour cela ce dernier renseigne par exemple son tour de tête. À l'heure actuelle, la simulation de distance est reproduite par un module de réverbération fonctionnant en HOA jusqu'à l'ordre 3.

Le casque 3D Sound One commercialisé en 2015 fonctionne à partir d'applications iOS ou Android, il permet le head tracking et l'utilisation de profils HRTF types présents dans la base de données produite par Orange Labs. L'algorithme permettra d'utiliser dans le futur des fonctions HRTF individualisées. Il est intéressant de noter que le head tracking se fait sur un espace 3D (horizontal et vertical).

Les autres systèmes d'écoute binaurale offrent quelques « presets » (un à trois) de valeurs de HRTF moyennées, c'est le cas des procédés Dolby Headphone, Fraunhofer (mp3D, MPEG Surround) et Beyer Headzone.

La qualité de rendu binaural sur casque dépend :

- du « head tracking » ou des micromouvements afin de solliciter les indices spectraux dynamiques ;
- de la personnalisation des HRTF ;
- de l'éducation au binaural ;
- de l'interaction potentielle avec les informations visuelles ;
- et de la préparation auditive.

Les fonctions HRTF varient en fonction des sujets et il est difficile d'obtenir des valeurs de HRTF moyennées satisfaisantes. L'utilisation de HRTF personnalisées donne alors de très bons résultats de localisation avec une bien meilleure externalisation. Voici des éléments de comparaison HRTF individualisées/non individualisées issus d'une synthèse de publications faite par Rozenn Nicol, Orange Labs.

HRTF INDIVIDUALISÉES/NON INDIVIDUALISÉES

HRTF non individualisées

- Externalisation pauvre.
- Localisation en élévation pauvre.
- Inversions frontal/arrière et bas/haut.
- Les sources frontales sont perçues au-dessus ou à l'arrière.

HRTF individualisées

- Le système d'audition ne fait pas la différence entre la source sonore virtuelle et la source réelle.
- Bonne externalisation.
- Une petite augmentation de l'erreur de localisation (élévation et inversions frontal/arrière) par rapport à la source réelle.

Il existe plusieurs méthodes pour obtenir des HRTF individualisées :

- la mesure personnalisée ;
- la méthode de calcul à partir d'un maillage de la tête (calcul par éléments finis de dimension quelques millimètres) ;
- la méthode de reconstruction reposant sur des approximations par sphères ou harmoniques sphériques avec interpolations spatiales ;
- par sélection à partir d'une base de données HRTF (le tour de tête donne l'ITD mais l'effet du pavillon est le plus dur à reproduire : les techniques de photos existent mais elles sont peu précises) ;
- par sélection à partir d'une base de données HRTF et transformation, la dimension du pavillon et son orientation étant les deux principaux paramètres.

Le projet BiLi devrait aboutir à des solutions pour l'amélioration de la qualité de restitution et la simplification du procédé de synthèse binaurale, tenant compte de la morphologie des individus.

3.5 Impression spatiale

3.5.1 Définition

De nombreuses publications en acoustique des salles étudient l'impression spatiale. Ces études nous permettent de caractériser à partir de critères bien définis les différents rendus sonores des salles.

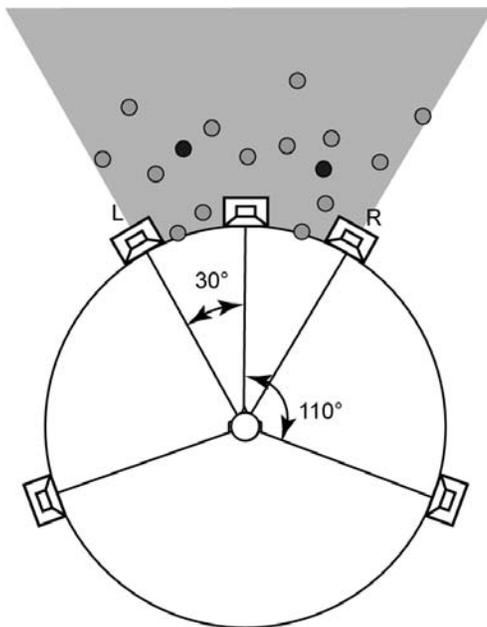
Voici la définition de l'**impression spatiale** que J. Blauert donne : c'est « l'image conceptuelle du type, de la taille et des propriétés d'une salle », réelle ou virtuelle. C'est la sensation d'espace, l'impression d'être entouré par les sons.

Cette impression repose sur différents critères définis par J.M. Potter. Des attributs perceptifs qui reprennent ces critères ont été développés dans l'étude de M. Deschamps dont voici la synthèse :

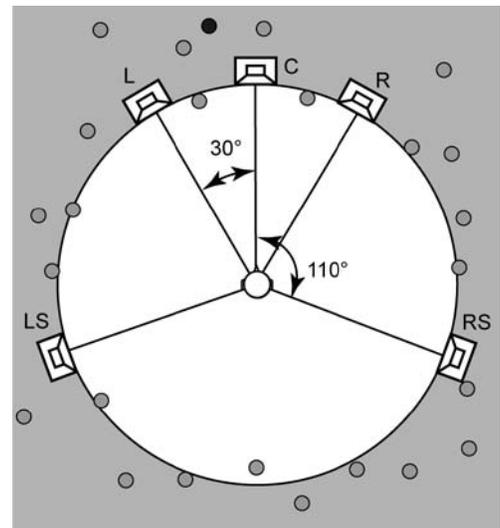
- la taille apparente de la salle ;
- « Spaciousness » : la largeur apparente des sources (ASW, Apparent Source Width) ; l'enveloppement de l'auditeur (LEV, Listener Envelopment) ;
- la réverbérance : rémanence temporelle des événements sonores résultant des réflexions tardives et de la réverbération.

Le paramètre « spaciousness » pourrait être traduit en français par spatialisation mais la signification n'est pas exactement la même. En effet, la notion de spatialisation induit une localisation des sources dans une salle plutôt qu'une quantité d'espace occupée par celles-ci.

La **largeur**, ou ASW, est un critère bien connu en acoustique des salles. Dans le contexte de perception, on parlera de largeur reposant tout d'abord sur la localisation externe d'un sujet, ce dernier pouvant s'étaler d'un point extrême depuis l'enceinte gauche dans le frontal, vers l'enceinte droite. D'autre part, en comparant la mono, la stéréo puis l'écoute sur les trois enceintes frontales du 5.1, la sensation d'espace ou d'ouverture est croissante.



- Réverbérance > 40 ms
- 20 ms < Premières réflexions < 40 ms
- 0 ms < Son direct < 10 ms



- Réverbérance > 40 ms
- 20 ms < Premières réflexions < 40 ms
- 0 ms < Son direct < 10 ms

Figures 3.28 et 3.29 – Distribution spatiale des réflexions dans une salle en stéréo (a) et en 5.1 (b),

© B. Périaux d'après G. Theile.

RÔLE DES DIFFÉRENTES COMPOSANTES DU SON

- Son direct : localisation.
- Réflexions latérales précoces entre 15-50 ms : distance et espace, largeur de la source.
- Réverbération : après 50 ms, enveloppement.

Champ des premières réflexions
signaux distincts corrélés aux deux oreilles
provenant des réflexions précoces
distinctes

Champ diffus
signaux diffus décorrélés aux deux oreilles
provenant des réflexions depuis toutes
les directions

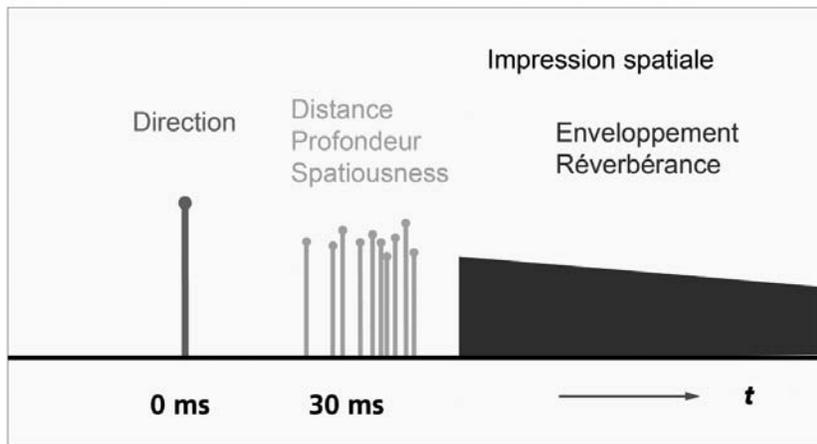


Figure 3.30 – Représentation du rôle de chaque réflexion, © B. Périaux d'après Günther Theile.

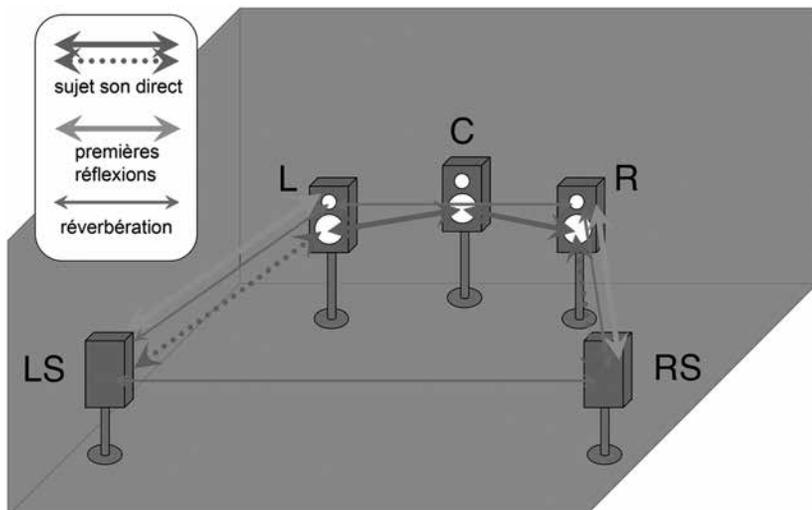


Figure 3.31 – Reproduction spatiale en 5.1, esthétique frontale, © B. Périaux d'après Günther Theile.

3.5.2 Enveloppement

Selon Arthur Harold Marshall, et telle que la décrit Magali Deschamps, la notion d'enveloppement définit la sensation que peut éprouver un auditeur dans une salle de concert : cette sensation correspondrait à la différence entre se sentir à l'intérieur de la musique, et la regarder comme à travers une fenêtre.

Le critère d'enveloppement repose principalement sur deux paramètres, la **décorrélacion interaurale**, différence entre le signal qui arrive à l'oreille gauche et celui qui arrive à l'oreille droite, et le **contenu basses fréquences** en dessous de 600 Hz. Plus les basses fréquences sont décorrélatées, plus la sensation d'enveloppement est grande.

Dans l'étude de l'acoustique des salles, on définit plusieurs critères afin de mesurer l'enveloppement, voici les principaux :

- **LF** : Lateral energy Fraction, rapport entre l'énergie des premières réflexions latérales et omnidirectionnelles.

Ce premier paramètre traduit l'importance des premières réflexions latérales précoces (arrivant dans les premières 80 ms) ;

- **IACC** : Interaural Cross Correlation.

Ce critère est lié à la décorrélation interaurale, il mesure la corrélation des signaux arrivant à l'oreille gauche et droite ;

- Rapport de l'énergie des réflexions frontales sur l'énergie des réflexions arrières.

Ce dernier critère met en avant le rapport frontal/arrière des réflexions.

D. Griesinger définit autrement trois critères pour caractériser l'impression spatiale : le CSI, l'ESI, le BSI. En voici la définition, traduite par Magali Deschamps :

- **CSI** (Continuous Spatial Impression) : cette impression survient lorsque le son direct d'une source continue (qui ne peut être divisée en différents événements sonores) interagit avec le son réfléchi latéralement, arrivant au moins 10 ms après le son direct. Le CSI dépend du rapport entre l'énergie sonore du plan médian et l'énergie du plan latéral, il est totalement enveloppant, le son semble venir de toutes les directions autour de l'auditeur ;
- **ESI** (Early Spatial Impression) : cette impression est liée aux différents événements sonores qui caractérisent le son direct, elle est générée par l'énergie latérale réfléchie arrivant dans les 50 ms suivant la fin d'un son de nature impulsive. Comme le CSI, l'ESI dépend du rapport entre l'énergie sonore du plan médian et l'énergie du plan latéral, mais il n'est pas enveloppant. Il correspond à l'impression spatiale d'une petite pièce, le son réfléchi semble venir de l'emplacement de la source sonore, et est attaché à celle-ci tout en l'élargissant ;
- **BSI** (Background Spatial Impression) : cette impression survient lorsque la source consiste en une série d'événements courts (notes courtes, phonèmes). Notre perception organise ces événements en un flux de premier plan, l'énergie arrivant dans les

interstices séparant ces événements est attribuée à un seul flux d'arrière-plan. Ces sons sont spatialement diffus et sont considérés comme enveloppants.

Nous pouvons caractériser l'enveloppement sous deux formes qui dépendent de l'esthétique de l'image souhaitée : la salle qui entoure l'auditeur (esthétique frontale) et le son direct et la salle qui enveloppe l'auditeur (esthétique équirépartie).

IMPRESSION SPATIALE

- Taille de la salle.
- « Spaciousness ».
- Largeur apparente des sources.
- Enveloppement.
- Réverbérance.

3.6 Attention auditive

3.6.1 Principe général

L'attention auditive est une notion importante dans l'étude du multicanal. En effet, face à l'augmentation des canaux et des nombreuses possibilités de localisation des sons, le cerveau va devoir traiter beaucoup d'informations auditives. Il va devoir valider la cohérence entre les signaux sonores qui se déroulent sur un cercle autour du spectateur et les informations visuelles suivies à l'écran, sur un angle de vision variable : 30° pour la TVHD, 50° environ pour le cinéma, 100° environ pour le format Super Hi-Vision 8K de la NHK, 180° pour les formats Imax, etc. Le latéral et l'arrière sont donc des zones non contrôlées par l'œil, et par conséquent tout son s'y trouvant est synonyme de danger. Le premier réflexe du spectateur sera de se retourner, automatiquement. On définira ce comportement comme **une réaction d'orientation réflexe de l'attention**.

Selon Nicolas Moreau, le traitement auditif se fait en trois étapes : l'étape sensorielle, l'étape figurative et l'étape cognitive.

L'étape sensorielle consiste à analyser les traits du stimulus entrant, les sons pouvant être associés forment un tout perceptif constituant un seul flux sonore. L'étape figurative structure ces flux auditifs en figure auditive validée et comparée avec les représentations auditives déjà acquises en mémoire. Enfin, l'étape cognitive confronte les représentations à l'aide des paramètres et « presets » liés au savoir, au passé et aux affects de l'auditeur.

Définissons maintenant ce qu'est la réaction d'orientation de l'attention auditive. L'oreille a un pouvoir séparateur qui permet de trier et sélectionner parmi les flux sonores coexistants : celui qui nous intéresse est la cible. Cela s'appelle l'effet « cocktail party ». Les autres flux sont considérés comme des distracteurs. À l'inverse de la cible, ils sont sous échantillonnés (inhibés), le nombre de décharges nerveuses étant

diminué. Il existe donc un filtre, situé entre le traitement sensoriel et le traitement perceptif de l'information, bloquant ainsi les distracteurs et dirigeant la cible vers le traitement perceptif. Le blocage du distracteur est remplacé par une atténuation du traitement. Nous disposons, pour traiter l'information auditive, d'un réservoir de ressources limité. L'élément cible mobilise une grande partie des ressources, les distracteurs ne bénéficiant que des ressources résiduelles, insuffisantes pour permettre un traitement profond. La réaction d'orientation a lieu lorsque le distracteur n'est pas inhibé et au contraire devient la cible de l'attention auditive, par le biais de différents stimuli sonores.

3.6.2 *Les causes de la réaction de l'orientation réflexe de l'attention*

La réaction d'orientation est gênante lorsqu'elle se porte vers la zone arrière ou sur le latéral car le spectateur risque de sortir du scénario du film. Selon N. Moreau, il existe trois types de stimuli pouvant provoquer cette réaction : le stimulus présentant une forte saillance physique (fréquentielle et dynamique), le stimulus comportant une saillance cognitive mettant en jeu des automatismes attentionnels cognitifs, et enfin le stimulus à caractère nouveau (nouveau spectrale, spatiale ou cognitive).

◆ **Saillance physique**

La saillance physique porte sur la caractéristique physique des signaux susceptibles d'attirer l'attention : leur dynamique et leur contenu en fréquence.

- Dynamique : les sons avec transitoires ou extinctions marqués attirent l'attention ;
- Fréquentielle : la zone sensible de l'oreille (1,5 kHz – 3,5 kHz) et les hautes fréquences synonymes de proximité (cf. perception de distance) sont des caractéristiques avec saillance.

Afin d'éviter une réaction d'orientation réflexe de l'attention, il existe peu d'espoir sur le traitement des sons à part comme des outils de traitement dynamique ou de traitement en fréquence, seule la coexistence des sons entre eux réellement importe (fusion frontale/arrière des sons spatialisés).

◆ **Saillance cognitive**

Parmi le flux de représentations qui occupent continuellement notre vie mentale, certaines sont susceptibles d'être activées sans que leur mention ait fait l'objet d'une intention préalable, ou d'une attention particulière. Il existe ainsi des automatismes cognitifs qui mettent en défaut les principes d'inhibition en échappant aux filtrages attentionnels mais l'arrêt de leur déroulement est coûteux et provoque une orientation de l'attention. Par exemple un son qui rappelle un souvenir dans l'affect du spectateur, une ambiance ou une musique, va provoquer le déroulement d'une pensée. L'arrêt de son déroulement est uniquement conditionné par l'intervention de l'attention. Pour provoquer cet arrêt, le sujet devra dépenser beaucoup de ressources au détriment des autres traitements en cours. La perception du message de la scène

sera en quelque sorte sous échantillonnée, sortant ainsi momentanément le spectateur du scénario.

◆ **Stimulus à caractère nouveau**

On appelle stimulus à caractère nouveau un stimulus sonore apparaissant pour la première fois. Cette notion est gérée par la mémoire sensorielle qui peut être effacée au bout d'un certain temps et un son déjà entendu peut conserver son caractère éveillant à sa deuxième apparition. Cela dépend de la durée qui sépare les apparitions du même son et du nombre de répétitions. La nouveauté peut être fréquentielle, spatiale ou cognitive. Lorsque nous répétons à l'identique deux fois le même stimulus sonore, le système s'adapte. On constate alors que les symptômes de la réaction d'orientation de l'attention s'affaiblissent jusqu'à disparaître. Un son de même localisation que le précédent éveillera moins l'attention. Il en va de même pour les timbres et les fréquences.

3.6.3 *La mémoire visuelle prépare l'identification auditive*

Claude Bailblé écrit : « le degré de prévisibilité modifie la réaction de l'auditeur à l'endroit du son émergent, la prise en compte attentionnelle n'est pas la même si la source est déjà rangée dans le preset cognitif suscité par le contexte, si la source est déjà recrutée dans la mémoire rapprochée des sons possibles, prêts à s'actualiser.

Par exemple :

- un ciel d'orage appelle le tonnerre ;
- une bagarre appelle des cris ;
- une discussion appelle des réparties mais aussi des surprises ».

La mémoire visuelle agit ainsi sur le caractère nouveau du stimulus sonore et nous sommes aussi habitués depuis le plus jeune âge à la spatialisation de certains sons comme l'orage qui peut éclater dans n'importe quelle direction.

3.6.4 *Comment gérer l'orientation attentionnelle du spectateur ?*

Le champ visuel est frontal, déployé sur 30° jusqu'à une centaine de degrés sur un champ fixe (dans le cadre du cinéma, de la TVHD), tandis que le champ sonore issu du format 5.1 associé occupe 360°. Dans la réalité, nous retrouvons cette différence spatiale entre l'auditif et le visuel, mais nous avons la possibilité de bouger la tête pour aller scruter d'autres paysages autour de nous.

Afin de décrire les incohérences dues à cette différence spatiale entre l'auditif et le visuel, nous définissons deux paramètres : la dislocalisation et la distraction.

◆ **La dislocalisation**

Claude Bailblé définit la notion de dislocalisation par l'absence de correspondance entre les provenances visuelles et auditives des événements. « Le spectateur est déboussolé

par la rotation incessante des directions sonores, les provenances visuelles ne correspondent plus aux provenances auditives. Les écarts de positions plus ou moins prononcés, voire contradictoires, empêchent la fusion des images et des sons ». Ce phénomène, présent en stéréo, s'accroît dans le format multicanal avec les possibilités étendues des directions du son.

◆ La distraction

Le phénomène de distraction a lieu lorsque le spectateur est gêné par des sources intruses présentes sur le latéral ou dans la zone arrière, ce dernier sort alors du scénario du film pendant un court instant ou bien il perd sa concentration dans son écoute.

- l'attention visuelle et l'attention auditive du spectateur sont dissociées
- l'auditeur peut être gêné par ces sources latérales ou arrière avec ou sans image

Selon Claude Bailblé, « les sources off trop incisives, dotées de transitoires d'attaque ou d'extinction trop marqués, déclenchent des réactions d'orientation involontaires, des coups d'œil instinctifs en direction de l'intrus auditif, l'œil reste tendu vers l'écran tandis que l'oreille est momentanément appelée en coulisses, sur les côtés ou en arrière ».

DISLOCALISATION ET DISTRACTION

- Dislocalisation : les provenances visuelles et auditives ne correspondent plus.
- Distraction : attention visuelle et auditive dissociées suite à des stimuli à caractère saillant sur le latéral ou à l'arrière, l'auditeur peut être aussi gêné par ces sources intruses dans un contenu sans image.

3.6.5 Quelques règles

Il est certainement utile de rappeler quelques bonnes règles dans la gestion des rapports son à l'image.

Le plan à l'image peut présenter la source sonore : un personnage parle, c'est un **son in**. Il est toujours placé au centre pour assurer la bonne fusion de l'image et du son.

Le plan peut cacher la source sonore : un personnage que l'on ne voit pas répond au précédent, **le son est hors champ**. Ce son est placé à proximité du centre, dans la zone frontale.

Le son peut être extérieur au plan et donc à l'histoire, comme la musique du film, une voix off : c'est **le son off ou extradiégétique**. Ce son peut être spatialisé.

Les ambiances qui précisent l'unité de lieu ou les effets qui renforcent l'action sont généralement spatialisées.

En théorie pour les sons in on peut distinguer des **valeurs de cadres sonores** : gros plans, plans rapprochés, plans d'ensemble. Ces valeurs de cadres ne sont pas nécessairement corrélées aux valeurs du cadre image. Le réalisateur peut introduire, par le son, un niveau de proximité avec l'action de la scène dont il nous distancie par l'image. La spatialisation suit le même sens, les valeurs de cadres sonores ne définissent pas toujours l'enveloppement. Généralement, le monteur son fait un choix pertinent des sources off localisées sur les côtés et à l'arrière en fonction de leur poids attentionnel (caractère nouveau, leurs caractéristiques dynamiques, fréquentielles), leur degré de prévisibilité, les contraintes esthétiques de l'œuvre, le poids attentionnel de l'image (scène avec des contraintes réalistes ou image virtuelle avec peu de contraintes).

Il est judicieux de jouer sur la coexistence des sons frontaux et arrière :

- accorder les sons en fréquence, en dynamique et en espace de sorte à constituer un seul flux auditif ;
- masquage attentionnel d'un son prosexigène (caractérise la capacité qu'ont certains stimuli d'attirer automatiquement l'attention) à l'arrière par une autre stimulation frontale ;
- habitude contre un stimulus à caractère nouveau, jouer sur la ressemblance des sons entre eux.

En dernier recours, traiter en fréquence, temporellement et en dynamique les sons arrière pour éviter toute distraction.

3.7 Ébauche de solutions pour l'image sonore multicanal

Après l'étude psycho acoustique de la spatialisation du son abordée dans ce chapitre, voici une ébauche de solutions pour l'analyse d'une image sonore spatialisée.

3.7.1 *Esthétique*

- Image réelle, image virtuelle ;
- Esthétique frontale, esthétique équirépartie.

3.7.2 *Localisation*

- Garder une dominance frontale des sources, importance de l'équilibre frontal/arrière en niveau ;
- Localisation interne des sources : effet, réalisme ou artefact ?
- Fonctions de transfert HRTF : deux flux auditifs perçus lorsqu'un son est placé entre les enceintes avant et arrière du 5.1, possibilité d'accorder en fréquence le frontal et l'arrière pour renforcer la localisation latérale d'une source ;
- Perception des sources latérales imprécises et instables ;

- Perception de la distance plus importante en multicanal qu'en stéréo compte tenu du champ réverbéré additionnel présent sur les canaux arrière.

3.7.3 *Impression spatiale*

- Repose sur la décorrélation interaurale : décorrélation des cinq canaux ;
- Repose sur la présence de basses fréquences ($f < 600$ Hz) ;
- Dépend de l'équilibre frontal/arrière en niveau ;
- Contenue généralement par la réverbération de salle réelle ou artificielle, ou l'ambiance du lieu (taille du lieu, réverbérance) ;
- Enveloppement constitué de son réverbéré, ou de son direct et son réverbéré combinés.

3.7.4 *Attention auditive*

- Le son in est placé dans le frontal, au centre pour une meilleure fusion de l'image et du son ;
- Le son off ou extradiégétique peut être spatialisé ;
- **Choix pertinent des sources off localisées sur les côtés et à l'arrière** en fonction de leur poids attentionnel (caractère nouveau, leurs caractéristiques dynamiques et fréquentielles), leur degré de prévisibilité, les contraintes esthétiques de l'œuvre, le poids attentionnel de l'image (scène avec des contraintes réalistes ou image virtuelle avec peu de contraintes) ;
- Jouer sur la coexistence des sons frontaux et arrière : fusion des sources frontales et arrière ;
- accorder les sons en fréquence, en dynamique et en espace de sorte à constituer un seul flux auditif ;
- masquage attentionnel d'un son prosexigène à l'arrière par une autre stimulation frontale : perte des traits prosexigènes ;
- habituation contre un stimulus à caractère nouveau : jouer sur la ressemblance des sons entre eux ;

Traiter en fréquence, temporellement (délai) et en dynamique (compresseur, limiteur) les sons arrière en dernier recours pour éviter toute distraction.

Chapitre 4

L'image sonore, les esthétiques, et la réalisation

La musique de demain sera spatiale et les sons donneront l'impression de décrire des trajectoires dans l'espace, de se situer dans un univers sonore en relief.

Edgar Varèse

POINTS ESSENTIELS

- Présentation des différents critères objectifs d'analyse de l'image sonore.
- Analyse d'images sonores.
- Comparaison de systèmes de prise de son, ambiances de salle, ambiances extérieures, systèmes principaux.
- Extraits 5.1 et binauraux.
- Esthétiques abordées par les différents genres : film, fiction, animation, musique, acousmatique et acousmonium, sport, documentaire, radio, jeu vidéo.
- Éléments de réalisation.

Ce chapitre étudie le langage développé pour la fabrication des images sonores en multicanal. Il traite dans un premier temps de l'analyse de l'image sonore par un ensemble de critères techniques et artistiques afin de développer un langage commun. Un certain

nombre d'exemples sonores aident à la compréhension et à l'apprentissage de cette méthode. Puis, l'auteur abordera l'écriture de contenus en définissant les différents genres, les esthétiques possibles et en exposant des éléments de réalisation pour l'espace.

4.1 Espace de création et espace de reproduction

Avant d'aborder l'image sonore, il est nécessaire de définir les deux notions d'espace que l'on perçoit à l'écoute d'une œuvre. On dissocie **espace de création** et **espace de reproduction**. Le premier est l'espace que l'on capte avec des microphones ou que l'on génère au mixage, c'est un espace lié au contenu. Il repose sur la dimension spatiale du lieu de captation, sur les caractéristiques du système de prise de son et sur l'impression spatiale apportée par le mixage. Il s'agit d'un espace indépendant, créé en amont du système de reproduction sonore. Mais pour percevoir cet espace de création, nous avons besoin d'un système de reproduction sonore capable de restituer les caractéristiques spatiales de notre contenu (le dispositif 5.1 ITU par exemple) la salle de cinéma, ou bien le casque. On définit alors **l'espace de reproduction** comme étant un référentiel porté par les caractéristiques acoustiques du lieu dans lequel on écoute le contenu, la distance qui sépare l'auditeur des enceintes les caractéristiques des enceintes ou du casque stéréo. Le casque peut restituer un contenu binaural issu d'une prise de son binaurale native par exemple ou bien un contenu spatialisé de type 5.1 accompagné d'un procédé de synthèse binaurale. Dans le premier cas, l'espace de reproduction est limité au casque stéréo seul. Dans le second cas, il repose sur le casque et sur l'effet de synthèse binaurale nécessaire pour la reproduction spatiale de la scène. Cette synthèse joue de façon virtuelle exactement le même rôle que le dispositif 5.1 sur enceintes, elle véhicule avec le casque plus ou moins bien l'espace de création. Ainsi, on pourra analyser les différents procédés de virtualisation binaurale par leur capacité à plus ou moins bien reproduire l'espace de création des contenus.

Plus généralement, l'espace évoqué par la suite fait référence à l'espace de création lié au contenu.

4.2 Présentation des critères objectifs d'analyse

L'analyse du chapitre 3 traitant des facteurs perceptifs met en évidence un certain nombre de critères objectifs qui facilitent l'analyse de l'image sonore en multicanal. Ce sont des critères techniques et esthétiques, directeurs dans la cohérence de l'image sonore et qui ne tiennent pas compte de la réalisation du produit audiovisuel ni du goût personnel de l'auditeur. Par exemple, la réalisation d'un produit audiovisuel peut nécessiter un traitement particulier des sources sonores sans profondeur, avec de l'internalisation, apportant une certaine intimité et une proximité avec l'auditeur. Cela peut être le cas d'une voix off, placée au sweet spot. L'image perçue de cette séquence sera une image fermée, mais répondant parfaitement aux désirs du réalisateur.

Les critères proposés dans ce chapitre permettent simplement de décrire l'image sonore en s'appuyant sur un langage commun pour le mixeur, le preneur de son, le monteur son, le réalisateur, le compositeur, etc.

Les critères font l'objet d'une quantification de 1 à 5 :

- 1 = mauvais, très faible ;
- 2 = pauvre, faible ;
- 3 = correct, moyen ;
- 4 = bon, grand ;
- 5 = excellent, très grand.

4.2.1 *Esthétique de l'image, réalisme ou effet*

L'image sonore est définie par Guy Laporte (preneur de son et expert de l'image sonore) comme étant la représentation mentale d'un sujet et de son acoustique. L'étude des critères exposée dans ce chapitre repose sur un grand nombre de publications, d'ouvrages, de recherches et d'analyses d'enregistrements. On retrouve en grande partie les paramètres de la stéréophonie définis par Guy Laporte, mais développés et complétés dans le contexte du son 3D.

L'image sonore peut être traitée à travers une esthétique réelle ou virtuelle, entre **réalisme** et **effet sonore**.

L'image réelle définit la reproduction d'une scène connue corrélée au lieu d'enregistrement, accompagnée des sensations naturelles d'espace et de localisation des sons. Elle est traitée dans la plupart des cas par une **esthétique frontale** car l'action se passe généralement devant nous.

On rencontre ce cas dans une scène de film, un concert de musique classique, un événement sportif ou un documentaire. Le son direct provient de la zone frontale et l'acoustique du lieu s'installe tout autour du spectateur. Les canaux arrière sont alors exploités pleinement pour la perception de salle ou plus généralement pour donner la sensation du lieu.

L'image sonore virtuelle définit un espace imaginaire qui, en multicanal, met en scène les sons tout autour de l'auditeur. L'enveloppement du spectateur est fait de sons directs et de son réverbérés, avec des possibilités d'internalisation que l'on ne peut expérimenter dans la réalité qu'en chambre sourde ou dans des lieux exigus. L'image virtuelle est généralement traitée par une **esthétique équirépartie**, elle présente le son direct et le son réverbéré tout autour de l'auditeur.

L'image sonore peut se décomposer en plusieurs familles conceptuelles : le dessin de l'image (lequel repose sur la construction, les équilibres et la dynamique), la couleur de l'image et la qualité du signal.

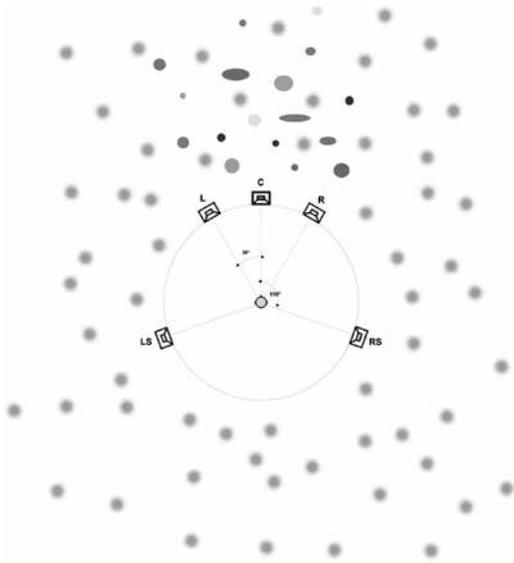


Figure 4.1 – Image sonore frontale, © B. Périaux.
En couleur sur dunod.com

Les points foncés représentent les sources directes,
les points gris flous représentent le champ réverbéré
(l'effet de salle ou l'ambiance).

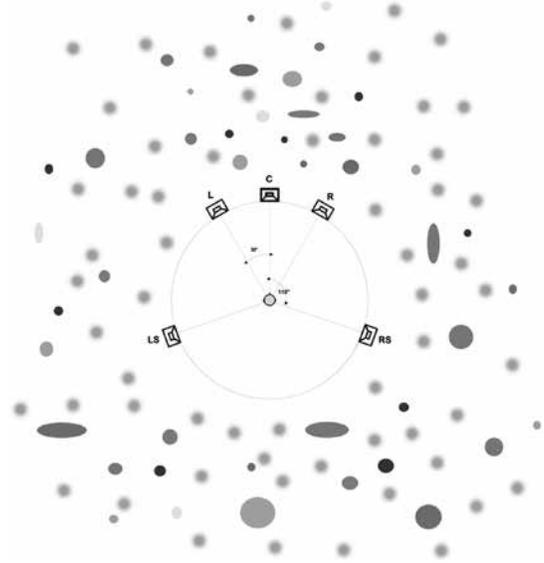


Figure 4.2 – Image sonore équirépartie,
© B. Périaux. En couleur sur dunod.com.

4.2.2 Construction de l'image

La construction de l'image traite essentiellement de la localisation des sons, de la sensation d'espace perçue et des équilibres qui font l'homogénéité du son.

◆ Localisation

Profondeur

La profondeur traduit la distance globale ressentie à l'écoute d'un enregistrement, elle repose sur la distance maximale que l'on perçoit, évaluée par le plan sonore le plus lointain ou le fond de la salle. On définit le plan sonore par un ensemble de sources perçues à égale distance de l'auditeur.

En multicanal, l'effet de profondeur prend une grande importance notamment avec l'utilisation des microphones arrières. Le plan sonore est circulaire, compte tenu de la disposition des enceintes sur le cercle défini par le dispositif 5.1 ITU. Dans le cas d'une esthétique frontale, les plans sonores prennent la forme d'arcs de cercle situés devant l'auditeur. On parle alors de plan sonore frontal.

Localisation latérale ou circulaire

Pour les images qui s'inscrivent dans une esthétique frontale, la localisation du son direct reste latérale, sur l'arc de cercle d'ouverture 60° défini par les trois enceintes L C R. Dans le cas de l'extra largeur en 5.1, l'arc de cercle peut s'ouvrir sur une centaine

de degrés ($\pm 50^\circ$) au maximum. Ce critère décrit le positionnement des sources, leur stabilité, leur mouvement et peut être quantifié.

Pour les images traitées en équirépartition, les sources sont placées sur le cercle : on parle alors de localisation circulaire.

Localisation en distance

La distance perçue dans une image multicanal est délimitée par le cercle des enceintes qui forme le premier plan sonore. On parle alors de localisation externe, l'auditeur semble pouvoir percevoir jusqu'à cinq ou six plans différents au maximum.

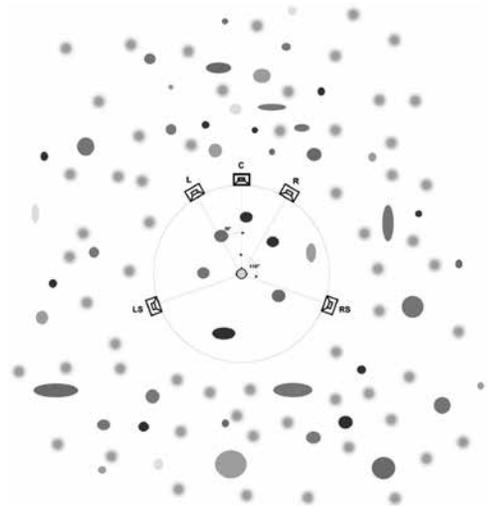
Dans le cas d'une esthétique frontale, les plans sonores définis par le rapport CD/CR s'installent sur l'arc de cercle LCR. Ces plans se transforment en cercles dans l'esthétique équirépartie. L'auditeur analyse le positionnement des sources sur ces plans, la stabilité des plans et le mouvement des objets dans la profondeur. On peut quantifier le critère.

Localisations internes

Au chapitre 3, nous avons vu qu'il existe un effet d'internalisation pour une image très peu décorrélée : les sources sont perçues dans notre cercle d'écoute, proches de la tête.

L'internalisation est une perception peu naturelle, mais elle crée des effets sonores intéressants dans le contexte des images virtuelles. Une voix placée tout autour de l'auditeur permet une proximité avec ce dernier.

Figure 4.3 – Image sonore équirépartie avec internalisation sur certaines sources, © B. Périaux. En couleur sur dunod.com



Relief

Le relief repose sur l'existence de différents plans sonores (ou cercles), il est lié aux caractéristiques du lieu. Plus le lieu est de grande taille, plus le relief est important, à condition d'utiliser des techniques de prise de son capables de restituer la profondeur.

Il ajoute un côté vivant, un certain degré de réalisme à l'image. Le critère est quantifié en fonction du nombre de plans sonores perçus.

Localisation en hauteur

Le dispositif 5.1 est plan, sur deux dimensions, il ne comporte pas d'enceinte de hauteur. Pour les installations son 3D ou pour une écoute binaurale, la localisation en hauteur prend tout son sens. L'auditeur analysera la précision de localisation, la stabilité et la linéarité du mouvement des sources.

Volume des composants en fonction de leur éloignement, la taille des objets

Le volume des composants prend une place plus importante en multicanal qu'en stéréo. Il s'agit d'analyser la place occupée par les différents éléments qui constituent le sujet de notre image, en fonction de leur éloignement. Par exemple, dans un enregistrement d'orchestre, certains instruments peuvent ne pas présenter un volume cohérent en fonction de leur place dans l'image.

La taille d'un objet peut aussi être exagérée ou diminuée en utilisant un système de captation multicanal plutôt qu'un autre. Prenons par exemple la restitution du son de l'eau déplacée lorsqu'on rame à bord d'une barque. Les mouvements peuvent être très exagérés ou bien diminués suivant le système de prise de son utilisé. Ainsi la taille de la barque perçue peut varier, on cherche plutôt une dimension réaliste, autrement le son devient un effet.

◆ **Impression spatiale**

Taille de la salle, dimension du lieu

Le volume de la salle ou du lieu est un paramètre important dans la perception de l'espace. Plus il est grand, plus l'auditeur perçoit un espace important.

Largeur apparente des sources : frontale, latérale, arrière

On définit par largeur apparente des sources, la distance maximale sur laquelle s'étend un secteur du multicanal : la zone frontale constituée des trois enceintes, la zone arrière ou les secteurs latéraux. Cette notion n'a de sens que si la perception reste externe. La largeur globale de l'image provient de la largeur du sujet et de celle générée par l'acoustique du lieu.

Dans le cas d'une esthétique frontale, la largeur globale de l'image repose essentiellement sur le sujet localisé sur les canaux L, C, R. Pour une esthétique équirépartie, la largeur provient des différents secteurs frontal, latéraux et arrière. La largeur et la taille d'un objet sont corrélés.

Enveloppement

L'enveloppement traduit la sensation d'être entouré par les sons. L'auditeur a la sensation d'être immergé dans un lieu, comme s'il se trouvait réellement dans ce lieu. Cette sensation est très fortement liée à la décorrélation interaurale et repose sur la présence de signal dans les canaux arrière, plus précisément sur le rapport en niveau entre le frontal et l'arrière (équilibre F/A). Un enveloppement optimal demande un rapport F/A bien équilibré, et une décorrélation importante des cinq canaux du multicanal. S'il existe trop de corrélation entre les canaux, l'espace se referme et tend vers une perception d'internalisation. L'enveloppement peut provenir de l'acoustique du lieu (réverbération de la salle, ambiance du lieu) ou bien de la combinaison entre son direct et acoustique.

Réverbérance

La réverbérance est définie par le prolongement temporel des événements, généré par la réverbération d'une salle ou d'un lieu, ou par des délais ajoutés dans un mixage. Ce prolongement temporel influe sur la perception d'espace.

Diffusion de la salle

La diffusion de la salle est une caractéristique qui traduit la densité des réflexions. La salle peut générer un grand nombre de réflexions ce qui donne un champ diffus très homogène, ou bien un nombre réduit de réflexions : on perçoit alors plutôt des délais.

◆ Les équilibres

Équilibre de construction

Homogénéité en construction

L'auditeur analyse la continuité de localisation, l'existence éventuelle de zones vides dans l'image, la focalisation possible vers certains canaux comme le centre, la présence de distorsions angulaires. L'homogénéité en construction repose sur la présentation des composants de l'image suivant la même échelle. Cette homogénéité repose en partie sur l'équilibre frontal/arrière et l'équilibre horizontal/hauteur (pour la perception 3D) définis ci-dessous.

L'auditeur étudie l'homogénéité de l'effet de salle, ou de l'ambiance du lieu. L'effet de salle en multicanal peut être très diffus, très homogène ou parfois polarisé vers les enceintes. La comparaison d'une croix IRT et d'un carré Hamasaki (voir chapitre 5) illustre les différentes images sonores obtenues : la première est très focalisée vers les enceintes, donnant des sources ponctuelles tandis que la seconde est très homogène et plutôt diffuse. Les exemples sonores sont donnés dans la partie web.

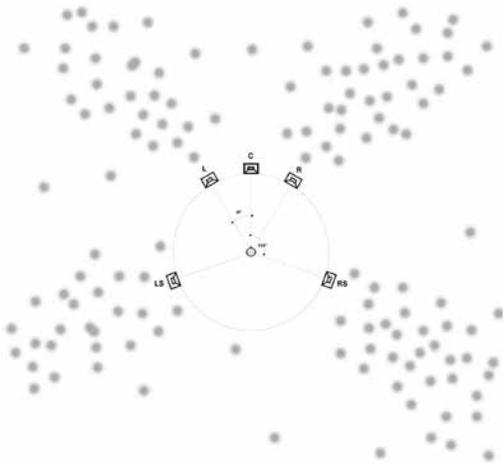


Figure 4.4 – Image sonore d'une ambiance croix IRT, © B. Périaux. En couleur sur dunod.com

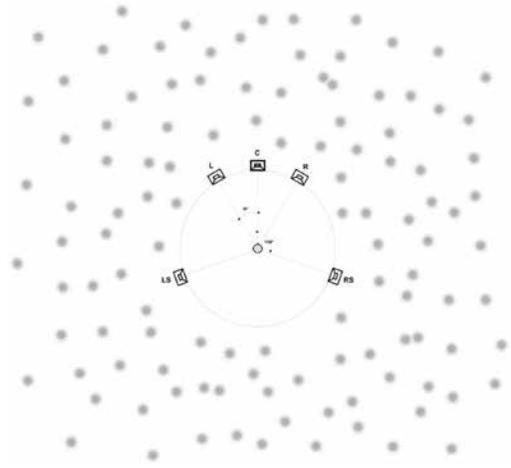


Figure 4.5 – Image sonore d'une ambiance carré Hamasaki, © B. Périaux. En couleur sur dunod.com

Équilibre Frontal/Arrière en niveau

L'équilibre frontal/arrière (Eq F/A) est un paramètre subtil, difficile à ajuster. Généralement, on recherche autant d'ambiance à l'avant qu'à l'arrière afin d'obtenir l'enveloppement optimal. Trop de son à l'arrière provoque une gêne auditive, allant jusqu'au renversement de la scène sonore. Et trop peu de son dans la zone arrière, ne permet pas l'enveloppement confortable de l'auditeur. À l'image, l'ajustement de ce critère se fait différemment : souvent on filtre les canaux arrière ou on diminue le niveau des canaux Ls et Rs pour ne pas attirer l'attention du spectateur vers l'arrière.

Équilibre Horizontal/Hauteur en niveau

On peut définir aussi l'équilibre horizontal/hauteur de la même façon pour les images son 3D. Son ajustement est délicat, trop de son alimentant les enceintes de hauteur provoque une gêne auditive.

Zone d'écoute

Il est vivement conseillé de se déplacer dans sa zone d'écoute et de changer l'orientation de sa tête, afin d'évaluer la robustesse de l'image sonore. En effet, la zone d'écoute est directement liée à la décorrélation de l'image sonore issue des systèmes de prise de son et des choix de mixage. Pour évaluer les caractéristiques de la zone d'écoute, l'auditeur évalue la taille de la zone dans laquelle l'écoute reste confortable. Il peut y avoir des variations fréquentielles et des changements dans la construction lorsque l'auditeur bouge la tête. Dans l'évaluation de ce critère, il faut prendre en compte la fragilité de l'image sonore multicanal inhérente au format 5.1 (image sonore particulièrement sensible au déplacement dans la profondeur, voir chapitre 3), le choix des enceintes (directivité) et le réglage du système d'écoute qui peut stabiliser l'image sonore (voir chapitre 2).

La dislocalisation

La dislocalisation caractérise des événements visuels et auditifs qui n'ont pas les mêmes provenances. Ce phénomène, présent en stéréo, s'accroît dans le format multicanal avec les possibilités décuplées des directions sonores. Citons, pour illustrer ce défaut, le cas d'un plan d'image serré sur un objet dont le son serait localisé sur l'enceinte gauche ou droite, ou bien le cas d'un son in placé sur les côtés ou à l'arrière. Les sources off peuvent être positionnées en dehors du canal central généralement sans créer de gêne auditive.

Équilibre spectral

L'équilibre spectral traduit la bonne ou mauvaise linéarité du spectre. On peut diviser le spectre en trois bandes de fréquences : les basses fréquences, les moyennes fréquences et les hautes fréquences. L'auditeur analysera l'équilibre spectral global de l'image ainsi que sa distribution spatiale. Afin d'obtenir une image sonore cohérente, les directions du multicanal doivent avoir le même équilibre spectral.

Équilibre CD/CR

L'équilibre champ direct sur champ réverbéré (Eq CD/CR) traduit la proportion de son direct et de réverbération ou d'ambiance. Le multicanal induit ici encore une contrainte spatiale logique. Afin d'obtenir une image cohérente, le rapport CD/CR doit être le même dans les différentes directions qui portent le sujet. Il arrive souvent qu'on obtienne des rapports CD/CR différents sur les canaux L, R et C d'un même plan sonore, pour des raisons de choix de systèmes de prise de son, d'orientation des capsules ou de choix de mixage.

Distraction

La distraction a lieu lorsqu'une source intrusive sur les côtés, à l'arrière ou en hauteur crée une certaine gêne auditive, avec ou sans présence de l'image. C'est le cas de certains mixages qui présenteraient des sons directs sur les côtés, à l'arrière ou en hauteur mal intégrés au frontal. Dans ce cas, les sources avec transitoires émergent trop et attirent l'attention. Ce peut être aussi la conséquence d'un choix de système de prise de son multicanal trop corrélée ou bien une captation où le son direct est très présent autour d'un ensemble cohérent de capsules.

4.2.3 *Couleur de l'image*

◆ **La restitution des timbres**

La qualité de restitution des timbres dans une image sonore reste assez approximative puisque la saisie se fait après passage du signal dans des transducteurs déformants, en particulier les enceintes. D'autre part, chaque source possédant un timbre qui lui est propre, l'appréciation ne peut se faire que par comparaison avec l'écoute directe de la source ou suivant une référence culturelle. Ainsi, la reproduction à l'identique est impossible.

On a en mémoire une référence de timbre sur une voix, un instrument ou un bruitage qui nous appartient. Pour caractériser le timbre, on utilise des adjectifs comme *dur, métallique, acide, agressif, strident, rond, doux, chaud, riche, plein, pauvre, creux, terne, clair*, etc. L'enveloppe temporelle des sources, la caractéristique des transitoires, la tenue des sons, l'extinction, influent sur le timbre.

La restitution des timbres reste une notion globale, peu liée à la spatialisation. Il peut cependant arriver qu'on obtienne une restitution de timbres différente entre le frontal et l'arrière suivant la configuration de prise de son ou du mixage.

◆ **La coloration de la salle, du lieu**

La salle, la réverbération artificielle, le lieu colorent les sources et peuvent les détimbrer selon le rapport CD/CR choisi, le volume du lieu ou la diffusion de la salle. La résonance de la salle ou de la réverbération artificielle met en valeur une partie du spectre du signal et crée souvent une émergence fréquentielle que le preneur de son ou le mixeur rééquilibre.

4.2.4 *Qualité du signal*

◆ **La dynamique**

Guy Laporte définit la dynamique comme étant la différence entre les passages les plus forts du programme et les passages les plus faibles. Elle évoque aussi la perception des nuances qui rend l'image sonore pleine de réalisme. Si elle véhicule le grandiose et la force dans les hauts niveaux, elle se doit également de conserver l'intelligibilité du message dans ses passages faibles tout en préservant la « propreté » du signal par un rapport signal/bruit correct. On étudiera la dynamique globale de l'image et également celle propre aux différentes parties du sujet.

Les contraintes domestiques et de voisinage rendent les grandes dynamiques peu exploitables, il est donc important d'aborder ce critère avec un certain réalisme.

◆ **La définition**

La définition peut être comparée à la notion de netteté de l'image photographique (ou de piqué). Elle désigne la pureté du signal sonore, sa finesse et la précision de ses transitoires avec laquelle on la perçoit.

L'auditeur analysera la définition de l'image dans sa globalité ainsi que son homogénéité spatiale.

◆ **L'intelligibilité**

L'intelligibilité désigne le degré de compréhension du signal. Sommes-nous en mesure de comprendre le texte chanté, parlé ou bien joué ? On parle aussi de séparation des timbres lorsque l'on étudie le spectre de l'image.

L'intelligibilité reste un paramètre global. La spatialisation améliore en général l'intelligibilité de l'image en offrant plus de place pour les différents composants et réduisant ainsi l'effet de masque.

L'auditeur analysera l'intelligibilité de l'image dans sa globalité, ainsi que son homogénéité spatiale.

◆ La transparence

La transparence est la capacité de mettre en relation immédiate les sources sonores et l'auditeur, en l'absence de tout autre élément. On peut parler de pureté ou d'absence de « voile » qui s'interpose entre la source et l'auditeur. L'analogie avec la photographie nous amène à comparer la transparence sonore à la qualité de transmission de la lumière. Cette transparence résulte des conditions de propagation de l'onde sonore (température, nature du milieu, célérité, etc.), de l'absence de micro bruits divers et de la ligne de transmission de la source jusqu'à l'auditeur, en passant par les microphones, les câbles, les convertisseurs, les supports d'enregistrement et le système de monitoring, auxquels il convient d'ajouter les conditions de propagation propres à la salle d'écoute.

L'auditeur analysera la transparence de l'image dans sa globalité ainsi que son homogénéité spatiale.

◆ Existence de microparasites

Ce dernier paramètre énumère l'ensemble des bruits qui peuvent altérer la qualité d'une image sonore : des micro-éléments d'origine acoustique ou électrique, *rumble*, saturations, distorsion harmonique, pleurage, ainsi que des artefacts très fréquemment présents sur les sons encodés.

Est proposée en **figure 4.6**, sous la forme d'un tableau, une synthèse de l'ensemble des critères abordés.

4.3 Application : analyse d'images sonores par les critères

La comparaison des systèmes de prise de son principaux met aisément en évidence des différences dans la construction des images sonores : la localisation des sources, les équilibres, l'impression spatiale. La comparaison des ambiances touche essentiellement à la notion d'enveloppement entre systèmes coïncidents et systèmes espacés.

Complément Web 4.1

Écoute d'extraits sonore 5.1 et binaural, comparaison des systèmes de prise de son multicanal

Esthétique		Image réelle ou virtuelle Esthétique: frontale (direct/ambiance) ou équirépartie	
Le dessin de l'image	Construction	Localisation	Localisation latérale ou circulaire Stabilité des sources Mouvements des sources (linéarité)
			Localisation en profondeur Stabilité des plans (ou cercles) Mouvements des sources (linéarité)
			Localisations internes ou proches de la tête Internalisation des sources
			Profondeur Distance maximale perçue
			Localisation en hauteur
			Relief Existence de différents plans
			Volume, taille des composants, des objets
	Impression spatiale	Taille de la salle	Largeur des sujets frontal, latéral, arrière
		Enveloppement (externe) Type: son direct et/ou salle	Réverbérance Prolongement temporel des sources, réverbérations, délais
		Équilibre de construction	Homogénéité en construction Présentation des composants sous la même échelle Continuité de localisation, zones vides Focalisation sur centre, zoom sur enceintes Distorsions angulaires
			Dislocalisation (rapport son/image) Correspondances visuelles et auditives
	Équilibre frontal/arrière Équilibre horizontal/hauteur Zone d'écoute Taille de la zone, variations latérale et en profondeur Robustesse de l'image		
	Éq.spectral	Équilibre spectral: BF, MF, HF	
	Éq.son direct / son réverbéré	CD/CR global, homogénéité spatiale du CD/CR	
Distraction	Réactions de l'orientation réflexe de l'attention dans les zones latérales et arrière		
Dynamique		Différence entre niveau maximum et niveau minimum Perception des nuances	
Couleur de l'image	Restitution des timbres	Caractérisation des timbres Enveloppe dynamique des sons	
	Coloration de la salle	Homogénéité en fréquence	
Qualité du signal	Définition	Globale, homogénéité spatiale	
	Intelligibilité	Globale, homogénéité spatiale	
	Transparence	Globale, homogénéité spatiale	

Figure 4.6 – Paramètres multicanal.

4.4 Genres, esthétiques et réalisation

4.4.1 Genres, esthétiques et éléments de spatialisation

Voici sous forme de tableau la synthèse de différents contenus spatialisés avec la définition du genre, les principaux éléments de spatialisation du son et le type d'esthétique.

Programme Type	Définition du genre	Éléments principaux de spatialisation, esthétique
Programmes avec image Cinéma, TV	Film, Fiction	FX, ambiances, musique, image virtuelle/réelle
	Documentaire	Ambiances, musique, image réelle/virtuelle
	Musique : concerts, documentaire musical	Ambiances, image réelle/virtuelle
	Sport	Ambiances, image réelle
	Animation	FX, ambiances, musique, image virtuelle
Programmes Radio	Documentaire	Ambiances, image réelle/virtuelle
	Musique : concert	Ambiances, image réelle/virtuelle
	Fiction	Image virtuelle, écriture spécifique
	Paysage sonore : plans séquences d'ambiance d'un lieu	Ambiances, image réelle
	Reportage	Ambiances, image réelle
	Nouvelles formes d'écriture : musique, poème mis en son, reconstitution non commentée d'un événement, adaptation théâtrale d'un texte, adaptation sonore d'un tableau	Image virtuelle, écriture spécifique
Musique supports	Album	Image virtuelle/image réelle
Jeux vidéo	Jeux de personnage : le joueur est dans la peau du personnage, contenu interactif	Déplacement du personnage, bruitages, ambiances, FX, musique
Scénographie	Installations, contenus sons et images interactifs	Spatialisation spécifique liée au lieu de diffusion
Réalité virtuelle	Simulation interactive immersive visuelle et sonore d'environnement réels ou imaginaires	Image réelle et virtuelle

4.4.2 Film, fiction, animation, documentaire, sport

Le film est le premier genre à profiter du son spatialisé, pour des raisons qui reposent à l'origine sur la problématique de diffusion en salle. La salle de cinéma implique une diffusion sur plusieurs canaux afin d'obtenir une meilleure couverture de la zone d'écoute. Avec ce type de diffusion s'est développé un langage spatial bien maîtrisé

des monteurs son et des mixeurs. Ce langage s'appuie avant tout sur le placement des dialogues, généralement dans l'enceinte frontale afin qu'ils soient intelligibles par tous et bien en rapport direct avec l'image. L'action est scénique. La musique extradiégétique, les ambiances et les effets constituent le décor généralement spatialisé.

Le film est généralement abordé par une esthétique frontale, parfois très spatialisée sur certaines séquences d'action. La spatialisation est traitée différemment suivant les scènes : il y a un temps pour l'enveloppement et un temps pour le frontal. Tout dépend de ce qui est traité, du type de film, du lieu, de l'image, de ses cadres et de son poids attentionnel... La règle partagée par les mixeurs est d'éviter de perturber un spectateur dans la vision d'une histoire, d'où l'existence des paramètres en rapport avec l'image abordée précédemment, **distraction** et **dislocalisation**.

L'ambiance est souvent utilisée pour créer des changements de lieux, comme le passage d'une scène intérieure à une scène extérieure. La voix-off, élément extradiégétique, peut être restituée très proche du spectateur, en interne, plus près que le plan défini par l'image. Mais on peut aussi se servir de la **spatialisation** pour **exprimer une émotion particulière du personnage que l'on voit** et établir des contrastes.

Les films d'animation proposent un univers imaginaire, dans lequel le mixeur peut pleinement spatialiser les effets et définir des mouvements sonores, en utilisant les canaux arrière qui marquent les déplacements à l'image.

Le documentaire est généralement spatialisé de la même façon, le spectateur suit l'histoire contée dans la zone frontale. Certaines séquences virtuelles sont plus équiréparties. Dans le documentaire animalier *Microcosmos*, certaines séquences reposent sur des images de synthèse, des schémas qui demandent une illustration sonore imaginaire, ou bien sur des scènes où le point de vue de la caméra plonge dans l'univers des insectes. L'accent est mis sur le point de vue de l'insecte : le spectateur glisse alors dans ce monde spatialisé et devient acteur.

Le documentaire musical *This is it* combine différentes esthétiques : les séquences très frontales des interviews, les répétitions réalistes du spectacle en salle avec le travail des ambiances et les vidéo clips où le son devient plus équiréparti et plus virtuel.

Le sport en multicanal est abordé par l'esthétique frontale. L'ambiance captée par un système de prise de son multicanal suspendu au niveau de la tribune ou placé sur le terrain constitue la spatialisation du son. Les sons directs sur les matchs de football ou rugby appelés « le zoom » sont disposés dans la zone frontale, ils précisent les frappes de balles, les échanges entre les joueurs et l'arbitre. Les commentaires sont généralement divergés entre centre physique et centre fantôme pour une meilleure homogénéité.

4.4.3 Musique

L'esthétique de la musique en multicanal s'appuie sur deux approches. La première est définie par l'esthétique frontale où le son direct est dans la zone frontale et le champ réverbéré est déployé tout autour de l'auditeur. C'est généralement le cas de la musique classique donnée en représentation frontale. L'opéra, le ballet, l'orchestre symphonique

s'inscrivent dans ce type d'esthétique sonore. L'image sonore est réelle, la référence est donnée par la perception de l'auditeur en salle que l'on s'efforce de reproduire.

L'ouverture de la scène sonore (l'angle sous lequel on voit le sujet) varie selon le genre abordé et suivant le type d'enregistrement, **concert** ou **disque** (sans représentation publique), elle place l'œuvre à la frontière entre le frontal et l'équiréparti. Souvent l'orchestre s'étale jusqu'aux enceintes arrière, sur un angle d'incidence de 220°, ce qui améliore considérablement la lisibilité de l'œuvre.

La seconde approche repose sur l'esthétique équirépartie, dans laquelle on dispose le champ direct et le champ réverbéré tout autour de l'auditeur. C'est le cas par exemple de la musique actuelle ou celui de la musique acousmatique définissant une image virtuelle. Dans ce contexte, la musique n'a plus cette représentation frontale, le compositeur sollicite toutes les directions du multicanal, sur des formats souvent même supérieurs au 5.1 tels que le format octophonique.

Les œuvres classiques contemporaines dont l'orchestration est très dense sont traitées en équirépartition : l'enregistrement reprend la disposition des sources dans la salle (notamment les œuvres *Répons* de Pierre Boulez et *On the transmigration of souls* de John Adams).

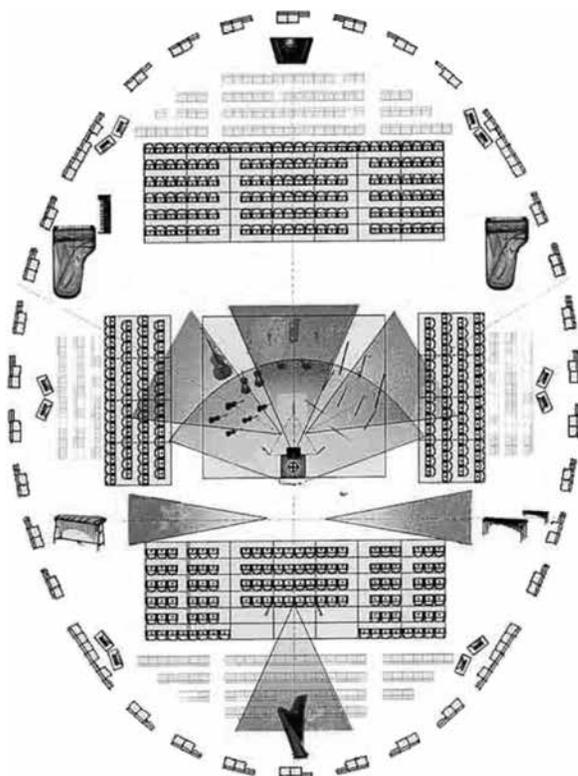
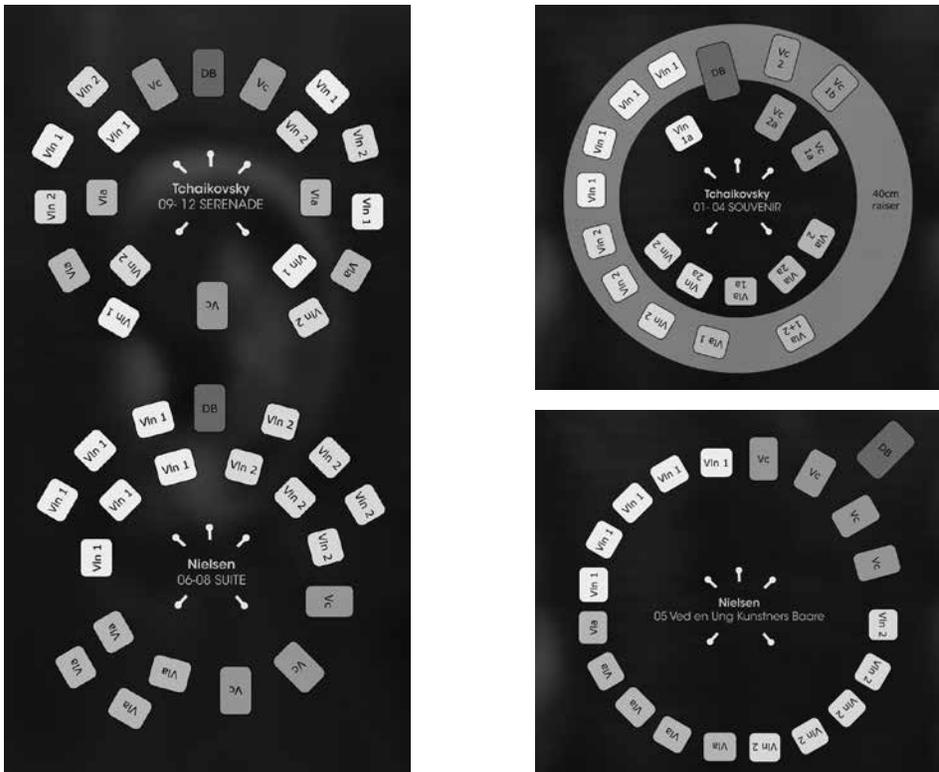


Figure 4.7 – Disposition des instruments,
Répons, P. Boulez, Cité de la musique,
© J. M. Lyzwa, CNSMDP.

Certains enregistrements d'orchestre exploitent aussi l'extra largeur en frontal sur $\pm 50^\circ$ d'ouverture. Le label norvégien 2L Lindberg Lyd, spécialisé dans l'enregistrement musical haute définition multicanal, illustre aussi cette esthétique équilibrée sur des projets comme *Souvenir* édité en 2012. L'orchestre de chambre est spatialisé tout autour de l'auditeur, exploitant aussi la zone arrière. Voici la représentation de l'orchestre de chambre pour les œuvres de Tchaïkovski *Sérénade pour cordes*, *Souvenir de Florence* et celles de Carl Nielsen *Suite et Ved en Ung Kunstners Baare*. Ces œuvres ont été enregistrées aux formats Auro-3D 9.1, 7.1 et 5.1 avec une disposition des instruments (violon 1, violon 2, alto, violoncelle, contrebasse) assez variable, mais utilisant l'ensemble de l'espace sur 360° (zone arrière comprise).



Figures 4.8, 4.9, 4.10, – Dispositions d'œuvres spatialisées pour orchestre autour du système de prise de son, extraits de *Souvenir*, Trondheim Solistene, production 2L, © Morten Lindberg 2L.

En ce qui concerne les musiques actuelles, l'album *Vespertine* de Björk sorti en 2001 est un exemple d'écriture réalisée à partir de bruitages et de sons musicaux pour un espace totalement virtuel. Le traitement interne de la voix de Björk repose sur une distribution corrélée sur les cinq canaux. L'effet apporte une certaine proximité avec l'auditeur et une bonne intelligibilité, l'utilisation de réverbération apporte la profondeur, l'enveloppement. On retrouve d'ailleurs ce type de traitement de voix dans toutes les œuvres de l'artiste. L'orchestration repose sur des sources, souvent



Figure 4.11 – Exemple de prise de son d'orchestre 3D du label 2L, ©2L

des bruitages, rythmées en mouvement, plutôt proches de l'auditeur, mélangées à des acoustiques profondes.

Souvent, les écritures 5.1 pour la musique actuelle sont équiréparties avec prédominance frontale. Les sources directes qui donnent le sens, le dialogue, sont disposées dans la zone frontale, c'est le cas aussi de certaines sources acoustiques qui attirent l'attention. Les sources directes dont la texture est enveloppante, sont placées à l'arrière et sur les côtés. Le mixeur spatialise souvent dans les DVD/Blu-ray de concerts des sources directes sur les côtés ou à l'arrière pour envelopper davantage le spectateur. Le concert de Peter Gabriel *Growing Up* (édition DVD 2003) en donne un exemple réussi. La version studio du même album *Up* (édition SACD) définit un espace qui alterne entre stéréo prédominante et 5.1 équiréparti, suivant la nature des sources : les instruments enveloppants avec peu de transitoires tels que les sons de synthèse sont situés dans la zone frontale et à l'arrière. Les sources comportant un dialogue comme la voix, les instruments solistes, sont dans le frontal. Les instruments à transitoires, les instruments électroniques et acoustiques entre le frontal et l'arrière, répartis avec symétrie.

L'album *Love* des Beatles est sorti en 2006 en DVD Audio à l'occasion de la représentation du Cirque du Soleil. Cet album a été réécrit pour le 5.1, il s'inscrit dans le même type d'esthétique. La spatialisation ici repose sur des ambiances naturelles rajoutées, les ambiances du public, les chœurs, les cordes qui exploitent tout l'espace ou la partie arrière. La voix se trouve toujours dans la zone frontale, souvent dans le canal central, parfois diverge gauche droite.

Diana Krall a sorti quelques albums en SACD ou DVD Audio mixés par Al Schmidt. Sur l'album de jazz orchestral *The look of love*, sorti en 2001, la voix est placée dans le frontal, répartie entre les canaux LCR, accompagnée de la batterie, de la contrebasse

et du piano. L'orchestre composé de cordes et de bois, accompagné de percussions, est disposé dans la zone arrière. Il existe un album plus expérimental de Diana Krall (*The girl in the other room*), édité en 2004, avec piano, batterie, contrebasse, guitare, orgue et chant. La musique joue ici avec l'internalisation de certains des instruments, ce qui donne à l'image sonore une certaine virtualité.

Dans un style différent, l'album *Playing the angel* de Depeche Mode édité en 2005, constitue une belle référence pour la réalisation d'albums studio pop-rock 5.1. La spatialisation est très cohérente, l'enveloppement est important et la gestion du centre pertinente. La voix est spatialisée en image fantôme L et R et le centre physique complète la partie instrumentale sur certains passages.

On trouve donc de nombreux exemples dans les différents genres musicaux (classique, jazz, variété, pop-rock), chaque genre trouvant sa propre esthétique. L'ouverture de la scène varie entre le frontal, l'extra largeur, le 220° qui étend le sujet jusqu'aux canaux arrière et l'équipartition où toutes les zones de l'espace sont sollicitées.

4.4.4 *L'Acousmatique et l'Acousmonium par Daniel Teruggi*

L'apparition de la musique concrète en 1948, réalisée à partir de sons enregistrés et modifiés avec des outils technologiques selon différentes techniques, remet en question les fondements de la musique et notamment les mécanismes de la création. Cette nouvelle manière de faire de la musique ouvre le possible sonore à tous les sons existants et met à l'œuvre le concept d'invention sonore, c'est-à-dire une approche consistant à fabriquer les sons qui serviront à composer la musique. Cette démarche change radicalement le processus habituel de composition fondé sur la combinatoire de sons produits à partir d'un nombre relativement restreint de sons générés par des instruments musicaux. Le compositeur fabrique des sons et il les agence dans des structures plus ou moins complexes fixées sur des supports (analogiques et ensuite numériques) pour être jouées en tant qu'objets musicaux finis.

Cette démarche change également de manière radicale la réception et la perception de la musique ; des mouvements et courants se créent utilisant la technologie de manière différente en fonction des sons ou de la finalité créatrice¹. La musique ne se compose plus comme avant, elle ne s'écoute plus de la même manière et les concerts sont d'une toute autre nature. Du point de vue de la perception elle pose le problème de la source

1. L'arrivée des technologies dans la production des sons et de la musique a créé différentes approches et courants, favorisant soit la musique enregistrée, soit la performance en temps réel ou associée à des instruments. L'appellation initiale de « musique concrète » a été donnée par Pierre Schaeffer en 1948 pour décrire cette musique faite à partir de l'écoute et l'assemblage d'objets sonores, (ces événements eurent lieu au Club d'Essai de la Radiodiffusion française). Peu de temps après, en 1953, arrivait la « musique électronique », développée à la radio de la WDR à Cologne, en Allemagne, musique qui utilisait des sons synthétiques générés par des oscillateurs, des générateurs de bruit et des filtres. D'autres appellations sont apparues par la suite : « musiques expérimentales », « live-electronics », « digital music », etc. Le terme plus générique de « musique électroacoustique » a commencé à s'appliquer vers 1960 pour décrire tout travail musical utilisant des technologies électroniques ou numériques. C'est à partir de 1972 que nous trouvons le concept de « *musique acousmatique* ».

sonore et son identification ; tout son est candidat à être musique c'est-à-dire susceptible de générer une réception de type musicale, mais comment s'opère l'écoute quand les sons entendus sont inouïs ou utilisés dans des contextes totalement inattendus ? Dès les années cinquante a été associé au processus de perception le concept d'écoute acousmatique. Ce mot très ancien décrit l'écoute d'un son sans référence visuelle ; Pythagore l'aurait introduit dans son enseignement au cinquième siècle avant JC en tant que catégorie pour y placer les « auditeurs » qui l'écoutaient derrière un rideau avant d'avoir accès à la catégorie de « mathématiciens ». Cette écoute de la source masquée est un des fondements de l'Acousmatique ; elle définit la situation dans laquelle nous entendons un signal sonore, en interprétons l'origine et essayons de comprendre l'information véhiculée. Ainsi notre perception effectue un double travail d'identification et d'interprétation basé sur notre expérience auditive. La musique concrète, ensuite acousmatique, propose un cadre d'écoute différent des musiques instrumentales, où les sources sont connues et identifiables ; l'écoute découvre, interprète, construit des cadres possibles de perception musicale et éprouve, comme dans toute musique, le plaisir des sons.

La restitution de la musique pose d'autres problèmes, de nouveaux problèmes. La source principale est le haut-parleur, donc le contexte du concert impose la présence d'un nombre variable de ces objets. Le premier concert de musique concrète en 1951 utilisait une invention appelée « pupitre de relief » (également appelé « potentiomètre d'espace ») conçu par Jacques Poullin (ingénieur travaillant avec Pierre Schaeffer), qui permet de contrôler l'intensité et le positionnement des sons en provenance d'une source monophonique sur quatre points de l'espace : avant-gauche, avant-droite, arrière-centre en haut et arrière-centre en bas, à partir d'une bobine que l'opérateur déplaçait à l'intérieur d'un champ magnétique. Ainsi naissait la première musique concrète « interprétée » en concert !²

La musique concrète s'est prolongée jusqu'à nos jours à travers le GRM de l'INA (Groupe de Recherches Musicales) qui, depuis 1958, travaille autour de la création musicale, la conception et le développement d'outils de traitement du son et des concerts spatialisés. Le concept de musique acousmatique appliqué à la création de musiques enregistrées commence à se répandre à partir de 1972 où le concept est proposé par François Bayle, compositeur et alors responsable du GRM pour décrire toutes les musiques dont la finalité est l'audition pure, sans médiation d'intermédiaires instrumentaux ou visuels. La musique acousmatique ne définit pas une esthétique particulière autre que celle du son inventé, transformé et mis dans une forme définitive et fixée, destinée à l'écoute. Environ vingt compositeurs travaillent tous les ans dans les studios de création du GRM et construisent ainsi des nouvelles perspectives pour la musique. L'objectif est de proposer un espace de création libre où le compositeur invente, expérimente sa pensée de la musique et aboutit à une œuvre qui sera donnée en concert.

2. Jacques Poullin, « L'apport des techniques d'enregistrement dans la fabrication de matières et de formes musicales nouvelles. Applications à la musique concrète », *L'Onde électrique*, vol. 34, n° 324, Paris, 1954, p. 282-291.

L'invention est le travail des compositeurs mais elle nécessite des outils, c'est pourquoi, dans cette approche, les outils de transformation et génération de sons sont indispensables pour fabriquer les éléments nécessaires aux musiques. Une démarche très souvent observée est celle de l'utilisation d'un nombre restreint de sources sonores, enregistrées et ensuite modifiées de nombreuses fois par des outils de modification spectrale, temporelles ou hybride associant des outils de synthèse. Ainsi tous les sons créés à partir des sources d'origine (tout en étant différents) conservent l'empreinte initiale de cette source et impriment aux sons créés une logique spectro-temporelle qui facilite leur intégration dans une construction musicale.

L'aspect probablement le plus intéressant est celui de la restitution des musiques en concert. Les œuvres sont produites dans des formats très divers ; nous avons vu les débuts monophoniques des premières musiques ; ensuite est arrivée, la stéréo, les quatre pistes, le huit-pistes, 16, 24 et le 5.1. Chaque format complexifie le processus de restitution, surtout si l'œuvre est donnée en public et si ce public est nombreux. Pour cela il faut des outils puissants et versatiles capables de s'adapter à des situations et des contextes acoustiques très différents. Dans la continuité, du concept d'Acousmatique, un outil appelé l'Acousmonium a été développé en 1974. C'est un outil composé d'un nombre variable de haut-parleurs distribués permettant un contrôle adaptable des musiques dans différents espaces acoustiques pouvant aller de quelques dizaines à plus de mille de personnes. Ses caractéristiques principales sont les suivantes :

- proposer une multiplicité de points de diffusion sonore distribués dans le lieu de projection, y compris dans le public grâce à des ensembles de haut-parleurs spécialisés appelés « arbres à sons » avec quatre ou six branches contenant chacune un haut-parleur de haute fréquence ;
- permettre une distribution complexe du spectre sonore : la plupart des haut-parleurs sont spécialisés dans des régions du spectre allant de l'infra grave à l'extrême aigu, d'autres ont une réponse constante sur tout le spectre ;
- offrir une modulation très souple de l'espace acoustique, depuis les haut-parleurs de proximité créant une image très nette du son jusqu'aux projecteurs lointains faisant intervenir toute l'acoustique du lieu ;
- assurer un contrôle précis de l'ensemble à travers une console de mixage avec laquelle les musiciens travaillent l'interprétation des œuvres en l'adaptant aux conditions acoustiques de la salle de concert.

Ce dispositif a permis de réaliser plus de six cents concerts depuis sa création et continue d'évoluer pour permettre de s'adapter aux formats mais surtout aux musiques. L'Acousmonium permet une « mise en espace » du son acousmatique. Le son devient source mobile, sa projection dirigée s'ouvre à l'interprétation.

4.4.5 *Radio*

Les œuvres radiophoniques écrites en multicanal sont le documentaire, le reportage, la fiction et le concert. Elles sont généralement très spatialisées, c'est le cas particulièrement

de la fiction radiophonique, où le réalisateur cherche souvent à rapprocher l'auditeur des objets sonores, dans un univers virtuel riche. La réalisation des fictions des radios suédoise SR et norvégienne NRK met en avant cette spatialisation par l'utilisation de nombreux espaces, de mouvements et d'effets. La fiction *Les sonnets de Shakespeare* produite par France Culture, abordée dans le chapitre 5, nous plonge aussi dans un univers où les sons tournent autour de nous : les voix anglaises et françaises des récitants se poursuivent dans des zones opposées de l'espace. L'effet d'internalisation y est souvent sollicité, en présentant des voix intérieures, proches de l'auditeur, soulignant le caractère intime et dramatique. L'auditeur devient véritablement acteur. Le réalisateur Claude Guerre souhaitait d'ailleurs que l'auditeur se place librement dans sa zone d'écoute et qu'il ne soit pas contraint de s'orienter vers le frontal. D'où la disparition, dans ce cas, de la polarisation du système d'écoute 5.1.

Le documentaire et le reportage multicanal en radio sont généralement des programmes dont l'esthétique est frontale, avec quelques séquences qui justifient un enveloppement de son direct. Ces scènes sont avant tout écrites et réalisées avec cette intention de spatialiser le son. *Le Singe Soleil* produit par France Inter est un documentaire qui comporte des scènes où l'on profite pleinement de l'enveloppement du son direct. La séquence des abeilles par exemple l'illustre bien : elle place l'auditeur dans la scène réelle du tournage, entouré par les abeilles, avec des effets d'internalisation où l'insecte rentre dans la zone d'écoute et frôle nos oreilles. Dans d'autres séquences, l'auditeur est immergé dans les ambiances riches et profondes du Gabon, tandis qu'une voix proche chuchote derrière lui l'histoire.

4.4.6 Jeu vidéo par Thierry Dilger

Le jeu vidéo est un exemple où le son multicanal sert essentiellement à l'interactivité. Le son est géré par le moteur audio directement en 5.1 à partir d'objets. Voici un exemple concret d'interactivité sonore avec l'analyse de la bande-son du jeu Batman.

Batman Arkham Asylum

Batman : Arkham Asylum est un jeu vidéo d'action-aventure produit par Rocksteady Studios et édité par Eidos Interactive, sorti en 2009 sur PlayStation 3, Xbox 360 et PC. Il met en scène le célèbre personnage de Batman. Peu après sa sortie, le jeu entre dans le livre Guinness des records avec la distinction de « jeu de super-héros le mieux noté de tous les temps par la presse spécialisée ».

Il a reçu les distinctions suivantes : meilleure musique et meilleur son (AceGamez 2009), meilleur Sound Design (Gamespot Readers' Choice 2009), meilleure intensité dramatique (Gamespy 2009), distinction d'excellence pour le son (IGN Reader's Choice 2009).

Sa suite, *Batman : Arkham City*, est sortie en 2011.

Audio Director : Nick Arundel

Audio Programmer : Antony Lloyd

Sound Designers : Lee Banyard et Andrew Riley

Ce jeu utilise le moteur sonore FMOD avec encodage Dolby Digital 5.1. On profite du son multicanal en utilisant les sorties mini jack du PC (ou la fibre optique via Dolby Digital Live), par fibre optique sur XBOX 360 et sur HDMI (et fibre) sur la PS3.

Présentation générale du jeu



Figure 4.12 – Batman : Arkham Asylum, © Eidos Interactive.

Dans ce jeu, nous dirigeons le personnage de Batman dans une vue 3D (à la « troisième personne », c'est-à-dire que nous voyons le corps de Batman en permanence devant nous). Nous avons la liberté de diriger librement la caméra (là où regarde Batman) ainsi que les déplacements (marcher, se baisser, voler, etc.). Il est aussi possible de porter des coups (coups de poing, coups de pied, coup rapide, coup puissant, etc.) et d'actionner des mécanismes (ouvrir la grille d'un conduit d'aération, par exemple).

Le « gameplay » (briques de jeu) est constitué principalement de phases d'exploration/énigme (trouver le passage jusqu'à la salle des gardes, actionner le circuit de ventilation, etc.) et de phases de combat (tuer des ennemis de plus en plus dangereux, soit de manière directe, soit de manière furtive).

L'histoire se déroule dans un asile de fous dont le Joker a pris possession. Batman tente de le capturer mais de nombreux pièges l'attendent. L'ambiance générale du jeu est assez oppressante : les décors sont sombres, le mobilier est cassé, il y a des gardes et des détenus, nous assistons à des scènes de torture, de nombreuses machines et mécanismes peuplent cet univers et les personnages que nous croisons sont tous dérangés.

Le son du jeu est composé de trois familles distinctes : Musique/Voix/SFX (bruitages et ambiances).

La musique est une musique orchestrale/électronique inspirée de la musique de Danny Elfman (compositeur de la B.O. du film Batman). Les voix sont typées dont celle de Batman, très froide, qui contraste avec celle du Joker, très excentrique et pleine d'humour noir. Les SFX sont composés de bruitages riches avec toute une panoplie de bruits d'actions (les différents types de coups, utilisation du grappin, tirs ennemis) et de textures sonores (ronronnement des machines, ventilation).

Les musiques

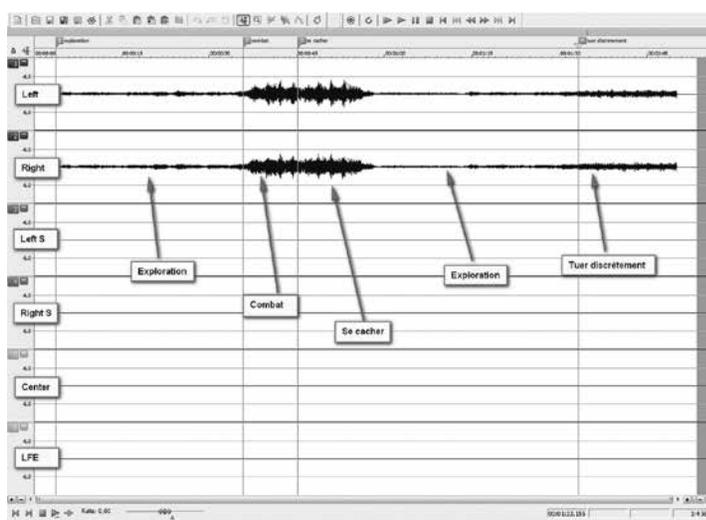


Figure 4.13 – Batman, les musiques, © T. Dilger.

La musique est omniprésente dans le jeu avec un mixage qui réagit de manière dynamique, ce qui permet d'obtenir une orchestration vivante. La musique extradiégétique souligne à la fois le côté héroïque de Batman et le monde sombre dans lequel il se trouve.

Elle est diffusée en stéréo pendant le jeu et n'utilise pas le canal Lfe. Une raison de ce choix étant de ne pas troubler les repères spatiaux du joueur (la position des ennemis) et ne pas masquer les effets spéciaux. Les trois états principaux sont :

État 1 : Exploration (en boucle – durée une minute)

C'est l'état calme du jeu. C'est le moment où l'on n'est pas en présence d'ennemis et que l'on cherche son chemin. On entend le thème principal du jeu très calme (orchestral) au cor solo, des violons, des chœurs, quelques cymbales d'orchestre, des nappes sourdes (comme des souffles), des sons de synthèse discrets et des bruitages de vent et de chocs lointains.

État 2 : Furtivité (en boucle – durée une minute)

C'est l'état qui se déclenche quand on tue discrètement un ennemi. Le thème est toujours présent mais son orchestration a changé. Maintenant, les violons sont plus présents et plus rythmés, les sons électroniques prennent plus de place. Cette partie musicale remplace doucement le premier état (utilisation de crossfade) puis reste environ deux cycles (deux minutes), même si l'on ne tue pas d'autres ennemis. Ainsi la tension est mise en place... et revient sur l'état 1.

État 3 : Combat (en boucle – durée une minute)

Les violons sont encore plus marqués et sont très rythmiques, le thème est maintenant soutenu par la section des cuivres, les percussions orchestrales font leur entrée (timbales) et les bruitages disparaissent. Cette partie se déclenche dès que Batman est repéré et s'estompe quand il est caché.

Les voix

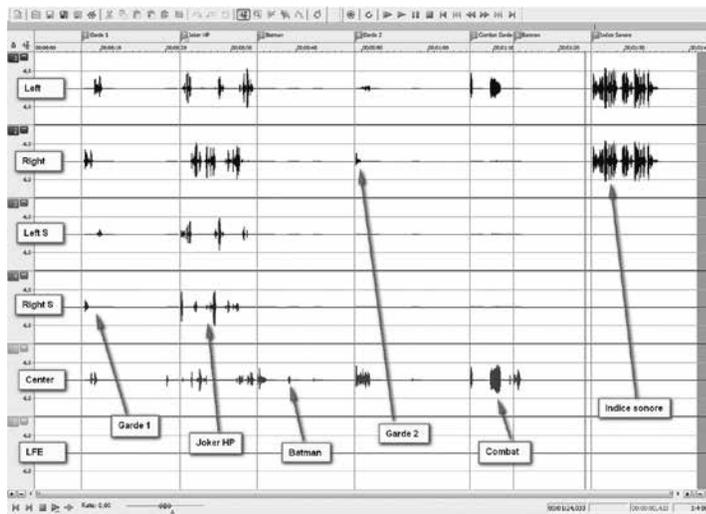


Figure 4.14 – Batman, les voix, © T. Dilger.

Les voix originales (US) sont très variées entre la voix sombre et monotone de Batman, la voix riieuse du Joker dans les haut-parleurs et sur les télévisions de l'asile, les remarques des gardes et les innombrables cris. À noter que dans un jeu de cette ampleur la localisation (traduction) est faite en treize langues.

La majorité des voix sont spatialisées en temps réel sur cinq canaux et bénéficient en plus d'une réverbération multicanal qui se transforme en fonction de la taille des différentes pièces.

La position d'un son dans l'espace 5.1 correspond à la position de la source dans le jeu (garde derrière soi dans les canaux arrière par exemple) et sa distance atténue son volume. Il n'y a pas d'utilisation du Lfe.

Layer 1 : La voix de Batman ainsi que les sons de réactions et son contact radio (une oreillette qui lui permet d'obtenir des informations) = canal central uniquement + réverbération sur les cinq canaux.

Layer 2 : La voix des gardes ainsi que leurs sons de réactions (suffocation en mourant) = spatialisés sur les cinq canaux, réverbération multicanal. La spatialisation est très précise et le moindre mouvement de tête de Batman fait bouger la voix des gardes (permet une bonne représentation de la scène spatiale et de l'endroit du danger).

Layer 3 : La voix du Joker que l'on entend souvent à travers les haut-parleurs de l'asile et les télévisions = spatialisées sur les cinq canaux avec comme point d'émission les haut-parleurs (et télévisions) accrochés dans le décor et utilisation de la réverbération multicanal.

Layer 4 : Les cinématiques (scènes narratives à regarder) ainsi que les « indices sonores » (sorte de cinématiques ne comportant que du son) = diffusés sur les trois enceintes frontales.

Les SFX (bruitages et ambiances)

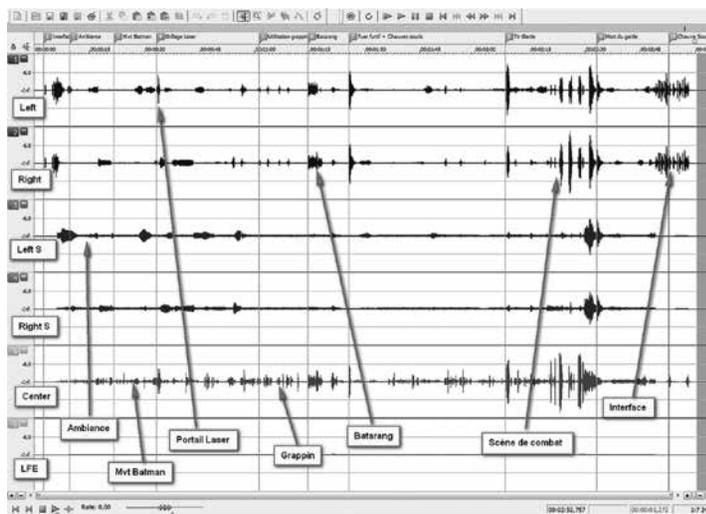


Figure 4.15 – Batman, les SFX, © T. Dilger.

Cette catégorie est la plus riche notamment pour sa diversité et les techniques de mixage spatial utilisées.

Cette fois-ci le canal Lfe est utilisé et renforce la présence des ambiances du jeu (Layer 2).

Layer 1 : Les pas et les bruits de déplacements de Batman ne sont pas mixés très fort (toujours présents). Les sonorités changent en fonction des matières rencontrées (plaque de métal, béton) = dans le canal central + réverbération 5.0.

Layer 2 : L'ambiance omniprésente dans les différents niveaux change subtilement en fonction de la position de Batman (souffle lointain, ronronnement des machines) = diffusée en statique en 5.1 (avec Lfe).

Layer 3 : Les éléments du décor comme les portails laser (grincement électronique), les machineries (grondement et cliquetis), les conduits d'aération (souffle) = spatialisés sur les cinq enceintes.

Layer 4 : Les différents coups que porte Batman (sorte de bruit de vêtements, cape) = placés dans le canal central uniquement + réverbération en 5.0.

Layer 5 : Utilisation des accessoires de Batman comme le batarang, la cape, le grappin, etc. = diffusés en statique sur les trois enceintes frontales.

Layer 6 : Passage en mode détection (pouvoir de vision spécial) = son de flash sur les cinq enceintes en statique en entrée et sur la centrale en sortie.

Layer 7 : Les tirs des gardes = spatialisés en 5.0 + réverbération avec les impacts de balle dans la centrale.

Layer 8 : L'interface du jeu (utilisation des menus) = en stéréo.

Mixage multicanal interactif

La nature imprédictible des interactions du joueur dans un monde virtuel rend la tâche du mixage extrêmement ardue. Dans le jeu Batman, rien n'empêche un joueur de réaliser plusieurs actions en simultané. Par exemple courir en lançant des batarangs vers un groupe d'ennemis qui nous tirent dessus et qui crient pendant que le Joker parle dans les haut-parleurs de la prison, tout en longeant une série d'énormes ventilateurs. Tout cela souligné de musique orchestrale dans un mixage 5.1 interactif !

C'est l'étape de l'intégration qui permet de relier les différents états du jeu avec les déclenchements sonores. Ici, le moteur FMOD est sollicité pour faire ce travail. Dans cet environnement, il faut déterminer des catégories et des sous-catégories sonores (par ex : « arme » puis « arme du joueur » puis « coup de poing »). Chaque famille créée possède des attributs (lecture en boucle, aléatoire, etc.) avec des réglages et des comportements spécifiques. On peut ainsi définir le volume de tous les coups portés par Batman et ne les assigner que dans le canal central.

Dans certaines situations, le joueur se retrouve avec des incohérences de mixage. Par exemple la hauteur (monter ou baisser la tête) n'a pas d'influence sur le son et on a du mal à associer le son du Joker avec les haut-parleurs accrochés. La réverbération 5.0 change radicalement quand on passe d'une pièce à l'autre, même si les portes sont grandes ouvertes. Les voix ne sont pas atténuées lorsqu'on se situe derrière un mur

(problème d'occlusion). Les distances d'atténuation sont parfois mal gérées et il arrive de pouvoir couper une parole en se reculant...

Il est donc nécessaire pour l'équipe son de tester un maximum de situations et, ainsi, pouvoir faire les ajustements nécessaires dans le moteur audio. Il arrive assez souvent en production que le manque de temps (ou de disponibilité d'un « programmeur son ») oblige l'équipe à faire confiance au moteur de rendu sonore et à ne pas l'adapter (utilisation des réglages par défaut).

La compatibilité stéréo du jeu via une réduction automatique par FMOD oblige aussi certains choix de travail, comme l'utilisation des enceintes frontales pour les cinématiques.

4.4.7 *Éléments de réalisation*

Après l'analyse des différents genres spacialisés, abordons maintenant la partie de la réalisation. Sont présentés ci-après les éléments de réalisation qui font sens pour la spatialisation d'une œuvre.

Le **scénario** doit toujours être compris, la spatialisation n'est là que pour aider à la compréhension du sens que l'on souhaite donner à l'œuvre. Si le sujet dont il traite est favorable au voyage du spectateur, à la découverte, au divertissement, l'espace trouvera alors tout son sens.

L'**esthétique** de l'œuvre, frontale ou équirépartie, est généralement associée au type d'image sonore, image réelle ou virtuelle, elle pose ainsi la dimension de la spatialisation. L'image réelle spatialisée vise à mettre l'auditeur en situation réelle, comme s'il vivait l'événement sur le lieu, tandis que l'image sonore virtuelle consiste à apporter un imaginaire riche en émotions.

Le **dialogue** donne le sens à l'œuvre, il est placé dans le frontal pour des raisons liées à la prédominance frontale de notre écoute et au rapport son à l'image. Les films d'animation sont souvent une exception : ils utilisent la divergence gauche et droite pour les dialogues mais c'est un genre où l'effet sonore est très sollicité. Tout **discours** est mieux perçu dans le frontal : le soliste dans une œuvre comme le dialogue d'un personnage. La sollicitation de l'arrière pour ce type de son apporte la notion d'effet sonore et peut distraire l'auditeur.

L'**information de lieu** est donnée par **les ambiances**, c'est le critère de spatialisation le plus sollicité. C'est pourquoi le lieu, s'il est riche en sons, est un sujet intéressant à réaliser en multicanal. On recherche aussi la mise en scène spatiale de séquences favorisant l'immersion de l'auditeur.

Les **sons IN** et les **sons hors champ** sont positionnés à proximité de l'écran, généralement centrés. **Les sons OFF** peuvent être divergés gauche droite, sur les côtés ou à l'arrière parfois, cette divergence peut préparer l'apparition d'un nouvel élément dans le scénario.

ÉLÉMENTS DE RÉALISATION POUR L'ESPACE

- Le scénario invite-t-il au voyage ?
- Les contraintes culturelles de l'œuvre.
- Le scénario, le sens.
- Image sonore réelle ou virtuelle (reproduction ou fiction).
- Esthétique frontale, esthétique équirépartie.
- Les éléments de discours : le dialogue, le soliste (en musique).
- L'information de lieu, les ambiances, intérieur/extérieur.
- Immersion de l'auditeur, enveloppement, mise en scène de séquences pour l'enveloppement de l'auditeur.
- Son à l'image : les sons IN, les sons hors champ, les sons OFF.
- Les cadres de l'image.
- Les cadres sonores des sons IN : gros plan, plan rapproché, plan d'ensemble.
- Le poids attentionnel de l'image.
- Le degré de prévisibilité d'une scène sonore, son spatialisé enregistré dans la mémoire auditive.
- Le point de vue donné à la scène, le déplacement du point de vue.
- L'existence d'actions se déroulant de façon simultanée à des endroits différents.
- La multitude de sources, la densité de son.
- Le confort d'écoute.
- Les sons attirant l'attention.
- Les sons favorisant l'enveloppement.
- Le sens de lecture de l'œuvre (positionnement d'écoute du spectateur ou de l'auditeur).
- Les éléments d'effet sonore.
- La source extradiégétique (musique, voix off).
- La proximité avec l'auditeur, internalisation.
- Les mouvements.
- L'effet de constellation, directions de l'espace mises en scène.
- L'étendue spectrale, la gestion du grave.
- La dynamique.
- Les changements d'espaces.
- L'espace pour coder l'émotion d'un personnage du récit, pour renforcer une émotion de l'auditeur ou pour traduire la vision sonore d'une personne.
- Le spectateur devient acteur, il s'identifie au personnage.
- La spatialisation permet de se situer et de se déplacer.

Les cadres de l'image imposent le point de vue, la spatialisation du son en résulte. Mais une corrélation des valeurs de **cadre image et de cadre sonore** des sons IN n'est pas systématique. Certains gros plans à l'image pourront « déborder » vers la zone arrière créant un certain enveloppement, d'autres resteront en frontal, centrés. **Le point de vue** peut solliciter la spatialisation des sons avec de l'enveloppement : des scènes spécifiques où le personnage entre dans un espace enveloppant IN permettent la spatialisation. Ce point de vue peut aussi être en mouvement.

Le poids attentionnel de l'image limite généralement la spatialisation des sons. C'est le cas des images avec les contraintes de la scène réelle composée de personnages dans un lieu par exemple : ce type de scènes nous ramène généralement dans l'espace frontal. Les images qui présentent peu de contrainte, telles que les séquences virtuelles constituées d'images de synthèse dans un documentaire ou un clip musical permettent beaucoup plus de liberté pour la spatialisation.

Le multicanal permet aussi le déroulement de **plusieurs actions en simultané** à différents endroits de l'espace : la fiction radio et certaines scènes de film exploitent cette possibilité.

La **multitude de sources, la densité du son** appellent naturellement à une spatialisation.

La spatialisation du son apporte ainsi un certain **confort d'écoute**.

Certaines sources ont des **caractéristiques enveloppantes**, elles s'installent aisément sur les côtés et à l'arrière. Elles peuvent être des sons naturellement spatialisés (enregistrés comme étant spatialisés dans la mémoire auditive), des sons visuellement préparés, ou des sons attirant peu l'attention.

Certains **sons attirent l'attention** tels que les sons à transitoires marqués, ceux dont le spectre est autour de 3 kHz, zone sensible de l'oreille, ou bien tout son à caractère nouveau. Ces sons, comme évoqué au chapitre 3, sont susceptibles de créer de la distraction s'ils sont placés sur les côtés, à l'arrière ou en hauteur. On peut les spatialiser autour de l'auditeur en travaillant la fusion frontal/arrière ou horizontal/hauteur des sons.

Le sens de lecture d'une œuvre spatialisée est généralement frontal. Sans l'image, l'écoute n'est pas toujours polarisée, la fiction radio est un cas à envisager où l'auditeur peut être placé dans n'importe quelle direction.

Une œuvre peut comporter des **contraintes culturelles** limitant la réalisation en multicanal. C'est souvent le cas de la musique actuelle où la stéréo a formaté le langage. C'est pourquoi il est indispensable d'écrire ou de réécrire l'œuvre en considérant les opportunités pour l'espace. Certains essais de spatialisation dans la musique classique qui impose une image réelle ne sont pas concluants car l'œuvre n'est pas respectée. Les instruments peuvent être séparés et l'harmonie écrite par ces instruments desservie.

Le caractère **extradiégétique** d'une source facilite sa spatialisation, c'est le cas de la voix off ou de la musique de film. La voix off impose un certain détachement du récit, la spatialisation peut nous aider à comprendre sons sens. La musique de film spatialisée renforce le scénario.

Le recours à l'**internalisation** apporte une certaine proximité avec l'auditeur et donne une dimension virtuelle à l'objet sonore : c'est un **effet sonore**.

L'utilisation du subwoofer permet d'étendre considérablement le spectre dans le **grave**, il agit sur l'impact des sons, sur les effets sonores, sur la dimension du son.

L'effet de **constellation** est défini par la sollicitation assez rapide des différentes directions de l'image sonore par de nombreuses sources localisables : il apporte un enveloppement saisissant.

Le mouvement dans la réalisation d'une œuvre multicanal se trouve en quelque sorte libéré : la fiction radio, le film d'animation, la musique acousmatique, le jeu vidéo sont des genres qui l'exploitent de façon quasi permanente.

La dynamique du son dans les formats spatialisés est considérable, les changements d'espaces sont très sollicités, notamment pour ponctuer l'enchaînement des scènes d'un film.

La spatialisation peut aussi **coder l'émotion** du personnage afin de la renforcer et de donner à l'auditeur la possibilité de s'identifier au personnage du récit : le spectateur devient acteur. L'espace permet aussi de coder la vision sonore d'une personne, la traduction sonore de ses pensées et de ses émotions. Cette représentation peut être naturellement portée par une image sonore virtuelle.

La spatialisation permet aussi tout simplement de **se situer et de se déplacer** dans le jeu vidéo de personnage par exemple ou dans les installations scénographiques.

Chapitre 5

La prise de son

En son multicanal, vous n'écoutez plus les échos d'un reportage, mais vous faites partie de l'expédition.

Guy Senaux

POINTS ESSENTIELS

- Rappels sur la prise de son stéréophonique.
- Principe de la prise de son multicanal.
- Systèmes principaux espacés : MMAD, OCTS, arbre Decca, arbre Fukada, INA5.
- Systèmes principaux matricés : double MS, Soundfield, sphère Schoeps KFM360.
- Systèmes d'ambiances : double ORTF, ORTF Surround, croix IRT, carré Hamasaki, omni square, holophone H2 pro, DPA 5100.
- Combinaison système principal, système d'ambiances et microphones de proximité.
- Haute résolution spatiale : Trinnov SRP, HOA.
- Prise de son binaurale.
- Prise de son en tournage : enregistreurs, pièges et astuces, méthodes.
- Applications : Film, documentaire à l'image, reportage/documentaire/fiction, radio, sport TVHD, musique TVHD/Film.

Nous définissons la prise de son multicanal d'une scène sonore par l'utilisation d'un système de captation prévu pour alimenter les enceintes du dispositif de restitution, généralement les cinq enceintes du 5.1. Nous utilisons donc un ensemble de microphones spatialisés, afin de capter l'acoustique d'une salle ou l'ambiance d'un lieu. La scène sonore doit être spatialisée dans une salle suffisamment réverbérante ou dans un environnement ayant des propriétés sonores enveloppantes afin d'obtenir une image sonore spatialisée intéressante. Lorsque les priorités spatiales du lieu sont limitées, la construction du sujet sur les trois canaux du frontal donne généralement des résultats satisfaisants.

Il existe un grand nombre de systèmes travaillant à trois, quatre ou cinq microphones. Il est important de faire la différence entre un système principal et un système d'ambiance. Le système principal est comparable au couple stéréophonique, il donne l'image sonore principale avec une spatialisé des sources et un rapport champ direct sur champ réverbéré adéquat, de telle sorte que les éléments sonores sont suffisamment timbrés. C'est une photo de la scène qui apporte l'information de localisation des sources et qui donne un point de vue. En général, on construit le reste de l'image sonore à partir de ce système principal.

Le système d'ambiance ne présente pas de sources directes, il pose en quelque sorte le décor, l'atmosphère d'une scène, en restituant une certaine distance. Ensuite, on travaille en général le reste de l'image en intégrant le son direct en frontal, ou parfois tout autour de l'auditeur, capté en mono, en stéréo ou à partir de trois ou cinq canaux.

5.1 Rappels sur la prise de son stéréophonique

5.1.1 *Différence d'intensité*

Le standard stéréophonique impose une différence d'intensité de l'ordre de 15 dB entre les deux enceintes pour obtenir la latéralisation totale d'une source. Ainsi, deux microphones coïncidents ayant une certaine directivité et formant entre eux un angle physique, génèrent une différence d'intensité. Les systèmes XY ou MS sont les principaux systèmes coïncidents utilisés, ils restituent une très bonne localisation, stable et précise. En revanche, l'impression d'espace, la profondeur et le relief obtenus par ces systèmes sont très limités, et du fait de la directivité des microphones, le rendu sonore dans les basses fréquences est filtré. Ces systèmes donnent en revanche une bonne compatibilité descendante (downmix mono).

5.1.2 *Différence de temps*

Une différence de temps d'environ 1,1 ms entre les deux enceintes de la stéréo permet la latéralisation totale d'une source. Ainsi, deux microphones distants entre eux, génèrent une différence de temps qui permet la latéralisation des sources, mais leur localisation est très imprécise : on localise les sons à transitoires mais beaucoup moins bien les sons tenus. Le système AB définit la famille à différence de temps, il permet d'obtenir une bonne impression d'espace, et la directivité omnidirectionnelle qui peut être

utilisée donne un bon équilibre spectral. En revanche, la compatibilité descendante n'est pas toujours de qualité, elle peut poser des problèmes de filtrage en peigne dus à la différence de phase entre les signaux.

5.1.3 Combinaison intensité et temps

On associe les deux types de familles pour conjuguer intensité et temps afin de profiter des avantages de chacune. Le couple ORTF (110°, 17 cm) nous donne ainsi un bon compromis entre localisation et impression d'espace. Il existe une infinité de couples utilisant la distance et l'angle, caractérisés par leur angle de prise de son. Ces possibilités sont définies par les abaques de Mike Williams présentées dans les figures 5.1 à 5.6. Les abaques définissent les angles de prise de son en fonction des angles physiques, des distances et des directivités.

Figure 5.1 – Les directivités cardioïdes, © M. Williams.

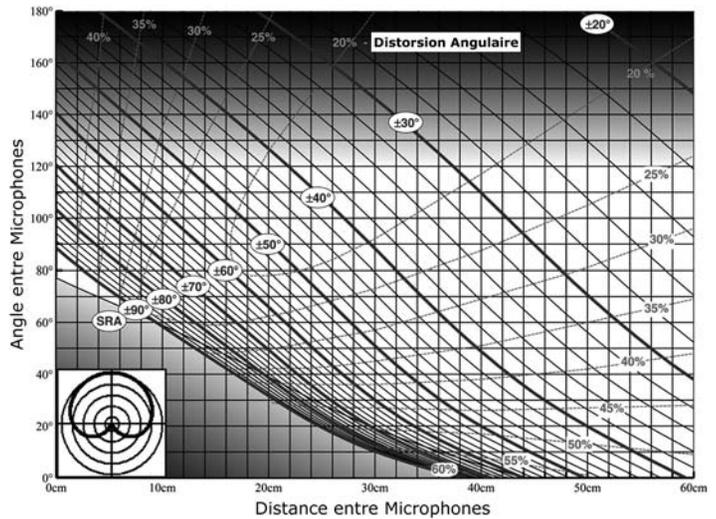


Figure 5.2 – Les directivités hypocardioides, © M. Williams.

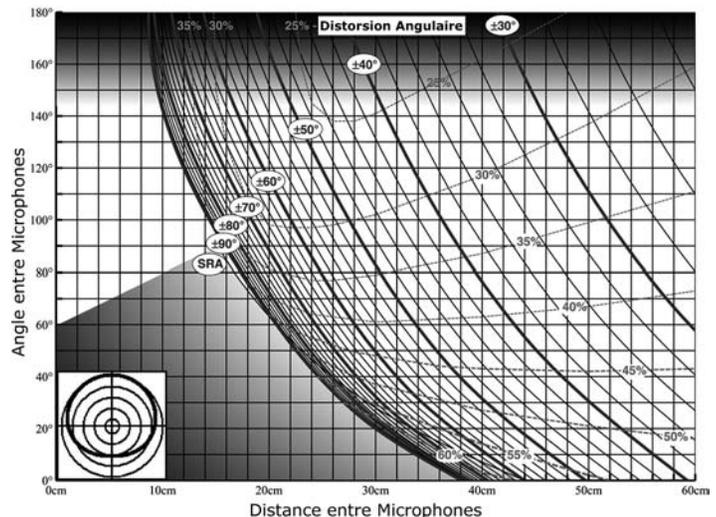


Figure 5.3 – Les directivités hypercardioïdes, © M. Williams.

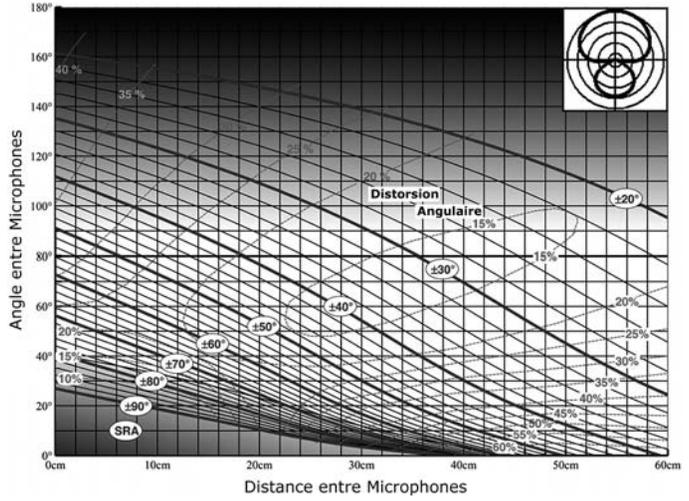


Figure 5.4 – Les directivités supercardioïdes, © M. Williams.

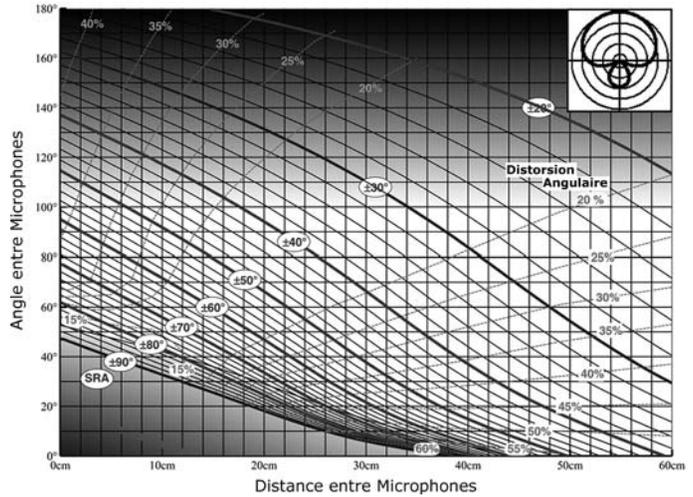
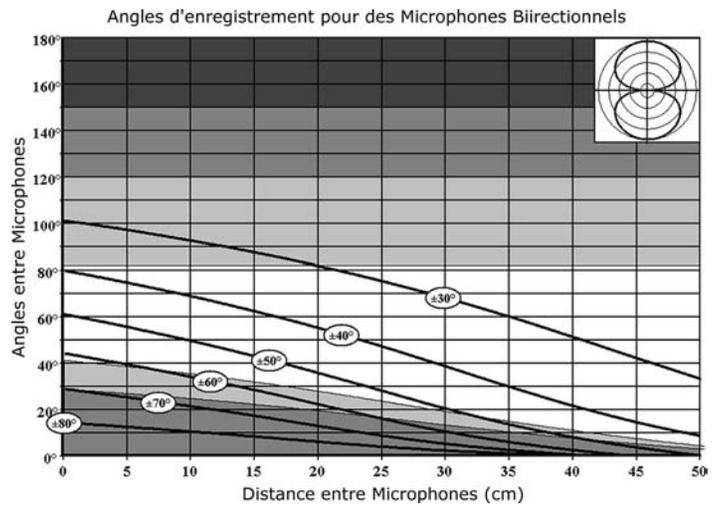


Figure 5.5 – Les directivités bidirectionnelles, © M. Williams.



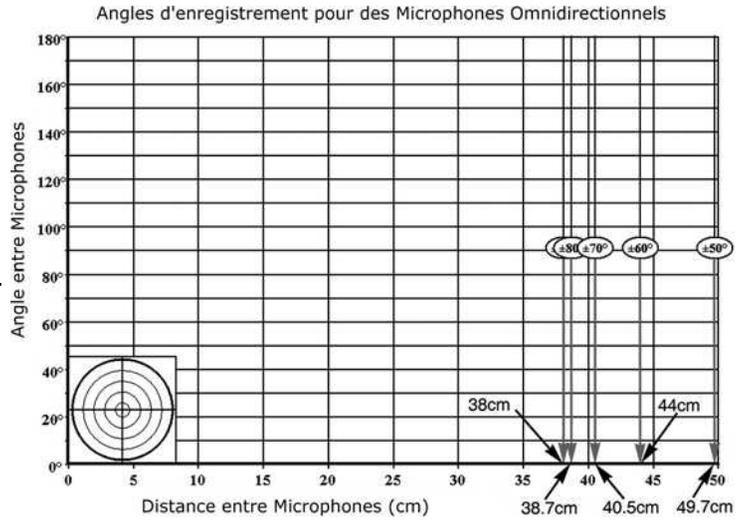


Figure 5.6 – Les directivités omnidirectionnelles, © M. Williams.

La **figure 5.7** représente l'angle de prise de son de systèmes à différence d'intensité (en ordonnées) et à différence de temps (en abscisses) conjuguées. Cet angle ici de 100° (−50°, +50°) est reproduit sur les 60° de l'espace stéréophonique par une infinité de couples. Les valeurs limites définies par le standard stéréophonique − 15 dB et 1,1 ms sont représentées.

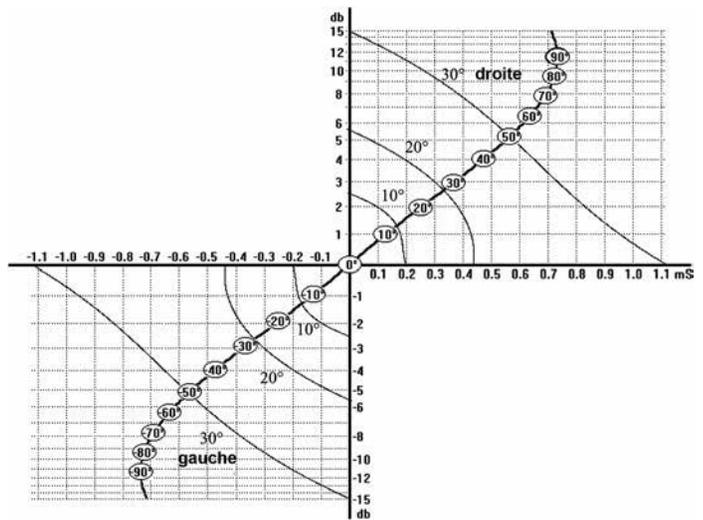


Figure 5.7 – Représentation de l'angle de reproduction stéréophonique (−30°, +30°) en fonction de l'angle de provenance de la source pour un angle de prise de son de 100°, © M. Williams.

Le couple AB omnidirectionnel avec un angle, souvent utilisé pour la musique classique, utilise la différence de temps, et la différence d'intensité dans les hautes fréquences, car un microphone omnidirectionnel devient directif à partir de 2 kHz.

5.1.4 Notion d'angle de prise de son

L'angle de prise de son est défini par les valeurs limites de différence d'intensité 15 dB et de différence de temps 1,1 ms, imposées par le standard stéréophonique. Il s'agit de l'angle dans lequel on peut disposer les sources, il est reproduit sur les 60° de la stéréophonie et, en dehors de cet angle, les sources ne peuvent être exploitées car elles se retrouvent superposées sur les enceintes. En général, le preneur de son fait correspondre l'angle d'incidence, l'angle sous lequel on voit le sujet, avec l'angle de prise de son. L'angle de prise de son varie de façon inversement proportionnelle avec l'angle physique entre les microphones et avec leur distance.

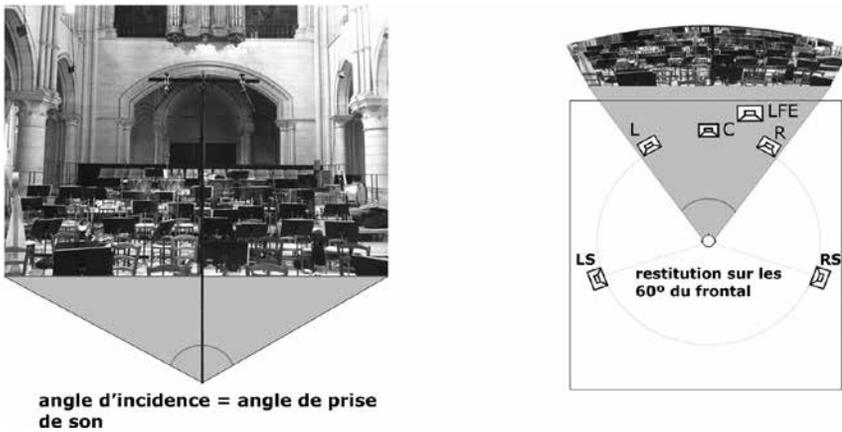


Figure 5.8 – Angle de prise de son et restitution, © B. Périaux.

5.1.5 Directivités

Les microphones que nous utilisons ont des lois panoramiques de captation d'ordre 0 pour l'omnidirectionnel et d'ordre 1 pour le bidirectionnel. Nous définissons alors, par combinaison, six directivités possibles : l'omnidirectionnel, l'hypocardiôïde, le cardiôïde, l'hypercardiôïde, le supercardiôïde et le bidirectionnel. Ces directivités sont définies par le schéma de la **figure 5.9**.

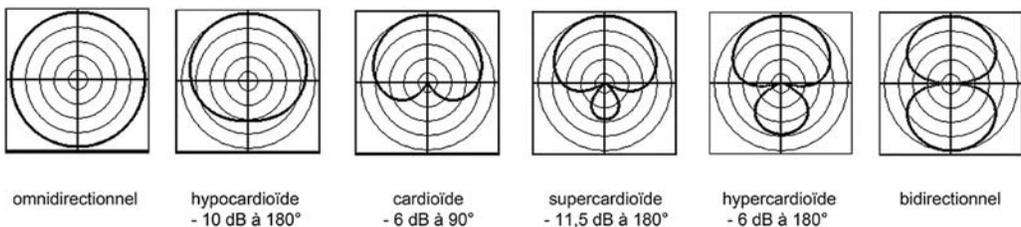


Figure 5.9 – Caractérisation des directivités omnidirectionnelle, hypocardiôïde, cardiôïde, supercardiôïde, hypercardiôïde, bidirectionnelle, © B. Périaux.

Il existe deux types de directivités hypocarديوïdes : la plus courante présente une atténuation de -10 dB à 180° , l'autre comporte une atténuation de -15 dB à 180° (capsule MK22 appelée cardioïde douce chez Schoeps par exemple). Nous considérerons par la suite l'hypocarديوïde avec -10 dB de sensibilité arrière. Nous verrons qu'il est possible de resserrer ces lois panoramiques en utilisant d'autres approches comme la haute résolution spatiale, fonctionnant à partir d'un réseau de microphones.

RAPPELS PRISE DE SON STÉRÉO

Systèmes à différence d'intensité : XY, MS.

- Latéralisation totale d'une source pour $\Delta I = 15$ dB.
- Avantages : bonne localisation, précise et stable, downmix mono de qualité.
- Inconvénients : impression d'espace et profondeur limitées.

Systèmes à différence de temps : AB parallèles.

- Latéralisation totale pour $\Delta t = 1,1$ ms.
- Avantages : bonne impression d'espace et profondeur importante, possibilité d'utiliser des microphones à directivités omnidirectionnelles améliorant le rendu spectral.
- Inconvénients : localisation imprécise, instable.

Systèmes intensité et temps : AB avec directivité et angle, AB omni avec angle.

- Bon compromis pour la localisation et l'impression d'espace.
- L'angle de prise de son varie de façon inversement proportionnelle avec l'angle physique et la distance entre les microphones.

5.2. Principe de la prise de son multicanal

Le principe de la prise de son multicanal repose sur deux notions : la localisation définie par la construction des différents secteurs de l'image sonore et l'obtention d'une bonne impression spatiale.

Si l'on considère un système constitué de cinq microphones, chaque paire de microphones adjacents se comporte comme un couple stéréophonique restituant un angle de prise de son sur l'image fantôme correspondante. La difficulté est d'optimiser la configuration pour obtenir un lien parfait entre les cinq images fantômes issues des cinq angles de prise de son, comme représenté sur la **figure 5.10**.

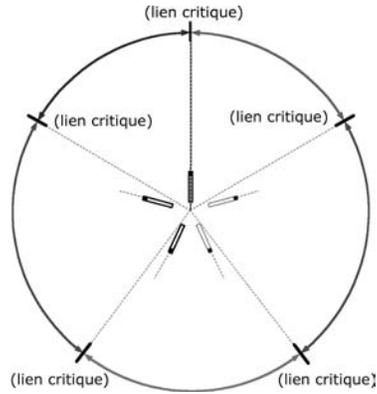


Figure 5.10 – Segmentation du champ sonore et liaison parfaite des secteurs, © G. Le Dû.

Ainsi optimisé, le dispositif ne présente aucune superposition de secteurs, ni aucune zone non couverte.

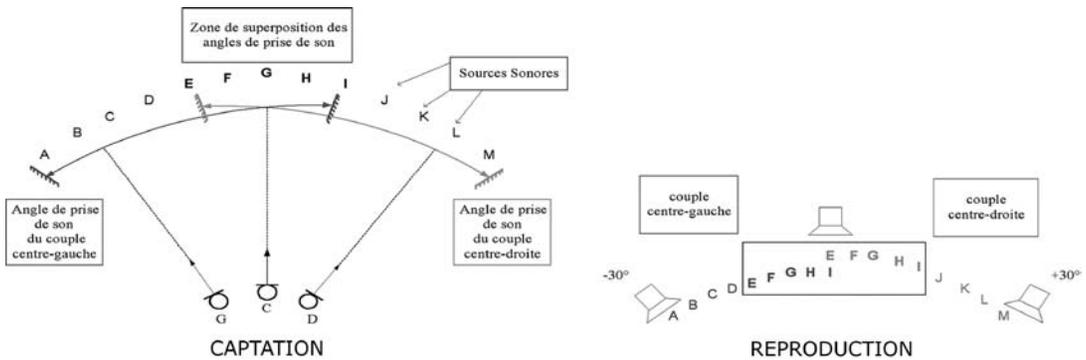


Figure 5.11 – Captation du triplet avant avec superposition des angles de prise de son, © G. Le Dû.

Comme l'indique la **figure 5.11**, s'il y a superposition des angles de prise de son LC et CR, correspondant respectivement aux couples de microphones gauche-centre et centre-droite, alors les sources EFGHI sont répétées à la restitution sur les trois enceintes frontales.

Si les deux angles de captation ne sont pas parfaitement adjacents et ne couvrent pas toute la zone frontale (voir **figure 5.12**), alors une zone non captée donnera une focalisation sur le centre avec entassement des sources FGH.

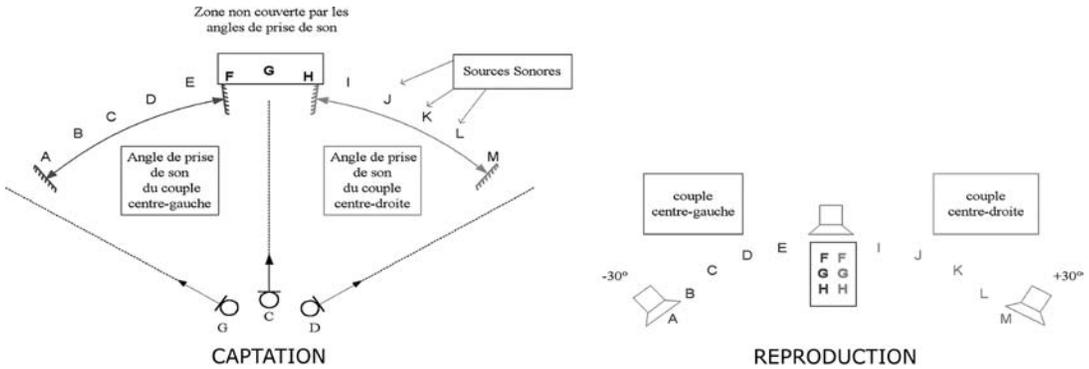


Figure 5.12 – Captation du triplet avant présentant une zone non couverte par les angles de prise de son, © G. Le Dù.

Il convient donc de choisir des systèmes cohérents pour obtenir un équilibre de construction homogène, tant pour un système principal qu'un système d'ambiance.

L'impression spatiale, comme définie au chapitre 3, fait appel principalement à la notion de décorrélation interaurale de temps. C'est pourquoi, un système de captation stéréo coïncident donne peu de sensation de profondeur et d'espace. On peut s'attendre à ce que les systèmes coïncidents en multicanal donnent de la même façon une sensation d'espace limitée.

5.3 Les systèmes principaux espacés

5.3.1 MMAD

Le système Multiphonic Microphone Array Design a été conçu par Mike Williams et Guillaume Le Dù. Il repose sur la théorie des angles de prise de son et comporte une infinité de configurations possibles publiées sous forme d'abaques.

◆ Présentation

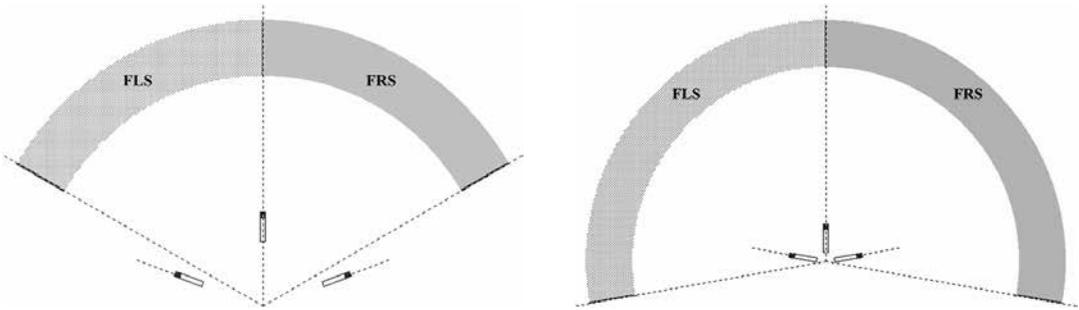
Le principe repose sur la segmentation parfaite du champ sonore, c'est-à-dire l'obtention d'un raccord parfait (« critical linking ») entre les cinq angles de prise de son pour couvrir les 360°, sans superposition ni zones non couvertes, avec une totale liberté dans le choix des différents angles de prise de son. La diaphonie acoustique entre les microphones opposés est négligeable par rapport aux couples stéréophoniques adjacents du fait des orientations opposées des microphones, de leurs distances et des directivités utilisées (cf. publication AES 118th AES Convention in Barcelona Preprint 6373, 2005).

Le système utilise les directivités cardioïdes, hypocarديوïdes, supercardioïdes, hypercardioïdes ou omnidirectionnelles au choix. Il existe aujourd'hui un support, « l'étoile de Williams », qui permet de travailler aux formats 4.0, 5.0 et 7.0 avec une configuration usuelle correspondant à des angles de prise de son de 72° pour le format 5.0 en cardioïdes, hypocarديوïdes, et omnidirectionnels. La configuration 4.0 repose sur quatre angles de captation égaux de 90°.

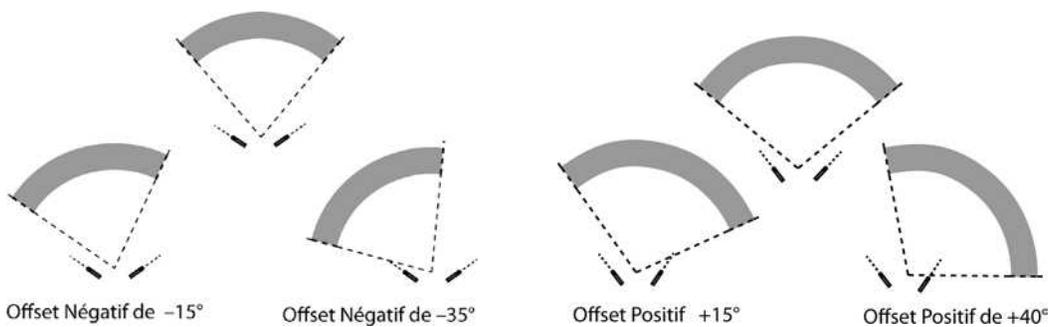
◆ Les techniques d'offsets

Pour obtenir cette liaison parfaite entre les segments, il est nécessaire de pouvoir **faire pivoter à volonté les angles de prise de son** des cinq couples dans le sens désiré, ceci afin de rendre ces angles de prise de son **asymétriques**.

Dans le triplet avant, on cherchera ainsi à aligner une extrémité de l'angle de prise de son avec l'axe du microphone central alors que dans les segments latéraux, les angles de prise de son pourront au besoin être totalement asymétriques.



Figures 5.13 et 5.14 – Liaison parfaite des segments frontaux gauche et droite, FLS : secteur frontal gauche, FRS : secteur frontal droite, © M. Williams.



Figures 5.15 et 5.16 – Angle de prise de son inférieur (5.15) et supérieur (5.16) à l'angle physique entre les microphones, © M. Williams

Les quatre types d'offsets

La **figure 5.17** représente le couple de microphones centre et droite, avec les axes d'intensité (en pointillés) et de temps (ligne perpendiculaire à l'axe qui rejoint les deux capsules et dont l'origine est située au centre) superposés. L'angle de prise de son est symétrique par rapport à cet axe.

On définit quatre types d'offsets :

Les offsets électroniques EO

EIO – Offset Intensité Électronique

ETO – Offset Temporel Électronique

Les offsets par position des microphones MPO

MPLO (ou MPIO) – Offset Intensité par position des microphones

MPTO – Offset Temporel par position des microphones

Les offsets d'intensité électronique correspondent à l'addition d'une différence d'intensité constante à la fonction différence de temps/intensité d'un couple de microphones (atténuation d'un micro). Les offsets temporels électroniques consistent à additionner une différence de temps constante à la fonction différence de temps/intensité d'un couple de microphones (retard électronique d'un micro).

Les **figures 5.18, 5.19 et 5.20** représentent des offsets par position des microphones, MPLO et MPTO :

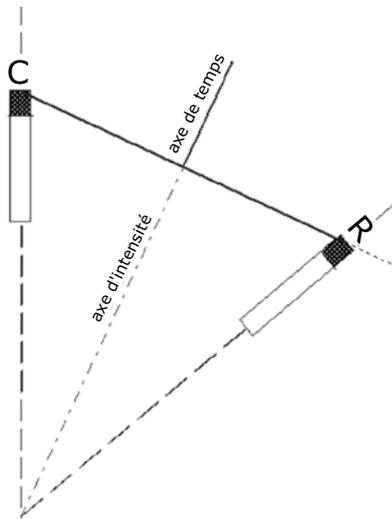


Figure 5.17 – Couple de microphones centre et droite, axes d'intensité et de temps superposés,

© M. Williams.

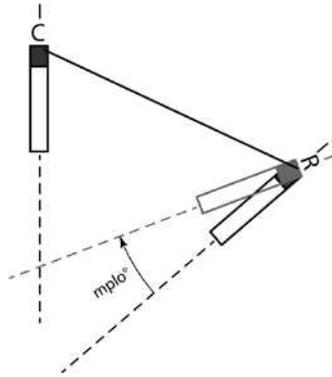


Figure 5.18 – MPLO : application d'un offset de position d'intensité en gris, © M. Williams.

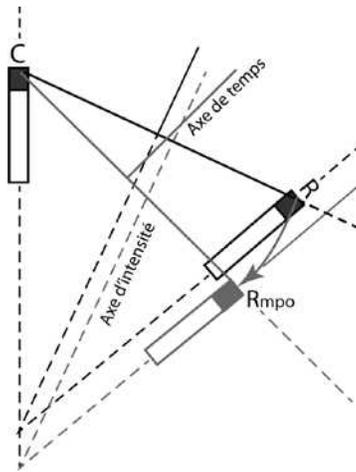


Figure 5.19 – MPTO : application d'un offset de position temporel en gris, © M. Williams.

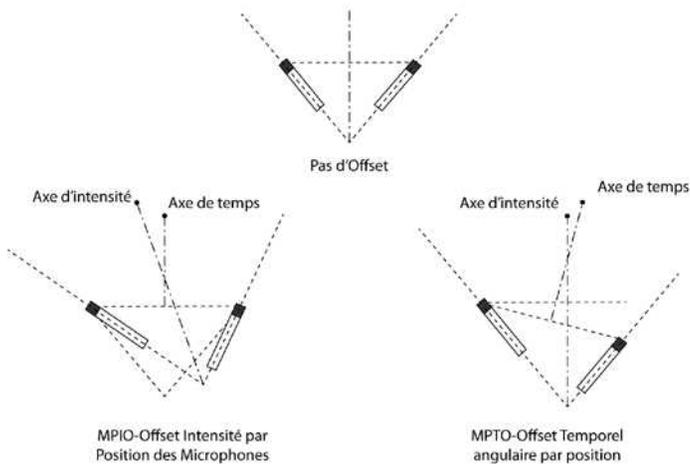


Figure 5.20 – MPTO et MPLO (ou MPIO), © M. Williams.

Dans les deux cas, les axes de temps et d'intensité ne sont plus superposés, le couple devient physiquement asymétrique. L'angle de prise de son a subi une rotation vers la droite ou vers la gauche, suivant les cas. Cette rotation de l'angle de prise de son nous permet d'établir le lien critique entre les secteurs couverts.

Les figures 5.21 et 5.22 illustrent la représentation sur le triplet :

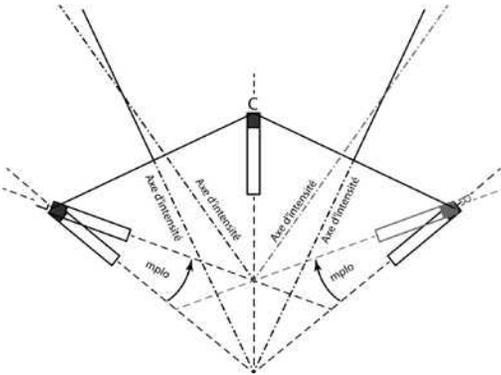


Figure 5.21 – Application d'offsets MPLO sur les deux couples gauche et droite du triplet, © M. Williams.

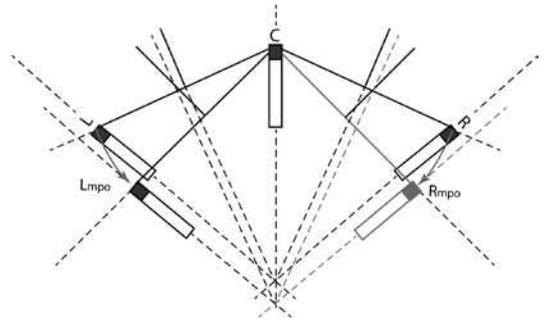


Figure 5.22 – Application d'offsets MPTO sur les deux couples gauche et droite du triplet, © M. Williams.

Les deux offsets des figures 5.21 et 5.22 sont équivalents et consistent à rendre le couple de microphones physiquement asymétrique.

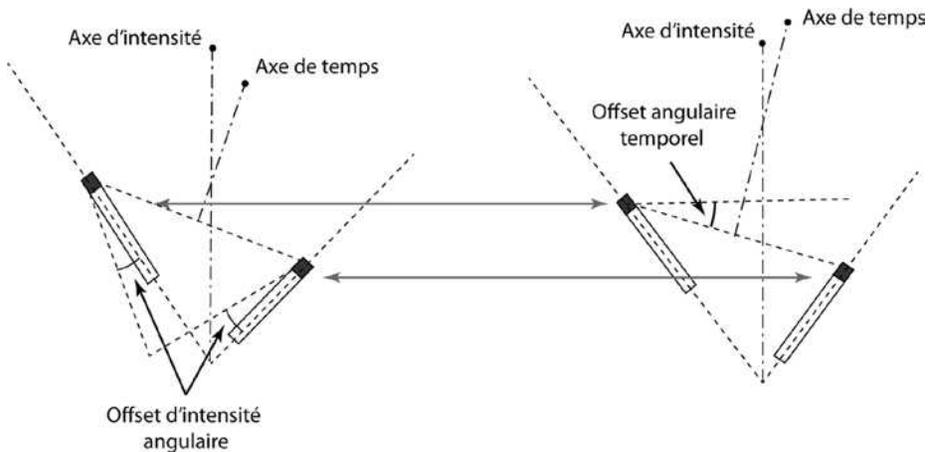


Figure 5.23 – Équivalence entre MPLO et MPTO, © M. Williams.

Les quatre types d'offsets produisent les mêmes effets : ils traduisent la courbe différence de temps / intensité du couple (en fonction de l'angle d'incidence de la source)

selon l'axe des ΔT ou l'axe des ΔI , ce qui entraîne un changement des extrémités de l'angle de prise de son (intersections avec les courbes de localisation stéréophonique) soit, plus concrètement, **une rotation de l'angle de prise de son**.

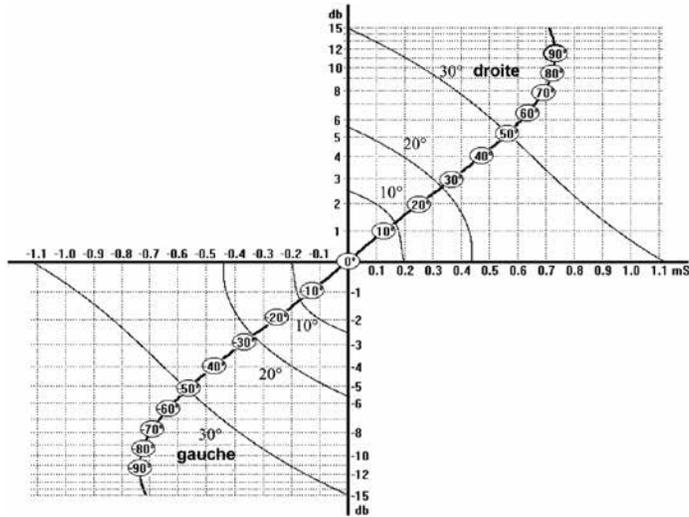


Figure 5.24 – Fonction différences d'intensité et de temps sans offset, angle de prise de son symétrique $-50^\circ / +50^\circ$, © M. Williams.

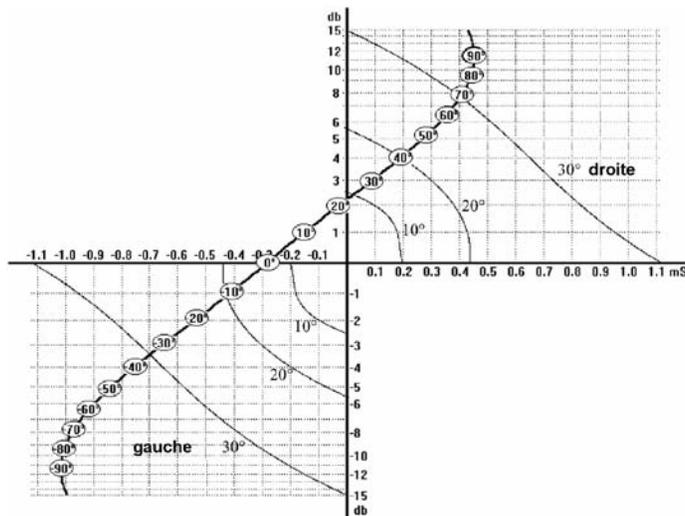


Figure 5.25 – Fonction différences d'intensité et de temps avec offset positif électronique de temps $-0,28$ ms, décalage selon l'axe Δt et angle de prise de son asymétrique $-35^\circ / +70^\circ$, © M. Williams.

◆ Mise au point d'un système complet

Le triplet avant

On définit en premier lieu les angles de prise de son des deux segments avant gauche et avant droit et on choisit une combinaison distance/angle entre les micros L, C et R qui les réalise. Ensuite, on utilise un offset par position de microphones (dissymétrisation des couples) pour réaliser la liaison parfaite des deux angles de prise de son.

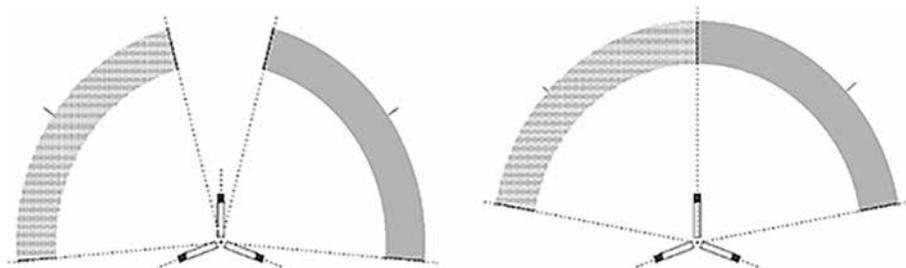


Figure 5.26 – Mise au point du triplet avant, © M. Williams.

Mise au point du segment arrière

On définit l'angle de prise de son arrière et on choisit une combinaison distance/angle qui le réalise (couple stéréo « classique »).

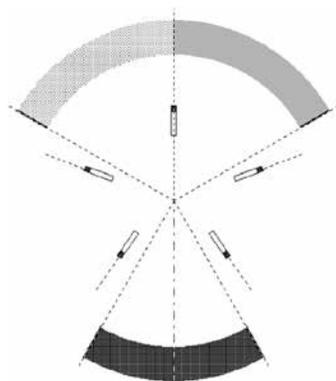


Figure 5.27 – Mise au point du segment arrière, © M. Williams.

Mise au point des segments latéraux

La valeur des angles de prise de son latéraux est imposée par les angles choisis pour l'avant et l'arrière (somme des angles = 360°). Pour réaliser ces angles de prise de son latéraux, on fait varier la distance entre le triplet avant et le couple arrière (les angles sont déjà tous déterminés).

Quand la bonne distance est ajustée, la liaison parfaite avec les segments latéraux reste encore à réaliser : on applique alors un offset électronique intensité ou temps sur le couple arrière, c'est-à-dire qu'on atténue ou on retarde le couple arrière par rapport au triplet avant, ou l'inverse, selon les cas.

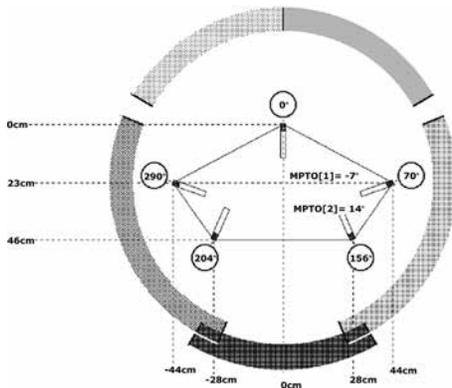


Figure 5.28 – Segments latéraux avec un offset temporel de position, inhérent au système, © M. Williams.

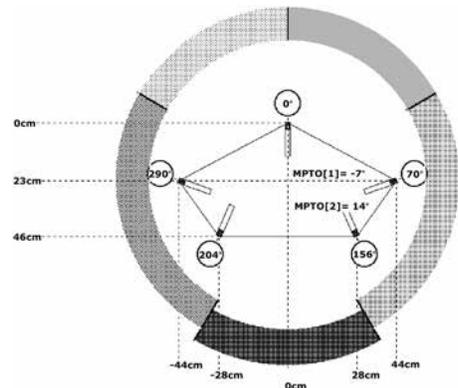


Figure 5.29 – Obtention de raccords latéraux parfaits grâce à un offset électronique sur les micros du couple arrière, © M. Williams.

◆ Caractéristiques générales

Les angles de couverture du système MMAD sont variables, et leur diffusion sur l'écoute 5.1 intègre des compressions en général dans le frontal et des extensions souvent sur les latéraux et dans la zone arrière. Il est cependant possible de compresser ou d'étendre à volonté des secteurs de l'espace sonore dans l'espace de reproduction, et grâce à la liaison parfaite des segments, de réaliser une extra largeur, c'est à dire de faire s'étendre l'image stéréophonique sur les côtés, au-delà des enceintes gauche et droite, sur une zone allant jusqu'à $\pm 50^\circ$.

La méthode est généralisable, le principe de ces systèmes reste valable quel que soit le nombre d'enceintes de diffusion.

Il existe une infinité de systèmes, comme pour les couples stéréophoniques, un nombre quasi infini de combinaisons distances/angles permet de réaliser les angles de prise de son désirés. Il est aussi utile, dans la démarche du preneur de son, de prendre en compte l'orientation des microphones par rapport aux sources pour obtenir le meilleur rendu.

Le système MMAD peut être utilisé avec n'importe quelle directivité, et cette souplesse convient parfaitement à l'approche du preneur de son qui consiste à faire le choix de la directivité en fonction des sources et de l'acoustique du lieu.

Pour la captation en salle, si l'impression de salle semble insuffisante, un système d'ambiances arrière peut être combiné au système MMAD.

◆ Le logiciel CAMAD (Computer Assisted Microphone Array Design)

La mise en œuvre manuelle d'un système MMAD est relativement complexe du fait de l'interaction d'un grand nombre de paramètres. Pour cette raison, un logiciel a été développé. Ce dernier permet de visualiser instantanément l'influence de chaque paramètre (distances, angles, offsets...) et de réaliser un système MMAD efficacement en suivant la méthode exposée précédemment.

◆ Configurations usuelles

Il existe des configurations usuelles qui fonctionnent pour la plupart des cas avec cinq angles physiques égaux de 72° , superposés aux angles de prise de son. Ces abaques n'utilisent pas d'offset électronique et sont proposés pour les cinq directivités cardioïdes, hypocarديوïdes, omnidirectionnelles, supercardioïdes et hypercardioïdes.

Complément Web 5.1

Pour aller plus loin : représentation des systèmes par le logiciel CAMAD, configurations usuelles.

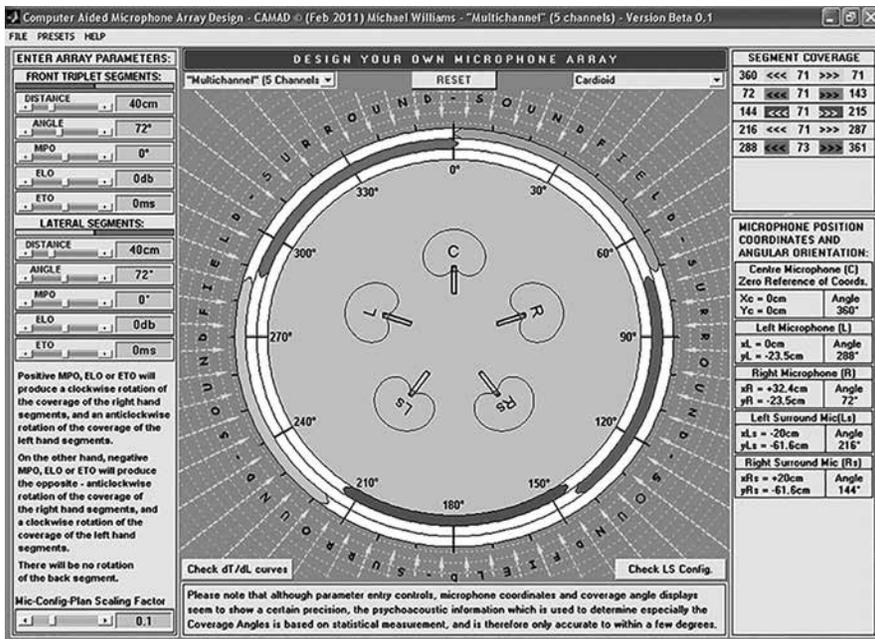


Figure 5.30 – Logiciel CAMAD programmé par M. Williams, MMAD 5 canaux cardioïdes 72° , © M. Williams. En couleur sur dunod.com

CONFIGURATIONS MMAD 5.0 ET 4.0

Omnidirectionnels

- 5.0 → distance d capsules = 63 cm ; angle PdS $\beta = 72^\circ$; rayon $r = 53,6$ cm.
- 4.0 → d capsules = 52 c ; mangle PdS $\beta = 90^\circ$; rayon $r = 36,8$ cm.

Hypocardioides

- 5.0 → d capsules = 46,8 cm ; angle PdS $\beta = 72^\circ$; rayon $r = 39,9$ cm.
- 4.0 → d capsules = 31,7 cm ; angle PdS $\beta = 90^\circ$; rayon $r = 22,5$ cm.

Cardioides

- 5.0 → d capsules = 39 cm ; angle PdS $\beta = 72^\circ$; rayon $r = 33,2$ cm.
- 4.0 → d capsules = 23,8 cm ; angle PdS $\beta = 90^\circ$; rayon $r = 16,9$ cm.

Supercardioides

- 5.0 → d capsules = 32,5 cm ; angle PdS $\beta = 72^\circ$; rayon $r = 27,7$ cm.
- 4.0 → d capsules = 16 cm ; angle PdS $\beta = 90^\circ$; rayon $r = 11,4$ cm.

Hypercardioides

- 5.0 → d capsules = 32,5 cm ; angle PdS $\beta = 72^\circ$; rayon $r = 27,7$ cm.
- 4.0 → d capsules = 16 cm ; angle PdS $\beta = 90^\circ$; rayon $r = 11,4$ cm.

◆ MMAD magic array

Le magic array est une configuration du MMAD directement compatible stéréo en gardant les microphones gauche et droite.

Lorsqu'on enregistre au format 5.1, on doit livrer un mixage stéréo qui, la plupart du temps, est dérivé du 5.1 par downmix pour une raison de diffusion et de temps de production. Le mixage 5.1 doit en effet être compatible stéréo par downmix pour la diffusion TVHD, ou le support DVD/Blu-ray. Le downmix se fait généralement par sommation des canaux L/R et C (voir chapitre 6). Il est généralement problématique d'abandonner le microphone central car la version LR du 5.0 n'est pas compatible, la stéréo obtenue est trop large, sauf dans le cas du Magic Array. Mike Williams a développé des configurations MMAD compatibles stéréo, il suffit pour cela d'abandonner le micro central et les microphones arrière afin d'utiliser la paire stéréo LR parfaitement compatible. Nous obtenons alors des ensembles pour lesquels le micro central est assez avancé par rapport aux quatre autres, la distance est de 1 à 2 m. Ces configurations sont parfaitement utilisables pour des captations suffisamment éloignées du sujet, notamment dans les grandes salles (opéras, orchestres), il posera plus de souci lorsqu'on est près des sources. La **figure 5.31** représente une configuration Magic Array 5.0 en cardioïdes dont l'avancement du

micro central est de 1,74 m, avec un délai de 4,2 ms et complété par une croix à l'arrière (24,5 cm entre les microphones, angle droit).

◆ High Order MMAD (6.0, 7.0, 8.0)

Comme précédemment évoqué dans les caractéristiques du MMAD, il est possible d'extrapoler le système 5 canaux à 6, 7 ou 8 canaux. Le système prend alors d'avantage d'espace et l'utilisation d'une directivité s'impose pour restituer une bonne localisation. En effet, une configuration 7 canaux en omnidirectionnels présente une diaphonie acoustique trop importante, la réjection de cette dernière n'est plus assurée par l'effet de précedence. Les configurations en 7 canaux divisent l'espace en segments égaux de 51°.

Complément Web 5.2

- Pour aller plus loin : représentation des systèmes par le logiciel CAMAD, High Order MMAD.

Est présenté en **figure 5.32** l'exemple d'un MMAD 8.0 magic array utilisant 8 microphones cardioïdes, dont les quatre centraux sont disposés en croix.

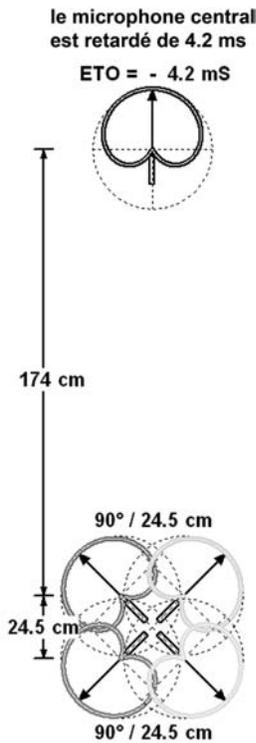


Figure 5.31 – Exemple d'un Magic Array MMAD 5.0 avec un angle de prise de son frontal de 90°, © M. Williams.

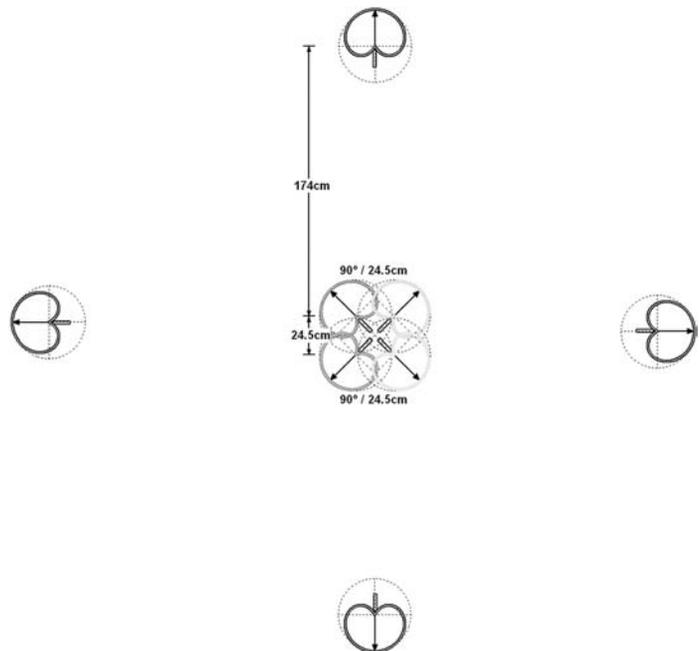


Figure 5.32 – Exemple d'une configuration MMAD 8.0 Magic Array en cardioïdes, © M. Williams.

◆ MMAD en 3D

Dans le contexte actuel des recherches sur la prise de son multicanal avec restitution de la hauteur (son 3D), il est possible de configurer un MMAD 3D. Le parti pris est de minimiser l'interaction entre le plan microphonique horizontal et celui de hauteur. L'autre possibilité serait de considérer l'interaction entre les deux plans, mais, comme abordé au chapitre 3, la perception de hauteur semble reposer plutôt sur les indices spectraux que sur des différences d'intensité et de temps. C'est pourquoi, il semble difficile de restituer des sources virtuelles avec hauteur, entre le plan horizontal et le plan d'élévation. Généralement, les microphones d'élévation apportent un intérêt pour la restitution de l'acoustique, sans présenter d'interaction avec les microphones horizontaux.

Le MMAD 3D dispose un système 4.0 additionnel placé à 1 m de hauteur par rapport au système horizontal, équipé de microphones supercardioides ou bidirectionnels, placés en croix, distants entre eux de 52 cm, orientés vers le haut, de telle sorte que la directivité annule la captation du son direct. Les capsules sont placées suivant les angles 0° , 90° , 180° , -90° . Le dispositif microphonique 4.0 de hauteur travaille en différence de temps simplement, les directivités étant orientées vers le haut. Cette configuration permet une restitution de l'acoustique sur les quatre canaux de hauteur LH, RH, RLH, RRH. Le dispositif de restitution peut être alors un format 11.1 composé d'un 7.0 pour l'horizontal (0° , 30° , 75° , 120°) et d'un 4.0 (0° , 90° , 180° , -90°) pour la hauteur. Pour une restitution Auro-3D 9.1, les capsules du système de hauteur sont placées aux angles -45° , -135° , 45° , 145° . Le système est encore en cours d'expérimentation et la question du dispositif de restitution reste posée car aucune norme pour la diffusion son 3D n'a vraiment été validée.



Figure 5.33 – Exemple d'une configuration MMAD 3D 9.0 cardioides pour la couche horizontale, bidirectionnels pour la couche élévation, le marché © B. Périaux.



Figure 5.34 – Essai d'une configuration MMAD 3D 9.0 hypocardioides pour la couche horizontale, cardioides pour la couche élévation, ensemble baroque (Catherine Jousselein, Xavier Terrasa) dans une église, © B. Périaux.

MMAD

- Choix de la directivité.
- Choix des angles de prise de son.
- Configurations usuelles 5.0 avec des angles physiques et de prise de son superposés de 72°.
- Compatible 4.0, 5.0, 6.0, 7.0, 8.0....
- Compatible stéréo avec la configuration Magic Array pour une distance de captation assez importante.

ANALYSE

- Système très homogène.
- Très bonne impression spatiale.
- Pas de contrainte sur l'équilibre spectral.
- Système modulable.
- Utilisation en salle de concert avec l'étoile, et en reportage avec le parapluie (voir ci-dessous).

◆ Supports MMAD

Le système MMAD a l'avantage de donner un excellent rendu spatial et une image très bien localisée, c'est en revanche un système complexe à mettre en place, car l'utilisateur doit respecter les angles et les distances entre les microphones suivant l'abaque qu'il choisit. C'est pourquoi, il est préférable d'opter pour des supports déjà configurés, utilisant les dispositions usuelles à 72° d'angle physique et de prise de son. L'étoile de Williams et le parapluie sont les supports commercialisés en France par la société Areitec. Le parapluie est prévu pour le tournage et le reportage, l'étoile est utilisée pour la captation en salle.

Le système de reportage, le parapluie, utilise une configuration de cinq microphones à 72°, compatible avec les microphones miniatures de type Schoeps CCM, en cardioïdes, hypocarديوïdes et supercardioïdes. Le système se monte très rapidement comme il peut se replier très facilement avec les bonnettes pour le transport.

L'étoile peut être équipée de CCM Schoeps ou de microphones traditionnels, elle comporte une suspension anti-vibrations développée en collaboration avec la société RYCOTE à partir du modèle Lyre de la série « InVision Broadcast ». Elle peut être suspendue ou fixé sur pied, avec différentes configurations :

- 4.0 cardioïdes ou hypocarديوïdes ;
- 5.0 cardioïdes, hypocarديوïdes ou omnidirectionnels ;
- 7.0 cardioïdes ou hypocarديوïdes.

Il existe un autre support expérimental, le « Pentapussy », ce dernier permet de réaliser toutes les abagues, on peut l'utiliser pour la formation par exemple ou pour la captation de musique de film avec l'utilisation de grosses capsules. On peut lire à la base du système les angles de positionnement des microphones et mesurer avec un mètre la distance entre les capsules.



Figure 5.35 – Étoile de Williams, version reportage cardioïdes Schoeps CCM4, © B. Périaux

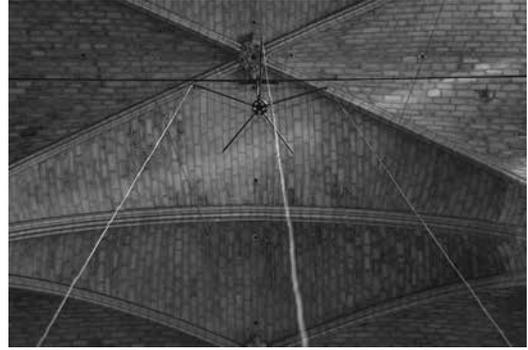


Figure 5.36 – Étoile de Williams, utilisée en salle de concert avec un système de suspension de microphones Schoeps CCM5 pour le frontal, © B. Périaux



Figure 5.37 – Parapluie MMAD équipé de Schoeps CCM4, © B. Périaux

5.3.2 OCT Surround

◆ Système OCT

Le système OCT (Optimized Cardioid Triangle) a été conçu par Günther Theile (IRT, Institut für Rundfunktechnik), il repose sur la théorie de la diaphonie acoustique minimisée. En effet, en théorie, un triplet de microphones frontaux crée trois images fantômes LC, CR et LR qui se superposent. G. Theile considère que cette diaphonie n'est pas négligeable.

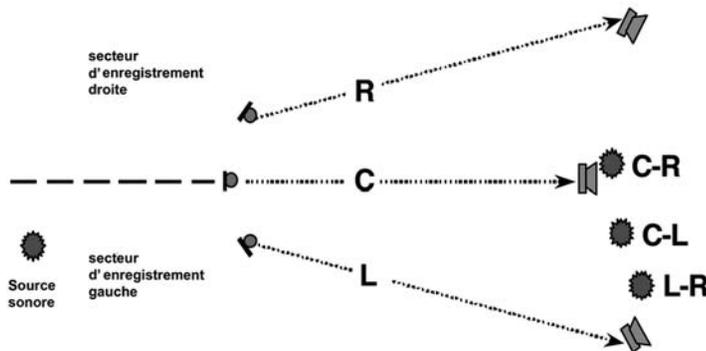


Figure 5.38 – Diaphonie acoustique obtenue par superposition de trois images fantômes, © G. Theile.

L'objectif est donc de limiter la diaphonie acoustique entre les microphones pour éviter l'apparition d'images fantômes multiples (pertes sur la précision en localisation, le timbre et la stabilité des sources avec un déplacement latéral de l'auditeur). Pour cela, on a recours à l'utilisation de microphones plus directifs (supercardioides) orientés de manière à rendre quasi nulle l'influence d'un micro sur le couple opposé (micro L sur le couple C-R par exemple). La disposition des microphones est celle de la **figure 5.39**.

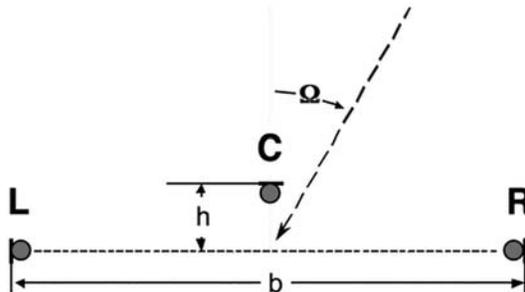


Figure 5.39 – Configuration OCT, © G. Theile.

Les microphones gauche et droite sont des supercardioïdes Schoeps CCM41, placés à $\pm 90^\circ$, et le microphone central est un cardioïde CCM4, avec un avancement constant de 8 cm.

L'IRT a développé un logiciel « Image Assistant » qui permet d'analyser la localisation des systèmes de prise de son. Le schéma ci-dessous représente la courbe de localisation pour une paire de microphones stéréo, dont l'angle de prise de son est de 100° .

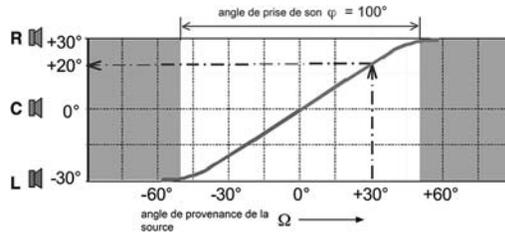


Figure 5.40 – Courbe de localisation pour un système de prise de son stéréo, © G. Theile, Image Assistant IRT.

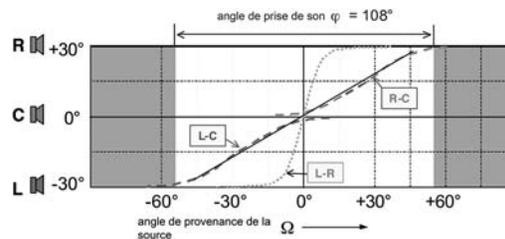


Figure 5.41 – Courbe de localisation de l'OCT70, © G. Theile, Image Assistant IRT.

L'angle d'enregistrement frontal est de 108° , l'image fantôme L-R est négligeable (de niveau 10 dB inférieur) par rapport aux images L-C et C-R, et la courbe résultante est linéaire.

En abscisse est représenté l'angle de provenance ω , et en ordonnée, l'angle de perception dans l'espace stéréophonique. Par exemple, ici, une source de provenance de 30° sur la droite, sera perçue à 20° sur le côté droit. Notons la linéarité de la courbe. Cette linéarité est aussi souhaitée pour l'utilisation des trois microphones, en l'absence d'image L-R, la diaphonie acoustique est optimisée.

L'angle de prise de son frontal du triplet dépend uniquement de la distance b entre les deux supercardioïdes, il est calculé comme nous l'indiquons dans la **figure 5.42**.

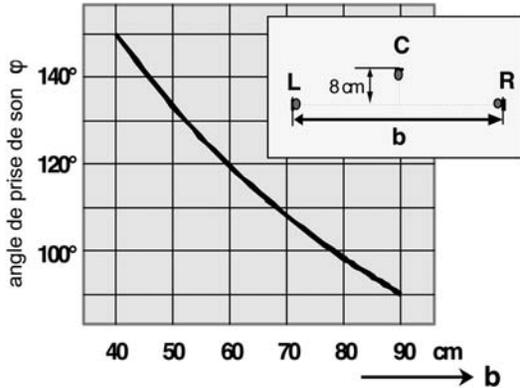


Figure 5.42 - Détermination de l'angle de prise de son frontal en fonction de la dimension de la base LR (b), © G. Theile.

Il existe un système optionnel de compensation des basses fréquences, dont la perte est inhérente aux supercardioïdes. Deux capteurs omnidirectionnels CCM2, placés à gauche et à droite, superposés aux supercardioïdes et filtrés au-dessus de 100 Hz complètent l'ensemble. On combine les signaux issus des supercardioïdes et des omnidirectionnels pour construire les extrémités gauche et droite de l'image.

◆ **OCT Surround**

L'OCT Surround complète l'OCT par deux microphones cardioïdes arrière, orientés à 180° et distants du triplet avant de 40 cm. La distance entre ces deux capteurs est égale à celle de la base, plus 10 cm de chaque côté.

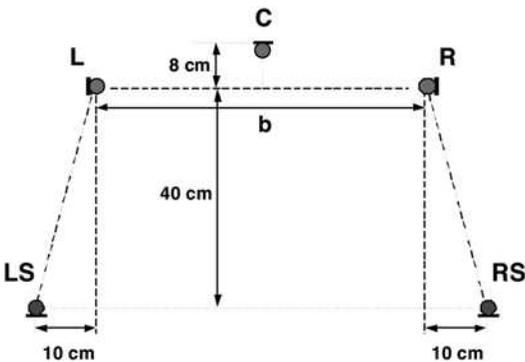


Figure 5.43 - OCT Surround, © G. Theile.

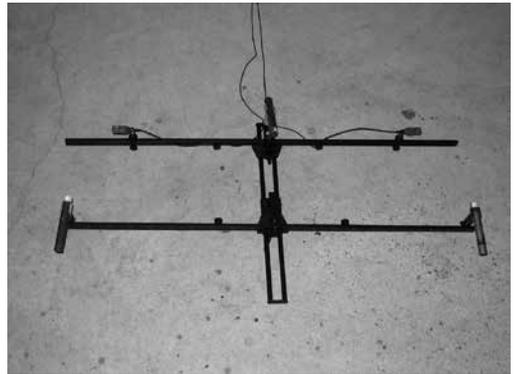


Figure 5.44 - Support Schoeps OCT Surround, © B. Périaux.

L'OCT est accompagné souvent d'un système d'ambiance en fond de salle qui remplace les deux microphones arrière, on utilise pour cela couramment une croix IRT ou un carré Hamasaki, pour renforcer l'impression de salle.

OCT SURROUND

- Micro central cardioïde avancé de 8 cm.
- Micros L et R supercardioïdes à $\pm 90^\circ$, distance b variable en fonction de l'angle de prise de son frontal souhaité.
- Micros arrière cardioïdes orientés à 180° , à 40 cm du triplet avant, distance entre eux = base b + 10 cm de chaque côté.
- Utilisation de deux omni pour le renfort des graves des canaux L et R.
- Utilisation fréquente de l'OCT avec un système d'ambiance Croix IRT ou Carré Hamasaki.

ANALYSE

- Très bonne localisation frontale, plans sonores bien restitués.
- Bonne impression spatiale.
- Manque de basses fréquences.
- Système peu modulable : orientations des micros fixes, directivités imposées.
- Utilisation pour la salle de concert.
- L'OCTS est plutôt utilisé pour des esthétiques frontales.

◆ OCT2

L'OCT2 est une variante de l'OCT pour laquelle le microphone central est avancé de 40 cm au lieu de 8 cm. Pour respecter rigoureusement les propriétés de localisation de l'OCT, le microphone central doit être retardé de 1 ms. Cette configuration se rapproche de l'arbre Decca (voir paragraphe suivant), la décorrélation entre les microphones est plus importante que pour l'OCT. Du fait de l'absence de diaphonie acoustique entre les microphones, l'OCT donne de bons résultats de downmix stéréo, ces propriétés de downmix sont encore améliorées par l'OCT2.

◆ OCT9 (3D)

L'OCT9 est le système proposé par Günther Theile pour la captation du son avec élévation. Le dispositif de restitution utilisé est l'Auro-3D 9.1. La base du système microphonique est constituée d'un OCT classique, elle est complétée par quatre microphones supercardioïdes placés à environ 1 m ou plus en hauteur, par rapport à la base horizontale. Ces quatre microphones Lh, Rh, Lsh et Rsh sont orientés vers le haut, et placés au-dessus des microphones L, R, Ls et Rs.

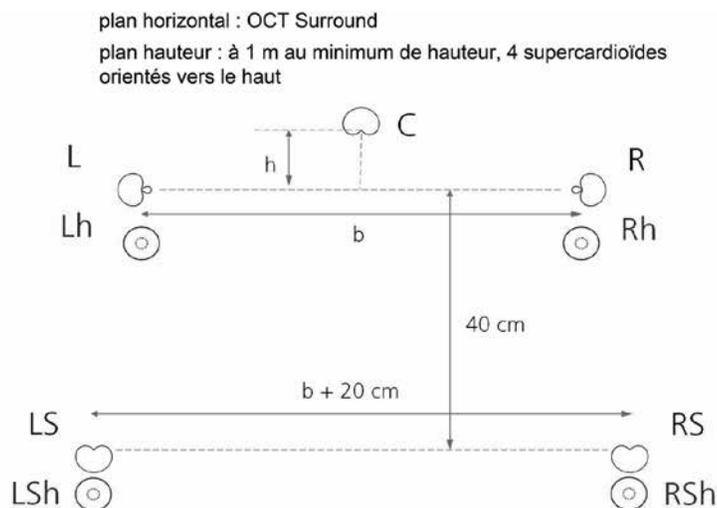


Figure 5.45 – OCT9 configuration pour le son 3D, © G. Theile.

5.3.3 Arbre Decca

◆ Triangle Decca

Le triangle Decca est le système utilisé pour les enregistrements d'orchestre par les preneurs de son du label Decca dans les années 1950. Le système était composé de trois microphones omnidirectionnels Neumann M50 largement espacés en triangle, le micro central étant replacé au centre de la stéréo. Ce système a par la suite été exploité lors des enregistrements d'orchestre pour le film, utilisant le micro central pour la création d'un centre physique, dès l'apparition du format Dolby Stereo. Il ne repose sur aucune théorie, il est constitué de trois microphones omnidirectionnels et sa configuration de base est celle de la **figure 5.46**.

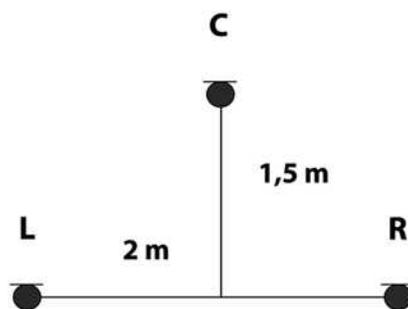


Figure 5.46 – Configuration Triangle Decca, © B. Périaux.

La base mesure environ 2 m et l'avancement du centre est de 1,5 m, les dimensions étant variables selon la source à enregistrer. En général, on donne un angle sur les microphones gauche et droite pour préciser la localisation.

L'approche du triangle Decca est de tirer parti de l'espace et du timbre reproduits par les systèmes AB omnidirectionnels et remplir le trou au centre, créé par le grand AB omni, grâce au micro central. Les grandes distances entre les microphones limitent les effets de diaphonie acoustique grâce à l'effet Haas.

L'analyse de la courbe de localisation du triangle Decca révèle la présence de l'image L-R générant une diaphonie acoustique. Cette image fantôme est d'un niveau comparable à celui des images L-C et C-R, mais elle est retardée d'environ 5 ms, ce qui améliore le rendu en localisation du système par l'effet de précedence. La diaphonie acoustique n'affecte donc pas la localisation des images fantômes. En revanche, le schéma traduit une non linéarité de localisation, laquelle repose alors sur les courbes L-C et C-R. Ces deux courbes montrent que toutes les sources situées dans un angle de provenance de $\pm 45^\circ$ sont reproduites au centre de l'image frontale ou à proximité, et celles qui sont en dehors du secteur $\pm 60^\circ$, sont reproduites sur les enceintes gauche et droite. La théorie prévoit donc une focalisation des sources sur le canal central, d'où la nécessité de baisser souvent le niveau du microphone central à l'écoute.

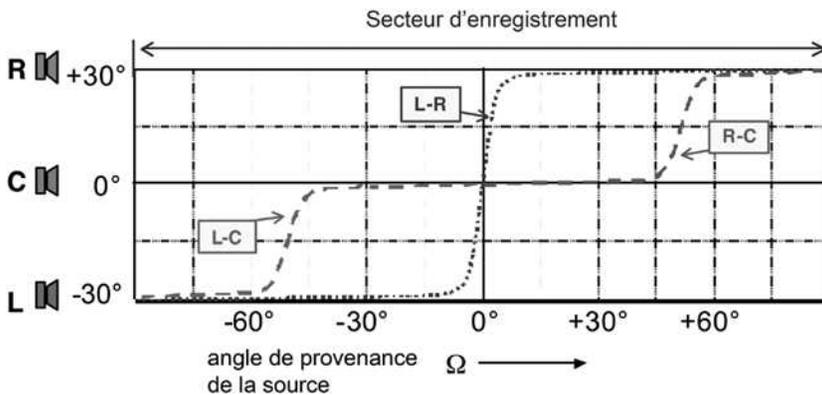


Figure 5.47 – Courbe de localisation du triangle Decca pour deux omnidirectionnels distants de 1,80 m et un angle de prise de son de 120° , © G. Theile. Le niveau de L-R est comparable à celui de L-C et C-R, mais L-R subit un retard d'environ 5 ms, ce qui réduit l'effet de diaphonie acoustique.

◆ Arbre Decca (Decca Tree)

L'arbre Decca complète le triangle Decca par l'ajout de deux microphones omnidirectionnels en configuration grand AB omni, distants du triplet avant, orientés vers le fond de salle ou vers l'orchestre. Cette configuration est très variable, on peut séparer largement les deux microphones arrière en prise de son d'orchestre pour le film.

Cet ensemble de cinq microphones omnidirectionnels restitue une très bonne impression d'espace avec un bon équilibre spectral, en revanche, pour préciser la localisation de l'image, les microphones d'appoints mono dirigés sont très sollicités.

DPA commercialise un support appelé S5 qui permet de configurer plusieurs configurations dont l'arbre decca, l'arbre fukada, le carré Hamasaki...

◆ AB large (3D)

La configuration AB large en 3D a été expérimentée lors des tests à la conférence de Detmold ICSA 2011. Elle repose sur une configuration arbre Decca, avec une distance de 0,5 à 2 m entre les microphones omnidirectionnels sur le plan horizontal. Les microphones de hauteur sont aussi omnidirectionnels, distants de plus d'1 m et orientés vers le haut. En fonction de la distance des microphones de hauteur, le dispositif est susceptible de modifier la localisation horizontale des sources et d'étendre ces dernières vers le haut.

plan horizontal : arbre decca, distance entre 0,5m et 2m
plan vertical : hauteur > 1m, capsules omni orientées vers le haut

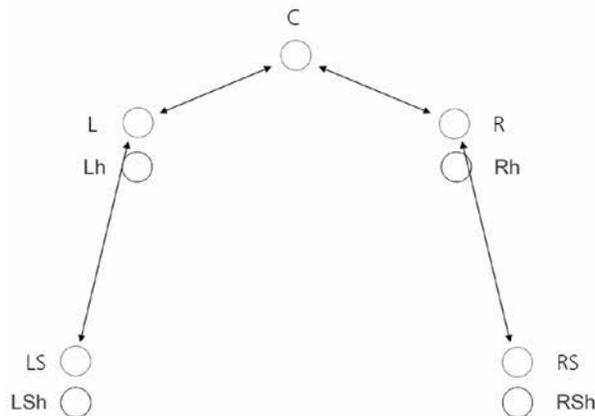


Figure 5.48 – AB large en 3D, © B. Periaux d'après ICSA 2011.

ARBRE DECCA

- 5 microphones omnidirectionnels optimisés à l'écoute.
- Frontal : triangle Decca, base environ 2 m, avancement du centre 1,5 m.
- Arrière : grand AB omni dirigé vers l'arrière ou vers l'orchestre.

ANALYSE

- Localisation imprécise, instable et non linéaire.
- Impose l'utilisation d'appoints pour préciser la localisation.
- Très bonne impression spatiale.
- Très bon équilibre spectral.
- Utilisation pour la musique orchestrale, en film particulièrement.

5.3.4 Arbre Fukada

L'arbre Fukada est une variante de l'arbre Decca qui utilise des directivités cardioïdes afin de réduire la captation du champ réverbéré au profit du champ direct. Fukada est le nom d'un ingénieur du son de la NHK. Les cinq microphones cardioïdes L, C, R, Ls, Rs sont complétés par deux microphones omnidirectionnels LL et RR, plus espacés de chaque côté, lesquels augmentent la largeur de l'orchestre et améliorent la liaison entre le frontal et l'arrière.

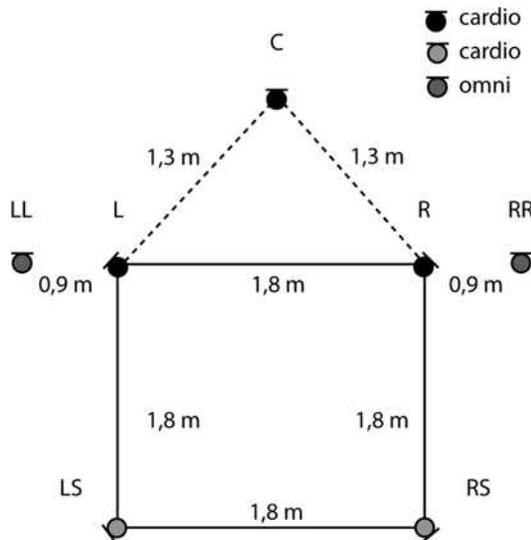


Figure 5.49 – Configuration d'un arbre Fukada, les angles entre les capsules L et R, et entre Ls et Rs sont variables, © B. Périaux.

ARBRE FUKADA

- Arbre Decca avec des directivités cardioïdes qui permettent d'adapter le rapport CD/CR capté.
- Utilisation pour la captation en salle.

5.3.5 INA5

Le système INA5 (Ideale Nieren Anordnung : arrangement idéal de cardioïdes) a été conçu par U. Herrmann et V. Henkels. C'est une disposition de cinq microphones cardioïdes donnant un angle de prise de son frontal de 180°. Le triplet avant (INA3) dispose les microphones à 25 cm de distance et à angle droit, ce qui met en évidence deux angles de prise de son de 90°. Les microphones arrière, distants de 53 cm par rapport aux microphones gauche et droite et placés à 150°, donnent des angles de captation de 60°. Nous obtenons ainsi trois angles de captation de 60° sur les côtés et à l'arrière.

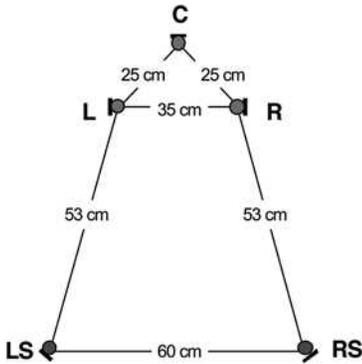


Figure 5.50 – Configuration du système INA5, © G. Theile.

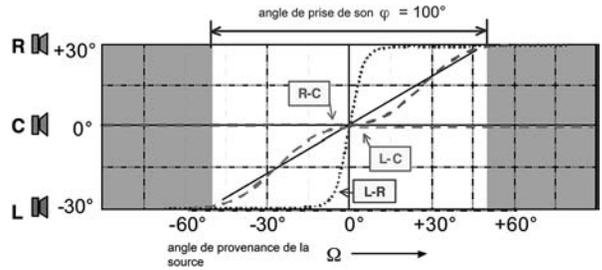


Figure 5.51 – Courbe de localisation du système INA5, © G. Theile. Le niveau de L-R est de 3 dB inférieur à celui de L-C et C-R, le retard de L-R est d'environ 1 ms, insuffisant pour éviter la diaphonie acoustique.

L'analyse de courbe de localisation dans l'espace frontal ci-dessus révèle l'effet non négligeable de la diaphonie acoustique, le niveau de l'image fantôme L-R est d'environ 3 dB inférieur à celui des images L-C et C-R. D'autre part, le retard de l'image L-R est de 1 ms, ce qui n'évite pas la combinaison des trois images fantôme pour la localisation des sources. L'angle d'enregistrement est de 100°, si l'on augmente ce dernier, le délai de l'image fantôme L-R diminue.

Computer Aided Microphone Array Design - CAMAD © (Feb 2011) Michael Williams - "Multichannel" (5 channels) - Version Beta 0.1

FILE PRESETS HELP

ENTER ARRAY PARAMETERS:

FRONT TRIPLET SEGMENTS:

DISTANCE: 25cm

ANGLE: 90°

MPO: 0°

ELO: 0db

ETO: 0ms

LATERAL SEGMENTS:

DISTANCE: 53cm

ANGLE: 60°

MPO: 39°

ELO: 0db

ETO: 0ms

Positive MPO, ELO or ETO will produce a clockwise rotation of the coverage of the right hand segments, and an anticlockwise rotation of the coverage of the left hand segments.

On the other hand, negative MPO, ELO or ETO will produce the opposite - anticlockwise rotation of the coverage of the right hand segments, and a clockwise rotation of the coverage of the left hand segments.

There will be no rotation of the back segment.

Mic-Config-Plan Scaling Factor: 0.1

DESIGN YOUR OWN MICROPHONE ARRAY

"Multichannel" (5 Channels) [RESET] Cardioid

Check dT/dL curves Check LS Config.

Please note that although parameter entry controls, microphone coordinates and coverage angle displays seem to show a certain precision, the psychoacoustic information which is used to determine especially the Coverage Angles is based on statistical measurement, and is therefore only accurate to within a few degrees.

SEGMENT COVERAGE

360	<<<	90	>>>	90
60	<<<	64	>>>	124
148	<<<	63	>>>	211
235	<<<	64	>>>	299
270	<<<	90	>>>	360

MICROPHONE POSITION COORDINATES AND ANGULAR ORIENTATION:

Centre Microphone (C)
Zero Reference of Coords.

Xc = 0cm	Angle 360°
Yc = 0cm	

Left Microphone (L)

xL = 0cm	Angle 270°
yL = -17.7cm	

Right Microphone (R)

xR = +17.7cm	Angle 90°
yR = -17.7cm	

Left Surround Mic (Ls)

xLs = -26cm	Angle 210°
yLs = -70cm	

Right Surround Mic (Rs)

xRs = +26cm	Angle 150°
yRs = -70cm	

Figure 5.52 – Courbe de localisation CAMAD pour INA5, © M. Williams. En couleur sur dunod.com

L'analyse du système par CAMAD présenté en **figure 5.52** révèle une superposition d'angles de captation sur les côtés.

La société Brauner a mis en application la théorie INA5 avec le système ATMOS ASM5 (Adjustable Surround Microphone). Le dispositif utilise des microphones à lampes Brauner VM1 double membrane dont la directivité peut se commander sur le préamplificateur (réglage continu entre la directivité omnidirectionnelle, cardioïde et bidirectionnelle). L'angle des capsules avant gauche et avant droite peut se paramétrer et la longueur des bras arrière est aussi réglable sur certains modèles. Il s'utilise avec le processeur Atmos 5.1 de SPL indispensable, ce dernier commande les directivités, les alimentations et préamplifie les signaux. Le processeur offre une matrice 5.1 avec la possibilité de diverger le centre, de créer un canal Lfe par bass management. Il est possible de brancher un casque multicanal (modèle co-développé par Studer et l'IRT, mais non commercialisé). Le processeur peut aussi alimenter et préamplifier des microphones standards en 48 V.



Figure 5.53 – Système Brauner ASM, © B. Périaux.

INA5 BRAUNER ASM

- 5 cardioïdes à la base, directivités variables sur l'Atmos (bidirectionnelle, cardioïde, omnidirectionnelle).
- Angle de prise de son frontal 180°, angles latéraux et arrière 60°.

ANALYSE

- Localisation : repliements sur les côtés, diaphonie acoustique.
- Très bonne restitution des timbres, homogène.
- Pas de contrainte sur l'équilibre spectral, directivités variables pour l'ASM.
- Système non modulable : pas d'orientation possible des capsules dans le plan vertical.
- Utilisation pour la musique en studio.

5.4 Les systèmes principaux matricés

5.4.1 Double MS

◆ Principe général

Le Double MS est un système à trois microphones, dérivé du MS auquel on a rajouté un microphone cardioïde orienté vers l'arrière. Le système dispose deux couples MS en frontal et à l'arrière, tête-bêche, avec partage du microphone bidirectionnel orienté vers le côté. Les microphones M. (avant et arrière) peuvent être cardioïdes ou hypercardioïdes, on trouve aussi des configurations avec un microphone canon pour l'avant. L'enregistrement se fait donc sur trois pistes que l'on dématrice avec le plug-in Double MS Tool développé par Schoeps.



Figure 5.54 – Double MS, © V. Magnier.

◆ Dématricage

Le plug-in Double MS Tool a l'avantage de pouvoir dématricer le signal en faisant le choix des directivités avant et arrière de façon indépendante, et en réglant l'angle entre les microphones virtuels ainsi obtenus. Le canal central est principalement alimenté par le microphone M. avant et sa directivité est variable selon le dosage du bidirectionnel. Il existe un boîtier de décodage pour le monitoring direct à la prise mais les enregistreurs permettent très facilement de décoder le signal dans les modes de monitoring. Il existe d'autres dématricateurs comme le Double MS Tool BF, codéveloppé par Schoeps et Illusonic, ainsi que le DMS2Five, développé par Daniel Courville.

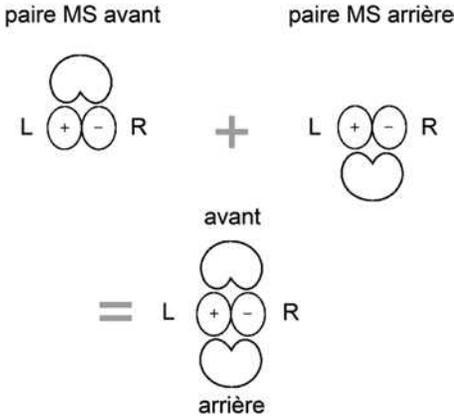


Figure 5.55 – Dématriçage du Double MS, © B. Périaux d'après Schoeps.

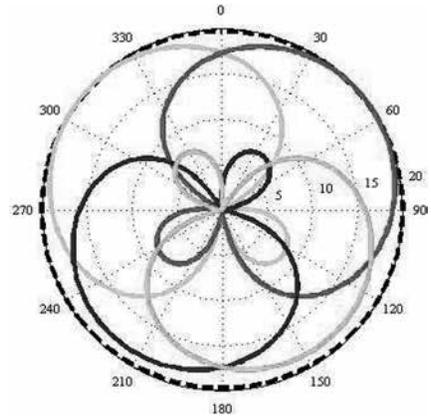


Figure 5.56 – Dématriçage du Double MS en 4 canaux, © Schoeps.

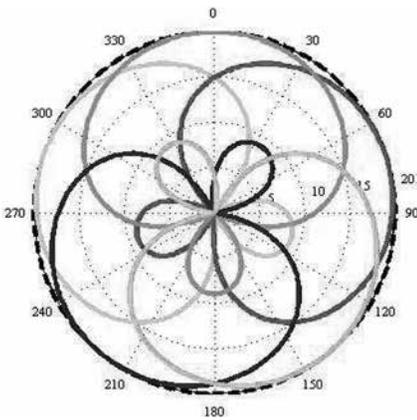


Figure 5.57 – Dématriçage du Double MS en 5 canaux, © Schoeps.

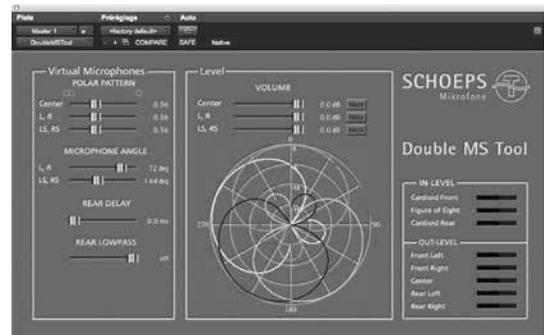
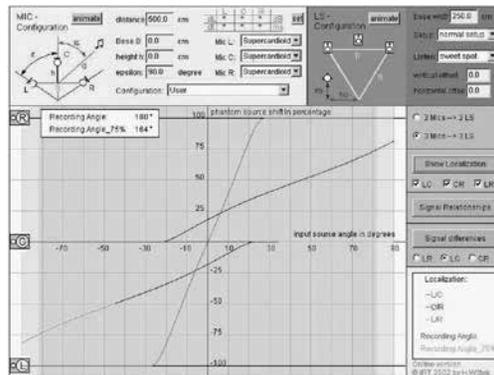
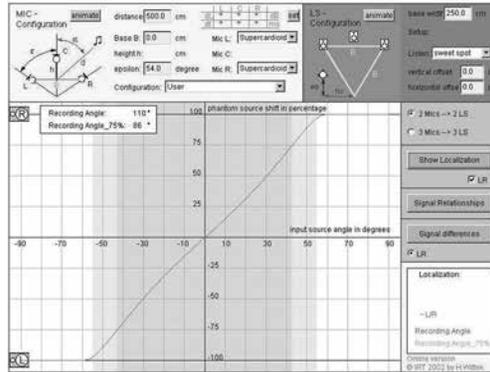


Figure 5.58 – Plug-ins Double MS Tool, © Schoeps.

CONFIGURER LE DMS TOOL DANS PRO TOOLS

Le plug-in DMS Tool se configure dans Pro Tools en insert sur une piste auxiliaire de format 5.1, dont l'entrée est un bus 5.1 créé dans le IO setup. Ce bus 5.1, que l'on peut nommer DMS, permet alors de router les trois pistes mono, sur lesquelles on place les prises cardio avant, bidirectionnel et cardio arrière. La piste cardio avant est envoyée à l'avant gauche, vers le bus DMS, en utilisant le pan-pot multicanal (ou directement vers le canal 1 du bus DMS), la piste bidirectionnelle est envoyée à l'avant-droite (canal 3), et la piste bidirectionnelle au centre (canal 2). L'utilisation des modes solo sur les pistes permet de vérifier que la configuration est bien faite, en identifiant le signal sur les entrées meter du plug-in.

La version DMS Tool BF intègre un algorithme qui permet d'améliorer la sélectivité des directivités, avec une meilleure séparation des canaux, ainsi qu'elle permet de gérer le champ diffus de façon indépendante du son direct avec des possibilités de décorrélation.



Figures 5.59 et 5.60 – Courbes de localisation frontale du Double MS avec des directivités supercardioides, © Helmut Witteck www.hautmikrofon.de

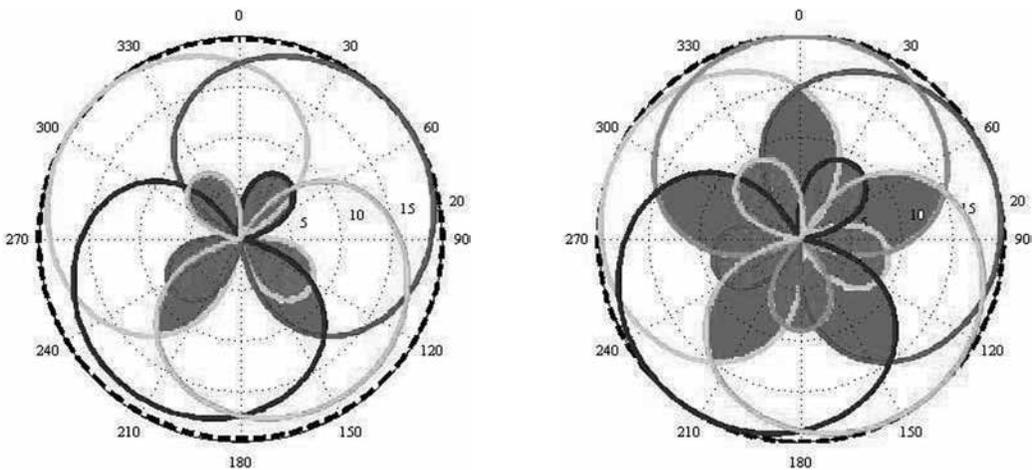
En couleur sur dunod.com La courbe a, donnée pour un décodage DMS 4 canaux, est linéaire. La courbe b, qui correspond au décodage DMS 5 canaux, avec centre, révèle une importante image L-R introduisant une certaine diaphonie acoustique.

◆ En pratique

Le système est léger, il peut être facilement perché. Il est équipé de capsules Schoeps CCM4V et CCM8 et d'une suspension avec bonnette anti-vent. C'est donc un système adapté pour le reportage, le documentaire ou le cinéma pour sa compacité, et aussi pour sa souplesse d'utilisation, car il laisse le choix au mixeur de n'utiliser que le MS stéréo avant voire que le M. sans compromis.

◆ Diaphonie acoustique

L'étude de la diaphonie acoustique du système met en évidence le phénomène d'internalisation, variable en fonction des réglages. La perception est interne, proche de la tête, mais le rendu sonore est ajustable dans certains cas grâce à l'utilisation du délai et du filtrage des canaux arrière. D'autre part, la version DMS Tool BF intègre un algorithme qui permet d'améliorer la sélectivité des directivités et la décorrélation du champ diffus. Le système ne donne pas une impression d'espace très importante du fait de la technologie coïncidente mais il reste très intéressant pour des applications dans des lieux de petite taille (intérieure voiture, monte-charge, intérieurs d'appartement) ou bien pour donner des effets d'internalisation. C'est d'autre part un des systèmes les plus pratiques pour l'enregistrement d'ITV en multicanal, parfaitement compatible mono.



Figures 5.61 et 5.62 – Diaphonie obtenue sur un système DMS 4 canaux et 5 canaux, © Schoeps.
En couleur sur dunod.com.

Les zones grises sont en phase, les zones rouges sont hors-phase.

◆ Double MS et Format B

En principe, les signaux d'un double MS peuvent être convertis en format B horizontal (voir paragraphe sur le Soundfield) par addition et soustraction. Les signaux W, X et Y du format B s'obtiennent par les opérations suivantes :

$$W = M. \text{ avant} + M. \text{ arrière}$$

$$X = M. \text{ avant} - M. \text{ arrière}$$

$$Y = S$$

Cette opération rend le double MS compatible avec un format ambisonique d'ordre 1, sans intégrer l'information verticale représentée par la composante z.

DOUBLE MS

- Un bidirectionnel orienté vers le côté, partagé par un cardio avant et un cardio arrière.
- Dématriçage avec le Double MS Tool, choix des directivités LR, C, LsRs et des angles entre les microphones virtuels obtenus.

ANALYSE

- Très bonne localisation circulaire, très peu de profondeur.
- Diaphonie acoustique.
- Impression spatiale limitée, internalisation des sources.
- Utilisation possible en reportage, documentaire, cinéma.

5.4.2 *Soundfield, système ambisonique d'ordre 1*

Le Soundfield est le système de prise de son multicanal le plus ancien, sa théorie a été développée par M. Gerzon qui dans les années 1970 s'est inspiré des recherches de A. Blumlein (système Blumlein, deux bidirectionnels à angle droit). La théorie a été mise en application par la société Calrec Audio en 1980, puis le brevet a été revendu à la société Soundfield en 1993. Aujourd'hui, la société anglaise Soundfield, rachetée en 2012 par TSL Professional Products Ltd., continue le développement en diversifiant les versions pour le tournage et pour les cars HD.

◆ **Principe général**

Le principe du système ambisonique est de capter en un point de l'espace la totalité des informations du champ acoustique. C'est une décomposition du champ sonore en harmoniques sphériques qui fonctionne en différence d'intensité et qui permet une restitution en 2D ou 3D indépendante du dispositif d'écoute.

Le système Soundfield est constitué d'un ensemble de quatre capsules hypocardiïdes quasi-coïncidentes, disposées en tétraèdre. Les capsules délivrent quatre composantes, appelées format A, au processeur d'entrée qui, par combinaisons linéaires, encode un format B. Ce format B est constitué des composantes W, X, Y et Z. Le processeur de sortie décode ensuite le format B en cinq directivités virtuelles cardioïdes (par défaut), quasi-coïncidentes, disposées pour une restitution en 5.1.



Figure 5.63 – Le Soundfield, tétraèdre régulier de capsules quasicoincidentes, © Soundfield.

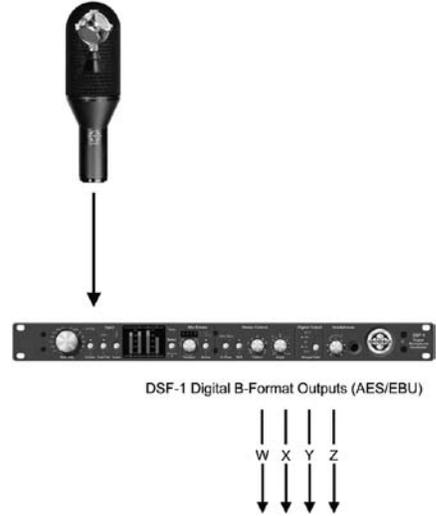


Figure 5.64 – Chaîne de captation Soundfield avec utilisation du Mark V, © Soundfield.

◆ **Format A et B**

Format A : signaux Lb, Lf, Rf, Rb issus du microphone, transmis au processeur d'entrée



Figure 5.65 – Format A, © Soundfield.

Lb – : left back down (arrière gauche bas)

Lf + : left front up (avant gauche haut)

Rf – : right front down (avant droite bas)

Rb + : right back up (arrière droite haut)

Format B : composantes W, X, Y, Z codées par le processeur à partir du format A, par combinaisons linéaires

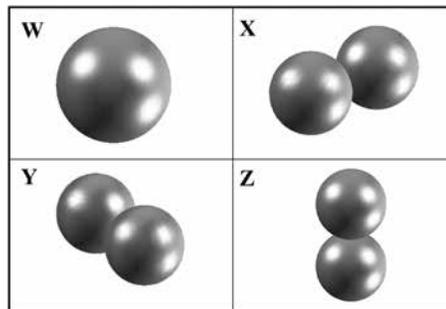


Figure 5.66 – Format B, décomposition du champ sonore en harmoniques sphériques W, X, Y et Z, © Soundfield.

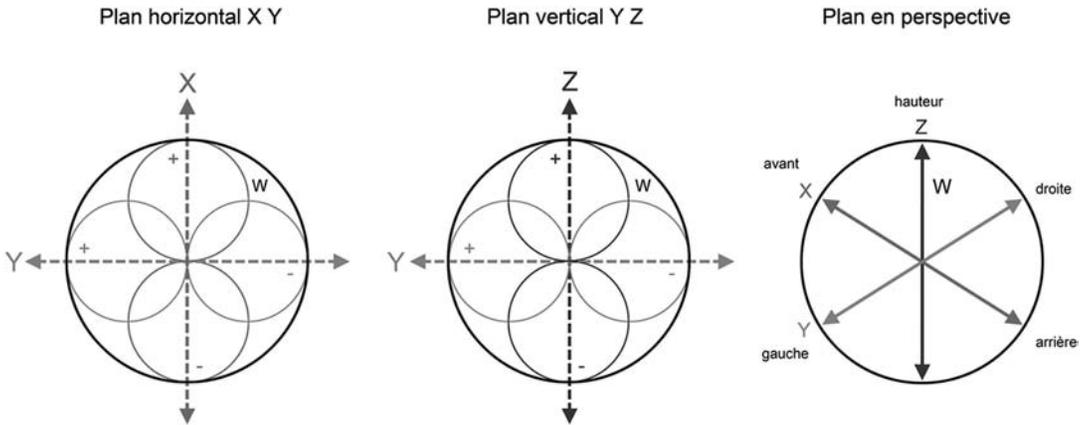
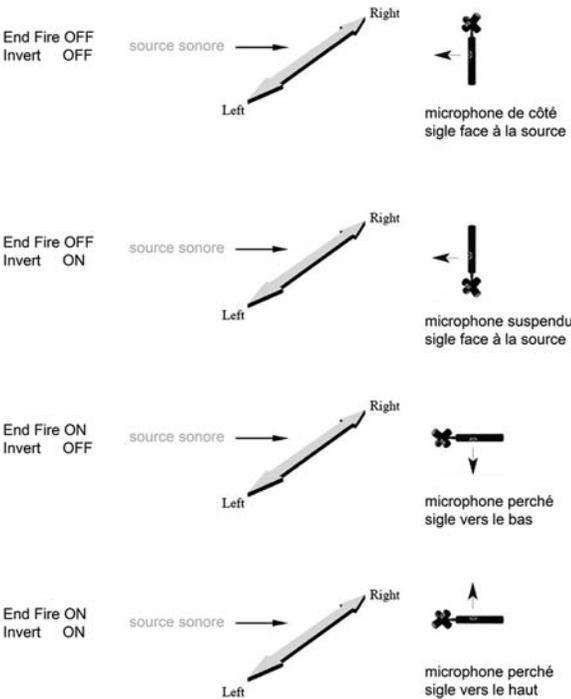


Figure 5.67 – Format B, représentation des composantes W, X, Y, Z, © B. Périaux d'après Soundfield.

- $W = Lb + Lf + Rf + Rb$ composante omni pression
- $X = Lf - Rb + Rf - Lb$ composante bi selon axe X vitesse
- $Y = Lf - Rb - Rf + Lb$ composante bi selon axe Y vitesse
- $Z = Lf - Lb + Rb - Rf$ composante bi selon axe vertical Z vitesse



◆ **En pratique**

Le microphone Soundfield peut être utilisé dans différentes configurations, sur pied, sur perche, ou suspendu. Il existe des paramètres de contrôle pour la disposition du microphone, ce sont les fonctions END et INV présentes sur le préamplificateur (le module d'entrée du MarkV ou le boîtier du ST450) et que l'on retrouve sur le décodeur (SP451 ou Surroundzone). Le paramètre END doit être actif si le microphone est disposé de façon axiale, et le paramètre INV gère l'orientation verticale (voir figure 5.68). Les fonctions « azimuth » et « elevation » contrôlent la rotation virtuelle du microphone, il est aussi possible de focaliser plus ou moins vers le champ direct grâce au paramètre « dominance » ou « zoom » (Back/

Figure 5.68 – Les différentes configurations de placement de la SPS200, compatibles avec tout microphone Soundfield, © B. Périaux d'après Soundfield.

Front). Ainsi, l'utilisateur peut réajuster l'orientation du microphone et préciser les informations de la scène sonore qui l'intéressent.

Le mark V, processeurs MkV, DSF1, DSF2 et décodeurs SP451, DSF3

Le mark V est le premier microphone Soundfield développé pour le Studio. Il est utilisé avec les processeurs d'entrée MkV processor ou DSF2 qui gèrent la préamplification du signal et l'encodage au format B. Ces processeurs utilisent une première section de réglages où l'on trouve les paramètres de gain, de contrôle de la disposition du microphone (« rotation », « INVERT », « END FIRE », « dominance » ou « zoom ») et un étage de dématricage stéréophonique paramétrable pourvue d'une sortie casque stéréo (angle des directivités virtuelles, directivités des microphones).

Le décodage du format B est réalisé par le SP451 (version originale avec entrées et sorties analogiques) ou par le DSF3 (version actuelle numérique).

La plupart des cars HD équipés d'un microphone Soundfield utilisent aujourd'hui les processeurs numériques DSF2 et DSF3.

Le décodeur SP451 (ou DSF3) propose à nouveau les traitements de la scène sonore définie par le format B (« rotate » rotation horizontale et verticale, « rear focus » directivité des microphones arrière). Il crée un ensemble virtuel de cinq directivités à partir du format B. Les cinq microphones virtuels du format 5.1 sont par défaut orientés à -45° pour le canal L, 0° pour C, $+45^\circ$ pour R, -135° pour Ls et $+135^\circ$ pour Rs. Le Sub (Lfe) est dérivé de la composante omnidirectionnelle W, filtrée à 120 Hz. Il est possible aussi de produire un format stéréo simultanément à partir du format B, en utilisant les sorties 7 et 8.

Lors du décodage, on paramètre la largeur frontale et arrière (« front width », « rear width ») ainsi que la directivité des microphones arrière (« rear focus ») entre omnidirectionnelle, cardioïde, et bidirectionnelle.

Les autres formats de production tels que le 6.1, le 7.1 ou le 8.0 sont compatibles.

Les microphones ST350, ST450 et le Surroundzone

Le microphone ST350 remplacé dernièrement par le ST450 est le microphone Soundfield de reportage, accompagné d'une bonnette anti-vent et d'un boîtier d'encodage au format B. Sur ce boîtier, on gère le gain, les paramètres « END » et « INV » et la sortie stéréo pour le monitoring au casque. En postproduction, le décodage se fait généralement avec le plug-in Surroundzone développé par Soundfield.



Figure 5.69 – Soundfield ST450, © B. Périaux.



Figure 5.70 – Soundfield ST450 et son boîtier, © Soundfield.

Le plug-in Surroundzone reprend les mêmes paramètres « END » et « INV », l'azimuth et l'élévation, le zoom, les largeurs frontales, arrière et permet le réglage de la directivité des capsules virtuelles arrière.



Figure 5.71 – Plug-in Surroundzone, © Soundfield.

L'Harpex constitue une autre solution de décodage format B, avec des possibilités de décorrélation complémentaires.

CONFIGURER LE SURROUNDZONE EN 5.1 DANS PRO TOOLS

Le plug-in Surroundzone se configure dans Pro Tools en insert sur une piste auxiliaire de format 5.1 dont l'entrée est un bus 5.1 créé dans le IO setup. Ce bus 5.1, que l'on peut nommer Soundfield, permet alors de router les trois pistes mono, sur lesquelles on place les signaux du format B, W, X, Y et Z. La piste W est envoyée vers le bus Soundfield, pan-potée vers l'avant gauche (ou directement vers le canal 1 du bus Soundfield), la piste X, envoyée vers le centre (canal 2), la piste Y, envoyée vers l'avant droite (canal 3) et la piste Z, envoyée vers le canal arrière gauche Ls (canal 4). La composante Z n'est utile que pour la restitution d'élévation, on peut donc l'abandonner pour une configuration en 5.1. L'utilisation des modes solo sur les pistes des signaux W, X, Y permet de vérifier que la configuration est bien faite, en identifiant le signal sur les entrées meter du plug-in. Le Surroundzone SPS200, compatible format A, se configure de la même façon. Les quatre pistes mono Lf, Rf, Lb et Rb sont envoyées respectivement vers le bus Soundfield en L, C, R et Ls.

Le SPS 200

Le SPS200 développé récemment est équipé de petites capsules, il permet d'enregistrer directement le format A avec le préamplificateur de son choix. DCAudiovisuel a réduit la taille du microphone pour qu'il puisse tenir dans une bonnette Zephyx. Son décodage est réalisé par le Surroundzone SPS200 qui encode en Format B puis décode en 5.1.



Figure 5.72 – Soundfield SPS200, © V. Magnier.

Le DPA-4

La société DPA a développé au début des années 1990 son propre système ambisonique d'ordre 1, le DPA-4, à partir de leurs capsules.

Le Core Sound TetraMic

Le TetraMic est un microphone ambisonique d'ordre 1 à petites capsules, développé par la société américaine Core Sound.

SOUNDFIELD

- Système quasi coïncident, 4 capsules hypocarديوïdes.
- Enregistrement directement au format B ou au format A (SPS200) suivi de l'encodage format B.
- Décodage du format B en 5.1, 6.1, 7.1, 8.0...

ANALYSE

- Très bonne localisation circulaire.
- Impression d'espace et profondeur très limitées, forte corrélation des 5 canaux.
- Indépendant du dispositif d'écoute multicanal.

5.4.3 Sphère Schoeps KFM 360

◆ Principe général

La sphère KFM360 est un autre système matricé, constitué d'une sphère stéréo de 18 cm de diamètre avec deux capteurs de pression et deux microphones bidirectionnels (CCM8) placés de chaque côté de la sphère et dirigés vers l'avant. L'angle de prise de son stéréophonique des deux microphones omnidirectionnels est de 120° tandis qu'il est de 90° pour la sphère Schoeps KFM6 (20 cm de diamètre).

La localisation latérale repose sur la différence d'intensité grâce au codage MS tandis que les localisations avant et arrière reposent sur de la différence de temps combinée à de la différence d'intensité, générée par la diffraction de la sphère (proche du principe de localisation binaurale naturelle).

Le processeur effectue une préaccentuation des aigus sur les signaux provenant des microphones omnidirectionnels et une préaccentuation des graves sur les signaux provenant des microphones bidirectionnels.



Figure 5.73 – Sphère Schoeps KFM360,
© B. Périaux.

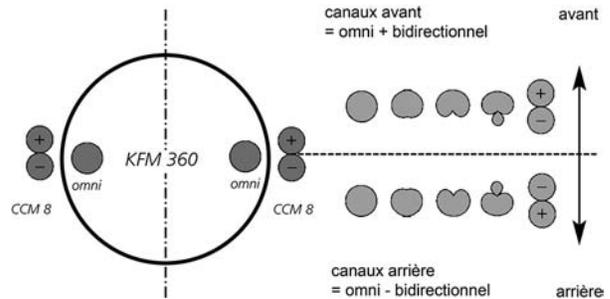


Figure 5.74 – Matriçages MS sur les côtés de la sphère,
© B. Périaux d'après Schoeps.

◆ Dématricage

Canaux avant gauche/droite

Les canaux avant gauche et droite sont obtenus par sommation des signaux issus des capteurs de pression et des bidirectionnels. Le dosage omnidirectionnels/bidirectionnels détermine la directivité des microphones virtuels ainsi créés (« FRONT »). La balance gauche/droite peut être modifiée.

Canaux arrière gauche/droite

Les canaux arrière gauche et droite sont obtenus par soustraction des signaux issus des capteurs de pression et des bidirectionnels. Le dosage détermine la directivité des microphones virtuels ainsi créés (« REAR »). Cette directivité arrière est indépendante de celle de l'avant.

Il est possible d'ajuster :

- Le niveau des canaux arrière par rapport à ceux de l'avant (« F/REAR BALANCE ») ;
- Le délai des canaux arrière par rapport à ceux de l'avant (« REAR DELAY/ms ») ;
- Le filtrage passe-bas des canaux arrière (« REAR SLOPE/Hz ») 12 dB/octave.

Canal central

Le canal central est créé par matriçage des canaux avant gauche et droite. Les signaux corrélés de ces canaux sont dirigés vers le canal central et soustraits des canaux gauche et droite. Le dosage du canal central est déterminé par le paramètre « FRONT PANORAMA ».

Canal Lfe

Le canal Lfe est créé à partir des signaux issus des capteurs de pression non égalisés par sommation et atténuation de 6 dB puis filtrage passe-bas à 80 Hz, avec une pente de 24 dB/octave.

SPHÈRE SHOEPS KFM360

- Système matricé double MS sur les latéraux.
- Couple AB omni avec diffraction de la sphère pour la latéralisation.
- Création électronique du centre, partie du signal corrélé avant gauche et droite.

ANALYSE

- Bonne localisation.
- Impression d'espace et profondeur limitées.
- Restitution des timbres limitée.
- Système complexe en exploitation, peu modulable.
- Utilisation pour la captation en salle de concert.

5.5 Systèmes d'ambiance

Les ambiances sont très couramment captées sur quatre canaux, restituées vers les enceintes L, R, Ls et Rs, laissant ainsi le centre disponible pour les dialogues par exemple. On trouve cependant des systèmes constitués de cinq microphones, avec aussi parfois une capsule pour la création d'un Lfe.

5.5.1 Double ORTF



Figure 5.75 – Double ORTF,
© B. Périaux.

Le système Double ORTF est plutôt considéré comme un système d'ambiance, bien qu'il soit parfois utilisé pour des fictions radiophoniques. Les deux couples sont disposés dos à dos, distants de 30 cm. L'image sonore restituée est cohérente, sans repliement d'espaces, ni zone non captée (voir les abaquages cardioïdes de Mike Williams, **figure 5.1**).

Ce système apporte une grande impression d'espace que l'on apprécie pour les ambiances extérieures avec profondeur. Il est possible de resserrer l'espace et de l'adapter au lieu en postproduction, avec l'utilisation d'un pan-pot. En revanche, son utilisation pour le son direct nécessite une mise en scène précise des sources et s'avère limitée car le système restitue des plans larges en proximité.

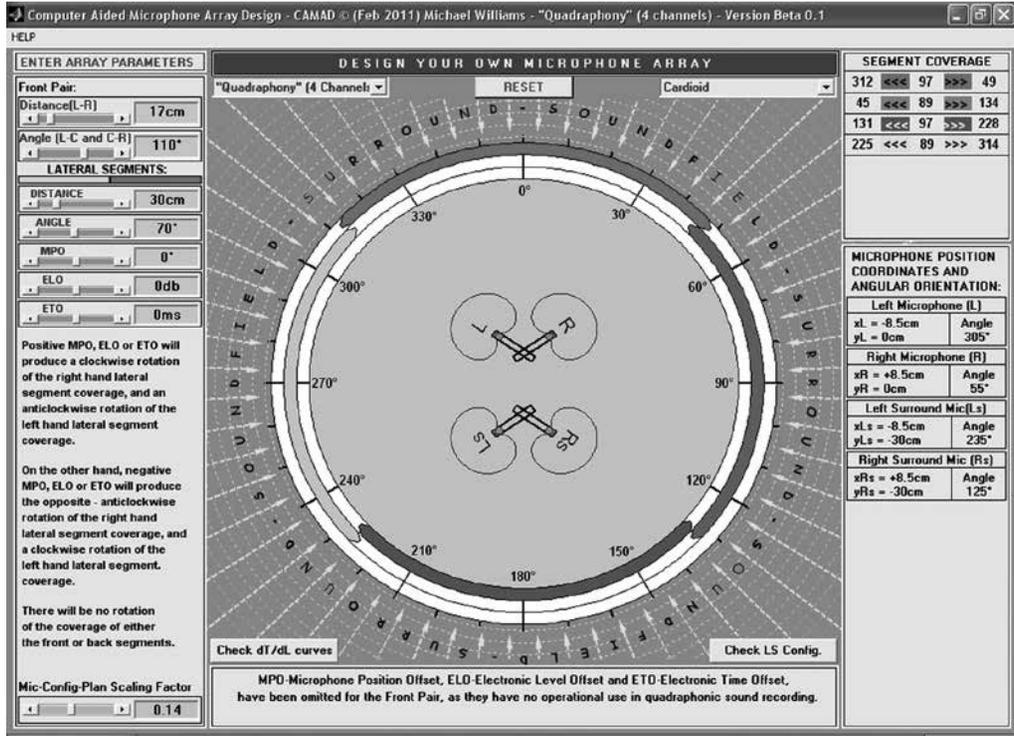


Figure 5.76 – Courbe de localisation CAMAD pour le Double ORTF, © M. Williams. En couleur sur dunod.com

DOUBLE ORTF

- 2 couples ORTF dos à dos, distants de 30 cm.
- Angles de PdS 90° sur chaque secteur.

ANALYSE

- Bonne localisation.
- Très bonne impression spatiale.
- Adapté pour les espaces extérieurs avec profondeur.
- Utilisation pour la captation d'ambiances en reportage, documentaire, et fiction radio.

5.5.2 ORTF Surround Schoeps

L'ORTF Surround est une modification du double ORTF développé par la société Schoeps. Il est plus compact grâce au choix des directivités supercardioides et peut être utilisé de ce fait pour le sport, le reportage et le documentaire. Il est constitué de quatre microphones supercardioides avec une distance gauche/droite de 10 cm et un angle physique de 100° pour l'avant, et une distance avant/arrière de 20 cm, et un angle

physique de 80°. Les angles de captation sont de 90°. La localisation de l'image sonore restituée est cohérente. Le système a sa propre cage anti-vent et son système de poils, on peut aussi l'équiper d'une bonnette Zephyx. Il est possible d'utiliser un système de chauffage des capsules pour les tournages en milieux humides.

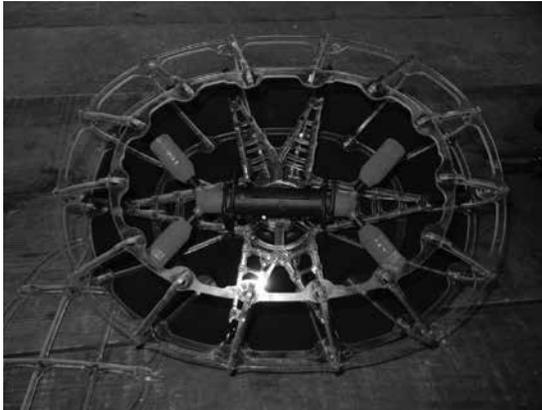


Figure 5.77 – ORTF Surround Schoeps, © V. Magnier.

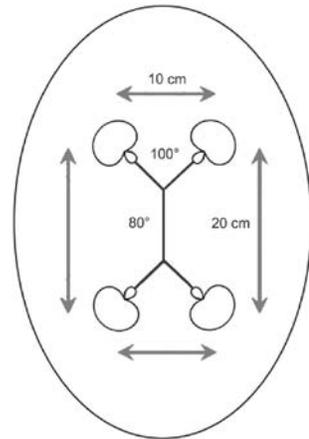


Figure 5.78 – ORTF Surround Schoeps, © B. Périaux d'après Schoeps.

Computer Aided Microphone Array Design - CAMAD © (Feb 2011) Michael Williams - "Quadraphony" (4 channels) - Version Beta 0.1

HELP

ENTER ARRAY PARAMETERS

Front Pair:
 Distance[L-R] 10cm
 Angle [L-C and C-R] 100°

LATERAL SEGMENTS:
 DISTANCE 20cm
 ANGLE 80°
 MPO 0°
 ELO 0db
 ETO 0ms

Positive MPO, ELO or ETO will produce a clockwise rotation of the right hand lateral segment coverage, and an anticlockwise rotation of the left hand lateral segment coverage.

On the other hand, negative MPO, ELO or ETO will produce the opposite - anticlockwise rotation of the right hand lateral segment coverage, and a clockwise rotation of the left hand lateral segment coverage.

There will be no rotation of the coverage of either the front or back segments.

Mic-Config-Plan Scaling Factor 0.14

DESIGN YOUR OWN MICROPHONE ARRAY

"Quadraphony" (4 Channels) RESET Supercardoid

SEGMENT COVERAGE

312	<<<	97	>>>	49
45	<<<	89	>>>	134
131	<<<	97	>>>	228
225	<<<	89	>>>	314

MICROPHONE POSITION COORDINATES AND ANGULAR ORIENTATION:

Left Microphone (L)	
xL = -5cm	Angle 310°
yL = 0cm	
Right Microphone (R)	
xR = +5cm	Angle 50°
yR = 0cm	
Left Surround Mic(Ls)	
xLs = -5cm	Angle 230°
yLs = -20cm	
Right Surround Mic (Rs)	
xRs = +5cm	Angle 130°
yRs = -20cm	

Check dt/dL curves Check LS Config.

MPO-Microphone Position Offset. ELO-Electronic Level Offset and ETO-Electronic Time Offset, have been omitted for the Front Pair, as they have no operational use in quadraphonic sound recording.

Figure 5.79 – Courbe de localisation CAMAD pour l'ORTF Surround Schoeps, © M. Williams. En couleur sur dunod.com

ORTF SURROUND SCHOEPS

- 4 supercardioïdes disposés en rectangle.
- 10 cm et 100° pour l'avant et l'arrière.
- 20 cm et 80° pour les latéraux.
- Angles de PdS 90° sur chaque secteur.

ANALYSE

- Bonne localisation.
- Bonne impression spatiale, espace plus resserré que le DORTF.
- Basses fréquences atténuées.
- Utilisation pour la captation d'ambiances en reportage, documentaire, fiction.

5.5.3 Croix IRT

La croix IRT est à l'origine un système conçu pour la captation d'ambiance en salle, elle se visse sur un pied. Elle est constituée de quatre microphones cardioïdes disposés en carré, distants de 25 cm, selon la recommandation J. Wuttke/Mike Williams, et formant entre eux des angles de 90°. Le système est parfaitement symétrique et offre quatre angles de prise de son de 90°. Le support Schoeps utilise des capsules CCM4.



Figure 5.80 – Croix IRT avec capsules Schoeps CCM4, © V. Magnier.

Il est possible de percher avec une croix si l'on ajoute un système de suspension. Autrement, la croix est souvent utilisée en complément d'un triplet avant.

Il existe deux autres versions de la croix IRT, l'une équipée de quatre microphones cardioïdes distants de 20 cm et l'autre équipée de microphones supercardioïdes, distants de 14 cm.

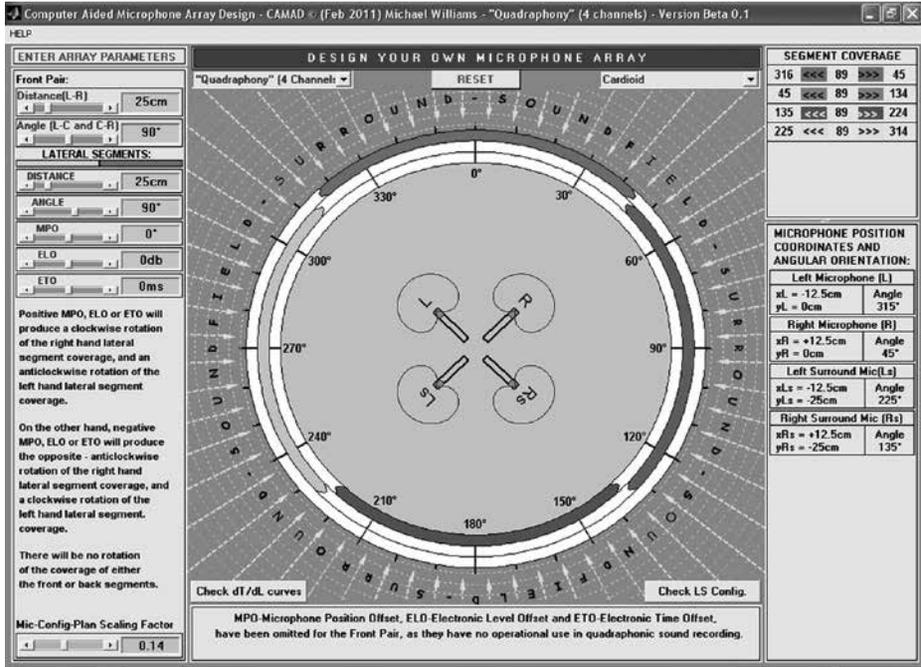


Figure 5.81 – Courbe de localisation CAMAD pour la Croix IRT 25 cm cardiïodes, © M. Williams. En couleur sur dunod.com

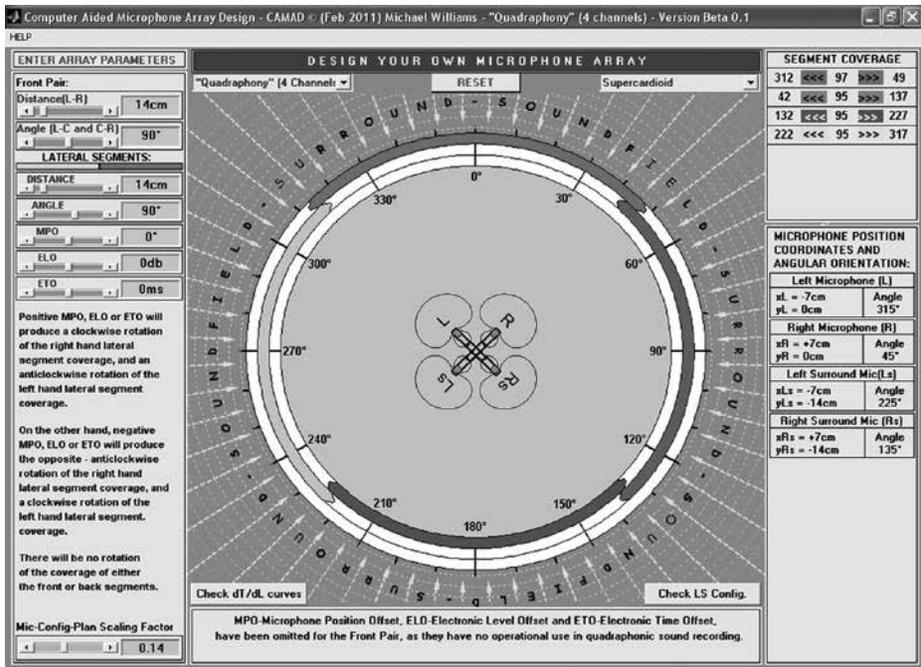


Figure 5.82 – Courbe de localisation CAMAD pour la Croix IRT 14 cm supercardiïodes, © M. Williams. En couleur sur dunod.com

CROIX IRT

- 4 cardioïdes disposés en croix, espacés de 25 cm, avec un angle droit.
- 4 angles de prise de son de 90°.

ANALYSE

- Bonne localisation.
- Très bonne impression spatiale.
- Bonne restitution des timbres.
- Utilisation pour la captation d'ambiances en salle de concert ainsi que pour les ambiances en reportage, documentaire, et fiction.

REMISE EN PHASE DES MICROPHONES EN SALLE

Lorsqu'on utilise des systèmes d'ambiance, les distances par rapport au système frontal étant considérables (10 à 30 m), il est indispensable de remettre en phase le système frontal et les microphones d'appoints par rapport au système le plus éloigné, c'est-à-dire le système d'ambiance. La **figure 5.83** compare les deux cas, avec et sans alignement temporel, pour un système frontal OCT, les appoints et un carré Hamasaki.

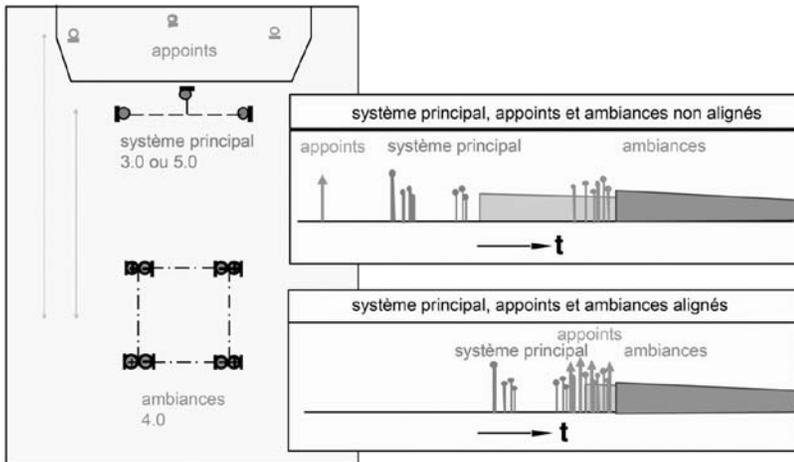


Figure 5.83 – Alignement temporel du système principal, des microphones d'appoint et du système d'ambiance, © B. Périaux d'après G. Theile.

5.5.4 Carré Hamasaki

Le Carré Hamasaki a été conçu par l'ex-ingénieur en chef de la NHK Kimio Hamasaki pour capter les ambiances dans une salle pour la captation des orchestres. Il repose sur

le placement en carré de 2 à 3 m de côté, de capsules bidirectionnelles, orientées vers les murs de la salle. Les microphones sont dirigés vers les canaux L, R, Ls, Rs comme toute ambiance multicanal. L'annulation de l'onde directe en direction de l'avant assure l'optimisation du champ réverbéré. Le système capte alors le champ sonore composé de réverbération et d'applaudissements du public, vers les murs de la salle ainsi que vers l'intérieur du carré. L'image sonore obtenue ne comporte que du champ diffus, elle est très homogène et apporte une grande impression spatiale. La taille de l'espace capté est variable, elle dépend de la taille du carré, plus on augmente la distance entre les microphones, plus on repousse les murs de la salle. Ce système est couramment utilisé pour les captations d'orchestres classiques à la NHK, à Radio France. Il est conseillé de noter alors la distance du système d'ambiance par rapport au système principal, et aux microphones d'appoints afin de remettre en phase les différents microphones.

CARRÉ HAMASAKI

- 4 bidirectionnels disposés en carré, dimension du côté de 2 à 3 m, les lobes positifs étant orientés vers les murs.

ANALYSE

- Pas de son direct.
- Très bonne impression spatiale.
- Construction du champ réverbéré très homogène.
- Basses fréquences atténuées.
- Utilisation pour la captation d'ambiances en salle de concert.



Figure 5.84 – Croix IRT et Carré Hamasaki en église, © B. Périaux.

5.5.5 *Omni square et variantes*

L'Omni square est une variante du carré Hamasaki exploitée par les ingénieurs du son de la NHK, il utilise des capsules de directivité omnidirectionnelle. On peut aussi utiliser des directivités hypocardiïdes comme représenté sur la photo ci-dessous.



Figure 5.85 – Carré hypocardiïdes suspendus, Auditorium de Bordeaux, © B. Périaux.

5.5.6 *Holophone H2 Pro*

L'Holophone H2 Pro a été conçu par la société canadienne Rising Sun Production. C'est un système non matricé, constitué de huit microphones omnidirectionnels DPA miniatures 4060 placés sur le contour d'une tête artificielle de forme ellipsoïde. Le principe repose sur les fonctions de transfert HRTF introduites par la tête artificielle simulant la tête humaine et sur la différence temporelle créée par les distances entre les microphones. La disposition des microphones sur le contour de la tête est une reproduction du système d'écoute 7.1 Imax, à l'échelle du système. Ce système, compatible 5.1, 6.1 cinéma et 7.1 Imax, est souvent utilisé pour des captations TVHD (événements sportifs). C'est un système robuste, très rapidement opérationnel.

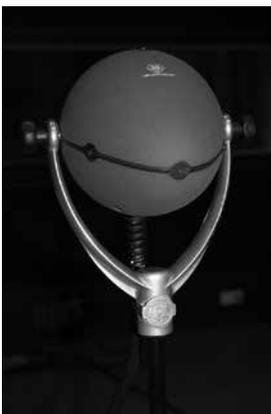


Figure 5.86 – Holophone H2 Pro, © B. Périaux.

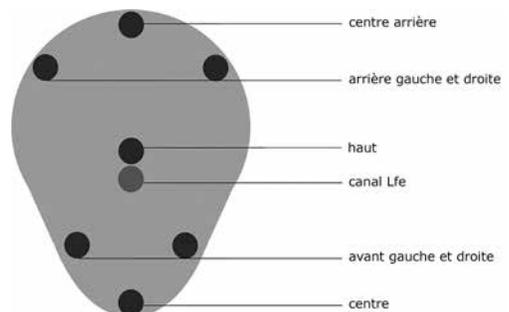


Figure 5.87 – Disposition des capsules, © B. Périaux d'après Rising Sun Production.

Voici l'ordre de branchement du H2 Pro : L, R, C, Lfe, Ls, Rs, Top, Center Rear.

Rising Sun Production a développé une gamme de systèmes plus petits, nommés H3 D et H4 super mini, ce dernier se fixe sur des caméras. D'autres modèles comme le PortaMic Pro et PortaMic 5.1, plus compacts, intègrent un encodeur Dolby Surround et véhiculent en sortie un signal stéréo LtRt.

HOLOPHONE H2 PRO

- 8 microphones omnidirectionnels DPA 4060 placés sur une tête hémisphérique.

ANALYSE

- Localisation plutôt imprécise.
- Impression spatiale peu favorable, image sonore avec internalisation.
- Équilibre spectral favorable aux basses fréquences.
- Peu sensible aux déplacements et au vent.
- Très rapidement opérationnel.
- Utilisation pour la TVHD, pour les programmes de sport.

5.5.7 DPA 5100

Le DPA 5100 fait partie des systèmes proposés pour la captation des événements sportifs, tout comme son concurrent l'Holophone, il est compact, léger, robuste et très rapidement opérationnel. À l'intérieur, on trouve cinq capsules DPA omnidirectionnelles. Les capsules omnidirectionnelles ont l'avantage d'être peu sensibles au vent et aux déplacements. Le triplet avant présente la particularité d'être coïncident, il permet alors un meilleur rendu par downmix stéréo et mono. Les capsules du triplet avant sont séparées par un bafflage acoustique, comparable au procédé du disque Jecklin, et sont montées sur des tubes à interférences, ce qui donne une certaine directivité aux capsules. Ce procédé, nommé « DipMic » (Directional Pressure Microphone), a été mis au point par les ingénieurs de DPA. Les capsules arrière, séparées de 18,5 cm, forment une sorte de couple AB et sont distantes du triplet avant. Le système propose un canal Lfe alimenté par la sommation L + R, atténuée de 10 dB pour la compatibilité du SUB en écoute ITU TVHD.



Figure 5.88 – Système DPA 5100, © B. Périaux.

DPA 5100

- 5 microphones omnidirectionnels DPA, coïncidents devant avec procédé DipMic et distants à l'arrière, disposés dans une forme revêtue de tissu.

ANALYSE

- Bonne localisation.
- Impression spatiale bonne, image sonore avec internalisation.
- Manque de transparence.
- Équilibre spectral favorable aux basses fréquences.
- Peu sensible aux déplacements et au vent.
- Très rapidement opérationnel.
- Utilisation pour la TVHD, pour les programmes de sport et le reportage radio.

5.6 La haute résolution spatiale

Ce contenu repose sur les excellents travaux de recherche de la société Trinnov Audio, www.trinnov.com

Tous les systèmes de prise de son étudiés jusqu'à présent utilisent des microphones traditionnels de directivités d'ordre 0 ou 1, allant de la figure omnidirectionnelle (ordre 0) à la figure bidirectionnelle (ordre 1). Ces systèmes sont dérivés de la prise de son stéréophonique, ils produisent des différences d'intensité et de temps entre les signaux captés suivant l'orientation des microphones et leurs distances respectives. Pour ce type de systèmes à résolution standard, le signal émis par une source est

diffusé par tous les canaux, par conséquent l'image fantôme correspondante est reproduite par un nombre élevé de haut-parleurs. La théorie OCT Surround de Günther Theile met justement en évidence la combinaison de ces images fantôme et le recours à l'optimisation de la diaphonie acoustique, grâce à l'utilisation de directivités supercardioides. L'utilisation de cinq enceintes implique théoriquement la combinaison de dix images fantôme. La superposition de ces images fantôme a pour conséquence un étalement spatial des sources apportant une certaine corrélation des canaux.

Le concept de la haute résolution spatiale s'inspire de la haute résolution temporelle (fréquence d'échantillonnage élevée) qui permet, grâce à une bande passante élevée, de suivre avec grande précision les variations du signal dans le temps. Les variations spatiales sont définies par la répartition angulaire du champ acoustique autour du microphone et pour suivre ces variations, le système d'enregistrement doit pouvoir enregistrer les fréquences spatiales élevées et donc avoir une bande passante spatiale considérable. La fréquence spatiale correspond à une figure de directivité dont la sélectivité des lobes détermine directement la précision spatiale. Plus la figure de directivité est sélective, plus la résolution spatiale qu'elle offre est élevée. Ces fréquences spatiales sont classées en fonction de leur sélectivité appelée l'**ordre**, voici ci-dessous cette classification, de l'ordre 0 à l'ordre 5.

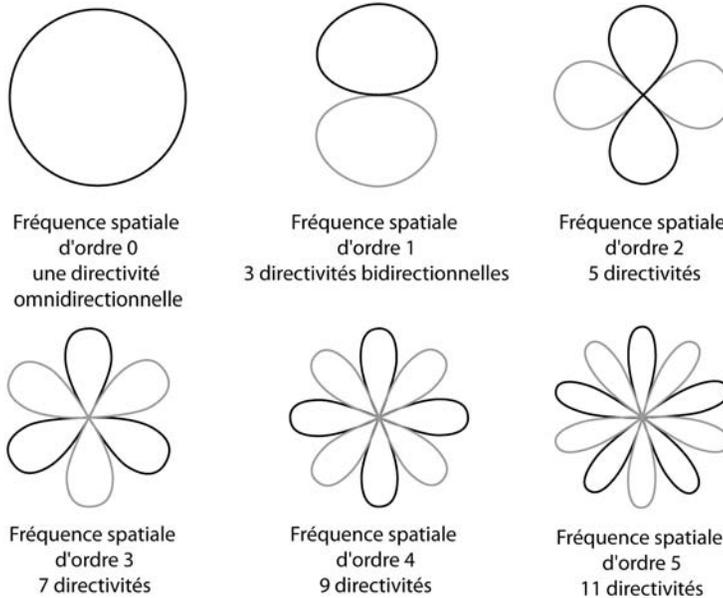


Figure 5.89 – Les directivités classées en fonction de l'ordre ou fréquences spatiales, les zones foncées représentent les lobes positifs, les zones claires, les lobes négatifs, © Trinnov Audio.

5.6.1 Haute résolution spatiale et Trinnov SRP

La précision spatiale obtenue avec un système d'ordre 5 est cinq fois supérieure à un système utilisant des directivités d'ordre 1. La **figure 5.90** compare les directivités cardioïdes et hypercardioïdes à l'ordre 1 et l'ordre 5.

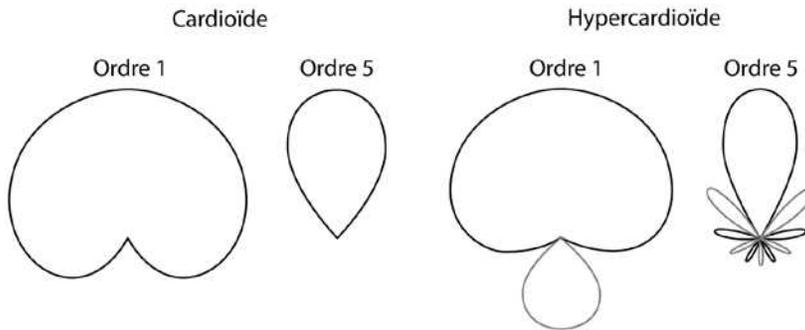


Figure 5.90 – Comparaison cardioïde et hypercardioïde ordre 1 et ordre 5, les zones foncées représentent les lobes positifs, les zones claires, les lobes négatifs, © Trinnov Audio.

Aujourd'hui, nous ne disposons pas de microphone d'une directivité aussi sélective que la précision obtenue à l'ordre 5. Si nous disposions de tels microphones, il serait alors simple de positionner nos microphones pour obtenir les directivités idéales adaptées au format 5.1. La loi panoramique idéale 5.0 est définie par cinq figures de directivités qui expriment l'intensité des canaux pour former une image fantôme optimale dans chacune des directions des enceintes.

Voici un comparatif donné par Trinnov entre les deux types de rendus de spatialisation multicanal obtenus avec un système coïncident à résolution standard, en différence d'intensité uniquement (la différence de temps n'est pas prise en compte), et avec un système à haute résolution type SRP (Surround Recording Platform) Trinnov Audio. Ce schéma illustre la diaphonie acoustique existante dans un système 5.1 ITU coïncident à résolution standard.

Sont représentées sur la **figure 5.92** les lois de panoramique d'intensité utilisées pour le mixage mono dirigé, le panoramique d'intensité d'ordre 5, et la loi « continuous surround panning » d'ordre 4 (P. Craven, 2003). Ces lois ont sensiblement la même forme générale.

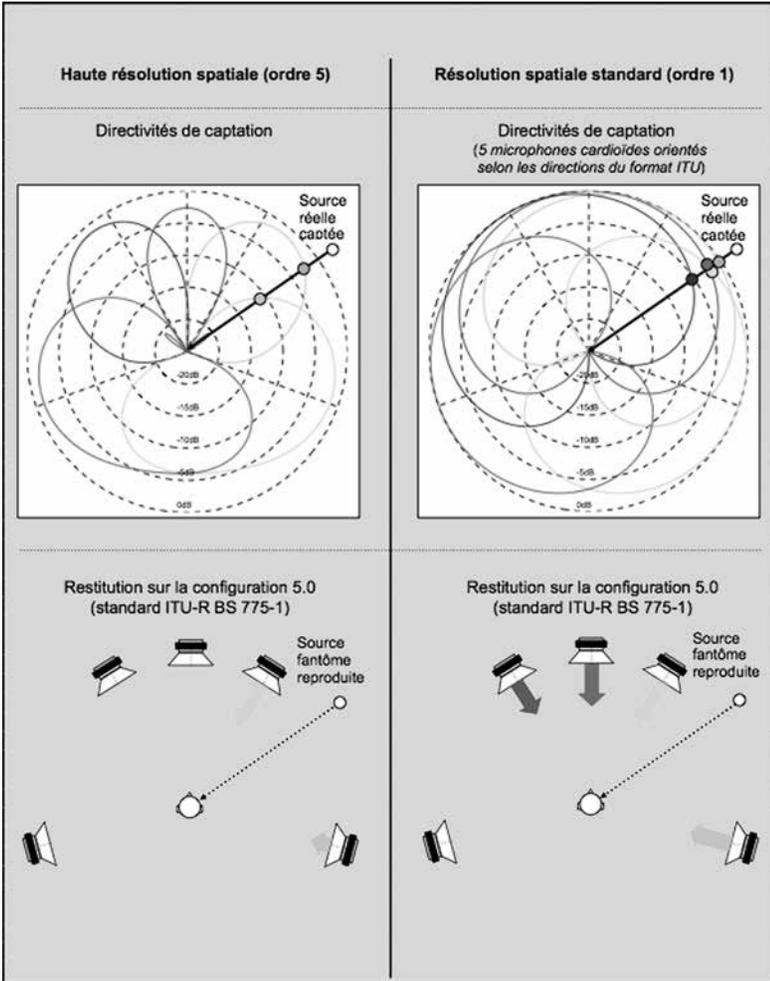


Figure 5.91 – Comparaison des prises de son 5.0 à haute résolution spatiale et à résolution spatiale standard, © Trinnov Audio. En couleur sur dunod.com

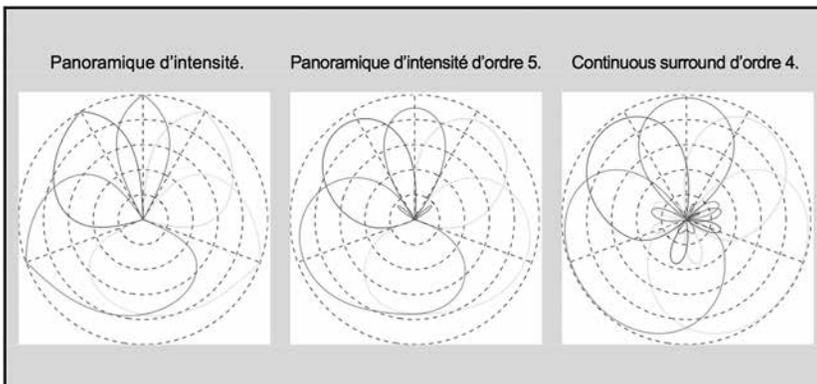


Figure 5.92 – Lois de panoramique équivalentes, panoramique d'intensité de source mono, panoramique d'intensité d'ordre 5, continuous surround panning d'ordre 4, © Trinnov Audio. En couleur sur dunod.com

La **figure 5.93** compare différentes lois de panoramique pour une restitution en 5.1 ITU, ordre 1 standard en cardioïdes coïncidents, ordre 5, et ordre supérieur à 5, avec leurs rendus sonores, insuffisant pour l'ordre 1, optimal pour l'ordre 5, trop sélectif pour les ordres supérieurs. La haute résolution spatiale repose sur la séparation optimale et non maximale des canaux. Une trop forte séparation des canaux ne permet plus le fonctionnement de l'image fantôme.

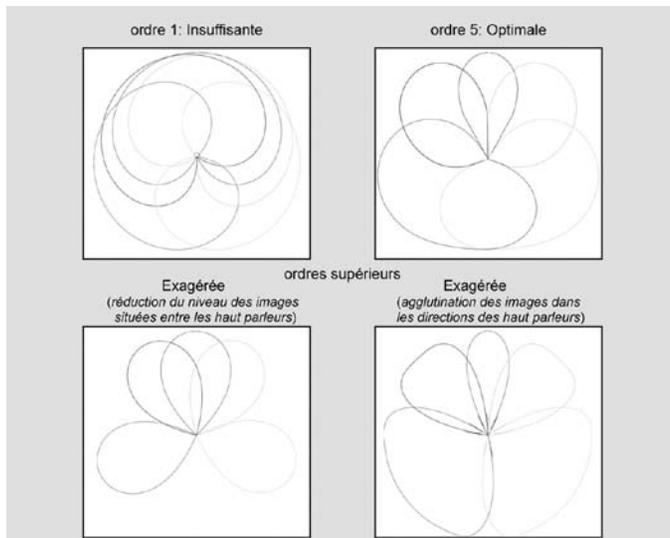


Figure 5.93 – Séparation optimale des canaux suivant l'ordre de sélectivité, © Trinnov Audio. En couleur sur dunod.com

On constate cependant que les lois panoramiques utilisées par les systèmes de prise de son multicanal d'ordre 1, abordés en début de chapitre (OCTS, MMAD, DMS, DORTE, Soundfield...) sont en général mixés à des microphones d'appoint qui travaillent avec des lois panoramiques d'intensité comparables à l'ordre 5 défini par le SRP. La résolution spatiale des images sonores ainsi combinées n'est en effet pas la même que celle obtenue pour un système principal unique travaillant à différence d'intensité seulement.

Ceci définit donc un système de prise de son 5.0 idéal qui reproduirait la meilleure image fantôme possible dans la même direction. Pour obtenir ce type de directivité 5.0 idéale, il est nécessaire de renoncer à l'approche traditionnelle qui consiste à utiliser **un microphone pour un haut-parleur**. Il faut s'orienter vers un procédé de matricage sur un réseau de microphones. Afin d'atteindre une qualité spatiale maximale, le SRP effectue un filtrage matriciel total. Chaque canal est obtenu en tirant profit de l'ensemble des microphones. Le signal de chaque microphone est filtré de manière spécifique, puis l'ensemble des signaux ainsi formés est combiné de manière à produire un canal optimal.

Les avantages d'un tel système d'enregistrement à haute résolution spatiale permettent une séparation optimale des canaux tout en gardant une certaine corrélation indispensable au fonctionnement de l'image fantôme. Le système augmente la

ponctualité des sources, car l’image fantôme sollicitée par la reproduction d’une source est fabriquée par la contribution de seulement deux canaux (voir **figure 5.94** la représentation des deux types de résolutions spatiales).

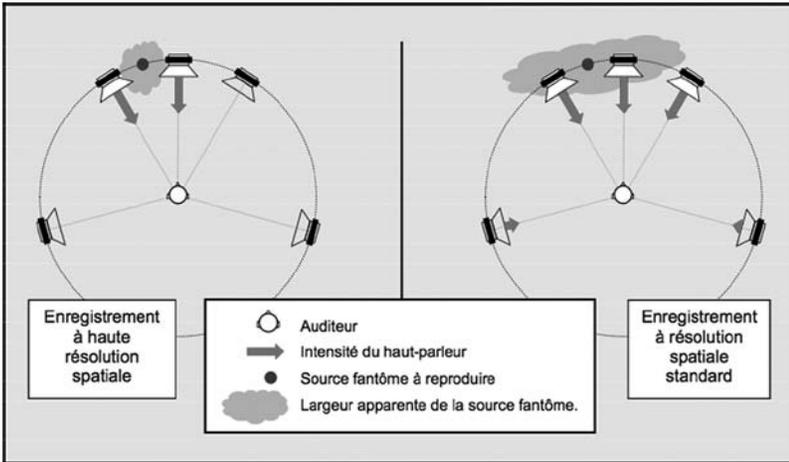


Figure 5.94 – Comparaison de la ponctualité des sources entre système à résolution spatiale standard et système à haute résolution spatiale, © Trinnov Audio.

Le système élargit aussi la zone d’écoute obtenue du fait de la contribution de deux canaux seulement pour la fabrication de l’image fantôme. D’autre part, les downmix stéréo ou mono sont de très bonne qualité, sans distorsion d’angle dans les images fantôme ni de modification de timbres, car les systèmes à haute résolution spatiale reposent sur un principe de coïncidence.

La haute résolution permet de contrôler finement toutes les directions de la sphère grâce aux 36 directivités ou fréquences spatiales illustrées par la **figure 5.95**.

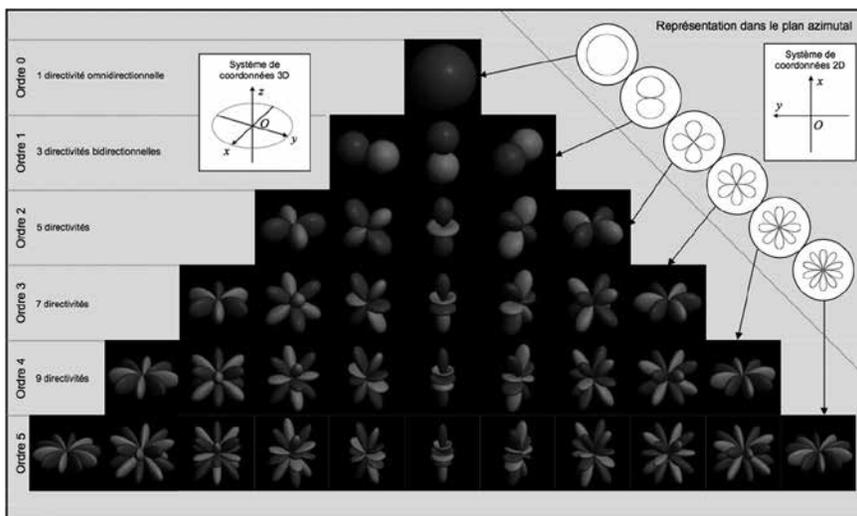


Figure 5.95 – Représentation des 36 fréquences spatiales jusqu’à l’ordre 5, © Trinnov Audio. En couleur sur Dunod.com

La perception d'élévation n'est pas envisagée par la prise de son avec le SRP, car les directivités virtuelles obtenues sont établies sur un plan horizontal, compatible avec le dispositif 5.1.

Principe d'échantillonnage spatial, recomposition des directivités sélectives

Comme les microphones usuels ne permettent pas d'obtenir des directivités très sélectives comme celles obtenues au degré 5, il convient de reconstituer ces directivités grâce à un filtrage matriciel. Le SRP fonctionne donc en deux étapes, une étape d'acquisition d'informations sur le champ acoustique à l'aide d'un réseau de microphones standards et une étape de traitement permettant de synthétiser les directivités souhaitées. L'étape d'acquisition est un échantillonnage spatial du champ acoustique, ce procédé est illustré par la **figure 5.96**.

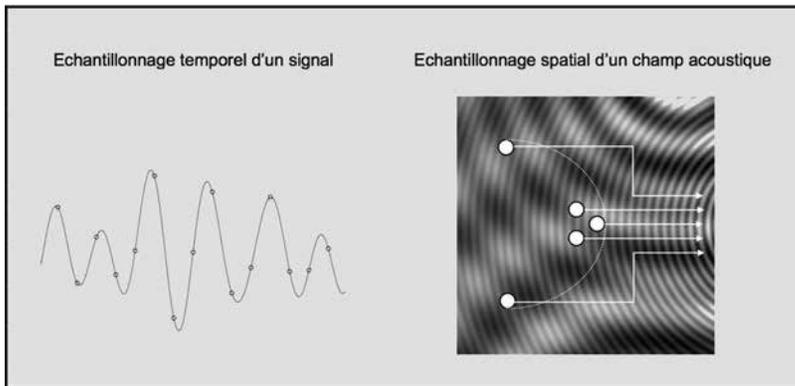


Figure 5.96 – Principe d'échantillonnage spatial du champ acoustique, © Trinnov Audio.

Le SRP est un réseau de huit microphones omnidirectionnels qui analysent les différences de temps sur huit points dont l'agencement est optimisé. L'approche du SRP est de réaliser un système de prise de son à haute résolution tout en utilisant un nombre de microphones **compatible avec l'économie de la production audiovisuelle**. Les travaux de Trinnov ont montré qu'un système non coïncident réalise une meilleure captation qu'un réseau coïncident. Le traitement est ensuite basé sur 40 filtres à 1 024 fréquences permettant une reconstitution optimale de chaque canal à partir des huit microphones omnidirectionnels. L'ensemble du système (réseau SRP + unité de traitement) se comporte comme cinq microphones coïncidents virtuels réalisant une prise de son à haute résolution spatiale. La capture se fait donc en différence de temps, codée par la suite en différence d'intensité, ce qui rend le système coïncident.

La **figure 5.97** représente les microphones virtuels obtenus, de directivités asymétriques, d'ordre 5 dans le frontal et d'ordre 2 à l'arrière.

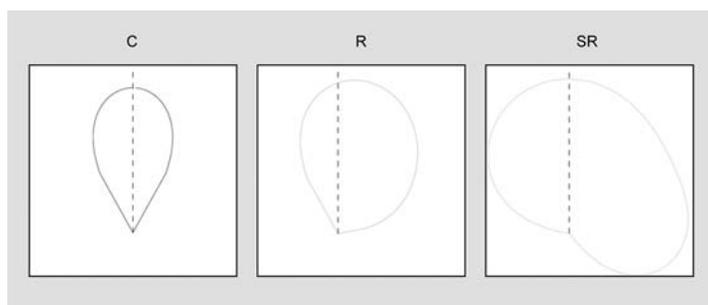


Figure 5.97 – Microphones virtuels obtenus, © Trinnov Audio.



Figure 5.98 – Trinnov Audio SRP, © B. Périaux.

En conclusion, le système SRP de Trinnov Audio est un système de captation à haute résolution spatiale d'ordre 5, il recrée cinq microphones virtuels coïncidents, fonctionnant dans un plan, pour une restitution 5.1. La captation en hauteur avec un tel système n'est pas envisagée. La sensation d'espace reste limitée compte tenu peut être de la coïncidence du système virtuel obtenu et la restitution des timbres nécessiterait d'être améliorée. En revanche la localisation et notamment la zone d'écoute sont d'une grande stabilité, bien supérieure aux systèmes à résolution standard. Enfin, les downmix sont de très bonne qualité.

5.6.2 High Order Ambisonics

Les technologies ambisoniques ont la faculté de reproduire ou de créer des environnements sonores en trois dimensions. Ces techniques ont été principalement développées par Michael Gerzon dans les années soixante-dix. L'ambisonie est une décomposition du champ de pression sonore sous formes de composantes (ou fréquences spatiales) harmoniques sphériques. Le Soundfield, étudié auparavant, est un format ambisonique d'ordre 1, il offre une décomposition du champ sonore en directivités bidirectionnelles suivant les axes des x , y , et z . L'extension du format ambisonique à des ordres supérieurs est appelée High Order Ambisonics (HOA), cette technologie a été développée conjointement par Orange Labs (Jérôme Daniel) et la société MH Acoustics au début des années 2000. Cette dernière permet d'améliorer la reproduction du champ sonore, notamment en élargissant la zone d'écoute. L'encodage spatial peut être envisagé sous deux aspects distincts : la simulation des caractéristiques spatiales du champ acoustique à partir de signaux audio monophoniques, il s'agit de synthèse de source virtuelle, ou l'enregistrement naturel par un dispositif microphonique. Puisque les signaux HOA sont une représentation physique d'une scène sonore indépendante du dispositif de restitution, le décodage spatial peut s'adapter à différents systèmes de diffusion (circulaire, sphérique, stéréophonique, 5.1, binaural etc.). Le dispositif HOA permet aussi un certain nombre de traitements possibles comme la rotation ou la distorsion de perspective.

Le système HOA est une sphère comportant généralement 32 microphones miniatures omnidirectionnels, d'autres systèmes fonctionnent avec 20 capteurs, tout dépend de l'ordre que l'on choisit. Voici les différents ordres utilisés en 3D, avec le nombre de capteurs et de composantes associés :

ordre 1 → 4 capteurs → 4 composantes (Soundfield)

ordre 2 → 12 capteurs → 9 composantes

ordre 3 → 20 capteurs → 16 composantes

ordre 4 → 32 capteurs → 25 composantes

L'ordre m du système HOA définit $(m + 1)^2$ composantes en 3D et $2m + 1$ composantes en 2D.

Il existe un modèle commercialisé par la société MH Acoustics nommé EigenMike, il est composé de 32 capteurs disposés sur une sphère de diamètre 8,4 cm et fonctionne à l'ordre 4.

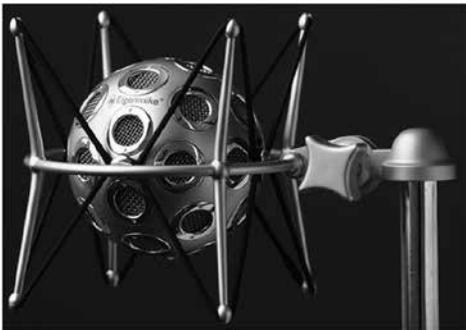


Figure 5.99 – La sphère EigenMike, MH Acoustics, © MH Acoustics.

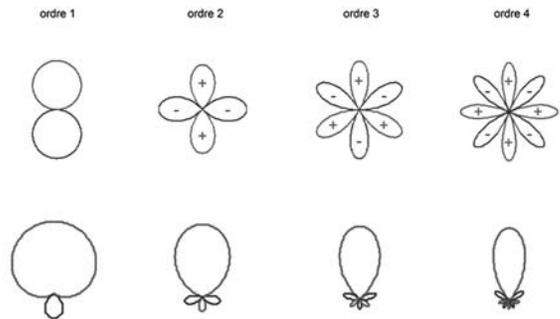


Figure 5.100 – Composantes HOA et directivité virtuelle obtenue, © J. Daniel.

Ce réseau de microphones effectue un encodage spatial du champ acoustique en un certain nombre de composantes ambisoniques selon l'axe des x, y, z , constituant le format B.

Lors du décodage spatial, ces différentes composantes sont matricées et égalisées afin de fabriquer un ensemble de directivités sélectives virtuelles compatibles avec le format souhaité, cinq directivités pour le 5.1 par exemple. Le décodage reconstitue un microphone virtuel par enceinte.

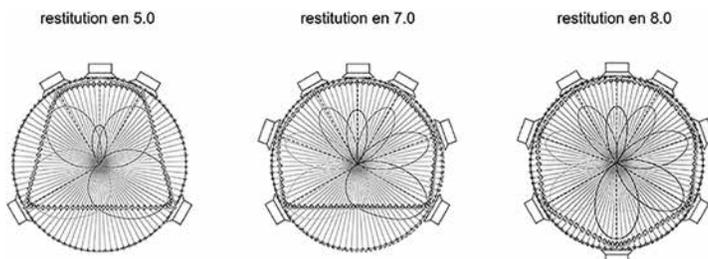


Figure 5.101 – Directivités virtuelles obtenues à l'ordre 4 pour des restitutions 5.0, 7.0 et 8.0, © J. Daniel.

Le système plutôt coïncident semble limiter la sensation d'espace (peu de perception de profondeur) et poserait des contraintes quant à la restitution des basses et hautes fréquences (dues à l'espacement des capteurs). Pour améliorer le rendu dans les basses fréquences, il faudrait augmenter la taille de la sphère, mais ceci augmenterait l'espacement des capteurs, donc limiterait le rendu dans les hautes fréquences. En revanche la localisation circulaire semble très satisfaisante, stable et le downmix obtenu est de bonne qualité.

5.7 Prise de son binaurale

On définit par prise de son binaurale, tout système de prise de son capable de restituer un rendu spatial au casque, avec sensation d'externalisation. Cette spatialisation qui se distingue de la stéréophonie au casque traditionnelle donnant une perception intracrânienne, reproduit les fonctions de transfert HRTF de notre tête. Aujourd'hui deux possibilités s'offrent au preneur de son : il peut utiliser des têtes artificielles, des bustes de mannequins ou bien placer des microphones miniatures dans le creux de ses oreilles et reproduire ainsi ses propres fonctions de transfert. Les résultats sont souvent meilleurs avec l'utilisation de têtes artificielles (tête Neumann KU100 par exemple) qu'avec une paire de microphones de type DPA 4060 par exemple, placée dans les oreilles, très sensible aux mouvements, et aux bruits de respiration.

Généralement, on perçoit dans cette approche d'avantage de son sur les côtés avec une localisation limitée sur le plan frontal. La perception de distance est en revanche très bien restituée et la localisation circulaire et en hauteur est très cohérente. La spatialisation des sources est réalisée sur le terrain, autour du système, ce qui nécessite une réalisation très précise.

La tête artificielle KU 100, représentée ci-dessous, est la réplique d'une tête humaine, avec un microphone omnidirectionnel placé dans chaque oreille.



Figure 5.102 – Tête artificielle Neumann KU100,
© B. Périaux.



Figure 5.103 – Tête artificielle
Neumann KU80, © B. Périaux.

Il existe d'autres systèmes tels que la tête artificielle Head Acoustics HMS3 ou HMS4 (celles-ci ne comportent pas de pavillons), le récent et original système sans tête Free Space Pro 3DIO Sound, le mannequin Kemar bien plus ancien constitué de fibre de verre ou le mannequin B & K 4100.



Figure 5.104 – Système Free Space Pro 3DIO Sound, © 3DIO.



Figure 5.105 – Mannequin Kemar, © GRAS, Sound and Vibration.

Voici un historique de la prise de son binaurale donné par Bernard Lagnel (preneur de son à Radio France) :

PETITE HISTOIRE DE LA PRISE DE SON BINAURALE, SYSTÈMES ANTHROPOMÉTRIQUES

- 1930 : Création du mannequin « OSCAR » par Barlett Jones de la Bell Systems Laboratory à Chicago, les capteurs sont placés en avant des oreilles.
- 1969 : OSKAR Dummy Head Stéréo de Sennheiser.
- 1974 : Dummy Head stéréo 2 de Sennheiser, MZK2002 et version microphones miniatures MKE2002.
- 1972 : Kemar, Gras.
- 1973 : Nuemann KU80, équipée de microphones KM83, tête égalisée pour une réponse linéaire en champ direct, les pavillons ne comportent pas de tragus.
- 1982 : Neumann KU81i, microphones KM83, tête égalisée pour une réponse linéaire en champ diffus, compatible avec hauts-parleurs, présence du tragus sur les pavillons.
- 1992 : Neumann KU100, tête égalisée pour une réponse linéaire en champ diffus, compatible avec hauts-parleurs, présence du tragus sur les pavillons.
- 2011 : Samrec type 2500s Southern Acoustics Co. Ltd Japon, sorte de clône de la tête MZK2002 ?
- 2013 : Free space pro 3DIO Sound.

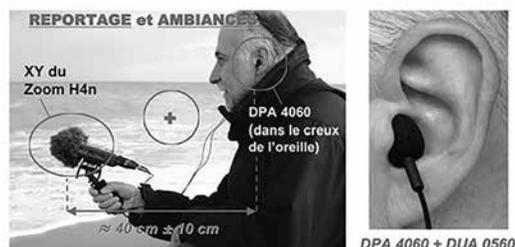
Ces systèmes de prise de son nécessitent une égalisation, laquelle est souvent contenue dans la tête, c'est le cas de la Neumann KU100. Cette égalisation dégrade quelque peu les indices spectraux mais minimise les artefacts de détimbrage, elle peut se faire pour

une réponse linéaire en champ direct (la méthode généralement utilisée en chambre anéchoïque) ou en champ diffus.

Voici le système expérimental de Bernard Lagnel qui utilise un couple AB avec des oreilles artificielles :

La prise de son binaurale native donne de bons résultats, c'est une forme de captation naturelle comparable au « système principal » défini en début de chapitre, impliquant donc une certaine distance avec la source et restituant une image générale. Elle contient les fonctions HRTF du système utilisé (tête artificielle ou tête du preneur de son) et donne une image spatialisée figée. Seule l'inversion gauche droite de la scène sonore est possible par inversion des canaux gauche et droite de la piste stéréo en postproduction, mais la rotation de la scène sonore n'est malheureusement pas possible. Elle peut être complétée par des microphones d'appoint traités alors par synthèse mais se pose alors la question des HRTF utilisées pour obtenir la meilleure compatibilité.

Pour améliorer la perception, il faudrait une individualisation des HRTF (possible avec l'utilisation de microphones dans le creux des oreilles mais valable que pour le preneur de son) et redonner à l'auditeur les indices dynamiques de localisation grâce à un système de « head tracking » (malheureusement quasi-impossible impossible en prise de son binaurale).



Figures 5.106, 5.107, 5.108 – Configuration binaurale de Bernard Lagnel, couple binaural avec oreilles artificielles, ce système est ici complété par un couple XY utilisé pour la précision des plans proches et pour la compatibilité 5.1 (système « plug and rec studio »), © B. Lagnel. La deuxième photo représente le système avec des DPA 4060 et le zoom H4n utilisé pour le reportage et les ambiances. La troisième photo représente « le cyclope » ou « Headgar », une enceinte Genelec 8020 B équipée d'oreilles artificielles, © B. Périaux.

Une autre approche consiste alors à encoder un flux binaural par synthèse :

- soit à partir d'une prise de son ou d'un mixage multicanal finalisés, on obtient alors un format binaural dont les directions restent quantifiées par le format multicanal ;
- soit par le mixage utilisant un logiciel ou un plug-in de synthèse, on obtient alors une image binaurale dont les directions sont quantifiées selon la finesse de la base HRTF ou du système d'écoute mesuré. Dans ce dernier cas, la spatialisation comporte généralement davantage de directions et de ce fait est plus naturelle.

5.8 Prise de son en tournage

Après cette étude théorique de la prise de son, abordons maintenant la présentation des enregistreurs portables compatibles multicanal. Nous exposerons ensuite quelques retours d'exploitation.

5.8.1 Les enregistreurs

Les enregistreurs multipistes portables se sont multipliés ces dernières années et permettent maintenant d'enregistrer en multicanal de façon « nomade ». En voici quelques-uns parmi les plus sérieux.

Les appareils sont présentés par le nombre de pistes qu'ils peuvent enregistrer et par le nombre d'entrées au niveau micro. Il va de soi qu'au-dessous de quatre pistes, les possibilités sont limitées. Quatre entrées au niveau microphone sont un minimum, et c'est même un peu juste : on ne peut utiliser que certains systèmes de prise de son et faire en même temps une prise de son monophonique, gérer des microphones d'appoint, et utiliser un système de prise de son multicanal devient alors impossible. D'où l'évolution du nombre d'entrées micro.

Complément Web 5.3

Liste d'enregistreurs utilisés pour la prise de son multicanal.

5.8.2 Les différents systèmes : pièges et astuces, avec la participation de Vincent Magnier

Voici un retour d'expérimentation sur la prise de son multicanal mettant en avant une méthode de travail, les pièges à éviter et quelques astuces.

Complément Web 5.4

Les différents systèmes, pièges et astuces.

5.9 Applications

5.9.1 Film

◆ **Gaël Nicolas : le point de vue d'un monteur son, utilisateur du format B**

Gaël Nicolas est monteur son pour le film, il pratique l'enregistrement ambisonique, en format B, depuis 2005, date à laquelle il achète son premier microphone Soundfield.

Les possibilités de décodage du format B, m'ont tout de suite amené à découvrir toute une perspective de recherche en matière sonore. Avec un son enregistré au Soundfield, je pouvais en créer 20 à 30 autres, selon le paramétrage du décodage, suivant la gestion de l'espace, des canaux avant et des canaux arrière.



Figure 5.109 – Gaël Nicolas, prise de son au Soundfield, © G. Nicolas.

Complément Web 5.5

Gaël Nicolas ITV.

◆ **La personne aux deux personnes, Thierry Lebon, mixeur film**

Thierry Lebon est mixeur film, il a expérimenté un système de captation multicanal pour l'enregistrement de voix off.

L'idée de cette prise de son 5.1 est venue du scénario. L'histoire du film « La personne aux deux personnes » met en scène Daniel Auteuil, un personnage introverti qui a un accident avec une personne qui entre littéralement dans son corps. Les vingt premières minutes du film reposent sur un dialogue entre les deux personnes, Daniel Auteuil et lui-même. L'image est faite de prises de vues subjectives qui sont à la place

du personnage, pendant ces images on entend la personne qui se trouve à l'intérieur de lui. On s'est imaginé que l'on pouvait faire une prise de son multicanal et pas seulement restituer une voix en mono, diffusée sur tous les haut-parleurs. C'est un procédé que j'avais déjà utilisé sur « Mission Cléopâtre » pour faire les enregistrements de la voix off de Pierre Tchernia, en 2001. Car pour moi, la voix off doit sortir de l'écran.



Figure 5.110 – Dispositif de captation voix narrative, « La personne aux deux personnes », C M149, L et R AKG C414, arrière Sony C48 et Lfe AKG C460, © Thierry Lebon.

Complément Web 5.6

Thierry Lebon ITV.

5.9.2 Documentaire à l'image

Documentaire animalier, Philippe Barbeau et Martine Todisco, preneurs de son

Philippe Barbeau et Martine Todisco, sont preneurs de son spécialisés dans le cinéma animalier. Ils ont travaillé notamment sur les longs métrages de Jacques Perrin « Microcosmos, le Peuple de l'Herbe », « Le Peuple Migrateur », « Le Peuple Singe », « Océans » Ce sont de fervents utilisateurs de systèmes de prise de son multicanal.

Nous avons commencé à nous intéresser au multicanal à travers la prise de son simultanée sur plusieurs canaux, mais sans penser enveloppement. En 2005 sur un long-métrage, « Les Animaux Amoureux », nous avons eu l'opportunité d'emprunter un Cantar et un Deva IV. Jusqu'alors, nous tournions sur Nagra 5 numérique. Le tournage débutait à l'été 2005, et commençait par le brame du cerf. L'idée première était de disposer plusieurs microphones sur les places de brame, et d'enregistrer

simultanément au lieu de passer d'un couple stéréo à l'autre, ceci, afin de ne pas rater un déplacement d'animal. Nous nous sommes rendu compte qu'il y avait mieux à faire : capter, non pas plusieurs points différents, mais un ensemble cohérent avant-arrière.



Figure 5.111 – Philippe Barbeau,
© P. Barbeau.



Figure 5.112 – Martine Todisco, © P. Barbeau.

Complément Web 5.7

Philippe Barbeau, Martine Todisco ITV.

5.9.3 Documentaire, reportage, fiction radiophonique par Guy Senaux, ingénieur du son Radio France



Figure 5.113 – Premier jour de tournage, *Les Bonobo*, Crespin, Christine Siméone, Boukavou, Guy Senaux, © On Voice.

Guy Senaux ingénieur du son à Radio France nous raconte ici les premiers documentaires, reportages et fictions radiophoniques produits en 5.1 à Radio France, de 1995 à 2002.

◆ **Le Cataclysme Sonore, France Inter**

Complément
Web 5.8

◆ **Le Singe Soleil, France Inter**



Figure 5.114 – Le Singe Soleil, d'après un dessin de Jonathan Kingdon.

Première émission de radio en son multicanal 5.1

Présentation

En octobre 1996, Jacques Santamaria, Directeur de France Inter, nous a permis d'explorer la spatialisation totale de sons purs en réalisant la première émission de radio, enregistrée en son multicanal aux normes 5.1, en acceptant, en cinq minutes, le projet de reportage de Robert Arnaut et Jacques Charreaux : le Singe Soleil.

Pour réaliser cette production, nous avons choisi le cadre de la forêt africaine, qui par la richesse de ses décors sonores, nous plonge dans un monde étonnant.

Nous sommes partis au cœur du Gabon, à 6 000 km de Paris, dans la forêt des Abeilles, forêt inexplorée par l'homme jusqu'en 1984. À cette date les chercheurs du laboratoire de primatologie et de biologie évolutive du CNRS ont découvert dans ce lieu unique au monde, une nouvelle espèce de singe, le Singe Soleil (cercopithèque avec une queue orange qui ne vit que dans la forêt des Abeilles, au centre du Gabon sur une superficie de 100 km sur 100 km). Il est semi-terrestre, c'est-à-dire qu'il marche à terre et ne grimpe dans les arbres que pour manger certains fruits et pour dormir. C'est une espèce entièrement protégée.

C'est à sa recherche que nous partons faire ce documentaire, en ayant pour but de mettre l'auditeur au cœur du voyage, au centre des cinq enceintes, à la place des microphones.

En son multicanal, vous n'écoutez plus les échos d'un reportage, mais vous faites réellement partie de l'expédition. Au milieu de ce décor sonore, l'auditeur devient acteur.

Choix du matériel

Pour des raisons de fret et de portage pendant nos différentes marches en forêt très humide, le poids a été un des éléments capitaux de notre sélection. Comme c'était en période de très grande humidité (près de 90 % et 30° de température au mois d'octobre), nous n'avons pas choisi le Nagra numérique à quatre pistes car il n'avait pas été utilisé dans de telles conditions et vu le coût de la mission, on ne pouvait pas prendre de risques.

Nous nous sommes dirigés vers le choix de :

- 2 couples ORTF Schoeps capsules MK5 ;
- 2 HHB Portadat 1000 TC, enregistreurs DAT stéréo, synchronisés par un générateur de time code (voir photo ci-dessous).



Figure 5.115 – Matériel de
Prise de son baptisé 4x4,

© Jean-François Bernard Sugy.

Prise de son des voix narratrices

Nous avons utilisé un couple de microphones séparés Schoeps, capsules MK5 en directivité cardioïde, fixés sur une barrette AKG (angle entre les deux microphones 45 à 60 degrés) pour éliminer le trou en très gros plan du couple ORTF. Avec un couple ORTF, les sons sont détimbrés dans le centre du couple dès que l'on s'approche de 18 à 15 cm.

Prise de son

En pratique la prise de son en multicanal est plus subtile que la prise de son en stéréo, mais tout à fait accessible. Le preneur de son est un metteur en scène sonore. Je détermine d'abord le couple avant de ma prise de son. Je place si cela est possible les personnages. Par exemple, si on enregistre deux personnes, l'une est placée à 1/3 à gauche et l'autre à 1/3 à droite. Autre exemple avec deux personnages : on place la personne que l'on interviewe devant et l'intervieweur derrière, avec le preneur de son sur le côté. Nous aurons un problème si nous devons marcher dans la forêt, car nous entendrons le preneur de son sur un côté. Dans ce cas-là, on donne la perche au producteur et le preneur de son se met derrière lui, à 2 mètres, comme cela on ne l'entend pas, ses pas étant couverts par ceux du producteur.

Analysons un autre exemple : la mise à l'eau d'une pirogue avec les cris des piroguiers et les clochettes fixées à l'avant de la pirogue qui permettent en écoutant leurs divers sons de connaître le tirant d'eau. Pour l'auditeur, je détermine une mise en place : je vais la faire venir par le centre de l'arrière, et la mise à l'eau se fera par devant. Si les indigènes ne suivent pas mon travelling sonore souhaité, je tourne mon système de prise de son pour qu'il en soit ainsi et pour ralentir leur déplacement.

L'écoute sélective doit se faire d'abord en écoutant sans casque, puis au casque, car après avoir déterminé le choix du positionnement des sons à enregistrer, il ne sera possible d'écouter au casque, lors de l'enregistrement, que le son provenant soit du couple avant, soit du couple arrière.

Les mouvements de perches doivent être adoucis car on déplace simultanément quatre microphones, ce qui accentue les risques de bruits parasites et les trépidations.

La proximité du preneur de son et des quatre micros, oblige celui-ci à ne faire aucun bruit (vêtements, déplacements, respirations, voix). Cette proximité est une grande contrainte, d'autant plus que les housses des PORTADAT ont des fermetures velcro qui ne sont pas particulièrement discrètes.

Dans la forêt équatoriale, il n'y a pas de grandes montées, mais des petites butes de 3 à 4 mètres de haut, très glissantes et l'air paraît rare. Lors des longs travellings dans la forêt, dans certains cas, il est nécessaire d'utiliser une perchette flexible de 30 cm plaçant le couple arrière au-dessus de la tête du preneur de son, afin de diminuer au maximum tous ses bruits de respirations.

Si j'ai choisi d'utiliser le couple ORTF, c'est que par ma longue pratique d'enregistrement de fictions, c'est le meilleur système de prise de son qui respecte la précision des plans sonores et la profondeur en plans moyens et en plans lointains. De plus, ce système 4x4 est très pratique pour les reportages avec la possibilité de séparer les deux couples, comme par exemple sur les pirogues ou pour la coupe d'un arbre Okoumé. Jacques Charreaux était à 40 mètres à la chute de l'arbre avec le couple arrière, et moi-même, avec le couple avant, au pied de l'arbre. On a synchronisé les deux enregistrements au mixage.



Figure 5.116 – Radeau d'Okoumé © Jacques Charreaux.

Précisons qu'entre deux enceintes espacées de 3 mètres, si la voix est enregistrée sur un seul microphone, elle a une dimension de 60 cm. Enregistrée en stéréo avec ce système de deux microphones, la voix a une dimension de 1,5 mètre à 2 mètres. Elle est d'autre part plus brillante, avec la résonance du studio qui lui rajoute une belle sonorité dans le haut médium.

À l'inverse du Cataclysme Sonore, le Singe Soleil est une émission « bio », car nous n'avons rajouté aucun effet sonore. Seulement de

temps en temps des corrections pour couper le léger vent caressant les capsules ou les trépidations de la perchette.

Séquence par séquence, on procède au mixage puis à la mise en place de chaque élément, dans le temps et dans un espace très calculé pour une émission de 56 minutes. Les emplacements étaient déjà choisis au montage en tenant compte des moments forts et des moments faibles, des moments en mouvements et des moments fixes, des moments très riches en sons et des moments plus légers.

Par le positionnement des sons, il faut toujours surprendre et étonner l'auditeur en son multicanal, en pensant toujours que l'auditeur est au centre du voyage, au centre des enceintes.

Anecdote

Robert Arnaut m'a dit « tu vas voir, à la chute de l'arbre, tous les oiseaux s'arrêtent de chanter, c'est le grand silence » Mais pas du tout ! Ils sont habitués au bruit de la tronçonneuse, car le forestier a une concession de 27 km² et coupe trois okoumés de 40 mètres par hectare. Tout l'isorel qui arrive en France vient du Gabon. Pour la prise de son de l'abattage, j'étais à la base de l'arbre avec le couple ORTF avant et Jacques Charreaux à 40 mètres plus loin à la chute, avec le couple ORTF arrière. Au mixage, grande discussion, car certains voulaient garder la vraie prise de son. Celle-ci restituait le son de la scie devant mais aussi à l'arrière, ce qui faussait complètement la mise en scène. Nous avons donc choisi de mettre le couple avant que sur le frontal avec une ambiance 4.0 de forêt fournie et, lors de la chute, on a basculé le couple vers l'arrière, sur le couple placé à la chute de l'arbre avec une ambiance 4.0 de forêt très calme. Pour certains, cela s'appelle tricher, pour d'autre cela s'appelle créer un décor sonore.

Remarques

Comme il est impossible d'enregistrer une pluie en forêt équatoriale car on mouille les microphones, et si on a un parapluie, on entend les gouttes sur le parapluie, j'ai trouvé une solution. J'ai enregistré à l'abri d'un petit toit du cabanon où nous dormions, la pluie tombant sur des grandes feuilles, avec le couple avant de mon système 4x4, le couple arrière captait la résonance de la pluie sur l'abri en bois.

Au mixage on a pris cet enregistrement (quatre pistes) et on a matricé la pluie du couple sur l'avant et une partie entre l'avant et l'arrière. On a pris cette même pluie 30 secondes plus loin et on a fait l'inverse. On l'a mise (sur 4 autres pistes) sur l'arrière en matricant une partie au milieu, entre l'arrière et l'avant. On a mixé les 8 pistes et comme cela on a formé un ciel et il pleut sur l'auditeur. Vous rajoutez un chasseur qui marche dans la forêt en se faisant un passage à coups de machette, et l'auditeur avance dans la forêt, sous la pluie, en ne se rendant pas compte que cette dernière est fixe et tombe toujours sur les mêmes feuilles.

Précisons que, dans le canal Lfe, nous n'avons pas mis volontairement toute l'émission. Nous avons choisi que les sons à fréquences basses : orage, intérieur de camion, intérieur de 4x4, passage d'un train dans un tunnel, etc.

Pour produire « cette émission bio » de 56 minutes, il aura fallu 10 jours de tournage, voyage compris, 24 heures d'enregistrements, 2 semaines de dérushage et montage



Figure 5.117 – Enregistrement des abeilles, © J. Charreaux.

et 4 semaines de mixage. Cette émission a été diffusée le 19 mai 1998 en version Dolby Surround sur France Inter dans l'émission Zinzin. À l'occasion du SATIS qui se déroulait pour la première fois dans la Maison Ronde de Radio France en 1998, le Singe Soleil a été encodé en DTS, avec un tirage numéroté de 200 cd.

Cette émission a obtenu le Prix Jean Thévenot en 1997, prix créé en 1984 à l'initiative des « Chasseurs de sons », émission culte pour les preneurs de son, sous l'égide de Paul Robert et de Dominique Calace de Ferluc.

Équipe

Producteurs Robert Arnaut et Jacques Charreaux, Prise de son et mixage Guy Senaux, Montage et mixage Philippe Carminati, Chargé de réalisation Emmanuel Geoffroy.

Complément Web 5.9

Bonobo : le hippie des Grands Singes, France Inter Interception.

Complément Web 5.10

Maître – maîtresse de ma passion, Les Sonnets de Shakespeare selon Pierre-Jean Jouve, France Culture.

5.9.4 Autres expériences Radio

Reportage France Culture *Interception* par Hervé Déjardin, ingénieur du son à Radio France

France Inter a renouvelé diverses expériences de reportages en 5.1 en 2011 avec son magazine *Interception*. Le premier de la série, « La Transhumance », est ici décrit.

Particularité de cet enregistrement : il ne doit pas ajouter de contraintes. La présence du preneur de son aux côtés du journaliste doit rester discrète. L'enregistrement doit donc se faire de manière souple et fiable.

Le dispositif technique décrit ci-après a été validé lors d'un reportage pour *Interception*, le magazine de la rédaction de France Inter. Un reportage de 45 minutes dans lequel les ambiances tiennent une place importante. Pour valider ce dispositif d'enregistrement

en 5.1, c'est un reportage sur la transhumance qui a été choisi, justement pour la richesse de ses ambiances : moutons, chiens de berger, paysage de montagne...



Figure 5.118 – Tournage sur *La Transhumance*, *Interception France Inter*, © P. Dervieux

Complément Web 5.11

Hervé Déjardin Tournage *Interception*.

5.9.5 Sport

Complément Web 5.12

Application Sport : football (VCF) et tennis (BBC, FTV).

5.9.6 Musique

Complément Web 5.13

Application Musique : film (*Requiem pour une tueuse*) et TVHD (*Passion selon Saint Matthieu*).

Chapitre 6

La postproduction

POINTS ESSENTIELS

- Généralités : monitoring, normes, montage, DAW, consoles.
- La mesure en 5.1.
- Le manager multicanal : comment gérer une piste 5.1 ?
- Bass Management, gestion du Lfe.
- Le pan-pot, gestion du centre et des canaux arrière.
- La réverbération.
- Le traitement dynamique.
- Downmix.
- Upmix.
- Méthodes de mixage et analyses.
- Synthèse binaurale.
- Exploitation des nouveaux formats, mixage objets : Auro-3D, WFS, Dolby Atmos.
- Outils de création sonore et spatialisation interactive.
- Applications : exemples de réalisation de mixages film, musique, habillages de chaînes HD.
- Points de vue de mixeurs film.

La spatialisation en multicanal fait appel aux mêmes règles qu'en stéréophonie, le mixeur réalise la balance des sources (les équilibres), il gère l'espace avec les réverbérations et les ambiances, il travaille leurs intelligibilités, leurs couleurs et affine l'évolution temporelle du mixage par l'automatisation. La seule grande différence consiste à gérer trois fois plus de données (en pistes, en routing, réverbérations et autres effets) et à créer une nouvelle image sonore. Il est très courant d'avoir des sessions de plus de cent pistes pour le film, la musique étant mixée en général à part, d'où la nécessité d'avoir une bonne organisation, de simplifier la postproduction par différentes étapes de montage et de mixage, et d'avoir une automatisation la plus souple possible. En effet, la multiplication des directions du 5.1 complexifie la gestion des sources, et une automatisation comme celle gérée par le Pro Tools, ou par d'autres logiciels équivalents, vont bien simplifier la démarche et faciliter le travail.

D'une façon générale, les outils 5.1 présentent quelques subtilités comme la divergence centrale, le bass management, le manager multicanal, la mesure des six canaux... Nous abordons dans ce chapitre ces différents outils indispensables à la gestion d'un mixage, des méthodes de travail sont exposées, des exemples de mixages sont analysés et enrichies par un retour d'expériences de mixeurs film professionnels.

6.1 Généralités

6.1.1 *Monitoring*

Le point de départ d'un mixage repose toujours sur le monitoring et ce dernier, en multicanal, n'est pas simple. Le chapitre 2 a introduit les différences fondamentales entre l'écoute cinéma et l'écoute ITU pour la TVHD, la radio, les supports DVD et Blu-ray. Rappelons que nous avons un niveau de référence de 85 dBC SPL en auditorium cinéma, et 79 dBC SPL pour une régie broadcast. Il n'y a pas vraiment de niveau de référence pour la musique ou la radio. L'écoute cinéma est sans bass management, en revanche, l'écoute broadcast nécessite un bass management, ainsi que tout produit diffusé sur une installation home-cinéma (DVD, SACD, Blu-ray...). Certains mixeurs en TVHD travaillent avec une écoute réglée avec bass management, mais il est intéressant d'avoir deux presets d'écoute, un avec et un sans, afin de mieux maîtriser le grave. En effet, mixer sans bass management conduirait forcément à des problèmes de compatibilité dans le grave, en général on obtient trop de basses fréquences ce qui nuit à l'intelligibilité du mixage. D'autre part, avoir un preset sans bass management permet de bien identifier le grave issu du canal Lfe seul et celui provenant de la sommation des canaux principaux. Il est aussi vivement souhaitable d'avoir un contrôleur de monitoring avec la fonction Solo, Mute sur les différents canaux, le calibrage d'un niveau de référence, la possibilité de « downmixer » en stéréo et en mono depuis le format 5.1 (certains contrôleurs proposent un downmix LtRt). Les autres fonctions telles que la possibilité d'activer un bass management avec une fréquence réglable, le réglage du gain du canal Lfe (0 dB ou +10 dB) et le recours possible au filtrage du canal Lfe à 120 Hz sont aussi très importantes, elles sont souvent assurées

par le processeur d'écoute. L'intégration d'encodeurs et de décodeurs (Dolby, DTS, simulateur de metadata DP570) peut aussi être associée au contrôleur de monitoring, certains modèles le proposent. Une sortie supplémentaire pour le metering 5.1 est à prévoir lorsqu'on utilise des modèles hardware.

La plupart des contrôleurs de monitoring sont intégrés dans la console de mixage, qu'elle soit analogique, numérique ou surface de contrôle. Mais il existe cependant des contrôleurs de monitoring 5.1 ou 7.1 indépendants, lesquels imposent un ordre des canaux, généralement ITU ou DTS. D'autre part les cartes son proposent progressivement un monitoring 5.1 ou 7.1 contrôlable par le DAW ou bien une application dédiée à la carte. Enfin, la fonction monitoring 5.1 peut proposer une écoute des stems séparés, suivant le DAW utilisé ou bien à l'aide d'un plug-in dédié au monitoring.

6.1.2 Normes sur l'ordre des canaux

Il existe différentes normes sur l'ordre des canaux, en voici une synthèse :

NORMES SUR L'ORDRE DES CANAUX

- ITU/SMPTE ou Dolby : L R C Lfe Ls Rs.
- Film : L C R Ls Rs Lfe.
- DTS : L R Ls Rs C Lfe.
- QuickTime : C L R Ls Rs Lfe.
- 7.1 SDDS : L Lc C Rc R Ls Rs Lfe.
- 7.1 HD : L C R Ls Rs Lb Rb Lfe.

Ces normes sont souvent liées à un environnement de travail, le cinéma, la TVHD, la Radio, le support, à un choix de codage, le Dolby, le DTS, le SDDS, l'AAC, à l'utilisation de certains logiciels ou périphériques dont l'ordre des canaux est imposé. Une certaine gymnastique d'esprit permettra au monteur et au mixeur de bien gérer le routing des bus dans les consoles, les DAW, les périphériques, les supports, les fichiers entrelacés, le monitoring. Il est en effet très rare de pouvoir garder le même ordre sur toute la chaîne de production.

6.1.3 Montage son en multicanal

Le montage constitue une étape où convergent différents formats de sons : mono, stéréo, LCR, quad, 5.0, 5.1. À cette étape, les sons seront répartis sur des pistes envoyées dans des stems séparés, avec une ébauche de spatialisation 5.1.

Le monteur travaille en salle de montage équipée en 5.1, il reçoit le fichier OMF ou AAF du monteur image et les rushes de la prise de son, issus d'enregistreurs Cantar, Deva, Sound Device et SXR4 (en deuxième enregistreur souvent pour des appoints). Les prises de son stéréo sont devenues quasi systématiques, elles sont synchrones grâce à ces enregistreurs multipistes, on trouve quelques fois des formats B Soundfield.

Le monteur organise sa session Pro Tools ou Pyramix avec des stems de monitoring de différents formats qu'il utilise pour son propre travail et livre ensuite sa ou ses sessions Pro Tools au mixeur avec des stems déjà regroupés, ces derniers seront réutilisés au mixage.

Le contenu s'organise sous forme de sons **directs** (synchrones), généralement les **dialogues**, les **post-synchro**, les **bruitages**, les **ambiances**, les **FX ou effets** et la **musique**.

Les sons **directs** sont composés de perches mono et HF, de sons stéréo synchrones parfois (couple MS, couple AB) sous forme de plans proches ou plans lointains. Le tout sera regroupé dans un montage d'environ 12 à 16 pistes mono et de 3 ou 4 pistes stéréo. Les sons VI sont constitués d'environ 4 pistes, des sons de déplacements, des bruits de portes ou de voitures qui permettent de garder la synchro pour la version internationale. Les directs peuvent être travaillés dans une session séparée dans le cas où il y a un monteur parole distinct et/ou un premix paroles. Ils peuvent aussi être intégrés à la session complète selon la méthode du monteur son. Il livre alors des pistes mono et stéréo séparées que le mixeur positionnera dans l'espace. En revanche, deux bus de monitoring DIALOGUES et VI sont utilisés pour faciliter le monitoring durant le montage.

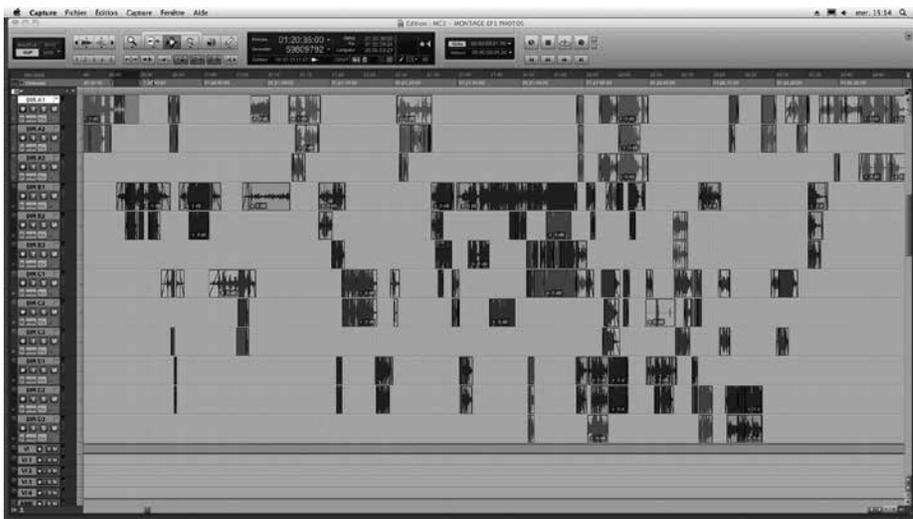


Figure 6.1 – Montage des directs et VI, © C. Richard. En couleur sur dunod.com

La **post-synchro** est composée de sons mono, et livrée sous formes de pistes mono séparées.

Les **bruitages** sont constitués de 12 à 16 pistes, au format mono, stéréo ou LCR quelques fois. Ce sont les présences, les pas et manipulations enregistrées en auditorium.

Les **ambiances** et **FX** sont organisés en stems au format 5.0 ou 5.1 (pour les FX), souvent décomposés par familles FX A, FX B, FX C, FX D... Ces stems sont constitués dans

l'automation, par routing, de telle sorte que le mixeur peut toujours accéder aux éléments séparés. Les ambiances et FX sont des sons additionnels, venant de sonothèques (personnelles ou commerciales), de prises de sons du monteur, et des sons seuls enregistrés au tournage. La quantité de ces sons seuls est très variable, allant de cinq sons à une centaine, ils sont indispensables pour la figuration, les séquences de voitures et doivent raccorder avec l'action du direct. Les ambiances sont généralement au format stéréo, bien que le format multicanal intéresse de plus en plus les monteurs. Quelques séquences spécifiques avec mouvements sont spatialisées. Quelques séquences spécifiques avec mouvements sont spatialisées dès cette étape.

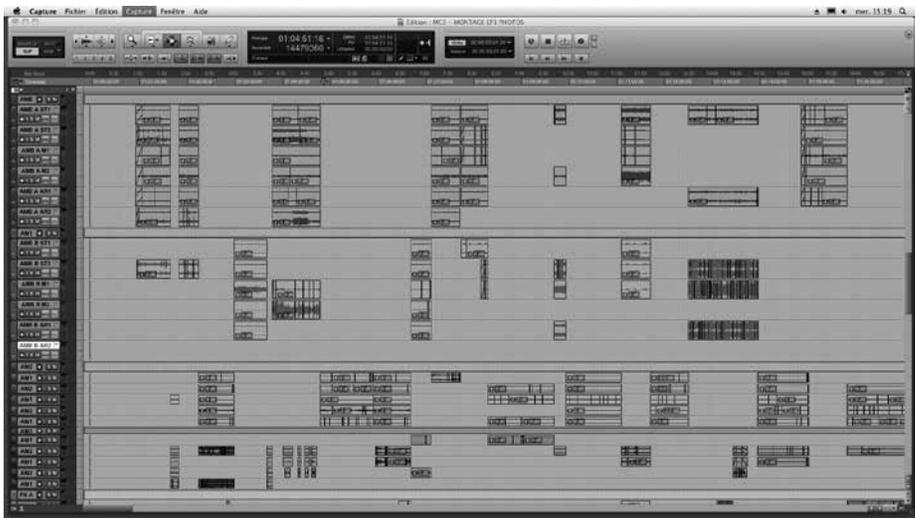


Figure 6.2 – Montage des ambiances, © C. Richard. En couleur sur dunod.com

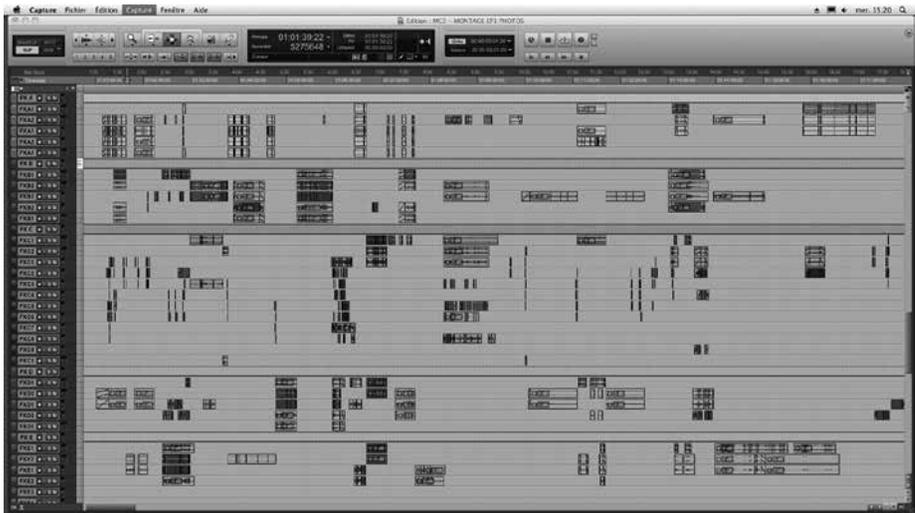


Figure 6.3 – Montage des FX, © C. Richard. En couleur sur dunod.com

La **musique** pendant le montage est constituée de musiques témoins, placées par le monteur image. Mais le monteur son récupère souvent des maquettes du compositeur, de qualité variable, celle-ci sont de bonnes indications de composition, ce qui évite les surprises. Au début du mixage, le monteur son récupère ensuite les musiques mixées en 5.1 et les monte comme il gère la réception de tout autre élément de postproduction. La musique est généralement constituée d'un full mix et de 3 stems majeurs ou plus séparés qui peuvent faciliter le mixage de certaines séquences très denses. Il existe donc au final un stem **musique** au format 5.1, composé de différents stems.

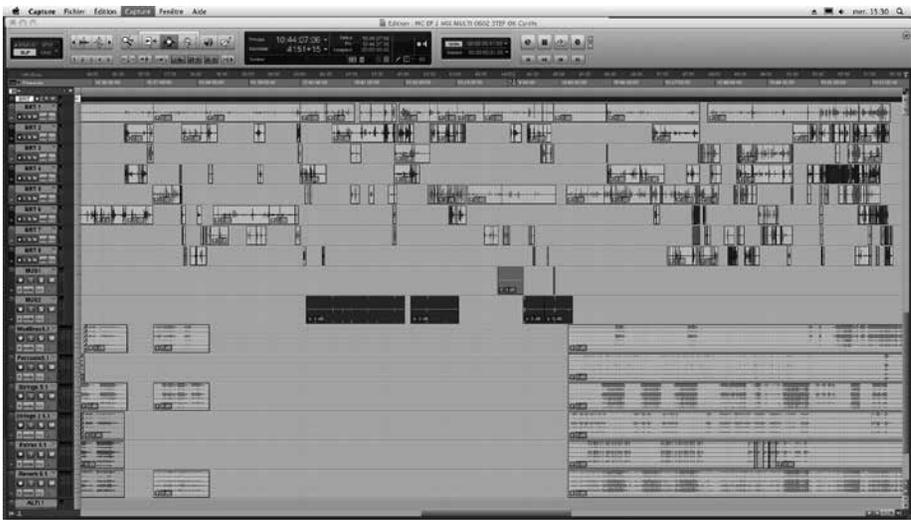


Figure 6.4 – Montage des bruitages et musiques, © C. Richard. En couleur sur dunod.com

Le monteur a recours à l'utilisation de réverbérations sur quelques éléments secondaires, par exemple des sons de portes lointaines, elles sont alors « processées » et séparées du son direct. Leurs formats sont LCR ou 5.0. Les réverbérations à convolution de type Altiverb sont généralement utilisées, elles permettent aussi d'utiliser des réponses impulsionnelles faites dans certains lieux.

Les chaînes de télévision HD comme Canal Plus demandent un 5.1 pour toutes les fictions, c'est aussi le cas des téléfilms pour d'autres chaînes HD. La série « Maison Close » est par exemple une fiction post-produite en 5.1, diffusée sur Canal Plus. La série est constituée de 8 épisodes de 52 min. Le montage de chaque épisode se fait sur 19 jours, directs et montage son, ce qui est supérieur à la moyenne pour cette catégorie de programmes en France. Ensuite, chaque épisode est mixé sur 6 jours. Un PAD Dolby E 5.1 pour la chaîne est fabriqué ainsi qu'une édition sur DVD. Le format multicanal apporte sans aucun doute une meilleure intelligibilité du signal, mais le monteur doit systématiquement penser, pendant le montage 5.1 du film, la version stéréo, ce qui suppose de fournir des sons de façon intelligente. Il procède très souvent à une réduction stéréo pour se rendre compte du résultat.

6.1.4 DAW

La gestion du multicanal dans une station audionumérique est tout d'abord conditionnée par la compatibilité des formats, les stations sont généralement toutes compatibles 5.1, certaines acceptent les formats 6.1, 7.1, ou 10.2. La compatibilité avec les formats supérieurs est rare mais certains logiciels de création sonore comme Reaper les gèrent. Il est courant de créer des pistes en 5.1, ceci facilite le montage et le mixage. Le réglage de l'ordre des canaux affichés sur une piste 5.1 n'est pas toujours possible, mais les chemins de sortie avec l'interface audio peuvent se régler afin d'adapter l'ordre aux différentes normes. L'import de fichiers entrelacés multicanal est bien géré par les logiciels qui sont compatibles avec le format entrelacé quelque soit l'ordre d'entrelacement, le logiciel reconnaît les canaux du multicanal, autrement, si ce n'est pas le cas, on doit respecter pour plus de commodité l'ordre des canaux du logiciel. Le logiciel doit pouvoir exporter des fichiers multicanal entrelacés, compatibles avec la norme demandée, mais dans ce cas également, l'ordre des canaux n'est pas toujours paramétrable. Le pan-pot multicanal est un outil primordial, il conditionne la spatialisation des sources, avec une palette d'options plus ou moins complètes et abouties (voir outil pan-pot) et intervient de ce fait souvent dans le choix du logiciel. Le monitoring multicanal intégré dans le logiciel est une option pratique et se présente sous forme d'un menu à part qui gère généralement le niveau d'écoute, les différentes options de downmix et quelques fois, la structure de mixage par stems est présente.

Complément Web 6.1

DAW et multicanal.

6.1.5 Consoles

La plupart des consoles de mixage gèrent aujourd'hui le multicanal, en 5.1, et en 7.1 pour certaines, ce qui suppose l'intégration d'un pan-pot dédié permettant d'acheminer le signal vers un bus multicanal et d'une section monitoring multicanal adaptée, comme vu précédemment dans les stations audionumériques.

Un autre point important est la gestion de la stéréo pendant le mixage 5.1 : la Duality chez SSL permet par exemple d'alimenter un départ stéréo issu du master bus 5.1 par downmix stéréo LoRo (voir définition section 6.9). Le mixeur peut donc mixer en 5.1 et il dispose en parallèle d'un bus de downmix stéréo LoRo paramétrable.

Une autre particularité de l'implémentation du multicanal dans les consoles, est la possibilité de gérer une piste au format 5.1 ou 7.1, contrôlée par un seul fader, ce qui simplifie le mixage surtout lorsque la session comporte beaucoup de pistes audio. Il est alors généralement possible de dégrouper les 6 canaux LCRLsRsLfe de la piste (mode « Spill » chez SSL C300, mode « Reveal » chez Lawo) et d'accéder individuellement à leurs faders. La particularité du mixage multicanal cinéma étant de travailler par

stems 5.1, l'implémentation de cette structure par stems dans la console simplifie alors considérablement le travail. La console System 5 Fusion par exemple offre une grande souplesse dans la gestion des stems, la surface de contrôle Nuage via Nuendo aussi.

6.2 La mesure

L'outil de mesure est bien abouti lorsqu'il met en évidence une représentation graphique en cohérence avec la perception de l'image sonore. En stéréo, le niveau électrique du signal est analysé par le crêtemètre qui mesure les quasi-crêtes (DIN 45406 QPPM 10 ms d'intégration) ou les crêtes instantanées (PPM fast), ou par le vumètre, plus en rapport avec le niveau perçu par l'oreille (300 ms d'intégration). Le corrélateur de phase donne l'information de corrélation des deux canaux. Nous développerons dans le chapitre 8 la nouvelle norme de mesure du loudness.

La mesure en multicanal intègre les six canaux du bus de mixage 5.1 et souvent la version stéréo (LtRt ou LoRo) pour le broadcast, soit les huit canaux d'un programme 5.1+2, tel qu'on le nomme dans le broadcast. Se pose alors la question de représentation graphique du niveau des six canaux. L'outil de base comporte en général six indicateurs de niveau (crêtes, VU), mais une représentation spatiale suivant la disposition des enceintes, avec le niveau de chaque bus est complémentaire. Généralement, chacun des segments des canaux principaux sont reliés entre eux pour former une forme de patateïde. Cette forme donne une idée assez précise de la répartition énergétique des canaux, et contient, par un code couleur, l'information de phase. L'utilisateur sait par la lecture de la forme s'il remplit les canaux de façon homogène ou non, et peut vérifier la construction du centre de l'image sonore par exemple.



Figure 6.5 – Représentation Surroundscope dans Pro Tools, © Avid.

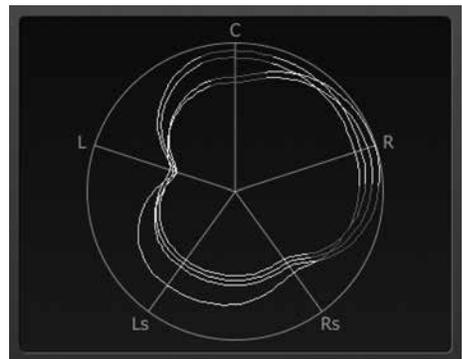


Figure 6.6 – Mesure dans Pyramix, © Merging Technologies.

La phase est représentée en général par un corrélateur que l'on règle en indiquant les deux canaux du multicanal mesurés. C'est le mode de mesure proposé par Surroundscope dans Pro Tools. Master Penguin Surround Meter indique toutes les informations de phase entre les canaux, elles sont au nombre de 10, tandis que Dk audio code l'information dans la représentation graphique Jellyfish en rouge, pour indiquer l'opposition de phase.

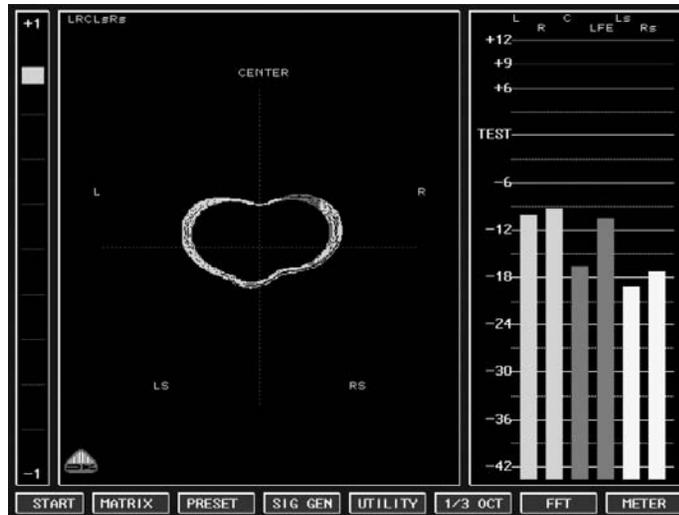


Figure 6.7 – Mesure Dk Audio, Jellyfish, © Dk Audio.

Comme évoqué dans le chapitre 3, il existe deux types de centres en multicanal, le centre fantôme issu des canaux L et R, et le hard center représenté par le bus C qui alimente l'enceinte centrale. Ces deux types de centres se combinent toujours, d'où la nécessité pour les outils de mesure avancés, d'afficher les niveaux des deux sortes. La mesure RTW suit ce principe, nous obtenons donc, dans l'espace frontal, deux représentations, celle de l'image fantôme LR, et celle de l'image frontale LCR.

Il est aussi intéressant de faire figurer le barycentre de l'image fantôme LR et celui de l'image frontale LCR (LC et CR). Cela s'appelle chez RTW le PSI (Phantom Source Indicator), il est représenté par la croix blanche et indique le barycentre de l'image fantôme, la localisation moyenne de la source. Nous pouvons lire aussi l'information de largeur du secteur

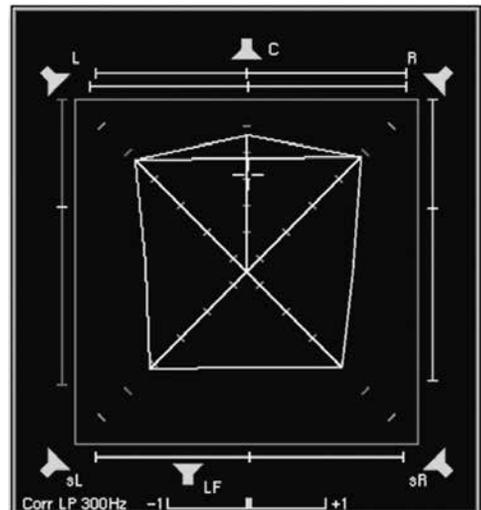


Figure 6.8 – Mesure RTW, © RTW.

(largeur du sujet), elle est représentée par l'étendu de la base du segment. Sa couleur associée, donne alors l'information de corrélation des canaux :

- le vert indique une corrélation entre +0,25 et +1 ;
- le jaune une corrélation entre -0,25 et +0,25 ;
- le rouge, une valeur entre -1 et -0,25.

Le PSI se retrouve sur chacune des paires adjacentes du surround, L/Ls, Ls/Rs, R/Rs. Lorsque nous considérons l'ensemble des canaux, nous retrouvons la patatoïde, chez RTW plutôt en forme de carré, cela s'appelle le TVI (Total Volume Indicator). Il donne l'étendu spatiale des canaux L R Ls Rs et intègre le centre C à part, pour une meilleure lecture en relation avec les canaux L et R. La forme du carré varie suivant la corrélation des canaux, les segments sont incurvés vers l'intérieur si c'est en opposition de phase, droits entre -0,25 et +0,25 et orientés vers l'extérieur pour une corrélation entre +0,25 et +1. Il existe une information très pertinente, le barycentre global de l'image, appelé DMI, Dominance Vector, il est représenté par la croix blanche centrale. Enfin, RTW a gradué l'échelle des 6 secteurs de la patatoïde, les positions rouges indiquent la valeur du niveau d'écoute de référence (85 dBC SPL ou 79 dBC SPL). Nous trouvons aussi un corrélateur de phase pour les canaux arrière Ls/Rs.

D'autres outils, notamment Nebula Surround Scope (Pure Analyzer) développé par Flux, représente la corrélation des canaux adjacents, il traduit par les petites valeurs (petites longueurs des arcs de cercles entre canaux adjacents) la sensation d'espace, l'enveloppement, ou bien la fermeture d'espace, soit l'immersion, par les grandes valeurs.

Nous verrons par la suite qu'un outil de mesure intègre maintenant la mesure true peak pour les crêtes mais aussi le loudness conformément à la norme EBU R128.

MESURE DU SIGNAL MULTICANAL

- Représentation graphique spatiale des niveaux des canaux.
- 8 indicateurs de niveau crête, VU, true peak (program 5.1+2).
- Mesure du loudness (gate R128, Dialog Level).
- Représentation de la phase par couleurs, par corrélateurs entre 2 canaux.
- Représentation des barycentres des images fantôme.
- PSI (phantom source indicator RTW).
- Indication de la largeur des images fantôme.
- Barycentre global DMI (dominance vector RTW).
- Indication du niveau d'écoute de référence 85 dBC SPL ou 79 dBC SPL.
- Analyse de la décorrélation des canaux de l'image sonore pour la mesure de l'enveloppement.

6.3 Manager multicanal

L'utilisation de pistes au format multicanal implique des outils de gestion d'espace multicanal, on peut les appeler des manager multicanal. Ils permettent de gérer l'équilibre en niveau des différents canaux, de modifier leurs phases et d'ajuster les délais. La rotation d'espace peut être utile pour modifier la polarisation de l'image, on peut alors décaler l'image sur la gauche ou vers la droite ou bien renverser complètement l'image vers l'arrière. Certains outils analysent la décorrélation de l'image sonore, il peut être intéressant de recorréler l'image, on utilise pour cela la divergence ou width/depth (largeur/profondeur), à ne pas confondre avec la divergence centrale. La divergence agit sur les différents secteurs de l'image (frontal, arrière, latéral), elle réinjecte par différences d'intensité une partie des canaux frontaux par exemple, vers les canaux arrière de telle sorte que l'on recorrèle les différents bus du mixage. On pourrait imaginer devoir recorréler un espace issu d'une prise de son multicanal un peu trop large, comme un système DORTF, ou bien vouloir réinjecter une partie des canaux arrière vers l'avant pour améliorer la compatibilité Dolby Surround du mixage.

Le changement de routing des différents canaux peut être pratique dans certains cas où l'ordre des canaux d'une réverbération ou d'un effet n'est pas conforme à celui de la session de mixage.

Le manager est aussi un outil de monitoring, comportant les fonctions solo, mute, ce qui permet d'isoler dans notre session de mixage un bus issu d'une piste ou d'un master 5.1. Certains outils manager sont capables de faire du bass management, l'utilisateur peut alors régler la fréquence du bass management et les pentes des filtres. Dans ce cas, l'outil permet d'ajuster le niveau d'écoute, avec des valeurs mémorisées, et de générer du bruit rose afin de calibrer l'écoute.

Enfin, l'outil peut permettre de « downmixer » vers des formats inférieurs stéréo et mono, avec ajustement des coefficients de downmix.

Dans Pro Tools, il n'existe pas aujourd'hui de manager par défaut, nous pouvons utiliser le plug-in multi-mono trim pour gérer les niveaux des différents canaux, autrement il faut installer des plug-ins comme le manager présent dans le bundle Waves Surround Tools. Ce Manager M360 est compatible avec différentes configurations pour l'écoute multicanal 5.1 (possibilité d'ajuster les angles entre les enceintes), il offre les fonctions suivantes :

- solo, mute, opposition de phase, niveaux des canaux ;
- bass management (« send to Sub », ou « get Lfe » appelé aussi « distribute » ou « extract ») ;
- ajustement rapide des niveaux du centre et des canaux arrière par pas de 3 dB ;

- adaptation du niveau du canal Lfe (0 dB, -10 dB, +10 dB) suivant la norme utilisée ;
- rotation de l'image sonore, différents modes de calcul (pair-wise, triple-wise) ;
- divergence vers les autres canaux (« width »), différents modes de calcul (balanced, front-stage, F/R stage, focus, mono diverge).

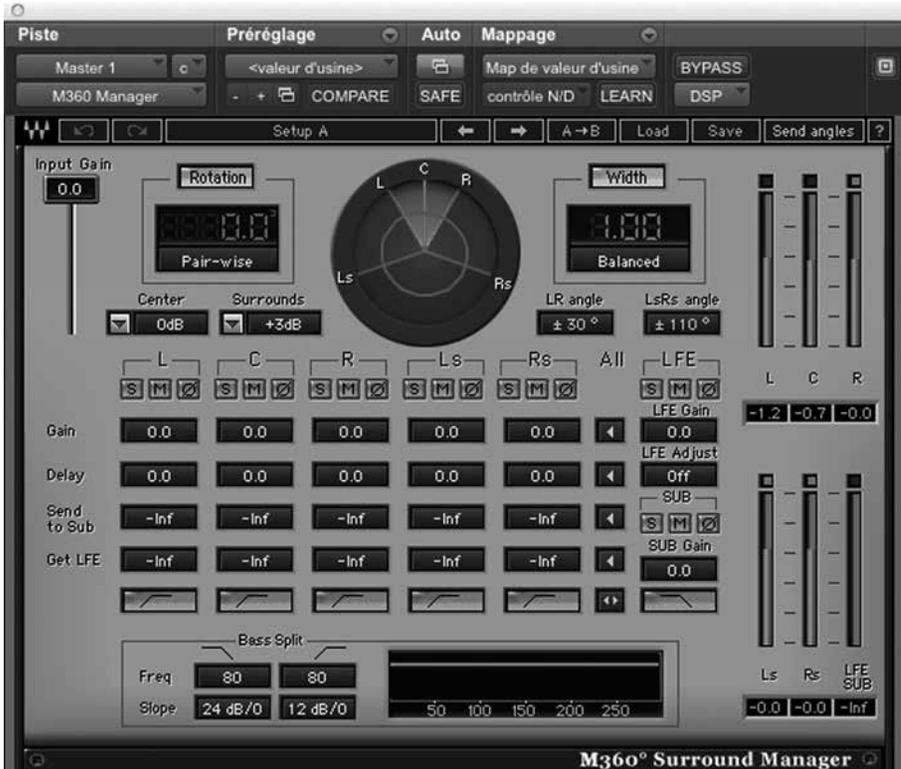


Figure 6.9 – Le manager Waves M360, © Waves.

Pyramix propose le « global channel routing », c'est une sorte de grille de routing basique mais très pratique. Il gère le niveau des différents canaux et permet aussi de changer le routing interne des canaux de la piste de mixage multicanal.

Le Toolbox est l'algorithme de management multicanal chez TC Electronic, présent dans la M6000. Cet outil est un manager complet, il propose toutes les fonctions citées auparavant. On y retrouve donc les fonctions de gestion du monitoring, avec trois niveaux d'écoute programmables, les solo, les mute, l'opposition de phase, le downmix stéréo avec limiteur, la calibration d'écoute, et le bass management.

La console Studer Vista 9 gère des pistes au format multicanal et offre les fonctions de Manager suivantes :

- Center level : gestion du niveau du canal C ;
- Center : gestion de la divergence centrale en % ;
- Front LR, Rear LR : répartition gauche – droite pour le frontal et l’arrière ;
- F-R : répartition frontal – arrière ;
- Front width, Rear width : ajustement de la largeur frontale et arrière en %, de 0 % (mono) à 200 % (super stéréo) ;
- F-R depth : ajustement de la profondeur frontale – arrière en %, de 0 % (pas de profondeur) à 200 % (super profondeur) ;
- Pan arround : rotation de l’image multicanal en degrés, $\pm 180^\circ$;
- Lfe level : gestion du niveau du canal Lfe.

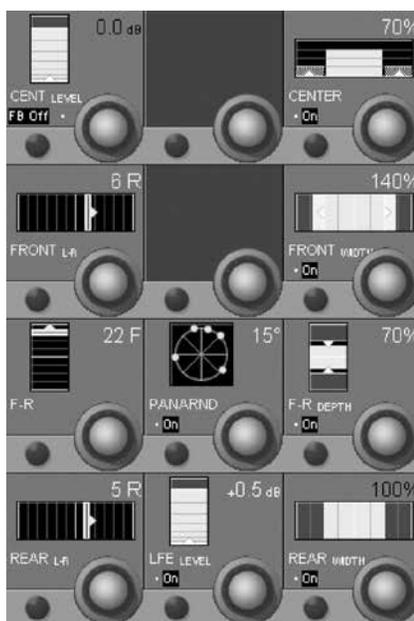


Figure 6.10 – Manager Studer Vista9, © Studer.

La possibilité de représenter sous forme de six faders les différents canaux d’une piste multicanal 5.1 (mode « spill » chez Euphonix ou « reveal » chez Lawo par exemple) facilite la gestion des niveaux.

L’hyperpan de Lawo permet de changer la répartition des différents canaux de la piste multicanal graphiquement, afin de resserrer un canal relativement aux autres, comme représenté sur le schéma ci-dessous. On peut gérer une piste multicanal comme un objet, et déplacer l’objet vers le frontal par exemple. Nuendo6 reprend aussi ce type de gestion objet avec le Surround Panner V5 qui fait office de manager de piste multicanal.

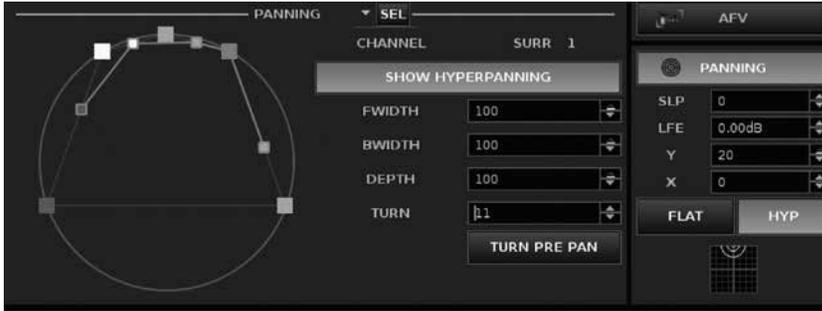


Figure 6.11 – Hyperpan Lawo, © Lawo.

MANAGER MULTICANAL

- Gestion des niveaux des canaux, délais, filtrage, opposition de phase, représentation des canaux d'une piste 5.1 ou 7.1 sous forme de faders.
- Monitoring : solo, mute des canaux.
- Rotation de l'image sonore multicanal.
- Divergence centrale.
- Recorrélation des canaux.
- Divergence vers les autres canaux ajustement des largeurs de chacun des secteurs (largeurs frontale et arrière, profondeur).
- Changement de routing des canaux.
- Bass management.
- Downmix avec limiteur ou non.

6.4 Gestion du Lfe, bass management

Le Lfe, low frequency effect, comme son nom l'indique définit le canal d'effet basse fréquence, sa bande de fréquence s'étend généralement de 20 Hz à 120 Hz. C'est un canal issu du mixage cinéma qui a trouvé par la suite son intérêt dans les autres applications TVHD, musique, radio. C'est un canal associé au subwoofer, qui interagit sur le corps de l'auditeur par les vibrations qu'il génère. Rappelons que son niveau est augmenté de +10 dB, dans la chaîne de monitoring des installations à l'image (cinéma, TVHD, DVD, Blu-ray). Il s'ajoute, dans la partie monitoring, à la sommation basse fréquence des canaux principaux, lorsque l'écoute est « bass managée ». Le filtrage du bass management est alors indépendant du filtrage du canal Lfe. Le canal Lfe se filtre en général à 120 Hz dans l'étape de mixage. Il peut être aussi filtré par un contrôleur de monitoring, par un encodeur, par le simulateur de metadata Dolby DP570 ou bien par un processeur d'écoute.

Pour le film, le Lfe est un canal à part, que l'on va solliciter pour renforcer les séquences d'action, la partie design sonore, les effets, il donne plus d'impact aux sons. On l'utilise pour la gestion des fréquences infrabasses, mais les graves du mixage ne doivent jamais reposer sur ce canal. Autrement dit, le mixage doit garder son spectre équilibré lorsque le Lfe est coupé. L'utilisation du canal Lfe en musique par exemple est beaucoup plus continue que dans le film. On l'associe en général à des instruments graves si bien qu'à chaque intervention, le Lfe est sollicité.

Le bass management est le procédé de redistribution du grave des canaux principaux vers le SUB, on le trouve dans la partie monitoring 5.1. On parle aussi de bass management lorsqu'il est question de gérer le grave d'un mixage 5.1. On peut alors redistribuer de la même façon le grave des canaux principaux du mixage vers le canal Lfe, afin de renforcer ce canal. Une application peut être l'optimisation du mixage afin de le rendre compatible avec l'écoute bass management (en dernier recours). Ce peut être aussi le moyen de générer un canal Lfe à partir d'un mixage 5.0. Ce mode de bass management est appelé « Mode Extract ». L'autre mode, appelé « Mode Distribute », permet de redistribuer le canal Lfe vers les canaux principaux, on déplace en quelque sorte le grave du Lfe vers les canaux principaux. On peut ainsi passer d'un format 5.1 à un format 5.0. On retrouve ces deux modes dans le Toolbox de la M6000.

BASS MANAGEMENT

Mode extract.

- Passage d'un 5.1 vers un 5.1 : le grave des canaux principaux est extrait et renvoyé vers le canal Lfe.
- Passage d'un 5.0 vers un 5.1 : fabrication d'un canal Lfe.

Mode distribute.

- Passage d'un 5.1 vers un 5.1 : le canal Lfe est redistribué vers les canaux principaux du mixage.
- Passage d'un 5.1 vers un 5.0 : le canal Lfe est supprimé et redistribué vers les canaux principaux.

CANAL LFE, QUEL FILTRAGE ?

- Bande passante du canal Lfe, généralement : 20 Hz – 120 Hz.
- Suivant les formats d'encodage et les applications voici ce qui est préconisé pour le filtrage du Lfe :

Dolby

- Dolby Digital : 120 Hz.
- Dolby Surround EX 6.1 : 120 Hz.
- Dolby True HD 7.1 : pleine bande.

DTS

- DTS Cinéma APT- X100 5.1, DTS ES Matrix 6.1.
- L, C, R : 20 Hz – 20 kHz.
- Ls, Rs : 80 Hz – 20 kHz, le contenu 20 Hz – 80 Hz est envoyé vers le Lfe à l'encodage.
- Subwoofer : 20 Hz – 80 Hz, canal Lfe filtré à 80 Hz et contenu Ls + Rs inférieur à 80 Hz dirigé vers le Lfe.
- DTS vidéo Coherent Acoustic.
- DTS Digital Surround : 120 Hz.
- DTS ES 6.1 : 120 Hz.
- DTS HD, DTS Master : pleine bande, filtre passe bas 100 Hz 60 dB/oct au décodage.
- MLP (DVD audio) : pleine bande.
- DSD/DXD (SACD) : pleine bande.
- AAC (MPEG2/MPEG4) : variable jusqu'à 1 kHz.
- MPEG2 BC : variable.
- SDDS : variable jusqu'à 330 Hz.

On utilise souvent le filtre Waves Lfe 360 qui peut s'insérer sur une piste 5.1, il n'agit que sur le canal Lfe, avec une pente à 60 dB par octave.

LFE ET BASS MANAGEMENT : UN COMPROMIS

Généralement, dans le cas où une même source alimente le canal Lfe et les canaux principaux, pour bien maîtriser le niveau de grave provenant du canal Lfe et celui issu du bass management, il est pratique dans Pro Tools de dupliquer la piste. L'une alimente les canaux principaux et génère le grave par bass management, l'autre est routée directement vers le canal Lfe. Ainsi, le travail des deux types de grave est indépendant et on peut aisément gérer le filtre passe haut sur la piste alimentant les canaux principaux et le niveau envoyé vers le canal Lfe

6.5 Pan-Pot, gestion du centre et des canaux arrière

6.5.1 Les lois de pan

◆ Pan-pot d'amplitude stéréo

La loi de pan d'amplitude stéréophonique est définie par une courbe présentant une atténuation de -3 dB sur les deux canaux gauche et droite, pour un signal positionné au pan-pot au centre de l'image stéréophonique. La source, « pan-potée » d'une voie à l'autre, garde alors un niveau constant. Il existe des lois de pan-pot à $-4,5$ dB et -6 dB, qui améliorent la compatibilité mono, mais ces lois font apparaître une perte de niveau pour un signal placé au centre de l'image. Ces lois travaillent en différence d'amplitude.

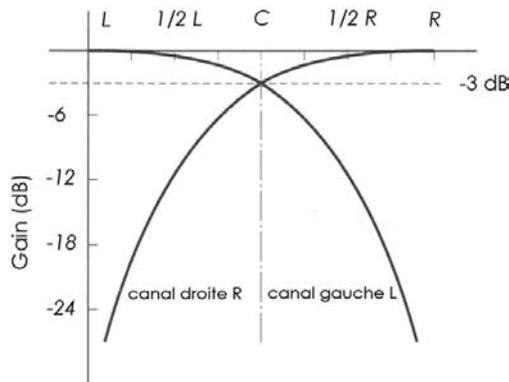


Figure 6.12 – Loi de pan stéréophonique à -3 dB, © B. Périaux d'après F. Rumsey.

◆ Amplitude, VBAP

La répartition d'un signal sur un nombre de haut-parleurs supérieur à deux présente quelques difficultés psychoacoustiques dont la distribution énergétique du signal, la précision de localisation des sources fantômes, l'écoute en dehors de l'axe central et les détimbrages possibles. Michael Gerzon, connu pour son travail de recherche sur le son ambisonique, définit une loi de pan de qualité par la bonne continuité de localisation entre les enceintes, sans trou ni focalisation vers une direction particulière.

Il existe différents procédés de pan-pot multicanal, le premier est une extension de la loi de pan-pot d'amplitude définie pour un format stéréo, c'est l'outil que la plupart des mixeurs utilisent dans les consoles et les DAW. On la définit par la loi de pan d'amplitude par paires d'enceintes (« pairwise amplitude panning »). Cette loi d'amplitude à -3 dB s'applique à chaque paire d'enceintes adjacentes, de telle sorte que la puissance est constante lorsque la source est « pan-potée » d'un canal à l'autre. La

perception de niveau reste alors la même. Cette première solution est assez efficace et son implémentation reste simple.

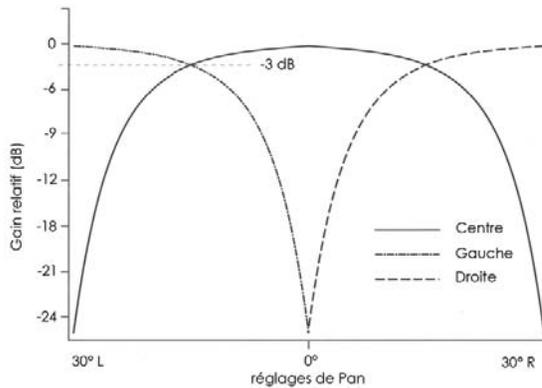


Figure 6.13 – Loi de pan d'amplitude par paires 3 canaux, © B. Périaux d'après F. Rumsey.

Le concept de pan-pot d'amplitude a été étendu à un ensemble de plusieurs haut-parleurs adjacents distribués sur un cercle ou une sphère au centre desquels se trouve l'auditeur, il s'agit du VBAP (Vector Base Amplitude Panning). La technologie VBAP nécessite de connaître par avance la position des haut-parleurs, et utilise les trois enceintes les plus proches de la position voulue de la source. Cette approche, développée par Pulkki en 1997, utilise la composante directionnelle des vecteurs de chacune des 2 ou 3 enceintes les plus proches de la source sonore.

Un autre procédé, le DBAP (Distance-Based Amplitude Panning) définit un pan-pot d'amplitude sur un ensemble de haut-parleurs dont la position n'a pas besoin d'être connue. Les facteurs de gains pour chaque haut-parleur sont établis sur un modèle d'atténuation de distance entre la position de la source sonore et chacune des enceintes.

◆ Pan-pot ambisonique

On trouve aussi des lois de pan reposant sur la théorie ambisonique à différents ordres, elles offrent une bonne stabilité d'image sonore, un placement des enceintes libre mais elles peuvent générer des oppositions de phase pour certaines positions de pan et la séparation des canaux semble assez pauvre. Le Spatialisateur de l'Ircam est un outil qui permet de spatialiser au format ambisonique ordres 1 et 2, en 2D et en 3D. La suite B2X plug-ins (développée par Daniel Courville) permet de spatialiser des sources jusqu'à l'ordre 5, le panner encode une source mono en un format B.

◆ Pan-pot avancé (amplitude, temps, filtrage)

Enfin, il existe des lois de pan qui intègrent des fonctions HRTF de têtes artificielles ou qui simulent des systèmes de prise de son multicanal comportant des différences d'intensité et de temps. Certains algorithmes simulent des fonctions HRTF qui annulent les chemins croisés issus de la diaphonie acoustique. Ainsi, pour un auditeur

placé précisément au sweet spot, une source positionnée sur l'enceinte gauche pourrait être perçue uniquement par l'oreille gauche, sa perception par l'oreille droite serait nettement amoindrie. Le Virtual Surround Panner que l'on trouve dans la console Studer Vista 9 repose sur ce type de fonctionnement. Le VSP intègre différents modèles de microphones, la sphère, le modèle ORTF, le modèle omni AB, utilisant des différences d'intensité, de temps, et de fréquence.

6.5.2 Le type de pan-pot

Aujourd'hui, toutes les consoles et les stations audionumériques sont équipées d'un système de routing multicanal 5.1 voire 7.1, accompagné d'un pan-pot. Le pan-pot multicanal gère la position de la source par un point que l'on place dans l'espace multicanal, lui-même représenté par une disposition d'enceintes sur un cercle ou un carré. Il existe différentes façons de déplacer le point définissant trois types de pan-pot : le pan-pot amplitude par coordonnées, le pan-pot amplitude piloté par joystick ou fenêtre tactile, et le pan-pot évolué utilisant des modèles psychoacoustiques avec des différences d'intensité, de temps et du filtrage.

◆ Pan-pot amplitude par coordonnées

L'utilisateur déplace la source par une commande de coordonnées spatiales.

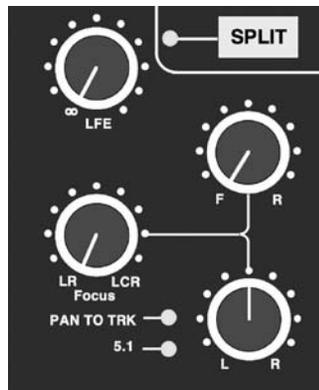


Figure 6.14 – Pan-pot type SSL Duality, © SSL

Ce premier type de pan-pot est celui que l'on rencontre sur la plupart des consoles et des logiciels. Il offre trois potentiomètres avec ou sans représentation graphique du point de localisation : un circulaire **PAN L/R**, pour gérer la répartition gauche/droite des sources, un circulaire LR/LCR appelé aussi **Divergence Centrale** ou Focus (définie par la suite) et un circulaire **PAN F/R** (front/rear) pour ajuster la disposition de la source en profondeur, entre le frontal et l'arrière. On trouve aussi un autre potentiomètre circulaire dédié au niveau de départ du canal **Lfe**. Ces quatre potentiomètres principaux, que l'on rencontre par exemple sur la console SSL Duality, constituent la base du fonctionnement.



Figure 6.15 – Pan-pot Pro Tools, © Avid

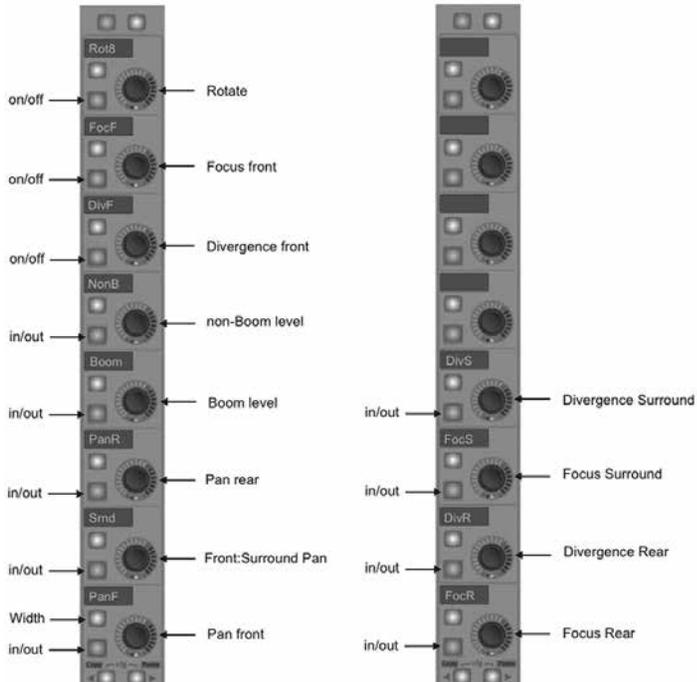


Figure 6.16 – Pan-pot type Euphonix Fusion S5, © Euphonix

On trouve ensuite des potentiomètres complémentaires, comme le « **Pan Front** » et le « **Pan Rear** » (exemple Pro Tools, Euphonix Fusion S5...), ces derniers ajustent indépendamment la répartition gauche/droite dans la zone frontale et dans la zone arrière. Le pan-pot par coordonnées permet de travailler avec des outils traditionnels et compatibles stéréo tels que la répartition gauche/droite. Composé de trois ou quatre potentiomètres, il s'implémente facilement sur les tranches de consoles. Le travail du mouvement des sources est simple : par exemple, le déplacement sur une diagonale avant gauche/arrière droite peut être réalisé avec les potentiomètres « Pan Front », « Pan Rear » et « Pan Front/Rear ». Dans ce cas, les « Pan Front » et « Pan Rear » définissent la diagonale du mouvement et l'utilisateur gère le mouvement avec la répartition frontal/arrière. Généralement, des ajustements de divergence (**divergence vers les autres canaux**) permettent aussi de diverger la source vers les autres canaux adjacents afin de délimiter les espaces.

Certains pan-pot proposent la fonction **rotation** (« Rotate » chez Euphonix, « Pan-around » chez Studer, « Turn » chez Lawo), celle-ci permet de faire tourner une source dans l'espace ou bien de faire subir une rotation complète à l'image sonore quand la piste est au format multicanal, dans ce cas les cinq points tournent ensemble. Certains pan-pot, comme celui de la console System 5 Fusion Euphonix, permettent de gérer indépendamment le niveau du canal Lfe et celui des canaux principaux, pour une piste 5.1 (« Boom level » et « Non-boom level »). On peut ainsi envoyer tout type de format de piste directement vers le canal Lfe.

Les consoles Studer et Lawo proposent les paramètres « **Front width** », « **Back width** » et « **Depth** », lesquels permettent de régler les largeurs frontales, arrière et la profondeur frontal/arrière. Ceci revient à gérer la divergence vers les autres canaux.



Figure 6.17 – Pan-pot type Studer Vista 9 « Multi-format Pan », © Studer

Chez Studer, la divergence centrale est gérée par le paramètre « Center », la divergence vers les autres canaux s'appelle « Diverge ».

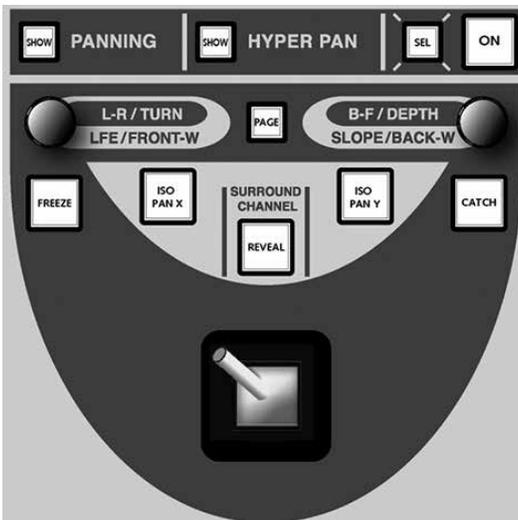


Figure 6.18 – Pan-pot type Lawo, © Lawo

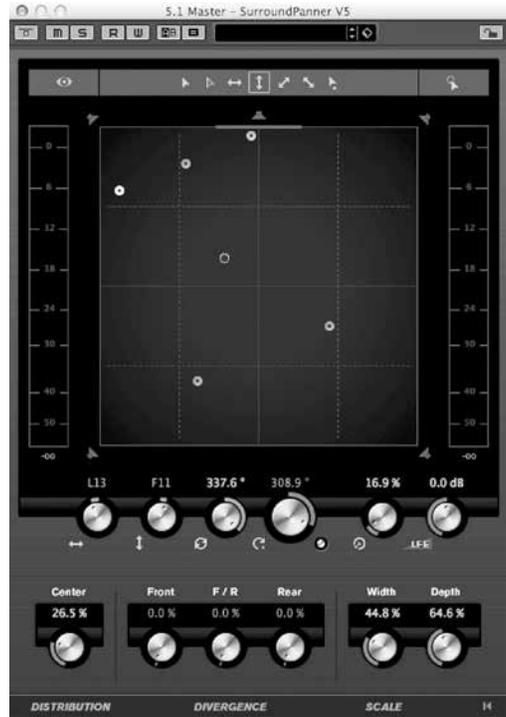


Figure 6.19 – Pan-pot Surround Panner V5, © Steinberg

Le Surround Panner V5 de Nuendo est à la fois un pan-pot et un manager très abouti. On retrouve la divergence centrale avec le paramètre « center » et la divergence vers les autres canaux (frontale, arrière et frontal/arrière) représentée par le terme « divergence ». Le groupe « scale » permet de **recorréler** l'espace pour un format multicanal 5.1, dans la largeur avec le paramètre « width » et dans la profondeur avec « depth » (les cinq sources se déploient plus ou moins dans l'espace). Les sources sont regroupées dans un objet que l'on déplace à l'aide d'une poignée de positionnement dont on peut régler le rayon par rapport au centre du carré, fonction « radius ». La fonction « rotate signal » permet de faire tourner les cinq sources autour de la poignée de positionnement. On peut restreindre les mouvements sur les diagonales, le latéral ou dans la profondeur, on peut couper les enceintes une par une, faire tourner l'objet constitué des cinq sources autour du point central grâce à la fonction « orbit center ». Il existe aussi une fonction contrechamp (« counter shot ») appréciable donnant la position inversée de l'objet à 180°.

◆ Pan-pot amplitude piloté par joystick ou fenêtre tactile

L'utilisateur déplace la source grâce à un joystick ou de façon digitale, sur une fenêtre tactile.

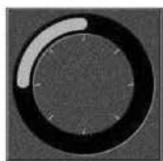


Figure 6.20 – Pan et « Ultipan » SSL C300, © SSL



Figure 6.21 – Pan-pot digital type Avid D-Control, © B. Périaux

Ce deuxième type de pan-pot utilise le joystick qui permet de diriger le point de localisation comme dans un jeu vidéo. Ceci permet un positionnement beaucoup plus naturel et plus réactif par rapport à notre perception. Le joystick, comme on le trouve sur les consoles Yamaha DM2000 ou surfaces de contrôle D-Control, s'accompagne de réglages qui permettent d'avoir à tout moment le contrôle du point (« Grab ») ou bien d'attraper le point lorsqu'on passe par-dessus. Le pan-pot de la console Euphonix System 5 Fusion permet de contrôler la décorrélation de l'espace, par un resserrement spatial, c'est une fonction assez rare et très utile. La SSL C300 HD offre un joystick mais aussi un mode alternatif appelé « Ultipan », lequel permet de faire tourner une source monophonique et de régler sa largeur sur le cercle. Le pan-pot des consoles Lawo permet de bloquer les coordonnées X ou Y du point, ce sont les fonctions « Iso Pan X » et « Iso Pan Y », elles facilitent le positionnement de la source en gardant par exemple les coordonnées X (gauche-droite) constantes.

◆ Pan-pot utilisant des modèles psychoacoustiques

Le modèle pan-pot psychoacoustique intègre l'acoustique du lieu dans lequel on souhaite placer la source, il peut comporter aussi des fonctions de filtrage pour simuler le modèle de perception de la tête. Dans ce cas, l'utilisateur gère le déplacement de la source par un pan-pot évolué dans une console, dans un plug-in spécifique ou bien dans une application autonome. Les modèles psychoacoustiques intègrent souvent la simulation acoustique avec les premières réflexions et le champ diffus.

L'« Imager » de Waves (360 Surround Tools) gère par exemple la disposition circulaire de la source déterminée par un angle, et sa distance simulée par les premières

réflexions. Une fois la position circulaire de la source validée, il est possible d'agir sur la divergence vers les autres canaux afin d'élargir la source localement sur le cercle. L'outil est complété par le champ diffus de la réverbération R360.

Le Virtual Surround Panner de la console Studer Vista 9 utilise différents modes de pan-pot avec l'intégration de fonctions HRTF. On trouve les modes suivants :

- HRTF : utilisation de courbes HRTF annulant les chemins croisés issus de la diaphonie acoustique ;
- Sphere : ce mode simule un microphone sphérique impliquant une différence d'amplitude, de temps et de fréquence (simulation de l'atténuation des fréquences aiguës par ombrage), c'est un modèle de restitution spatial **binaural** ;
- User (MST Microphone Simulation Tool) : ce mode permet de « designer » son propre modèle microphonique.



Figure 6.22 – Surround tools Imager Waves, © Waves

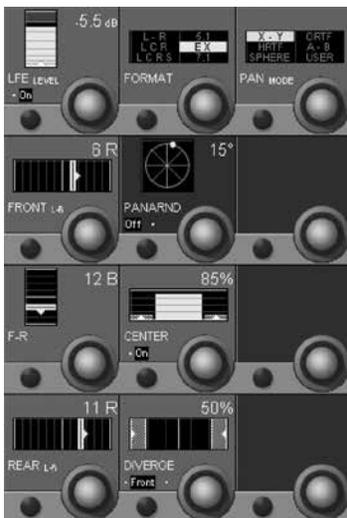
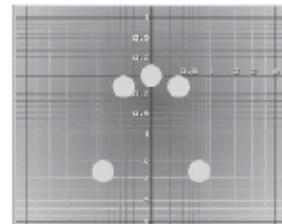


Figure 6.23 – VSP Virtual Surround Panner Studer, © Studer.

● source sonore (positionnement par pan et contrôle de distance)



simulation du placement des microphones

acoustique de pièce virtuelle

Figure 6.24 – VSP Virtual Surround Panner Studer, MST, © Studer.

L'Ircam, associé à Flux, a développé sous forme de plug-in une version du Spatialisateur. Cet outil repose sur un modèle perceptif avancé, il permet de spatialiser des sources en prenant en compte un grand nombre de paramètres acoustiques, comme la variation du niveau sonore avec la distance, l'absorption des fréquences aiguës qui croît avec la distance, la directivité de la source, sa présence, son orientation par rapport à l'auditeur (rotation de la source autour de son axe), l'effet Doppler, le positionnement des enceintes de restitution, la présence de la salle, son enveloppement etc.



Figure 6.25 – Le Spatialisateur de l'Ircam, © Ircam/Flux.

Il comporte une section de spatialisation intégrant le son direct et le champ des premières réflexions ainsi qu'un module de réverbération par synthèse. Il peut spatialiser jusqu'à 8 sources mono, les formats de sorties possibles étant la stéréo, le 5.1, le 7.1, le 8.0, et le format binaural avec simulation de fonctions HRTF.

La première application de ce type de logiciel est de spatialiser des sources pour une restitution multicanal sur enceintes compatible jusqu'au format 8.0. Les lois de pan utilisées sont nombreuses : par défaut la loi de pan d'amplitude par paires d'enceintes, le VBAP 2D et 3D, le DBAP, le format B 2D et 3D, l'ambisonique ordre 2 2D et 3D, la stéréo XY, le MS, et le modèle AB.

L'autre application est de synthétiser une version spatialisée pour le casque (traitement binaural) ou un format transaural (spatialisation à partir d'une écoute stéréo).

L'Ircam en collaboration avec le CNSMDP travaille sur une évolution du Spatialisateur afin d'améliorer le rendu sonore en 5.1 (localisations latérales et arrière) grâce à des procédés transauraux et binauraux. Le downmix stéréo est fait au format transaural, il préserve ainsi une partie de l'information d'espace à l'arrière. L'écoute de ce format transaural nécessite alors un placement de l'auditeur au sweet spot. Cette évolution est intégrée dans un module pour le Spat dans Max MSP et développée sous forme d'application autonome appelée Transpan.

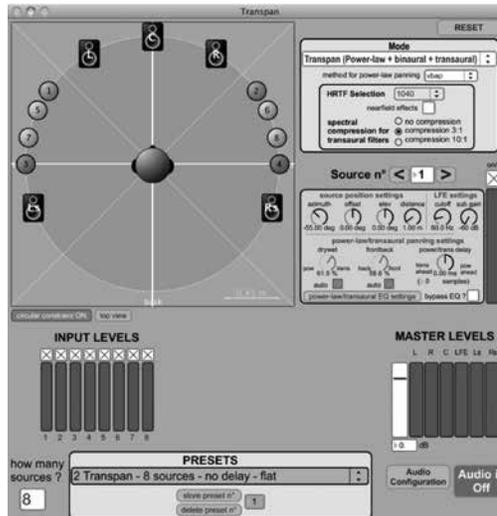


Figure 6.26 – Application Transpan, © Ircam/CNSMDP.

Longcat Technologies, société française spécialisée dans les technologies audio 3D, développe notamment un logiciel de synthèse AudioStage, dont les applications sont innovantes. C'est une application de modélisation numérique pour l'écriture 3D de scènes sonores animées. L'auditeur se déplace dans la modélisation de l'environnement, son point de vue change.

6.5.3 Gestion du centre, divergence centrale

On trouve la fonction divergence centrale, exprimée en pourcentage, dans la plupart des pan-pot, celle-ci permet d'équilibrer entre centre fantôme et centre physique le type de centre dans l'espace frontal. Avec une valeur de divergence de 100 %, une source centrée est envoyée dans l'enceinte centrale seule, elle est donc perçue comme réelle, c'est un centre physique. À l'inverse, une valeur de divergence de 0 % donne un centre fantôme issu des enceintes gauche et droite. Lorsqu'il positionne des sources de part et d'autre du centre, l'utilisateur peut ajuster la valeur de divergence centrale, il agit alors sur la combinaison entre les deux types de centres. Pour les positionnements des sources sur les enceintes L et R, cela n'a pas d'influence car aucun centre n'est généré.

On utilise souvent le centre physique pour les dialogues des films, ainsi la voix provient de l'écran, donc des comédiens et reste intelligible quel que soit le placement des spectateurs dans la salle. Toutefois, il est courant de diverger quelque peu les voix pour les programmes qui s'y prêtent, comme le sport ou les films d'animation. Dans le film « Ratatouille » version Blu-ray, la voix off est spatialisée sur les trois canaux frontaux, tandis que les voix In sont au centre ou quelque peu divergées gauche droite.

Le DVD musical utilise plutôt le centre physique, tandis que les programmes musicaux sans image ou les programmes radio sollicitent le centre fantôme. Une valeur de divergence de 30 à 40 % par exemple, nous permet d'élargir la source en intégrant la

voie centrale. Apparaissent alors les problèmes de diaphonie acoustique, perçus sous forme de filtrage en peigne. Il est possible de les réduire en retardant les canaux L et R, méthode proposée par Bernard Lagnel, chef opérateur son à Radio France. Ce dernier retarde les canaux L et R d'environ 0,7 ms lorsqu'il souhaite générer un centre à partir d'une captation 4 canaux L R Ls Rs. Il apparaît logique par conséquent d'adjoindre à cet outil divergence centrale le délai des trois enceintes, afin de décorrélérer la source mono sur les trois canaux LCR.

Dès lors qu'on utilise un triplet LCR pour la captation de spectacles (opéras, orchestres, théâtre, comédies musicales...), le routing du canal central se fait sans divergence, en revanche, les microphones de proximité ou appoints peuvent utiliser la divergence centrale. Par exemple, un microphone appoint sur une contrebasse disposée au centre de l'orchestre s'intégrera plus facilement dans l'ensemble avec de la divergence centrale. Dirigé vers le centre physique, l'instrument sera perçu de façon trop ponctuelle. Pour le mixage d'images sonores virtuelles (fictions radio, musiques acousmatiques et actuelles), on considère souvent le centre comme une direction sans réel lien avec les autres, il apporte simplement une diversification des provenances.

Dans le contexte du film, la musique est mixée en amont au format 5.1, elle est livrée au mixeur sous forme de stems. On doit considérer le fait qu'au mixage final, le centre d'un stem musique peut être considérablement baissé ou même coupé, afin d'améliorer l'intelligibilité des dialogues. D'où la nécessité de ne pas placer de source seulement dans le centre et de la diverger.

LE CENTRE EN MULTICANAL

- Dialogues film : la plupart du temps en centre physique.
- Dialogues film d'animation, sport TVHD : parfois divergés.
- Musique : utilisation du centre fantôme plus important que le centre physique pour exploiter la largeur de l'image.
- La voie centrale peut être considérée comme une voie indépendante sans réel lien avec les autres canaux (images virtuelles).

6.5.4 Divergence vers les autres canaux, *Width* ou *Depth*

La divergence vers les autres canaux est présente dans la plupart des pan-pots, représentée par le paramètre « Divergence » ou « Width » et « Depth » (largeur et profondeur). Elle permet de délimiter la spatialisation des sources dans une partie de l'espace multicanal. On peut resserrer l'espace frontal de telle sorte que les sources ne soient pas trop espacées (voir **Figure 6.27**). Le placement d'une piste dialogue légèrement divergée, par exemple, peut utiliser ce type d'ajustement, dans ce cas on sollicite les canaux L et R, cela revient à utiliser la divergence centrale. Autre application, on peut aussi resserrer un ensemble de sources déjà positionnées dans l'espace, sans retoucher

leurs placements relatifs, simplement avec cette divergence. Il suffit alors de trouver le bon paramétrage de divergence et de copier la valeur sur toutes les pistes concernées.



Figure 6.27 – Resserrement de l'espace frontal par Divergence dans Pro tools, © Avid.

Par continuité sur tous les canaux du multicanal, lorsqu'on « pan-pote » les sources en utilisant la divergence, on recorrèle l'image sonore, l'impression spatiale est alors diminuée. Si l'on diverge au maximum la source, elle alimente alors les canaux du multicanal avec le même signal.



Figure 6.28 – Resserrement de l'espace complet par Divergence dans Pro tools, © Avid.

6.5.5 Gestion des canaux arrière

Les canaux arrière sont utilisés principalement pour apporter l'effet d'impression spatiale à l'image sonore et sont indissociables de ce fait des canaux frontaux : il est souvent efficace de construire l'enveloppement du mixage avec des stems au format quad ou 5.1, des ambiances, des FX, des musiques. Cet enveloppement repose alors soit sur une acoustique ou une ambiance, soit sur du son direct (non réverbéré) pour les contenus qui le permettent. Les canaux arrière sont aussi souvent utilisés pour le mouvement des sources constituant les FX.

6.6 Réverbération

La réverbération en multicanal permet de localiser une source dans l'espace par le champ des premières réflexions qu'elle synthétise. Le champ diffus créé est constant dans toute la salle, donc réparti de la même façon, avec décorrélation, sur les différents canaux L, C, R, Ls, Rs du multicanal.

La réverbération en mixage multicanal peut se gérer de deux façons différentes. La première consiste à travailler avec plusieurs réverbérations stéréo que l'on spatialise dans l'image multicanal : on place généralement une réverbération dans l'espace frontal et une seconde dans l'espace arrière. Cette méthode permet de bien maîtriser le dosage de la réverbération du mixage et particulièrement son placement, elle est tout à fait appliquée au travail d'une image virtuelle ou une image où la scène est dynamique (les sources, les personnages bougent et les lieux changent).

Les images réelles, en revanche, utilisent plutôt une seule réverbération multicanal au format 5.0, afin d'obtenir une meilleure homogénéité. On peut alimenter la réverbération par un format 5.0 ou 5.1 (le Lfe n'est pas réverbéré, l'utilisation d'un chemin 5.1 simplifie le routing), ce qui consiste à gérer la spatialisation avant la réverbération : un départ auxiliaire 5.1 permet alors de spatialiser la source grâce à un pan-pot 5.1. Il est courant de relier le pan d'envoi de réverbération et celui de la piste principale, cette fonction s'appelle « Follow main pan » dans Pro Tools et permet de déplacer la source directe dans l'espace et gérer en parallèle son placement dans la réverbération.

On peut aussi alimenter l'unité de réverbération en stéréo, ce qui suffit dans certains cas. La latéralisation gauche-droite se fait alors en stéréo dans le départ aux, et le placement frontal/arrière ainsi que l'ajustement gauche/droite se paramètrent sous forme de sources à placer dans la réverbération (voir M6000, algorithme VSS 5.1 source).

On peut ensuite combiner les deux méthodes, réverbération multicanal et plusieurs réverbérations stéréo disposées à l'avant et à l'arrière.

Le paramétrage d'une réverbération multicanal se fait selon les critères habituels utilisés en stéréo : un temps de réverbération TR global qui peut s'ajuster par bandes de fréquences (« Damping »), une taille de salle reliée à un type de premières réflexions, le paramétrage de la densité des réflexions, un filtre en sortie et éventuellement en entrée, le réglage d'un pre delay, et des effets de modulation. Ensuite, il est possible de modifier par coefficients multiplicateurs le TR de chaque canal, et rallonger au besoin la réverbération dans certains secteurs de l'espace : on obtient donc cinq valeurs de TR à ajuster. On peut filtrer les canaux frontaux et arrière indépendamment, par défaut les ajustements sont reliés entre eux. Le « Damping » du TR peut aussi, dans certaines réverbérations, être paramétré de façon différente entre le frontal et l'arrière. On trouve parfois un paramètre intéressant, l'équilibre frontal/arrière (chez Waves, « Front/Rear »), il permet de changer rapidement la répartition en niveaux frontal/arrière de la réverbération, autrement on le gère avec le niveau des différents canaux (souvent LR, C, LsRs).

Le positionnement des sources dans l'espace en interne dans la réverbération est proposé par certains algorithmes, c'est le cas du Spatialisateur de l'Ircam, ou de

l'algorithme VSS 5.1 source de la TC M6000. L'utilisateur positionne les sources mono dans la réverbération, il peut aussi agir sur les caractéristiques des premières réflexions de chacune des sources.

PARAMÈTRES D'UNE RÉVERBÉRATION MULTICANAL

- TR global, coefficients multiplicateurs par canal L, C, R, Ls, Rs.
- Damping : TR différent par bandes de fréquences, frontal/arrière.
- Taille, type de premières réflexions.
- Filtrage en sortie des canaux frontaux et arrière.
- Équilibre frontal/arrière en niveau.
- Positionnement des sources dans l'espace pour certains algorithmes, caractérisation des premières réflexions pour chaque source.

6.6.1 Réverbération à convolution

Les réverbérations à convolution utilisent la mesure de réponse impulsionnelle de la salle et se paramètrent suivant le type de captation. Une impulsion est diffusée dans la salle en mono ou en stéréo, et on utilise un système de captation multicanal généralement au format quad, avec différentes directivités et différentes distances de placement.



Figure 6.29 – Altiverb 7, © Audio Ease.

Les réglages possibles sur les réverbérations à convolution étaient au départ plus réduits que sur les réverbérations à synthèse, mais ils ont bien évolué depuis. Le TR est ajustable en général en pourcentages ou en rapport multiplicateur (le rapport multiplie le TR défini par la réponse impulsionnelle), on trouve la taille de la salle, la répartition en niveau des canaux (LR, arrière, centre, Lfe), le « Damping », la position des haut-parleurs dans la salle, le filtrage en sortie. Sur certains plug-ins, il existe un paramètre (« Decorr » dans l'IR360) qui permet de décorrélérer les différents canaux

afin d'augmenter l'impression d'espace. On peut aussi paramétrer dans certains cas l'enveloppe temporelle de la réverbération (« Reverb envelope ») et créer des cluster échos (c'est le cas de l'IR360).

TC Electronic M6000

Algorithme VSS-6.1 Reverb : réverbération d'un mixage

Le TC M6000 propose deux algorithmes de réverbération. Le premier, le VSS- 6.1 reverb, permet de réverbérer un mixage ou un stem compatible jusqu'au 6.1. Cet algorithme ne modifie pas la localisation de l'image sonore, mais permet d'ajouter un champ réverbéré au mixage. On peut éditer le TR global, ainsi que les coefficients multiplicateurs pour chacun des canaux, le pre delay, la taille de la salle, le type de premières réflexions, le filtrage des canaux frontaux et arrière, avec la possibilité de « linker » les deux espaces, la modulation de la réverbération qui permet de générer des variations temporelles de l'enveloppe de la réverbération, simulant des effets réalistes de salles.

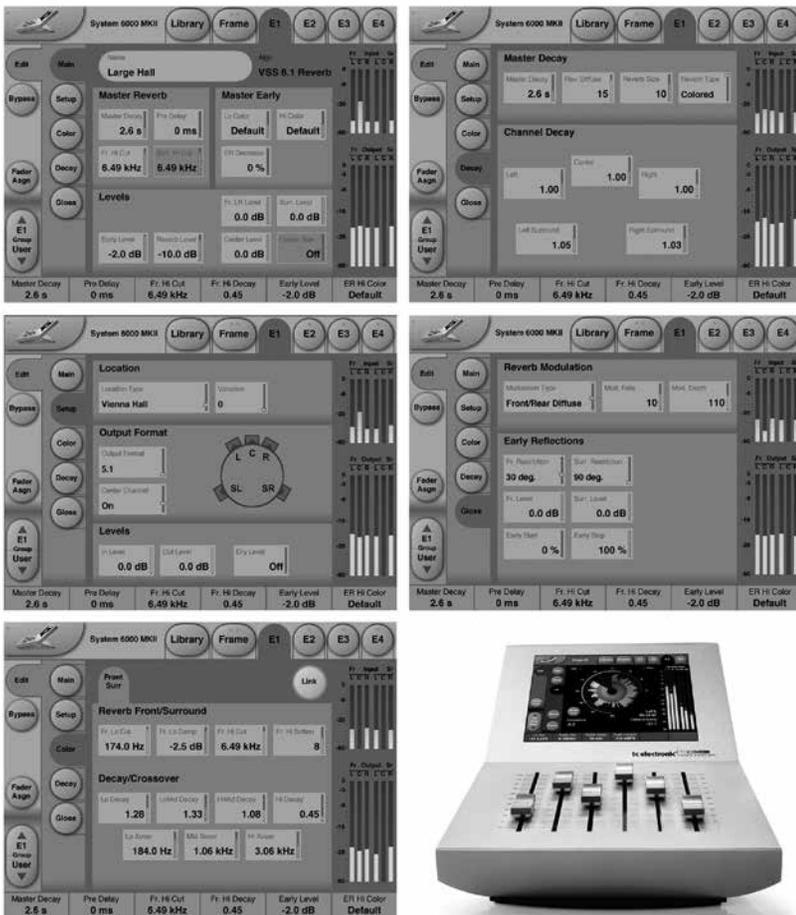


Figure 6.30 – TC M6000 VSS-6.1 Reverb, © B. Périaux d'après TC electronic.

Algorithme VSS-5.1 Source : spatialisation de sources

L'autre algorithme, VSS-5.1 source, est plus adapté à la réverbération de sources séparées, et propose en entrée quatre sources S1, S2, S3 et S4 au format mono. L'utilisateur décide de la spatialisation de ces quatre sources dans la salle suivant des positions prédéfinies proposées par la réverbération. Cet algorithme prend en compte le champ des premières réflexions, variable en fonction de la localisation de la source, le champ diffus étant constant. Il est possible aussi de changer le point d'écoute de l'auditeur afin de se rapprocher ou bien de s'éloigner des sources frontales. Les caractéristiques de chacune des sources peuvent être paramétrées de façon indépendante : les caractéristiques temporelles (démarrage et extinction), le filtrage des premières réflexions, le niveau du champ des premières réflexions et du champ diffus. Le TR peut être réglé pour chacun des canaux et il est possible d'ajuster la direction des premières réflexions dans l'espace frontal et dans l'espace arrière, les valeurs sont données en degrés :

Espace frontal : 22°, 30°, 45°, 90°

Espace arrière : 90°, 115°

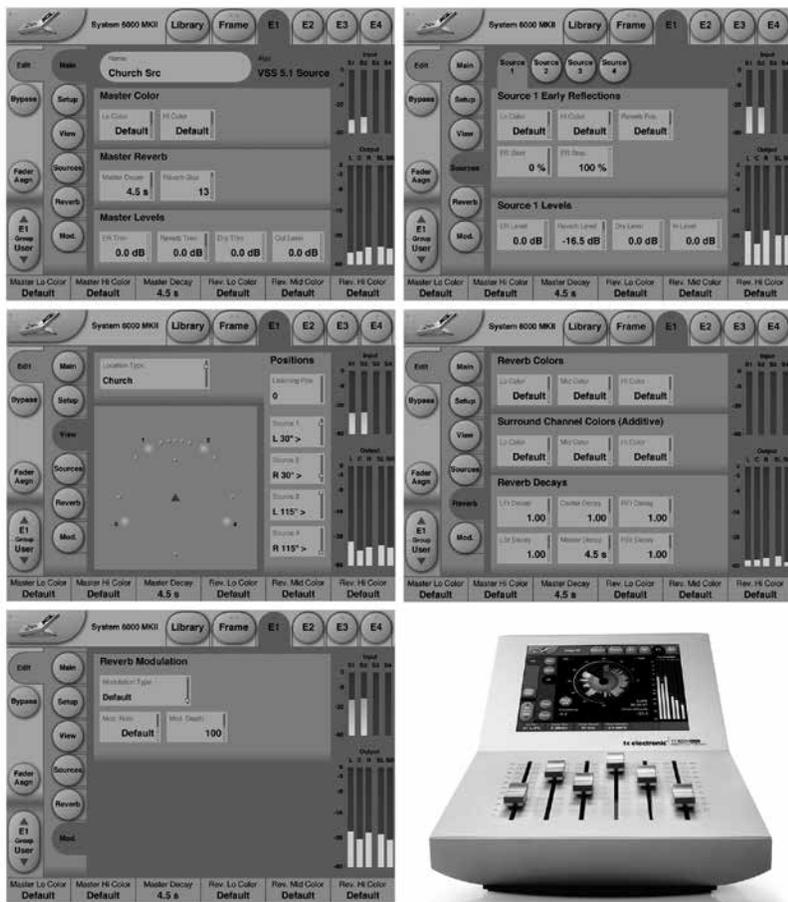


Figure 6.31 – TC M6000 VSS-5.1 Source, © B. Périaux d'après TC electronic.

6.6.2 Lexicon 960 L, Surround PCM96

Lexicon utilise globalement la même approche, la 960 L permet de réverbérer un mixage 5.0. L'algorithme dispose d'un pan-pot en entrée comme en sortie, l'utilisateur réglera généralement le pan-pot dans la position standard du canal (L, R, C, Ls, Rs). On retrouve le paramètre temps de réverbération « MidTR » de la réverbération Lexicon 480 avec le coefficient multiplicateur dans les basses fréquences, les niveaux de sortie par canal (C, LR, LsRs), le filtrage en sortie pour le frontal et l'arrière, le paramétrage de l'enveloppe temporelle propre à Lexicon, avec le « Shape » et le « Spread » (pour les programmes Chamber, l'utilisateur peut ajuster indépendamment les deux paramètres pour le frontal et l'arrière). La PCM 96 Surround reprend les mêmes paramètres.

6.7 Traitement Dynamique

6.7.1 Compresseurs multicanal

La plupart des consoles qui gèrent le 5.1 intègrent un module de traitement dynamique des Bus Master 5.1, comme les consoles stéréo de type SSL ou Neve, proposent un compresseur des Bus Master Stéréo. La compression fonctionne alors à partir d'un « side chain ». S'il existe un seul « side chain », les différents canaux du multicanal sont traités de la même façon, s'il existe plusieurs « side chain », alors le traitement des canaux peut être différent. L'analyse se fait généralement sur la somme des canaux ou bien sur le niveau maximal d'un des canaux.

Le Master compresseur de la console SSL Duality fonctionne par exemple sur le même modèle que la 4000, avec un « side chain » que l'utilisateur va alimenter comme il le souhaite par les canaux LCR, il peut intégrer les canaux LsRs et/ou le canal Lfe. Ce n'est pas toujours souhaitable d'associer le Lfe, ce dernier étant assez indépendant des autres canaux, mis à part le filtrage, on ne le traite généralement pas. La fenêtre d'analyse est unique, alimentée par les différents canaux (mode somme ou signal maximal d'un canal) et le module de traitement 5.1 est piloté par cette fenêtre avec les paramètres usuels : le seuil de compression, le rapport, le temps d'attaque, le temps de retour, le gain de rattrapage. Ce type de compresseur est illustré par la **figure 6.32**.

AMS-Neve a sorti récemment un compresseur analogique 5.1, le 8051 issu de sa console 88RS, celui-ci travaille de la même façon (l'analyse se fait sur le niveau maximal d'un des canaux) mais avec deux « side chain » indépendants A et B. Ce type de compresseur offre une plus grande maîtrise de la



Figure 6.32 – Compresseur Master 5.1 analogique type SSL Duality, © SSL.

compression grâce aux deux « side chain », deux compressions indépendantes peuvent être réalisées (voir **figure 6.33**). Il est aussi possible de filtrer le side chain par canal (filtre coupe-bas, « sidechain filter ») afin de rendre la compression sélective en fonction de la fréquence.

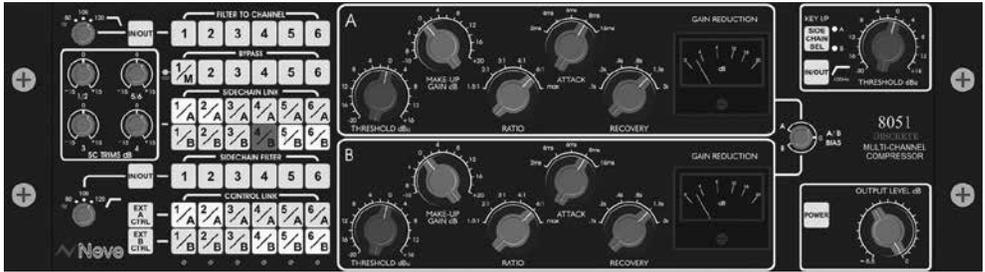


Figure 6.33 – Compresseur Master 5.1 analogique type Neve 8051 comportant deux « side chain », © Neve.

Pour obtenir une plus grande maîtrise dans la compression d'un programme 5.1 avec davantage de « side chain », il faut se tourner vers des outils numériques comme la TC M6000 ou d'autres plug-ins. L'étape de mastering 5.1 peut nécessiter ce type d'outils.

La M6000 présente deux types de compresseurs multicanal 5.1 multibandes, le MD 5.1 et le MDX 5.1. Le paramétrage des « side chain » fonctionne de la même façon pour les deux types de programmes : l'algorithme repose sur un système de quatre « side chain » :

- S1 pour L et R ;
- S2 pour C ;
- S3 pour Ls et Rs ;
- Lfe SC pour le canal Lfe.

Les différents canaux du signal multicanal peuvent ainsi être indépendamment compressés. Cela peut être intéressant de traiter le centre et les canaux arrière de façon différente des canaux gauche et droite, pour améliorer l'intelligibilité des dialogues (généralement au centre) et des ambiances (à l'arrière). Par défaut, le « side chain » est alimenté par le canal lui-même, les canaux L et R alimentent ainsi le « side chain » SC1, le centre alimente le « side chain » S2, les canaux Ls et Rs, le « side chain » SC3 et le canal Lfe, le « side chain » Lfe SC. Par ailleurs, le « side chain » peut être alimenté par des signaux auxiliaires supplémentaires (voir page « Side Chain Feed », SC1 add 1, SC1 add2, SC1 add3, SC2 add1 etc.) qui se somment au signal de base défini dans la page « Side Chain Control » ou bien par le niveau maximal du signal auxiliaire. Nous retrouvons alors les deux modes possibles de fonctionnement du « side chain » défini auparavant (somme des canaux ou niveau maximal des canaux).

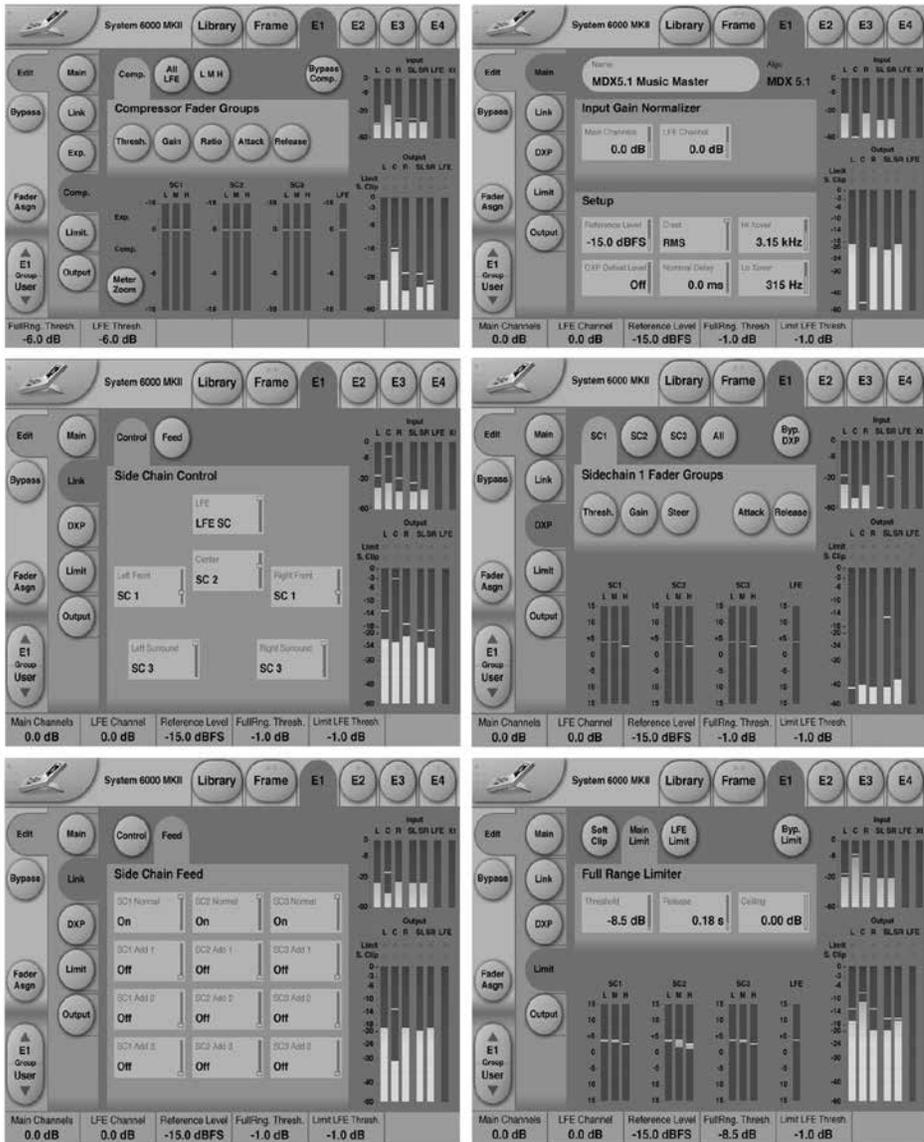


Figure 6.34 – Compresseurs MD5.1, MDX 5.1 TC M6000, © B. Périaux d'après TC electronic. Le premier schéma représente l'algorithme MD 5.1, les autres figures représentent l'algorithme MDX 5.1.

Les deux algorithmes MD 5.1 et MDX 5.1 fonctionnent en multibandes, la compression peut être modifiée de façon relative sur 3 bandes de fréquences définies dans la page principale. Le MD 5.1 intègre un compresseur, un expandeur, et un limiteur. Le MDX 5.1 utilise l'algorithme de compression-expansion DXP propre à TC electronic qui fonctionne avec des pentes de courbes de compression pilotées par le paramètre « steer », un seuil d'expansion réglé par le « defeat level ». Le module DXP est complété par un limiteur.

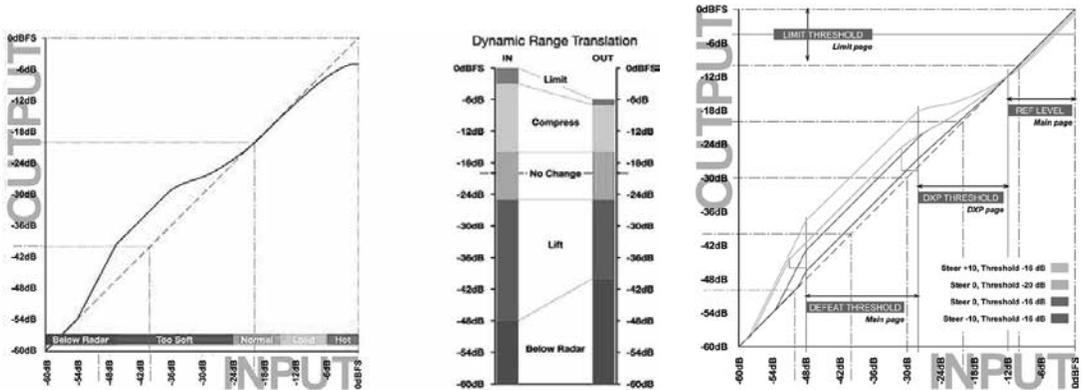


Figure 6.35 et 6.36 – Principe de fonctionnement du compresseur MDX 5.1, module DXP, © TC electronic.

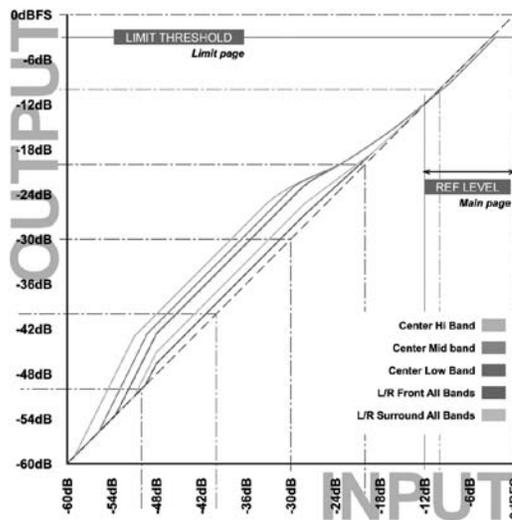


Figure 6.37 – Exemple de configuration du compresseur MDX 5.1 traitant le canal central de façon différente, © TC electronic.

Le plug-in McDSP Channel G Surround fonctionne de la même façon que les algorithmes de la TC M6000, il comporte quatre compresseurs, un pour les canaux L et R, un pour le centre C, un pour les canaux Ls et RS et un pour le canal Lfe. L'utilisateur règle le « key » pour chacun des compresseurs, ce peut être LR, C, LsRs ou Lfe, avec la possibilité de les combiner. On peut ensuite « linker » les différents compresseurs, on agit alors sur les paramètres de control du Master : le seuil, le rapport, les temps d'attaque et de retour, le gain de rattrapage, le « knee ». La compression se fera de la même façon pour les canaux liés, mais à partir de « key » différents ou non, suivant les réglages.

Le compresseur C360 Surround Tools Waves propose aussi quatre compresseurs, avec différents modes de « link », le schéma ci-dessous représente le mode de « link » usuel LR, C, LsRs et Lfe indépendants.



Figure 6.38 – Compresseur C360 Surround Tools Waves, © Waves.

6.7.2 Limiteurs multicanal

Le limiteur au format 5.1 est indispensable pour le traitement du mixage final, afin de ne pas dépasser le niveau crête maximal. On l'insère généralement sous forme de plug-in, en traitement hardware ou sur un master de console numérique. Notons qu'il devient impératif de pouvoir contrôler les crêtes en mesure true peak, afin de ne pas dépasser les -3 dBTP pour la diffusion broadcast. La version multicanal de ces limiteurs n'a pas de spécificité particulière par rapport au format stéréo si ce n'est le mode « link » entre les différents canaux, identique au mode « link » d'un compresseur multicanal.



Figure 6.39 – Pure limiter Flux, © Flux.



Figure 6.40 – Limiteurs Broadcast dBTP Elixir Flux, © Flux.

6.8 Autres effets

6.8.1 Autopan

L'autopan en multicanal permet de simplifier le travail d'automation que l'on pourrait gérer par des courbes de pan-pot. Il existe un autopan développé par Trillium Lane Labs (Avid) qui permet de faire tourner une source mono sur le cercle ou sur un arc de cercle, dans un sens ou dans l'autre, à vitesse réglable (la fréquence). Le calage au tempo est possible, la distance du cercle ou plus précisément la décorrélation du mouvement est paramétrable, le mouvement peut être continu ou bien quantifié sur des points prédéfinis dans l'espace. D'autres fonctions complémentaires sont proposées par le GRM Space (INA).

6.8.2 Délais

Il n'existe pas aujourd'hui de délai vraiment optimisé pour le 5.1. On trouve cependant le GRM Tools Shuffling (INA), un outil très créatif qui permet de « designer » le son en le découpant en fragments de durées variables. Le délai est paramétrable, ainsi que les temps de montée et d'extinction des fragments (rapide, lent) qui génèrent des transitoires plus ou moins marqués.

La TC M6000 propose un programme de délai circulaire, « reflector 6 ». Le son est délayé sur plusieurs points localisés sur le cercle d'écoute.

6.9 Downmix

Le downmix est le procédé de conversion d'un format multicanal vers un format inférieur : passage du 5.1 vers la stéréo ou la mono, ou passage d'un 7.1 vers un 5.1. Cette étape est importante car elle est associée automatiquement à la compatibilité stéréo d'un format 5.1 Dolby Digital ou DTS, en diffusion TVHD, ou sur supports DVD ou Blu-ray. En effet, la tendance est de proposer un format 5.1 Dolby ou DTS sur les supports et de ne plus intégrer de format stéréo. Le mixeur doit donc gérer la compatibilité stéréo de son mixage 5.1, cette compatibilité sera portée par des coefficients de réduction dans les formats Dolby Digital et DTS. Ceci fait l'objet d'une metadata qu'on appelle Downmix, intégré dans les codages Dolby Digital, Dolby E, DTS...

Voici ci-dessous la fabrication d'un downmix LoRo (downmix en phase), par simple sommation des différents canaux :

DOWNMIX LoRo

- $Lo = L + C (-3 \text{ dB}) + Ls (-3 \text{ dB}) \rightarrow \text{Left only.}$
- $Ro = R + C (-3 \text{ dB}) + Rs (-3 \text{ dB}) \rightarrow \text{Right only.}$

Les coefficients (-3 dB, -3 dB) sont ceux proposés par défaut, mais le mixeur peut les ajuster.

Le canal LFE est généralement abandonné dans les downmix.

On trouvera dans le chapitre 8, les downmix LtRt associés à la diffusion Dolby Digital, pour la compatibilité avec le format Dolby Surround.

Il existe donc des outils de downmix associés aux bus 5.1 des consoles ou sous forme de plug-ins. Le plug-in Neyrinck Soundcode Stereo par exemple permet de « downmixer » dans Pro Tools un mixage 5.1, l'ajustement des coefficients se faisant par pas de 0,1 dB. L'intégration du canal Lfe est possible, ce qui permet de réintégrer du grave supplémentaire dans la stéréo. Ce plug-in ne simule pas vraiment un downmix de diffusion, il permet simplement de fabriquer une stéréo, LoRo ou LtRt, à partir d'un 5.1. Le LtRt peut être décodé afin de valider l'encodage.



Figure 6.41 – Downmix LoRo SoundCode Stereo Neyrinck, © Neyrinck Audio.

Les consoles numériques intègrent une fonction downmix paramétrable, avec des options intéressantes comme le downmix LtRt proposé par Studer. Il est possible de mémoriser plusieurs presets de downmix sur le bus master, ou dans certains cas, directement sur les stems, comme le propose la console Euphonix Système 5 Fusion.

6.10 Upmix

L'opération inverse, celle qui consiste à passer d'un format mono ou stéréo à un format multicanal s'appelle l'upmix. L'utilisation d'un tel procédé se justifie lorsque l'on souhaite spatialiser une ambiance stéréo, par exemple, ou un mixage dont les pistes

séparées n'existent pas (des archives par exemple) ou bien ne sont plus exploitables. C'est aussi le cas malheureusement lorsqu'on ne dispose pas assez de budget pour envisager un mixage 5.1 et surtout dans des applications de diffusion (voir chapitre 8).

Il existe de nombreux algorithmes qui, par analyse de la phase, le délai entre les deux canaux de la stéréo et leur contenu fréquentiel, produisent une image sonore spatialisée. L'information hors phase est restituée dans les canaux arrière, l'information en phase, au centre, et le contenu restant alimente la stéréo du multicanal. C'est le principe du décodage Dolby Surround. Cette analyse de la corrélation des signaux est le principe de base des algorithmes d'upmix, ensuite d'autres modules additionnels peuvent être intégrés comme l'ajout de réverbération, la spatialisation dans l'espace multicanal par différence d'intensité et de temps de la stéréo, ou la séparation du champ direct et du champ réverbéré d'un signal.

L'algorithme unwrap du TC M6000 (ou DB8) intègre un module de réverbération dont le mode de décorrélation des canaux est paramétrable. Il repose sur l'analyse en phase, délais, fréquence des deux canaux de la stéréo. On peut ajuster la quantité en pourcentage de la stéréo traitée, 0 % signifie qu'on laisse la stéréo d'entrée en sortie LR du multicanal sans traitement. Une utilisation de 60 % permet de fabriquer un centre, tout en préservant la largeur du mix original.

Autre exemple, l'Upmix5 développé par Arkamys pour Pyramix, ce dernier n'ajoute pas de réverbération.

L'algorithme AMBIT, que l'on trouve dans les consoles Lawo, sépare le procédé d'upmix en deux modules, le module « Upmix » pour le traitement frontal et le module « Spatialize » pour la création des canaux arrière.

Il existe d'autres algorithmes d'upmix : Isostem, Penteo, Soundfield...

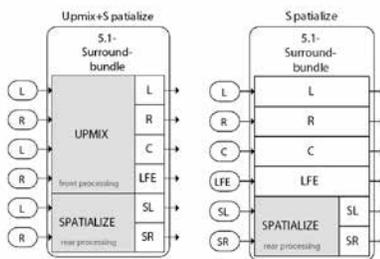


Figure 6.42 – AMBIT Lawo, © Lawo.

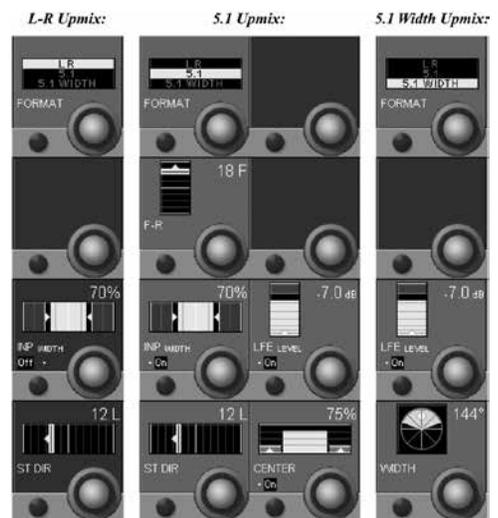


Figure 6.43 – Upmix Vista9 Studer, © Studer.

6.11 Méthode de mixage et analyses

6.11.1 Mixage film : les stems

Le mixage film se décompose en plusieurs étapes de prémixages appelés stems. Ces derniers sont généralement constitués des dialogues, des bruitages, des FX, des ambiances et de la musique. Le mixeur récupère de la session de montage des pistes séparées de ces différents éléments, il réalise le mixage de la VO puis le mixage de la VI, la version du mixage sans les dialogues (de la langue originale). Du mixage film sont dérivées les versions supports (DVD ou Blu-ray) et TV. On réduit généralement la dynamique pour être compatible avec la diffusion domestique, on vérifie la compatibilité du mixage avec l'écoute ITU 5.1 et les compatibilités stéréo (downmix LtRt généralement) et mono. Le mixage TV doit être compatible avec la norme EBU R128 (voir chapitre 8).

Les ambiances, les effets et la musique sont généralement spatialisés en 5.1. Les stems dialogues et bruitages sont généralement au format LCR avec réverbération ou LCR et réverbération 5.0 séparée ou 5.0 avec réverbération, plusieurs méthodes existent. Séparer le son direct de la réverbération permet de garder le contrôle du rapport CD/CR.

Voici une représentation de la construction des différents stems du film et des différents formats de diffusion cinéma, TV et supports.

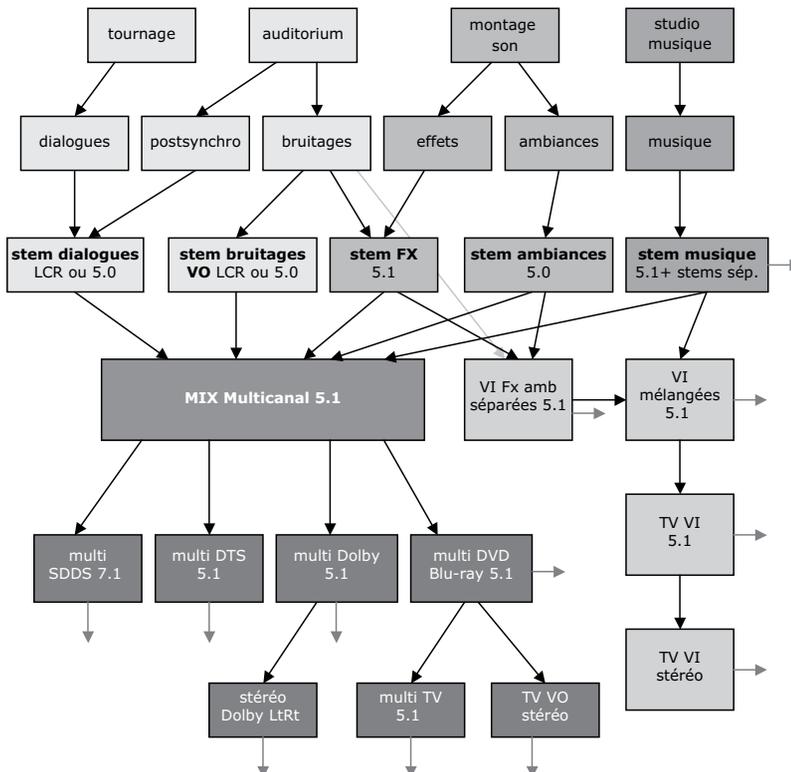


Figure 6.44 – Stems Film et formats de diffusion, © B. Périaux d'après un schéma de William Flageollet.

6.11.2 *Mixage Musique*

Pour la musique de film, on procède également par stems. Le mixeur film demande généralement au mixeur musique des premix pour mieux insérer la musique dans le scénario du film. Pour la musique orchestrale, il convient de séparer les cordes, les bois, les cuivres, les percussions, les chœurs, les solistes et les programmations. Ainsi, l'intégration de la musique se fera avec plus de souplesse notamment pour les solistes comportant la notion de dialogue musical qui peut nuire à l'intelligibilité des dialogues. Pour une raison de commodité, il est conseillé de fournir des stems en 5.1 même si le canal Lfe n'est pas toujours sollicité. D'autre part, le centre peut poser des problèmes de compatibilité avec le reste des éléments sonores de la bande son. Il est conseillé alors de ne pas trop fournir le canal central de la musique et diverger les instruments centrés par anticipation. Le mixeur film peut baisser le centre du stem musique, ou tout simplement le supprimer, s'il va contre l'intelligibilité des dialogues.

Pour une création musicale sans rapport à l'image, il est aussi intéressant d'aborder le mixage par stems 5.1, cette méthode nous donnera une bonne homogénéité dans la spatialisation.

6.11.3 *Classement des sources*

Dans l'analyse des sources du mixage, il semble primordial de distinguer les éléments qui vont donner le discours, de ceux qui vont apporter l'enveloppement (les sources qui par leur nature ont un intérêt à être spatialisées sans créer de gêne auditive). Certaines sources travaillent ensemble et peuvent s'opposer spatialement.

- le discours en frontal : la voix, un instrument de musique soliste ou tout autre élément mélodique, ce qui attire l'attention ;
- l'enveloppement : les ambiances, l'acoustique, la musique, les FX, des sources larges sans transitoires, les sources qui n'attirent pas l'attention ;
- l'homogénéité, la fusion frontale/arrière des sources : des sons qui travaillent ensemble ;
- l'acoustique : souvent au format multicanal pour donner l'effet d'une seule salle ou d'un même lieu, apporte une homogénéité.

Comme évoqué au chapitre 3, le choix des sources off localisées dans les zones latérales et à l'arrière se fait en fonction de leur poids attentionnel (le caractère nouveau de la source, leurs caractéristiques dynamiques, le contenu fréquentiel), leur degré de prévisibilité, les contraintes esthétiques de l'œuvre, le poids attentionnel de l'image (scène avec des contraintes réalistes ou image virtuelle avec peu de contraintes, des images abstraites).

6.12 La synthèse binaurale

La synthèse binaurale est abordée de deux façons différentes. La première consiste à spatialiser les sources dans un espace 3D, ce qui revient à mixer en binaural, la seconde permet d'encoder un mixage multicanal généralement 5.1 en binaural 2D. Le premier cas offre toutes les directions contenues dans les HRTF. Le second cas est une **compatibilité** du mixage 5.1 au casque, comme le proposent certains casques multicanal ou lecteurs binauraux pour la diffusion. Dans ce cas, la quantification de l'espace est limitée au format du mixage multicanal, d'où l'intérêt de l'approche mixage objets qui permet d'aborder tous les formats. Pour réaliser une image sonore spatialisée au casque compatible 5.1 ITU, il convient de placer les sources des canaux LCRLsRsLfe suivant la norme ITU, en faisant le choix d'une distance d'écoute. Cette distance peut être contenue dans la réponse impulsionnelle de l'algorithme, ou bien elle peut être générée par un module d'effet de salle, qui s'ajoute au module de localisation HRTF (la distance de mesure des HRTF est d'environ 2 m en chambre anéchoïque, voir chapitre 3).

Complément Web 6.2

Outils de synthèse binaurale.

6.13 Exploitation des nouveaux formats, mixage objets

6.13.1 Auro-3D, 22.2 NHK

L'Auro-3D comme le 22.2 de la NHK font partie des formats plutôt dédiés à la salle de cinéma ou à la grande salle de diffusion, mais s'inscrivent dans les techniques de spatialisation de la stéréophonie définies au premier chapitre. Ces méthodes de spatialisation usuelles associent en général un canal de mixage à un haut parleur. La production de ces formats est possible grâce à l'existence de partenariats entre industriels et chercheurs. L'Auro-3D est compatible avec les consoles AMS-Neve 88D ou DFC, lesquelles offrent un système de bus 14 canaux, compatible jusqu'au format Auro-3D 13.1. L'utilisateur utilise deux joysticks pour placer une source dans l'espace. Il existe aussi des outils compatibles Pro Tools et Nuendo, comme l'Auro-Panner, un pan-pot compatible 13.1 ou l'Auro-verb, une réverbération 3D.

Les ingénieurs du son de la NHK utilisent deux Wiimote de console Nintendo pour placer les sources dans l'espace, une pour l'horizontal, l'autre pour le vertical, sur des consoles Fairlight dont on peut configurer les bus surround.

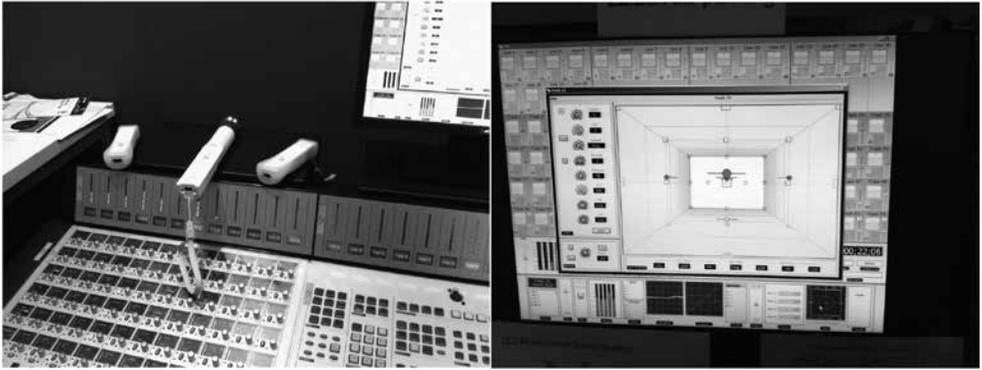


Figure 6.45 – Spatialisation 22.2 NHK, démonstration IBC, © B. Périaux.

Spatial Audio Designer (New Audio Technology) est un pan-pot sous forme de plug-in qui permet de gérer des formats supérieurs au 7.1, il est notamment compatible Auro-3D 13.1, DTS, 22.2, sans limite. Il gère un système de bus interne défini par le format choisi (9.1, 11.1, 13.1, 22.2...). L'outil permet de détourner les possibilités de bus multicanal d'un DAW qui généralement est limité au 7.1. Les sources sont repérées par un code couleur, et deux représentations graphiques permettent de placer les sources dans l'espace, une vue de dessus pour l'horizontal et une vue de face pour le vertical. Cet outil, développé par Tom Ammermann, intègre aussi les divergences centrales des secteurs qui comportent un centre.

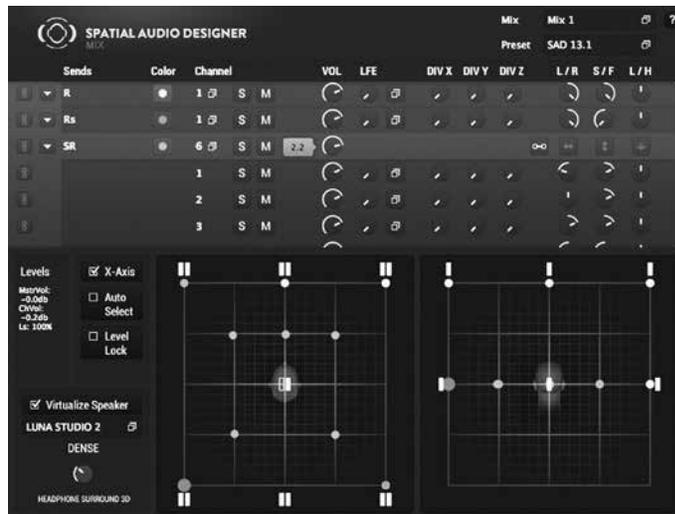


Figure 6.46 – Spatial Audio Designer, © New Audio Technology.

Le logiciel Reaper (Cockos) offre un pan-pot très complet appelé ReaSurround. Il permet de gérer des formats allant jusqu'à 64 canaux prenant en compte l'élévation. Il existe

des presets mais il est possible d'éditer son propre format en ajustant les paramètres influences, lesquels permettent de définir la corrélation de chaque enceinte. Le schéma ci-dessous représente le pan-pot 9.1 (type Auro-3D).

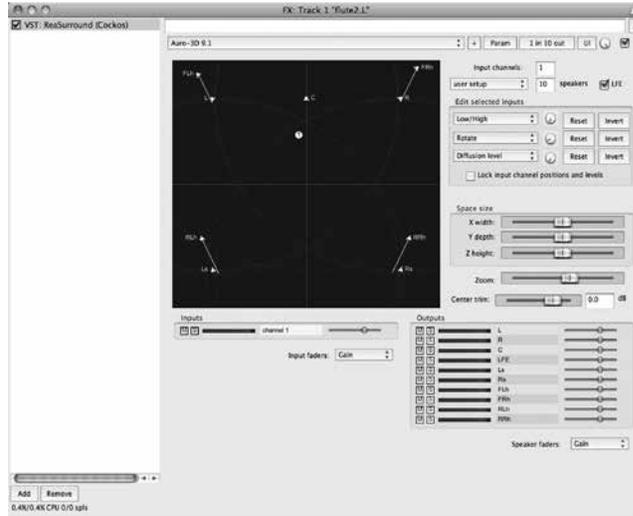


Figure 6.47 – Pan-pot ReaSurround 9.1 Auro-3D dans Reaper, © Cockos.

Auro Technologies propose aussi des outils 3D dont l'Auro-3D Panner qui travaille avec le moteur de mixage 3D Auro-3D mixing engine.



Figure 6.48 – Auro-3D Panner, © Auro Technologies.

Le GRM Space (INA) est un panner compatible avec de nombreux formats multicanal que l'utilisateur peut éditer, il propose de nombreuses fonctions complémentaires comme l'autopan, l'édition de trajectoires, la gestion des sources regroupées en objets.

6.13.2 Exploitation de la WFS

Autre format en pleine évolution, la WFS (voir chapitre 1 et 2), qui équipe des salles de spectacles, quelques théâtres et certaines salles de cinéma. Ce système utilise le mixage objet.

L'utilisation pratique de la WFS suppose l'utilisation d'un processeur dédié qui se place généralement à la suite de l'environnement de mixage (table de mixage, DAW...), en amont des canaux d'amplification et à raison d'un canal d'amplification par enceinte. Le processeur Wave 1 de Sonic Emotion permet de traiter en temps réel et avec une latence réduite (environ 5 ms) 32 entrées qui sont autant de sources virtuelles dont la position peut être manipulée en temps réel. Le processeur se charge alors de calculer les signaux de sortie pouvant alimenter jusqu'à 32 ou 64 haut-parleurs. Plusieurs processeurs peuvent être couplés pour adresser un plus grand nombre de haut-parleurs lorsque la situation le requiert.

L'environnement de mixage doit être configuré de façon à sortir un canal audio séparé par tranche, groupe, bus ou piste de console ou DAW. La présence d'interfaces multicanal largement répandues, avec lesquelles le Wave 1 est compatible (ADAT, MADI), rend cette opération très aisée. À chaque tranche ou piste correspond ainsi une source dont la position spatiale peut être manipulée soit sous forme d'automation à l'aide de plugin dédiés (VST, AU, AAX) soit sous forme de rappels de preset stockés dans le processeur Wave 1 (interface Wave Performer), soit enfin dans des environnements interactifs (Max MSP, interface Open Sound Control, dispositifs de tracking, interface Wave Performer).



Figure 6.49 – WavePerformer, outil de visualisation/manipulation des sources virtuelles en exploitation, © Sonic Emotion.

La manipulation des sources et la communication avec le processeur Wave 1 repose sur un protocole réseau propriétaire qui permet d'informer de façon instantanée tous les éléments du système lorsque qu'un changement intervient. Une modification de la position d'une source dans un plugin sera non seulement communiquée au processeur mais aussi à toutes les interfaces présentes sur le réseau, notamment la partie visualisation de l'interface WavePerformer.

La description de la scène sonore repose ainsi sur un format objet, largement indépendant du système de reproduction physique utilisé. Une description spatiale passant uniquement par les positions de sources sonores rend aussi le système compatible avec les formats multicanal usuels (stéréo, 5.1, 7.1...). Pour cela, les sources virtuelles associées aux canaux du format sont positionnées dans les directions requises (ex : $-30/+30$ degrés pour la stéréo) et forment ainsi des haut-parleurs virtuels diffusant le contenu multicanal. La WFS synthétise alors des ondes planes, éloignant virtuellement les haut-parleurs au delà des limites du lieu. Chaque haut-parleur virtuel est ainsi perçu dans la direction requise par le format dans l'intégralité de la pièce. La zone d'écoute pour la diffusion de formats multicanal standards est ainsi largement étendue et ne se limite plus au seul sweet spot de la diffusion avec un haut-parleur par canal.

Au delà de ce mode « rétro-compatibilité », l'utilisation de haut-parleurs virtuels peut être utile dans le mixage WFS afin de regrouper un certain nombre de sources issues de la captation, qui forment plutôt l'arrière plan et ne nécessitent pas un rendu exact de leur position. On les regroupe alors dans un bus multicanal classique qui sera diffusé sur un ensemble de haut-parleurs virtuels.

Les informations d'effet de salle peuvent être gérées soit en interne dans l'outil WavePerformer (développements en cours d'outils de gestion du régime de premières réflexions et de la réverbération tardive), soit par le biais d'un traitement externe mettant en jeu des unités de réverbération traditionnelles, lesquelles alimentent autant de haut-parleurs virtuels que nécessaire.

Outre les possibilités créatives qu'offre le positionnement précis d'objets sonores dans l'espace, la WFS permet de réconcilier perception visuelle et sonore d'un espace scénique, en assurant la mise en cohérence visuelle et sonore des sources.

Pour la mise en place et le réglage d'un système WFS dans un environnement existant, l'installateur a à sa disposition un outil appelé WaveDesigner.

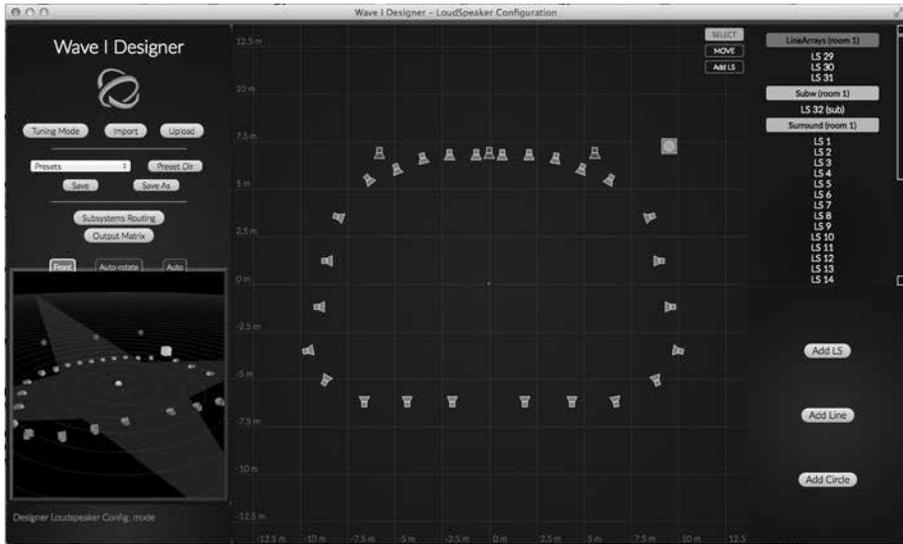


Figure 6.50 – WaveDesigner, outil de configuration/installation du système WFS, © Sonic Emotion.

La fonction première de cet outil est de positionner les haut-parleurs dans l'espace et de transmettre ces informations au processeur qui se charge alors de calculer les coefficients de filtrage nécessaire pour le rendu WFS. Un même mixage peut être diffusé sur des configurations d'enceintes différentes ; lorsque les lieux d'accueil ne permettent pas de placer les enceintes exactement de la même façon, l'outil WaveDesigner, qui prend en compte les nouvelles positions d'enceintes, permet d'adapter parfaitement le mixage au nouveau lieu, dans un temps très court.

Le WaveDesigner met aussi à disposition des outils de calibration du système avec un égaliseur paramétrique (8 filtres peak, low et high shelf, passe haut et passe bas), une commande de niveau, un limiteur et la possibilité d'appliquer des « cross over » pour l'alimentation des systèmes multi voies. Une évolution récente du processeur permet de plus de déclarer jusqu'à 4 sous-systèmes. Ceux-ci peuvent correspondre par exemple à différents éléments du système de diffusion (façade, rappels dans la salle) ou bien des systèmes secondaires (bar, cabine de mixage) avec des réglages de niveau et retard indépendant pour chaque source de chaque sous-système.

6.13.3 WFS, Iosono Spatial Audio Workstation

Iosono (Barco) développe des technologies WFS, dont voici le plug-in développé pour Nuendo, Spatial Audio Workstation :

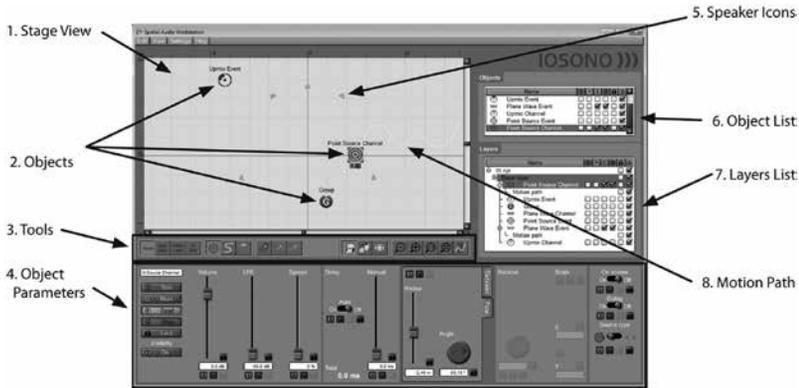


Figure 6.51 – Interface du plug-in Spatial Audio Workstation, Iosono, © Iosono.

Cette technologie orientée mixage objet permet de spatialiser des sources sans la contrainte du système de diffusion. Le système gère la position et le déplacement des objets. L'utilisateur indique la configuration multicanal des enceintes dont tiendra compte le mixage objet. Cette configuration n'affecte en rien le mixage objet élaboré, elle indique simplement comment la position des objets est présentée à la sortie audio du DAW. Cela permet aussi de travailler dans différentes configurations d'enceintes surround sans modifier les données de pan-pot ni l'automation. Le plug-in peut travailler avec le processeur de diffusion WFS Iosono.

Le « stage view » détermine la scène sonore de l'auditeur, avec la représentation des enceintes (« speaker icons »). Le **figure 6.51** représente différents types d'objets, chacun pouvant être déplacé de façon indépendante par rapport aux autres.

La **figure 6.52** présente un exemple de spatialisation d'une source par le procédé d'upmix :

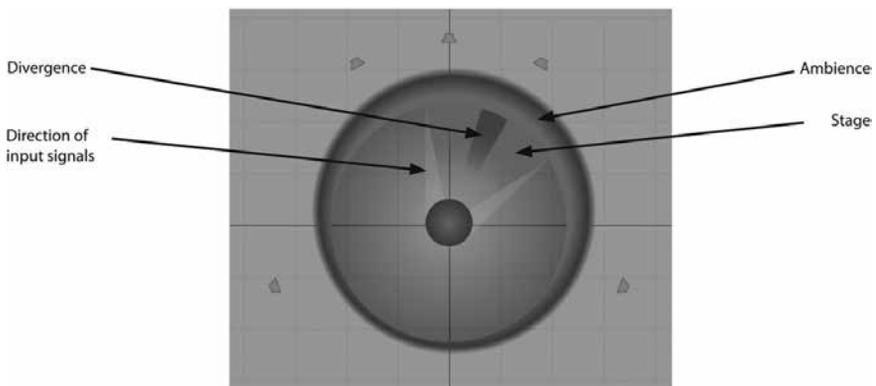


Figure 6.52 – Paramètres d'upmix, © Iosono.

6.13.4 Exploitation Dolby Atmos

Dolby Atmos est un procédé pour la salle de cinéma qui améliore considérablement la spatialisation dans les salles, notamment en décorrélant les sources sur les côtés et à l'arrière. L'écoute est plus symétrique et se rapproche en quelque sorte de la norme 5.1 ITU, quant à la précision de restitution. Il combine le mixage objets et le mixage multicanal traditionnel, ceci afin d'obtenir la meilleure compatibilité avec les différentes salles. L'« objet » s'oppose au « bed » dans le mode de spatialisation. Le « bed » est le mixage multicanal adapté à un dispositif d'écoute tandis que l'« objet » est une source audio accompagnée de l'information de spatialisation. Ce dernier est spatialisé à la diffusion en tenant compte du dispositif. En effet, comme évoqué au chapitre 2, chaque salle de cinéma a ses propres caractéristiques (nombre de haut-parleurs, leur disposition, le volume et la forme de la salle...), ainsi le processeur s'adapte à toutes les configurations, le nombre maximal de haut-parleurs étant de 64. Notons que les subwoofers placés à l'arrière sont alimentés par bass management des canaux surround, d'où la nécessité d'avoir un mixage compatible bass management.

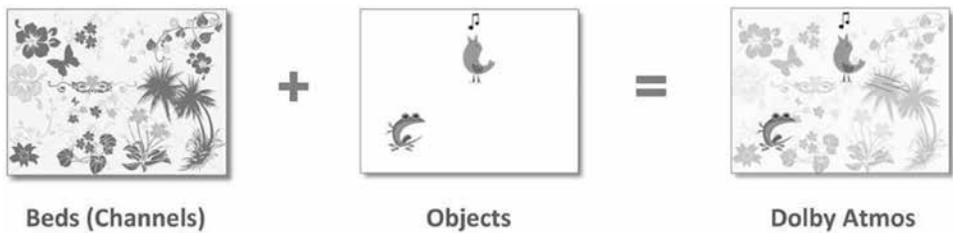


Figure 6.53 – Combinaison objets et « beds », © Dolby.

Un « bed » est compatible 9.1 (7.1 plus 2 canaux de hauteur). Dolby Atmos peut gérer jusqu'à 128 pistes réparties en « beds » et « objets », soit 10 canaux pour le « bed » et 118 canaux pour les objets.

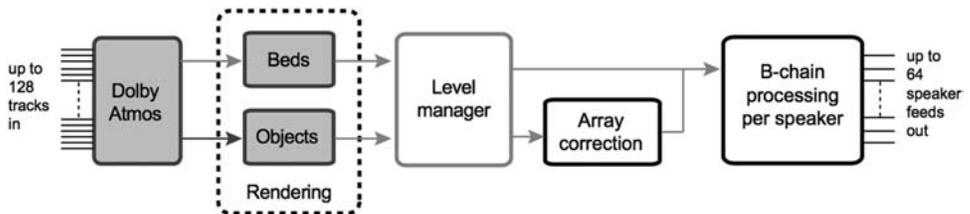


Figure 6.54 – Procédé de rendu sonore Dolby Atmos, © Dolby.

La postproduction est gérée par le RMU (rendering and mastering unit), l'unité de rendu et de mastering Dolby, que l'on contrôle par l'application Dolby Monitor. Le mixeur gère la spatialisation des objets par l'intermédiaire de metadata générées par le plug-in Avid Panner inséré sur chaque piste objet. AMS Neve, Euphonix et Harrison intègrent aussi des outils de spatialisation dans leurs consoles développés conjointement avec Dolby.

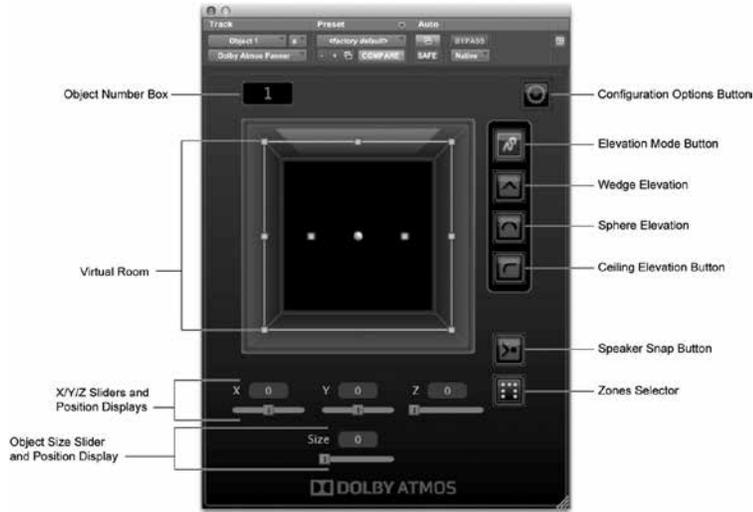


Figure 6.55 – Panner Dolby Atmos, © Dolby.

Le RMU permet à tout moment de simuler des downmix en Dolby Surround 7.1, en 5.1 et en stéréo LtRt ou LoRo.

Dans la phase de mixage, un stem sera donc composé d'un bed principal orienté « channel » et des objets accompagnés de metadata de spatialisation.

La chaîne de production se décompose comme suit, en étapes de montage/premix, mixage final et mastering :

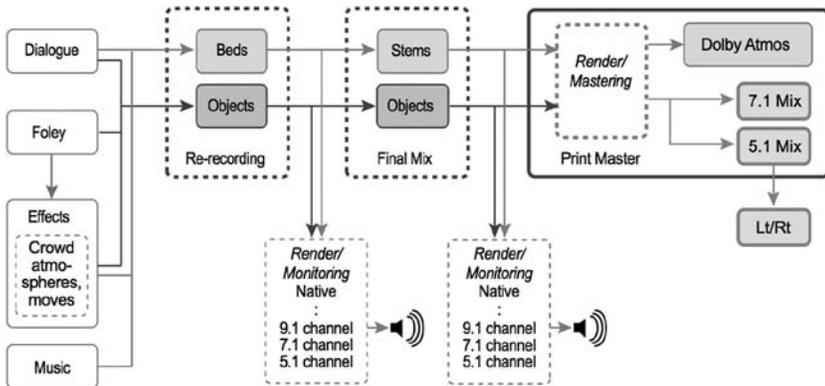


Figure 6.56 – Workflow de production Dolby Atmos, © Dolby.

La technologie Dolby Atmos s'intègre dans le workflow de postproduction sans ajout excessif de temps et de coût, elle est compatible avec le DCP et le processeur assure automatiquement par downmix la compatibilité 7.1/5.1 avec les salles traditionnelles.

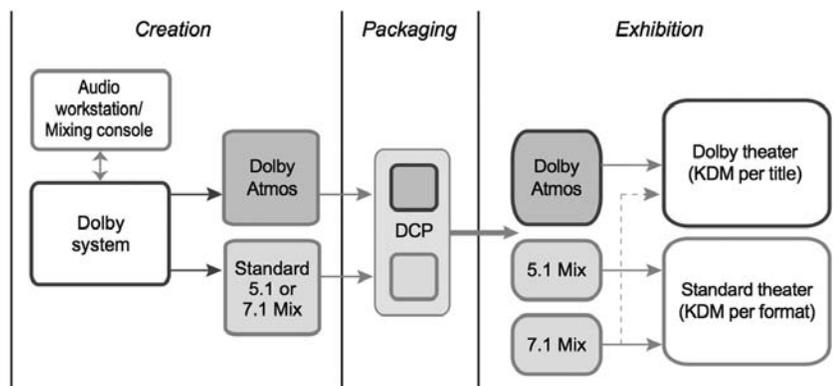


Figure 6.57 – Production et diffusion Dolby Atmos, © Dolby.

6.14 Création sonore et interactivité avec la participation de Thierry Dilger

La création sonore est un domaine où la spatialisation des sons est très courante, si bien que la norme 5.1 est souvent largement dépassée. La restitution en 3D est aujourd'hui très sollicitée par les compositeurs/sound designers. L'interactivité du contenu augmente d'autre part l'intérêt pour l'espace, de plus en plus d'expérimentations sonores spatialisées sont mises en place pour la scénographie.

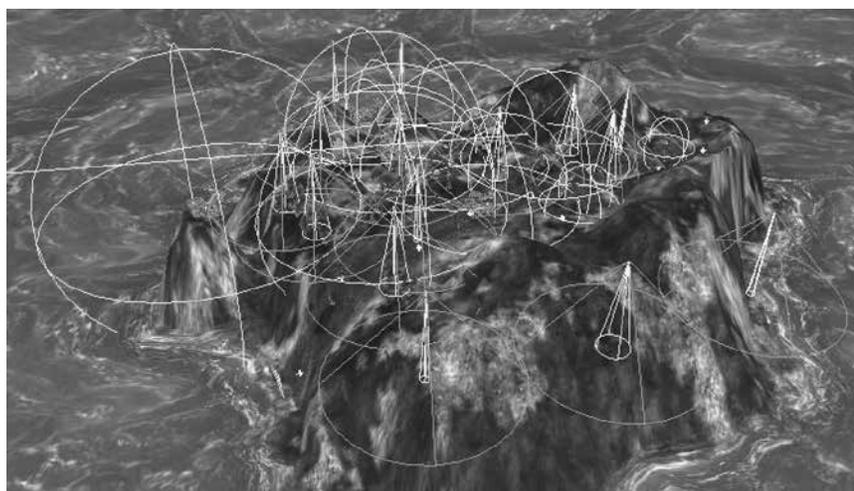


Figure 6.58 – Sound Island, Virtools technology, © Thierry Dilger.

Les outils de création de contenus interactifs sont passés, en quelques années, du stade de la programmation en ligne de codes à celui d'un environnement graphique très élaboré.

Le sound designer (responsable de la qualité du son interactif) possède aujourd'hui des logiciels qui lui permettent de mixer et d'ajuster les comportements sonores (condition de déclenchement, position dans l'espace 3D, effets...) pendant que l'application tourne. Ces outils se sont rapprochés dans leur forme de ceux que connaissent les ingénieurs du son et certains utilisent aussi les surfaces de contrôle avec la prise en charge de l'automation (enregistrement des mouvements). Le sound designer est donc moins sollicité sur la programmation et plus sur le mixage en temps réel.

En multicanal la création interactive a hérité des outils et méthodes issues du cinéma avec les « panner » façon Pro Tools mais aussi et surtout des outils et méthodes issues de la conception graphique 3D. C'est ainsi que de nombreuses créations 5.1 interactives utilisent des sources en mono traduites dans le système surround par la position du « sound emitter » (celui qui produit le son) vis à vis du « listener » (celui qui écoute, en général le joueur). Le rapport de distance dans l'espace 3D détermine directement l'atténuation du volume, la position latérale (azimuth) est retransmise fidèlement dans le système de diffusion, et tout cela en temps réel.

Dans le visuel ci-dessus, vous pouvez observer une carte qui possède 29 « sound emitter », certains étant statiques (position fixe), d'autres dynamiques (en mouvement) comme les mouettes circulant le long du rivage. Ajoutez à cela le fait que le joueur est complètement libre de se déplacer sur cette île et vous comprendrez l'extrême difficulté de faire une mise en scène sonore spatiale intéressante, cohérente et surtout bien mixée.

C'est pourquoi, même si les outils de créations sonores interactifs se standardisent de plus en plus, il reste nécessaire de pouvoir les adapter aux différents cas de figures, voire même de concevoir ses propres outils et méthodes de travail. Aujourd'hui encore, les sound designers les plus créatifs restent ceux qui ont des compétences en audio, mais aussi en programmation.

Complément Web 6.3

Outils pour la création sonore et l'interactivité.

6.15 Applications

6.15.1 *Film*

◆ **Il faut sauver le soldat Ryan (Saving Private Ryan)**

Le Film « Il faut sauver le soldat Ryan », réalisé par Steven Spielberg et mixé par Gary Rydstrom, est connu pour sa première séquence longue d'une vingtaine de minutes, mettant en scène le débarquement allié du 6 juin 1944 sur la plage d'Omaha beach en Normandie. La scène est très réaliste et choquante par sa brutalité. Le film a reçu l'Oscar du meilleur son et du meilleur mixage en 1998. Cette scène met en évidence le

contraste souligné par le son, entre la vision externe du personnage principal, avec une partie FX et ambiances très fournie, constitués d'impacts de balles, d'explosions d'une grande dureté réparties tout autour du spectateur et son ressenti intérieur, exprimé par le point de vue de la caméra qui plonge dans l'eau, et autres courtes séquences où l'on perd les sons des tirs pour entrer dans le point de vue subjectif du soldat.

L'analyse de Gary Rydstrom exposée dans « Surround Sound » de Tomlinson Holman, explique la réalisation son du film qui consiste à placer le spectateur au milieu du débarquement, avec le point de vue claustrophobique du soldat tremblant de peur. Ainsi, le son par son immersion intense, nous fait glisser habilement dans la peau du personnage. Le son spatialisé donne des informations sur l'orientation, l'échelle, l'émotion, données que la caméra n'a pas besoin de préciser. Bien sûr, la bande son est caractérisée par la librairie sonore dense des différents tirs de mitraillettes, des sons isolés de tirs, de balles, d'artillerie, de bateaux, de tanks, d'explosions, de débris.

Comme l'évoque le mixeur, lorsque l'on traite un son in, il suffit de le placer non loin de l'image, dans les canaux frontaux, mais la majorité des sources qui constituent la bande son du film sont des sons off, ce qui offre une grande liberté de spatialisation. L'auteur évoque quelques points intéressants sur l'intérêt de spatialiser un son : la séparation des canaux permet d'éviter l'effet de masque, l'orientation du spectateur est plus précise, les possibilités de dynamique sont bien plus grandes, et le mouvement des sons est possible. Notons aussi l'utilisation des différents points de vue, le point de vue sur l'extérieur de la scène et celui portant sur le ressenti du personnage.

La spatialisation permet d'éviter l'effet de masque

Dans son récit, le mixeur décrit sa façon de gérer l'effet de masque et la symétrie des sons, par exemple lorsqu'il place une explosion basses fréquences dans le frontal, il positionne en symétrie un bruit grave de bateau à l'arrière. Quand un bruit de débris de sable venant d'une explosion est placé sur le côté gauche, un son de passage de balle ayant le même contenu fréquentiel est symétriquement opposé sur la droite ou à l'arrière. Ainsi, on retrouve ces règles qui régissent la cohérence de l'image sonore, des sons qui convergent ensemble, par leur ressemblance, mais disposés dans des zones de l'espace qui s'opposent, formant une sorte de fusion.

Le son spatialisé oriente le spectateur

Le son spatialisé oriente le spectateur : ici, la défense allemande est sur l'arrière droite tandis que les troupes américaines sont dans le frontal, à gauche. Ainsi la caméra s'ouvre à d'autres champs, des détails, au-delà de l'information de localisation.

Les changements d'espaces, dynamique du multicanal

L'auteur évoque les possibilités de dynamique du multicanal qui servent au contraste dont le film a besoin. Le réalisateur coupe cette longue scène de guerre par des brèves

séquences : la caméra qui plonge sous l'eau ou les scènes où le personnage perd momentanément l'ouïe, symbolisé par un sifflement et un resserrement spatial vers la stéréo. Les changements d'espaces offrent ces possibilités de dynamique, s'ajoutant aux variations de niveaux et aux fréquences contrastées.

Le mouvement des sons, une façon de renforcer l'émotion

Le spectateur est très attentif aux changements de localisation des sons, ainsi la séquence comporte de nombreux passages de balles sifflantes, très souvent de l'arrière vers l'avant, précédant l'impact de la balle à l'écran. Les échos des tirs se répondent dans des zones opposées. Ces mouvements nous donnent l'impression d'être dans l'action, ils nous apportent la proximité avec le sujet traité et nous véhiculent l'émotion du personnage en danger de mort. Le mouvement des sons des explosions, des cris des soldats accompagnent aussi le mouvement de la caméra qui explore le champ autour du spectateur.

Les limites des canaux surround

Dans la configuration cinéma, les sons « pan-potés » à l'arrière perdent une partie de leurs fréquences, ils perdent de leur qualité et de leur poids comparés aux canaux avant.

Les canaux arrière au cinéma sont envoyés vers un ensemble de haut-parleurs, qui ne restituent pas un point précis de localisation. Le son restitué est plutôt ressenti comme une ambiance (le Dolby Atmos permettrait cette décorrélation).

Le mixage multicanal repose sur une utilisation intelligente des canaux surround pour envelopper le spectateur sans le distraire et le sortir du scénario du film. Il y a un temps pour l'enveloppement et un temps sans surround. Comme il existe pour l'image des artifices comme la mise au point, la profondeur de champ, les mouvements de caméra ou le découpage, destinés à orienter l'attention du spectateur, il existe des techniques identiques dans le son, comme l'utilisation ou non des canaux arrière.

◆ Didier Lozahic, mixeur film, interview

Didier Lozahic est mixeur film et musique, il aborde le multicanal par le mixage de musiques de film. Son premier grand projet pour le film a été « GoldenEye », dont la musique a été composée par Eric Serra. Il a ensuite évolué vers le mixage film avec notamment sa collaboration avec Luc Besson.

Pour la plupart des projets sur lesquels je travaille, j'ai souvent l'avantage d'enregistrer la musique, de la mixer et de mixer également le film. La véritable première expérience pour moi a été « Le Cinquième élément », je me suis retrouvé dans un auditorium, à Pacific Ocean Post situé à Los Angeles, extrêmement bien équipé en 5.1, dans une configuration cinéma. Il y avait 57 cues de musique à mixer, bien remplis, parfois jusqu'à 64 pistes, ce qui était assez inhabituel à l'époque, et nous avons une vingtaine de jours pour mixer l'ensemble du projet. Lors d'un rendez-vous avec le head mixer du

film, Chris Jenkins, nous avons échangé sur la façon d'amener la musique dans le film, j'étais à l'époque assez inexpérimenté dans l'immersion même du Cinéma. Je suis alors parti du principe où je disposais de six canaux que je pouvais pleinement exploiter, je ne souhaitais pas mixer avec une utilisation dominante de l'image fantôme L-R, et un peu de réverbération placée à l'arrière, qui souvent fait la blague. Je souhaitais gérer un vrai centre physique, avec une cohérence spatiale forte. Je me suis battu pour imposer cette vision, même avec le compositeur qui a l'habitude de faire des maquettes très précises mais qui pensait à l'époque « stéréo élargie ».

Complément Web 6.4

Didier Lozahic ITV.

◆ **Thierry Lebon, mixeur film, interview**

Thierry Lebon est mixeur film et musique, il travaille au début de sa carrière dans l'enregistrement musique, directeur du studio Merjithur. Collaborations : « Sur la piste du Marsupilami », « Un monstre à Paris », « Asterix et Obelix : mission Cléopâtre »

En dix ans, le mixage a bien évolué, les monteurs son arrivent avec des stems 5.1 déjà spatialisés. Comme le temps de mixage se réduit progressivement, on n'a plus le temps de spatialiser les sources au mixage, on n'a pas d'ailleurs le temps de dématricer les formats B, c'est fait au montage son. Mais 80 % des cas cela reste des ambiances stéréo, on utilise souvent deux ambiances stéréo captées dans le même lieu.

Concernant la méthode de travail, je fais toujours un premix paroles, à partir de pistes mono, stéréo et je fabrique alors trois stems : un stem dialogues directs avec à part les post synchros et les sons directs qui pourraient servir à la VI, un stem pour les foules 5.1 avec des paroles françaises, il n'alimentera pas la VI, et un stem constitué des réverbérations 5.0, issues des directs mais séparées. Les bruitages sont composés de 8 à 48 pistes, au format mono ou stéréo, je les prémixe sur un stem 5.1 avec une réverbération 5.0, généralement une Lexicon 960. Pour les musiques, je reçois un mixage complet 5.1, ou des stems 5.1 séparés.

Complément Web 6.5

Thierry Lebon ITV.

◆ **Vincent Arnardi, mixeur film, interview**

Vincent Arnardi est mixeur film, il a collaboré sur un grand nombre de films tels que « Revolver », « Alexandre », « Le fabuleux destin d'Amélie Poulain ».

Avec les technologies actuelles et le nombre de pistes grandissantes à gérer, la nouvelle génération de monteurs son préparent d'une manière plus sophistiquée les sources sonores destinées au mixage.

La multiplicité et la demande sonore étant plus fortes de la part des réalisateurs, l'approche est beaucoup plus précise, donc le travail de la bande son est bien avancée à l'issue du montage son pour qu'au moment du mixage, le mixeur se concentre sur d'autres problématiques : l'harmonie des sons les uns par rapport aux autres.

La colonne vertébrale d'un film est le dialogue, le récit, il se trouve le plus au centre de la salle. Il arrive que l'on spatialise certaines sources dialoguées et on met en scène des espaces, des mouvements. L'aspect émotionnel est donné par la musique. Les ambiances et les FX constituent l'habillage, le costume, ils donnent les profondeurs de champ et s'associent avec la musique aussi dans le contexte émotionnel.

**Complément
Web 6.6**

Vincent Arnardi ITV.

6.15.2 *TVHD : habillage sonore Arte HD 2008*

**Complément
Web 6.7**

Thème et moutons Arte 2008.

6.15.3 *Musique de film : Requiem pour une tueuse*

**Complément
Web 6.8**

Musique *Requiem pour une tueuse*.

Chapitre 7

Le codage et la diffusion

POINTS ESSENTIELS

- Principes du codage : codage entropique et codage perceptif.
- Échantillonnage et quantification.
- Modèles psychoacoustiques.
- Méthodes d'évaluation subjective.
- Codages MPEG, Dolby, DTS.
- Supports : DVD, SACD, Blu-ray.
- Diffusion : DVB, DAB, DCP, jeu vidéo, streaming.

7.1 Introduction

7.1.1 Définition

Ce chapitre traite du codage numérique, qui est une transformation du signal audio afin de réduire son débit et faciliter sa transmission ou son stockage.

Le format PCM a été le premier codage couramment utilisé et adopté par la norme CD audio. Cette simple numérisation du son est déjà un codage puisqu'en définissant une

fréquence d'échantillonnage, on limite la fréquence la plus haute et en définissant la quantification, on impose une dynamique maximale.

L'avantage incontestable du numérique réside dans la possibilité de conserver, transmettre et dupliquer indéfiniment sans dégradation.

7.1.2 *Pourquoi réduire le débit ou la taille ?*

Quel que soit le support, physique ou virtuel, la taille du contenu ou la bande passante nécessaire à sa transmission, ont un coût et il est généralement souhaitable d'en réduire le débit, pour des raisons économiques ou pratiques.

7.2 Principes

Deux voies peuvent s'associer : le codage entropique qui réduit le débit mais ne supprime aucune information et le codage perceptif conçu pour supprimer les informations inaudibles.

7.2.1 *Codage entropique*

Quand un signal audio est numérisé sous format PCM, une partie de l'information numérique est inutile ou redondante : parce que la nature du signal audio n'est jamais totalement aléatoire et parce que le canal complet de 0 à FS/2 n'est pas complètement occupé.

Des informations peuvent être supprimées sans dégradation par un codage alors appelé codage entropique qui doit satisfaire aux exigences suivantes :

- permettre de reconstituer le signal originel ;
- être robuste aux erreurs de transmission ;
- contrôler et limiter le débit maximal.

La réduction de taille possible par ces formats « lossless » est limitée à environ 50 % pour les signaux musicaux habituels. Une réduction de débit plus importante nécessite des codages perceptifs.

7.2.2 *Codage perceptif, enlever l'inaudible*

Dans le signal audio enregistré et numérisé, une partie du signal codé est inaudible. Par exemple, pour une quantification linéaire, un signal faible, qu'il soit seul ou en présence d'un signal fort est codé de la même façon. Pourtant en présence du signal fort, certains signaux faibles sont masqués. Les codages perceptuels permettent de réduire la taille ou le débit en supprimant ces informations non perceptibles. La complexité et l'évolution de ce type de codage viennent des améliorations de modélisation du système auditif pour supprimer le maximum d'information en maîtrisant la dégradation audible.

◆ Paramètres de codage

Qualité

La qualité auditive est liée à différentes caractéristiques techniques dont principalement le débit. La mesure de la qualité est abordée ultérieurement et concerne le codage perceptif et non le codage entropique.

Débit

Le débit, donné en bits par seconde, (kbit/s ou Mbps) est une caractéristique souvent utilisée pour qualifier un codage.

Complexité

Les complexités du codage et du décodage sont des éléments importants des systèmes matériels ou logiciels. Les utilisations étant asymétriques, un codeur et des millions de décodeurs pour une diffusion radiophonique ou télévisée, c'est la simplicité du décodage qui est favorisée. En principe, les spécifications ne définissent que le décodage, ce qui permet les progrès du codage en restant compatible avec les décodeurs en utilisation.

Latence

La latence est un délai de traitement, présent au codage et au décodage. Selon l'exploitation, les impératifs de latences sont différents : une conversation téléphonique exige une latence très faible. Par contre le délai de codage d'un signal audio pour un support DVD ou Blu-ray n'a pas d'importance.

Robustesse

Elle caractérise la résistance du codage aux erreurs et pertes lors des diffusions ou transferts. Rajouter des codes pour éviter les erreurs lors de la reconstruction du signal est toujours un compromis entre débit et robustesse.

7.2.3 Les bases du codage

◆ Traitement du signal

Le signal acoustique perçu par notre appareil auditif est de nature mécanique : les vibrations des molécules d'air font vibrer le tympan. Les caractéristiques principales en sont :

- la bande passante perçue entre 15 Hz et 20 kHz environ ;
- la dynamique du signal, estimée entre 0 dB et 130 dB SPL, représente l'écart entre le signal le plus faible perceptible et le niveau au seuil de la douleur.

Ces aspects sont associés aux deux principales caractéristiques de la numérisation du signal : l'échantillonnage et la quantification.

Numériser le signal consiste à mesurer la valeur instantanée du signal à un temps donné et le temps entre deux mesures successives détermine la fréquence d'échantillonnage.

Échantillonnage

Pour conserver toute l'information d'un canal, le théorème de Nyquist-Shannon exige que la fréquence d'échantillonnage soit supérieure au double de la fréquence maximale du signal.

Pour éviter des phénomènes de repliement de spectre et les distorsions en découlant, il faut strictement limiter la fréquence du signal avant numérisation. Compte tenu de la pente finie du filtrage, la bande passante utile est sensiblement inférieure à la fréquence de Nyquist qui est définie à la moitié de la fréquence d'échantillonnage.

Les fréquences d'échantillonnage du CD (44 100 Hz) ou du DVD (48 000 Hz) permettent des bandes passantes théoriques respectives de 22 050 Hz ou 24 000 Hz qui satisfont pleinement le critère de la bande passante audible.

Quantification

Chaque valeur instantanée du signal, mesurée à la fréquence d'échantillonnage, est représentée par une valeur de quantification. Pour un codage en N bits, on peut ainsi coder 2^N valeurs.

En quantification uniforme, les intervalles sont égaux et la dynamique se calcule par $20 \cdot \log(2Q)$ soit $6.02 \cdot Q$, donc 96 dB pour 16 bits et 144 dB pour 24 bits.

Dither et Noise-shaping

Les techniques de « dithering » et de « noise-shaping » permettent d'améliorer considérablement les performances objectives et subjectives du codage.

Par troncature lors de la numérisation, un signal de niveau inférieur au premier bit de quantification disparaît. Mais en rajoutant un autre signal de niveau un peu plus élevé, l'ensemble dépassera le seuil de quantification et après numérisation, on retrouvera en partie le premier signal ! Le signal de dither ajouté est généralement un bruit et la technologie du noise-shaping répartit ce bruit dans les hautes fréquences, moins audibles. Ces technologies sont utiles lors d'opérations sur le message numérique : réduction de la quantification, de 24 à 16 bits par exemple, diminution du gain, mais sont à appliquer avec précaution afin de limiter le niveau du bruit rajouté. Des techniques similaires sont directement appliquées dans certains convertisseurs A/D et D/A.

Complément Web 7.1

Effet de l'ajout du signal de dither.

◆ Transformation temps-fréquence

Le signal acoustique évolue temporellement mais l'analyse faite par le système auditif, dont la membrane basilaire de l'oreille interne, est plutôt de nature fréquentielle. Certaines performances de l'oreille étant plus faciles à caractériser dans un domaine fréquentiel, des outils mathématiques, les transformées, permettent de passer d'un domaine à l'autre. Dans le cas du codage audio numérique, les transformées sont basées soit sur des banques de filtres soit sur les transformées de Fourier discrètes (DFT). Ces transformées permettent de calculer le spectre d'un signal stationnaire ou non. Généralement on scinde le signal en blocs en imaginant que ceux-ci se répètent périodiquement. Les transformées « modifiées » utilisent une propriété de traitement de signal appelée TDAC (time-domain aliasing cancellation) qui permet de calculer sur des fenêtres consécutives se recouvrant et, sous certaines conditions, d'éviter d'introduire des erreurs dues au repliement temporel.

7.2.4 Bases de psychoacoustique

◆ Les seuils d'audibilité

La réponse de l'oreille n'est linéaire ni en fréquence ni en niveau, ce qui est constaté sur les courbes d'audibilité en **figure 7.1**.

Ces réponses sont approximatives et représentent une moyenne qui ne correspond pas forcément à la perception individuelle.

DÉFINITIONS

- dB SPL : mesure physique ayant pour référence à 0 dB une pression sonore de 20 micropascals.
- Phones : mesure du niveau sonore perçu référencé à 1 kHz. Un son de 60 phones à 300 Hz est perçu au même niveau qu'un son de 60 phones à 1 kHz alors que leurs niveaux SPL sont très différents.
- Sones : représentation du niveau sonore perçu. Un son de 2 sones est perçu comme deux fois fort qu'un son de 1 sone. Un sone correspond à un son à 40 phones donc une pression SPL de 40 dB SPL à 1 kHz.

◆ Masquage temporel

Le masquage temporel est un processus lié à l'oreille interne qui rend inaudibles les signaux faibles précédant ou suivant un signal fort. Le temps de post-masquage peut s'étendre à 100 ms, alors que le prémasquage est limité à environ 20 ms.

◆ Masquage spectral

On parle de masquage fréquentiel ou de masquage simultané quand un son n'est pas perçu parce qu'il est couvert par un son de niveau plus élevé mais simultané.

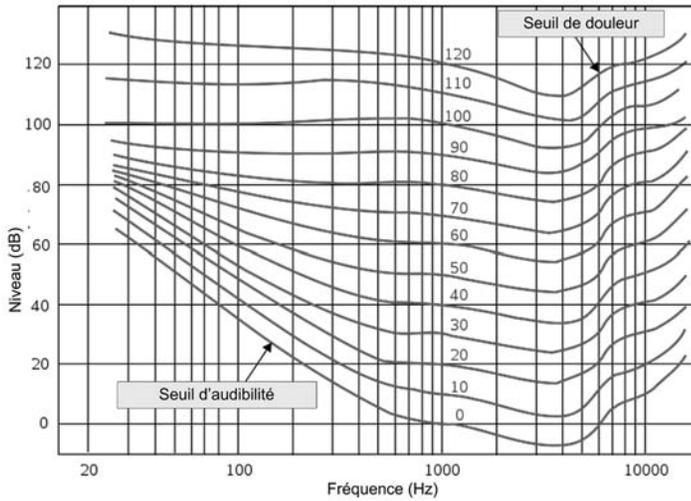


Figure 7.1 – Courbes d'isophonie © Wikimedia Commons.

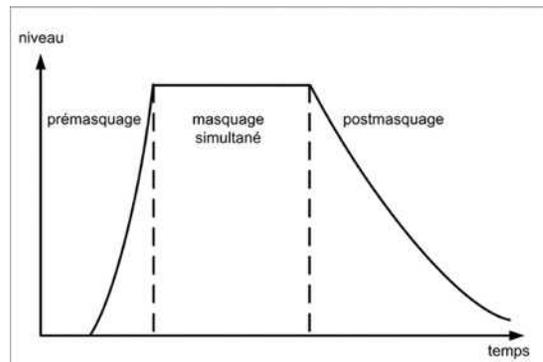


Figure 7.2 – Masquage temporel, © J.-L. Ohl.

Le seuil de masquage, différence minimale de niveau entre le son masquant et le son masqué, dépend d'un grand nombre de paramètres dont le contenu spectral. Par la membrane basilaire et ses cils, l'oreille interne effectue un filtrage, une transformation temps vers fréquence, qui induit la notion de « bande critique ».

Cette notion de « Bande Critique » est fondamentale en psychoacoustique : quand une tonalité est présente, les autres fréquences situées dans la bande critique sont masquées.

Sa largeur fréquentielle a d'abord été exprimée en Barks par Zwicker. Moore a approfondi et précisé ces courbes en exprimant la bande de masquage en ERB (Equivalent

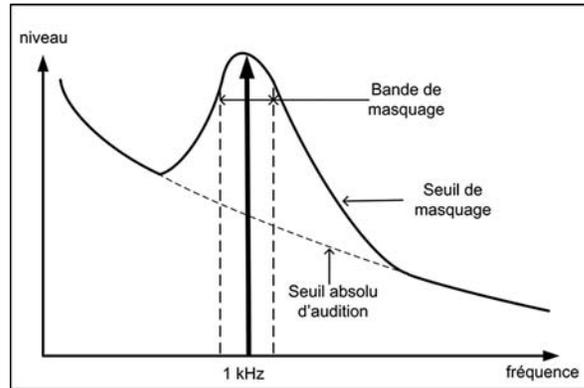


Figure 7.3 – Largeur de bande critique © J.-L. Ohl

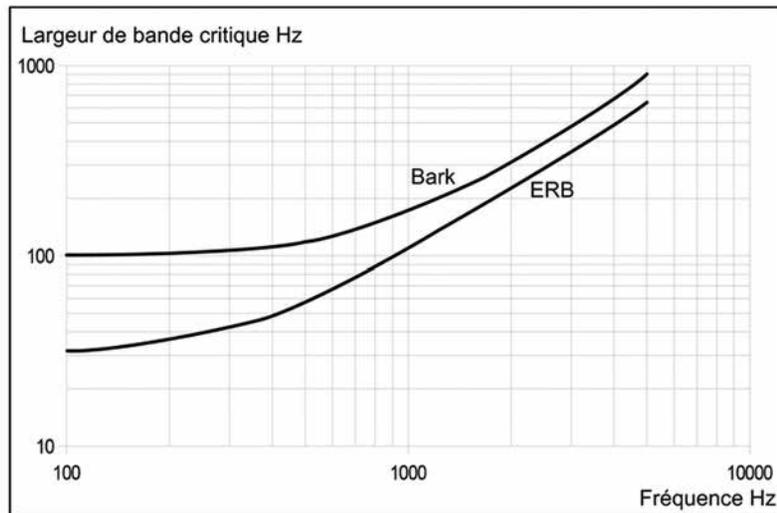


Figure 7.4 – Masquage fréquentiel, © J.-L. Ohl.

Rectangular Bandwidth). Dans les moyennes fréquences, les bandes présentent une largeur d'environ 1/6e d'octave. Dans le registre grave, les performances sont bien moindres.

7.2.5 Modèles psychoacoustiques

Un modèle psychoacoustique est une base de données qui intègre les caractéristiques de l'audition, principalement les phénomènes de seuils et de masquage pour les appliquer au traitement du signal de codage. Le signal est coupé en blocs temporels filtrés et analysés fréquentiellement. Chaque bloc est alors requantifié en appliquant les paramètres déterminés par le modèle.

La taille de la fenêtre temporelle doit répondre à deux exigences contradictoires : augmenter la taille pour assurer une bonne résolution fréquentielle mais la limiter pour

suivre fidèlement les signaux courts (percussifs, par exemple). En effet, les erreurs sont définies par la notion de bruit de codage, réparti sur la durée du bloc. Ce bruit dans un bloc court est plus proche temporellement du signal et donc mieux masqué. L'effet typique dû à une longueur de bloc trop importante est le pré-écho, une erreur ressemblant au signal à coder mais sans bénéficier du prémasquage.

Dans les codages évolués, la taille des blocs est asservie au signal, des blocs longs étant plus efficaces pour la réduction de débit et des blocs plus courts sont utilisés si l'audio a un caractère transitoire (128 échantillons donc 3 ms au minimum dans le cas de l'AAC). Les erreurs sont toujours présentes mais mieux masquées subjectivement parce que plus proches temporellement du signal.

Le spectre instantané extrait du signal est comparé aux courbes de masquage fréquentiel et le contenu sous le seuil n'est pas codé. Le signal restant, n'est codé qu'au-dessus de la courbe de masquage, en 4 ou 5 bits seulement au lieu des 16 ou 20 de départ.

La préparation des éléments est un aspect primordial du codage : si les signaux sont bruités, une partie du débit sera utilisée pour coder le bruit de nature aléatoire. Il est donc important, en audio comme en vidéo, de « nettoyer » les éléments avant toute compression.

Un aspect des modèles psychoacoustiques est lié à la perception spatiale : la localisation résulte des différences entre les signaux à chaque oreille, des variations de ces signaux lors des mouvements de la tête et des schémas cognitifs. Les différences fondamentales entre les signaux dichotiques sont celles liées à la phase interaurale dans les basses fréquences (jusqu'à environ 700 Hz), à l'écart d'amplitude et aux écarts temporels entre enveloppes des signaux (délai de groupe) au-dessus de ces fréquences.

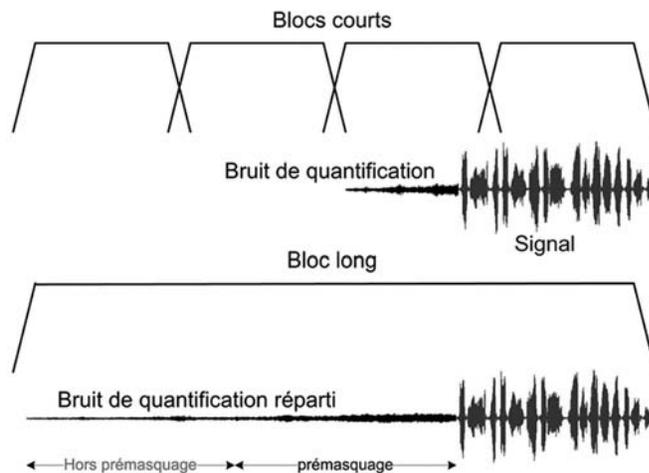


Figure 7.5 – Influence de la durée des blocs de quantification, © J.-L. Ohl.

MODÈLE PSYCHOACOUSTIQUE

- Seuils d'audibilité.
- Limite dans les hautes fréquences.
- Masquage temporel.
- Masquage simultané et fréquentiel.

Fonctions du système auditif

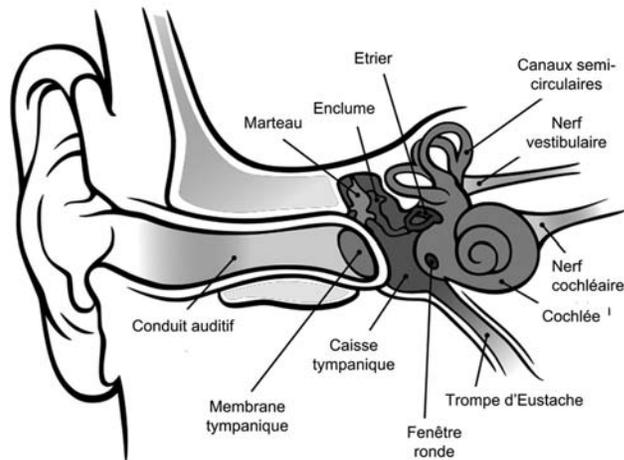


Figure 7.6 – Schéma de l'appareil auditif, © Wikimedia Commons.

Appareil auditif	Organes	Traitement du signal
Oreille externe	pavillon, conduit auditif	filtrage HRTF pour la localisation
Oreille moyenne	tympan, osselets, fenêtrée ovale	réduction de la dynamique
Oreille interne	cavité en colimaçon, cochlée, membrane basilaire, cellules ciliées	transformée temps-fréquence, filtrage en bandes critiques
Système nerveux	nerf auditif, noyau olivaire, colliculi, cortex	seuil de détection, filtrage, verrouillage de phase, localisation, signification

Figure 7.7 – Principales fonctions de l'appareil auditif, © J. L. Ohl.

◆ Quantification et codage entropique

Quantification et densité

Une quantification faite avec un dictionnaire de dimension unitaire est appelée quantification scalaire, ce qui est le cas en audio. Sinon il s'agit d'une quantification vectorielle. La quantification scalaire est uniforme quand les pas de quantification (niveaux) sont égaux.

Si la répartition de la densité des niveaux du signal est prévisible, il est plus efficace d'utiliser une quantification non linéaire. Les codages A-law et μ -law des télécommunications, utilisent une compression de la dynamique du signal par un codage en 8 bits sur une échelle logarithmique pour coder les 16 bits PCM. En cas d'estimation erronée de la répartition, la dégradation peut être importante.

Complément Web 7.2

Quantification non linéaire.

◆ Codage entropique

Les informations d'un signal numérisé n'occupent jamais la totalité du canal de transmission, une partie est inutile ou redondante. Le codage entropique propose de réduire le débit de codage sans perdre d'information, le décodage retrouvant totalement les informations d'origine.

Une technique utilisée en codage entropique est celle des codes de Huffman où les caractères les plus fréquents sont codés avec moins de signes. Appliqué à la langue française, cela revient à coder la lettre "e" avec le code le plus court et la lettre "w" avec un codage long. L'apparition du "e" étant plus fréquente que celle du "w", la taille du texte codé est réduite.

Le codage LZW (Lempel-Ziv-Welch) est comparable à celui de Huffman mais avec un dictionnaire adaptatif redéfini au fur et à mesure selon la fréquence d'apparition des caractères.

Prédiction linéaire

Un signal audio n'est pas aléatoire et il peut être partiellement prédit en se référant aux échantillons précédents. La prédiction linéaire est un codage différentiel par quantification vectorielle et seule l'erreur résiduelle reste à coder.

Cette technologie est efficace sur les signaux musicaux et particulièrement sur la voix, signal constitué de formants (base des voyelles qui sont des signaux sinusoïdaux modulés en fréquence et en niveau), et des consonnes produites par obstruction du passage de l'air, donc des signaux d'une nature plus bruitée et plus aléatoire. On code

ainsi économiquement en termes de débit les formants auxquels on rajoute les bruits des consonnes.

Banques de filtres, codage par sous-bandes

Dans ce type de codage, le signal est filtré par bandes de fréquences puis codé indépendamment dans chaque bande. Compte tenu du masquage fréquentiel, un signal fort va masquer les signaux plus faibles de la même bande. Ceci permet de réduire considérablement la quantification par bande : au lieu de 16 ou 24 bits, il est possible de réduire la quantification à 4 ou 6 bits.

Souvent utilisés pour ces applications de filtrage par sous-bandes, les filtres miroirs en quadrature (PQMF) sont implémentés en traitement du signal numérique pour séparer la bande passante par filtre passe-bas et filtre passe-haut complémentaires. En mp3, ATRAC et MPEG4-AAC-SSR, ces filtres PQMF sont suivis d'une transformée MDCT, on parle alors de filtrage hybride.

Reconstruction de Bande Spectrale (SBR)

Cette technologie supprime une partie du spectre des hautes fréquences, procurant une nette réduction de débit. Au décodage, un spectre reconstitué complète les harmoniques manquantes.

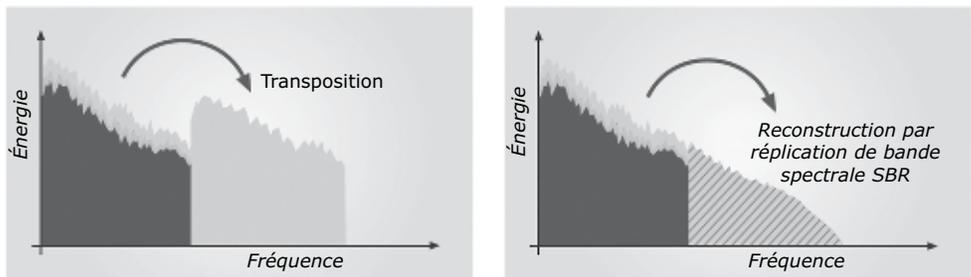


Figure 7.8 – Reconstruction de bandes spectrales, © Coding Technologies.

Réservoir (buffer)

Le signal audio étant peu prévisible, il est difficile de connaître par avance le débit instantané nécessaire. Un réservoir tampon est permis d'absorber les pics de débit et se libère à rythme constant. Ainsi le nombre de bits par trame du signal est variable mais la moyenne reste constante.

Common bit pool

Dans le cas de signaux multicanal, des algorithmes permettent de répartir les données selon le besoin de chaque canal.

◆ Codage spatial (Spatial Audio Coding)

Quand plusieurs canaux sont codés, la corrélation entre les canaux permet des réductions de débit. L'ensemble de ces techniques est souvent appelé «Joint coding» et «Joint stéréo» pour 2 canaux. Les formats MPEG Surround reposent sur les technologies SAC.

Codage M/S

Dans un signal stéréophonique, on peut décomposer les signaux en signal commun et en signal différentiel. Avec G pour gauche, D pour droite, M, signifiant «Mitte» (milieu) et S pour «Seite» (côté) :

$$M. = (G + D)/2 \text{ et } S = (G-D)/2$$

$$\text{est équivalent à : } G = M/2 + S/2 \text{ et } D = M/2 - S/2$$

Ce codage M/S réduit sensiblement le débit quand le niveau du signal commun M est bien plus important que celui du signal de différence S. Un calcul similaire peut s'appliquer en multicanal.

Joint Frequency encoding

Au-dessus de 5 kHz, l'oreille est moins performante. La direction apparente est liée à l'enveloppe des signaux : à partir de 2 kHz, la perception binaurale est surtout liée à l'ILD (Interaural Level Difference). Il est possible de coder le signal comme un signal monophonique sommé, avec un coefficient par canal. Lors du décodage, le signal mono est envoyé à chaque canal selon le coefficient. Pour deux canaux, on parle de stéréo intensity.

◆ Codages paramétriques

Pour réduire le débit, les techniques de codage paramétrique décomposent le signal en éléments décrits par des modèles et des paramètres. Les éléments perceptifs sont modélisés plutôt que les seuls paramètres physiques : un signal de base simple est codé en associant des paramètres complémentaires décrivant les éléments psychoacoustiques plus complexes.

Codage sinusoïdal

Le modèle le plus usité est le codage sinusoïdal où le signal est représenté par des composantes sinusoïdales, dont on définit les fréquences et amplitudes instantanées. Un modèle hybride rajoute des transitoires et du bruit. Par exemple, le modèle MPEG4-paramétrique (HILN) définit les composantes sinusoïdales et les composantes harmoniques par leurs fréquences et amplitudes, par l'enveloppe spectrale des harmoniques et le bruit par son amplitude et son enveloppe spectrale.

Autre exemple, les méthodes de la stéréo paramétrique permettent d'encoder plusieurs canaux en un seul flux monophonique associé à des informations liées à la localisation spatiale :

- différence d'intensité entre canaux (ILD) ;
- différence de phase entre canaux (ITD) ;
- cohérence (intercorrélation) entre canaux (ICC InterChannel Cross Correlation).

Grâce à l'information de cohérence, la stéréo paramétrique est utilisable sur l'ensemble du spectre, contrairement à la stéréo d'intensité. Une technologie similaire à l'ICC mais appliquée en multicanal est appelée BCC (Binaural Cue Coding). Le signal commun monophonique est encodé et accompagné des paramètres d'ILD, ITD et ICC permettant de reconstituer les informations de spatialisation. Ces paramètres supplémentaires sont transmis dans les données auxiliaires. Les codages dits « hybrides » associent filtres, transformées et codage paramétrique.

MPEG Surround

Le codage paramétrique est une des bases du MPEG Surround qui permet de coder et décoder par étapes en séparant de façon matricielle les informations de base, puis en ajoutant ces paramètres de directionnalité (side information), tout en conservant une compatibilité simple entre formats.

Complément Web 7.3

Codage paramétrique des informations directionnelles du multicanal.

◆ Débit variable

Constant Bit Rate et Variable Bit Rate

Le débit d'un flux encodé peut être soit constant (CBR) soit variable (VBR). Un débit constant ne s'adapte pas à la complexité du message et le canal d'information n'est pas utilisé de façon optimale. Un débit variable suit la difficulté de codage du message en conservant comme cible une valeur moyenne de débit. Généralement, un « réservoir » assure une constance dans le flux des données, en adaptant le débit instantané. En débit variable, la valeur nominale est fixée mais certains codages permettent de définir les débits minimal et maximal et la taille du réservoir.

7.3 Qualité subjective

Tout codage est un compromis entre différents paramètres mais surtout entre débit et qualité sonore. L'évaluation de la qualité est délicate, les tests d'écoute étant subjectifs et donc complexes, longs et coûteux. La **mesure objective de la qualité subjective** propose de remplacer les écoutes par des mesures objectives dont il faut vérifier la pertinence et les limites.

7.3.1 Méthodes d'évaluation perceptuelle

Selon la qualité du codage, il existe différentes méthodes d'évaluation adaptées à l'importance des défauts. Les défauts perçus facilement sont classés en distinguant les défauts inacceptables et ceux bénins. Pour un codage proche de la perfection, c'est plutôt la présence d'un défaut et son audibilité éventuelle qui est étudiée.

Les principales méthodes d'évaluation sont :

- méthodes par détermination de qualité absolue ;
- méthodes par dégradation ;
- méthodes par détection de différence.

Guides généraux			
BS. 1283-1 Guide des recommandations IUT-R sur l'évaluation subjective de la qualité du son		BS.1284-1 Méthodes générales d'évaluation subjective de la qualité du son	
Méthodes de tests d'écoute			
BS. 1116-1 Méthodes d'évaluation subjective des faibles dégradations dans les systèmes audio, y compris les systèmes sonores multivoies	BS.1285 Méthodes de présélection pour l'évaluation subjective des faibles dégradations dans les systèmes audio	BS.1534 Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage	BS.1679 Évaluation subjective de la qualité audio dans les applications d'imagerie numérique sur grand écran destinées à la présentation en salle.
Modèle objectif			
BS.1387-1 Méthode de mesure objective de la qualité du son perçu			
Audio-visuel			
BT.500-13 Méthodologie d'évaluation subjective de la qualité des images de télévision	BS.775-1 Système de son stéréophonique multicanal avec ou sans image associée	BS.1286 Méthodes d'évaluation subjective des systèmes audio avec image d'accompagnement	

Figure 7.9 – Normes ITU-R pour l'évaluation subjective, © Nick Zacharov.

Complément Web 7.4

Quelles normes pour quel besoin ?

◆ Méthodes en double aveugle

Double aveugle signifie que ni l'auditeur, ni le testeur ne connaissent l'identité du stimulus écouté à l'instant. Ce qui élimine le biais du testeur influençant, même inconsciemment, le résultat.

◆ Méthodes pour dégradations légères

La norme ITU-R BS.1116 (double blind triple stimulus with hidden reference) propose une méthode pour analyser les dégradations faibles de la qualité. Il est proposé trois stimuli à l'écoute, A étant le stimulus d'origine non dégradé et B ou C étant aléatoirement identique à A. Il s'agit de noter la qualité perçue de B et C sur l'échelle de dégradation présentée ci-dessous. Il est intéressant de consulter les normes parce qu'elles présentent de façon exhaustive les conditions pour mener correctement les essais.

ITU BS-116

Dégradation	Echelle
Imperceptible	5.0
Perceptible mais non gênant	4.0
Légèrement gênant	3.0
Gênant	2.0
Très gênant	1.0

IEC 60268-13

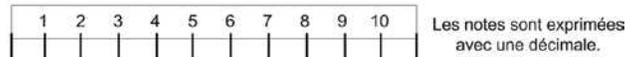


Figure 7.11 – Échelle de dégradation de la norme ITU-R BS 1116 et IEC 60268-13, © J.-L. Ohl à partir de ITU-R BS 1116 et IEC 60268-13.

Une autre méthode utilisée quand il s'agit de déterminer si la dégradation est perceptible, est la méthode ABX en double-aveugle. Deux sources, A étant le stimulus d'origine non codé et B le codage sont comparées.

Principe du test : X est choisi au hasard, étant A ou B. L'auditeur écoute et bascule entre A, B et X. Ensuite l'auditeur décide si « X est A » ou « X est B ». Puis une nouvelle séquence est démarrée avec X choisi au hasard. Un score suffisant est nécessaire pour obtenir un résultat statistiquement valide : par exemple 15 réussites sur 16 essais correspond à une probabilité de réussite « par hasard » inférieure à 1/1000. Dans cette méthode, certains critères objectifs sont à respecter : niveaux alignés à moins de 0,2 dB, commutations imperceptibles et identiques.

Complément Web 7.5

Interface d'un logiciel de test ABX.

7.3.2 Mesures perceptuelles

◆ **Mesure objective de la qualité subjective**

Les modèles psychoacoustiques permettant de coder en supprimant les éléments inaudibles d'un signal, peuvent aussi être utilisés dans le but de vérifier qu'un codage n'a pas supprimé une partie audible, ou encore pour mesurer la qualité subjective. La complexité de l'organisation et la difficulté des écoutes subjectives incitent à l'utilisation des tests objectifs qui deviennent plus performants.

◆ **Norme ITU-R BS.1387**

Comme les méthodes de mesure habituelles (distorsions, bruit...) ne sont pas utilisables pour un signal avec codage perceptif, il a été nécessaire de définir des normes permettant une mesure objective de la qualité, si possible en temps réel, pour remplacer les mesures physiques et les écoutes subjectives.

Le modèle cognitif calcule une note différentielle objective (ODG, Objective Difference Grade) à partir des dégradations mesurées :

- Réduction de bande passante ;
- Rapport bruit/masque ;
- Distorsions linéaires, distorsions non linéaires et bruit ;
- Structure harmonique du signal d'erreur.

Une échelle de dégradation entre 0 (dégradation imperceptible) et -4 (dégradation gênante) est utilisée pour cette note ODG.

7.3.3 Considérations pratiques

Les conditions d'écoute permettant des tests efficaces et valides sont précisées dans les normes dont l'ITU-R BS.1116-1 pour les tests subjectifs.

◆ **Choix des auditeurs**

Surtout pour les dégradations faibles, il est plus efficace de choisir des auditeurs entraînés : les défauts sont plus rapidement détectés et avec une meilleure fiabilité.

◆ **Choix des extraits**

Certains types sonores sont connus pour mettre en évidence les défauts : le diapason, le clavecin, le glockenspiel, les applaudissements sont parmi les séquences réputées difficiles à coder comme certains extraits SQAM (Sound Quality Assessment Material), disponible sur le site internet de l'EBU, associés au document explicatif EBU Tech 3253. Le site hydrogenaudio.io propose aussi des liens vers des séquences délicates pour le codage.

7.3.4 Comparaison de codage

Les comparaisons peuvent se faire à partir de fichiers ou de flux. L'écoute comparative de codage nécessite absolument de s'assurer de l'égalité des niveaux, qui n'est pas garantie par les décodeurs. Il est souvent plus simple de reconvertir les fichiers en format PCM, d'aligner les niveaux et de comparer les fichiers décodés. S'il s'agit de flux, il est préférable d'utiliser des décodeurs séparés afin d'éviter le temps de latence lors de la commutation entre formats.

7.3.5 Artefacts typiques de codage

Sans essayer d'être exhaustif, on peut présenter quelques problèmes subjectifs liés aux codages perceptifs.

Défaut	Causes possibles	Exemple de signaux de test
Pertes dans l'aigu	Filtrage passe-bas	Clavecin, cornemuse
Pré-écho	Taille de la fenêtre de codage	Caisse claire, castagnettes
Distorsions des transitoires, chirp	Taille de la fenêtre de codage	Caisse claire, castagnettes, applaudissements
Coloration de timbre		Clarinette, voix, glockenspiel
phasing		Bruit de pluie
Distorsions spatiales (resserrement)	Traitement M/S	Orchestre classique
Pompage		Piano
Bruit	Erreurs par démasquage binaural	Source quasi-mono

Figure 7.12 – Tableau des principaux défauts de codage.

◆ Évaluation en stéréo ou en multicanal

En raison du phénomène de décoloration binaurale, la détection des artefacts est quelquefois plus aisée en monophonie mais certains artefacts de distorsions spatiales (le resserrement de l'image sonore par exemple) ou l'apparition d'artefacts dans les canaux adjacents, ne sont mis en évidence que lors d'écoutes stéréophoniques ou multicanal.

7.3.6 Codage en cascade

Certains codages ne sont pas étudiés pour une utilisation en cascade qui peut alors engendrer une dégradation importante. Cette pratique est pourtant fréquente dans en diffusion radio ou télévisée, mais est difficile à éviter parce que le workflow est figé et que les fichiers ou les flux ne portent pas les informations des codages précédents. Dans un schéma de diffusion, entre captation, transport, contribution, montage et diffusion, il est fréquent qu'un flux ou fichier audio subisse 4 ou 5 codages successifs.

7.3.7 Canaux discrets et canaux matricés

Un contenu multicanal peut être présenté de deux façons différentes : canaux indépendants (discrets) ou canaux matricés.

◆ **Matriçage**

Le terme matriçage est employé dans deux cas très différents :

- par exemple en codage M/S, les canaux audio sans compression sont transmis par le même nombre de canaux et il n'y a pas de gain en débit lors de cette opération ;
- dans l'autre cas, on parle de matriçage pour coder un nombre supérieur de canaux dans un format ou un flux prévu pour en contenir moins. Ce type de codage est généralement conçu pour être compatible avec le format d'origine qui doit pouvoir être retrouvé sans dégradation si le dématriçage n'est pas disponible. Presque tous les médias en bandes, disques analogiques ou numériques... ont proposé des versions multicanal matricées.

7.4 Codage sans pertes

7.4.1 PCM

Nous avons vu précédemment le format PCM défini par la fréquence d'échantillonnage et la quantification utilisée. La quantification est généralement linéaire, on peut alors parler de LPCM, Linear Pulse Code Modulation.

Complément Web 7.6

Type de flux PCM selon l'utilisation.

7.4.2 Lossless

En appliquant les méthodes de compressions basées sur un codage entropique, différents formats sont apparus pour remplacer le PCM en ajoutant des éléments complémentaires (tags). Ces formats qualifiés de PPCM Packed Pulse Coded Modulation, proposent les avantages suivants par rapport au LPCM :

- une réduction de débit allant jusqu'à 50 % ;
- un format plus robuste aux transferts et aux erreurs ;
- des données auxiliaires utilisables pour décrire le contenu ;
- une adaptation à la diffusion en flux (streaming).

Parmi ces formats, on peut citer FLAC, Apple Lossless, Wavpack, Dolby True HD (MLP), DTS HD-Master Audio, Merging LRC.

Si les technologies « Lossless » sont plus fréquentes en mode fichier, les codages de flux doivent eux satisfaire à la contrainte d'encodage en temps réel tout en assurant une constance de débit.

7.4.3 DSD

Le Super Audio CD proposé par Sony s'appuie sur une technologie d'échantillonnage en format DSD (**Direct Stream Digital**) de type One-Bit à 64Fs soit 2,88 MHz (64 x 44,1 kHz), appelé aussi DSD64. Ce qui permet une bande passante extrême et l'avantage d'un filtrage simple lors de la conversion numérique-analogique, au détriment des performances en bruit et dynamique.

La production de SA-CD impose des convertisseurs et des stations de montage audio-numérique particulières travaillant en DSD et non en PCM. Pour pallier à cette limitation, la société Merging propose le format DXD Digital eXtreme Definition, en 24 bits, échantillonné à 8Fs soit 353 kHz et permettant un travail d'édition pour DSD, sans perte en résolution.

7.5 Codages perceptifs

7.5.1 Famille MPEG Audio

Le projet européen EUREKA 147 est un projet mené à partir de 1986 pour développer la diffusion audionumérique DAB (Digital Audio Broadcast). Parmi les propositions issues de ces recherches, le codage MUSICAM (Masking pattern adapted Universal Subband Integrated Coding And Multiplexing) ainsi que le codage ASPEC (Adaptive Spectral Perceptual Entropy Coding) ont mis en commun certaines technologies pour être adoptés et normalisés par le groupe MPEG pour des applications de stockage et de diffusion.

De nombreux supports, Vidéo CD, DVD, Blu-ray ainsi que la diffusion télévisée et radiodiffusée utilisent ces normes de codage de la vidéo et de l'audio. Puis ces normes ont évolué :

- MPEG-1 Audio, norme ISO/IEC 11172-3, 1992 ;
- MPEG-2 Audio, norme ISO/IEC 13818-3, 1994 ;
- MPEG-2 Advanced Audio Coding, AAC, norme ISO/IEC 13818-7, 1997 ;
- MPEG-4 Audio Version 1, norme ISO/IEC 14496-3, 1999 ;
- MPEG-4 Audio Version 2, parametric audio, norme ISO/IEC 14496-3/AMD1, 1999 ;
- HE-AAC Spectral Band Replication (SBR), norme ISO/IEC 14496-3, 2003 ;
- HE-AAC v2 Parametric coding et parametric stereo PS for high quality audio, norme ISO/IEC 14496-3, 2004 ;
- MPEG-D (MPEG Surround), Spatial audio coding, norme ISO/IEC 23003-1, 2005.

◆ Couches et profils

Les formats MPEG sont des boîtes à outils permettant de choisir un format avec de nombreuses options de codage selon l'usage. Les profils et les niveaux présentent un catalogue homogène et déterminent des configurations du format ce qui permet de simplifier les décodeurs qui n'ont pas à supporter toute les variantes.

Les profils

- Avec des appellations comme Simple Profile, Main Profile, High Profile, ils définissent un degré de complexité du codeur et du décodeur donc les algorithmes et options de codage possibles ;
- Les profils visent des applications spécifiques : HD, applications mobiles, etc.

Les niveaux (« levels »)

Les niveaux définissent un ensemble de paramètres de codage : ainsi pour l'audio, le type de débit, les valeurs de débit maximal, les fréquences d'échantillonnage...

Les appellations des niveaux suivent aussi une terminologie définie, comme : Low Level, Main level et High Level.

La notation généralement utilisée pour présenter une combinaison de profil et de niveau est par exemple pour le DVD, MP@ML signifiant « Main Profile at Main Level ».

◆ MPEG 1

MPEG1 Layer 1 est une variante LC Low Complexity du codec MUSICAM à faible latence et optimisé pour un débit de l'ordre de 192 kbit/s, utilisé pour la diffusion satellite.

MPEG1 Layer 2 est issu de l'optimisation de MUSICAM, avec filtrage polyphase à 32 sous-bandes, trame fixe de 1152 échantillons, et débit de 32 à 384 kbit/s Il s'agit du codage retenu pour le DAB en Europe.

MPEG1 Layer 3 repose sur le codage entropique ASPEC, complété par des filtres hybrides et transformées MDCT, ainsi que l'implémentation de la technologie du Joint Stéréo Coding. L'adoption du modèle Psychoacoustic II a permis l'utilisation de deux fenêtres : courte de 256 échantillons ou longue de 2048. Avec un débit allant de 32 à 320 kbit/s, il est surtout connu sous le nom de mp3 et a évolué constamment en améliorant sa qualité. Il s'agit d'un codec très répandu : lecteurs portables, sites de téléchargement...

◆ MPEG2

A partir du MPEG1 layer 2, le format MPEG2, normalisé en 1995 par l'ISO/IEC 13818-part3, a ajouté une extension multicanal avec matricage pour conserver une compatibilité de lecture avec les décodeurs MPEG1. Le codage de 6 canaux est possible avec la

voie LFE qui est échantillonnée à $F_s/96$ soit 500 Hz au lieu de 48 kHz pour les autres canaux.

Ce format fait partie de la norme DVD mais les codeurs MPEG2 multicanal n'ayant jamais été disponibles pour les studios d'authoring, le format Dolby Digital a été généralement utilisé.

Un autre format normalisé pour le MPEG2 en ISO/IEC 13818-part7 mais abandonnant la compatibilité avec les décodeurs MPEG1, est l'AAC Advanced Audio Coding, aussi appelé MPEG2-NBC (non backward compatible). Cette non-compatibilité a permis une amélioration importante : un codage AAC de 5 canaux à 320 kbit/s est équivalent en qualité à un codage MPEG2 compatible MPEG1 layer3 à 640 kbit/s. Le but de l'AAC, en plus de l'amélioration qualitative, est d'offrir un ensemble de profils afin de répondre à des applications diverses depuis 8 kbit/s.

◆ MPEG4-AAC

Dans la norme MPEG4, le format de codage audio est une évolution de l'AAC qui supporte un échantillonnage de 8 à 96 kHz, des débits de 16 à 576 kbit/s et jusqu'à 48 canaux audio.

L'apport de l'AAC repose sur les technologies suivantes :

- filtrage MDCT ;
- deux possibilités de trames temporelles : courte de 128 et longue de 1 024 échantillons ;
- temporal noise shaping ;
- backward adaptive linear prediction ;
- stéréo d'intensité ;
- joint stéréo par bande de fréquence ;
- stéréo M/S ;
- noiseless coding ;
- bitstream multiplexing ;
- Support des droits DRM.

Main Profile : version complète mais sans le module de contrôle de gain, fournit la meilleure qualité.

Low-complexity (LC) Profile : sans l'outil de prédiction et sans la complexité du « temporal noise shaping ».

Sample-rate Scaleable (SRS) Profile : le décodage le plus simple, version LC avec contrôle de gain.

Nomenclature de l'AAC

Compte tenu du nombre important de variantes de l'AAC, une codification a été suggérée :

Version	profil	codage	Utilisation
AAC	AAC	AAC-LowComplexity	Qualité en haut débit
AAC+	HE-AAC	AAC-LC + SBR + Parametric Stéréo	
eAAC+	HE-AACv2	AAC-LC + SBR + Parametric Stéréo	TV et radio numérique
HD-AAC	HD-AAC	AAC-LC + Scalable Lossless Coding	Blu-Ray

Figure 7.13 – Tableau de codification AAC, © J.-L. Ohl.

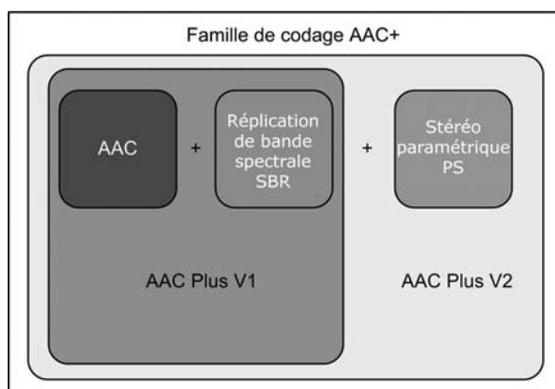


Figure 7.14 – Schéma de la famille AAC +, © J.-L. Ohl.

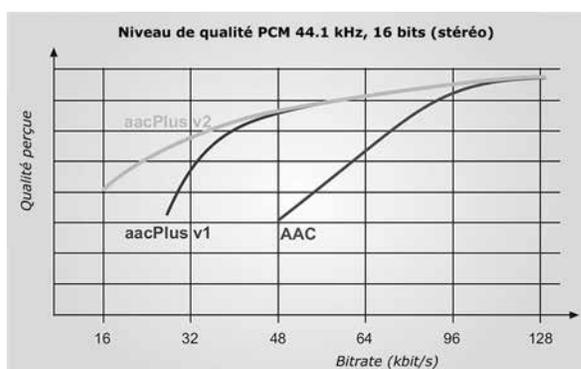


Figure 7.15 – AAC, qualité comparée selon le débit, © Coding Technologies.

Complément Web 7.7

Évaluation comparative de versions AAC.

Ce codage eAAC + est compatible par un système de transcodage, avec le DTS et Dolby Digital, codecs que l'on retrouve dans les amplificateurs AV, ce qui permet aux radios de diffuser un format 5.1 compatible avec le parc grand public existant.

◆ MPEG Surround, MPEG-H 3D

Abordées dans le paragraphe des codages paramétriques, les technologies du MPEG Surround complètent les codages MPEG1 layer2, AAC et HE-AAC pour étendre ces formats en multicanal jusqu'au 7.1 en restant compatible avec les décodeurs stéréo. Le débit supplémentaire pour le codage paramétrique des canaux ajoutés est réduit ce qui rend simple l'évolution des transmissions radiophoniques ou télévisées vers ce format.

**Complément
Web 7.8**

Décodage MPEG Surround en stéréo et multicanal.

**Complément
Web 7.9**

MPEG-H 3D.

7.5.2 Famille Dolby

◆ Dolby Surround

Principe du Dolby Surround

Le format multicanal Dolby Surround matricé en stéréo est aussi nommé LtRt (Left total, Right total). Dolby Surround en broadcast, il est appelé Dolby Stéréo en cinéma. Le procédé repose sur un matriçage 4-2-4 par lequel le mixage LCRS est matricé en stéréo LtRt, ce qui permet de stocker et diffuser par les format stéréo habituels du cinéma, TV (Nicam, câble, satellite) et radio FM. L'encodage 4-2 s'accompagne souvent du procédé de réduction de bruit Dolby SR (Spectral Recording).

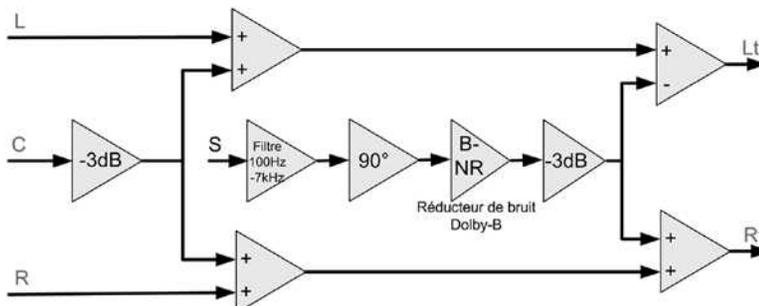


Figure 7.16 – Encodage Dolby Surround 4-2-4, © J.-L. Ohi d'après Dolby.

Pour le cinéma, ce signal est couché en analogique sur la pellicule, puis lu et décodé en LCRS par un processeur cinéma Dolby CP.

L'algorithme est figé et ne modifie pas directement les canaux L et R mais ajoute le canal du centre à -3 dB. Le canal arrière Surround, limité entre 100 Hz et 7 kHz et déphasé de 90° (« surround phase shift »), subit une réduction de bruit Dolby B puis est sommé dans les canaux de diffusion LtRt, en opposition de phase.

◆ Dolby Pro Logic

Le décodage Dolby Pro Logic est fréquemment utilisé dans les amplificateurs AV, dans les téléviseurs et les récepteurs satellite. Dans le domaine professionnel, c'est le processeur analogique SDU-4 qui fut utilisé pour le film et on le trouve en diffusion TVHD dans les processeurs numériques DP570 et DP564. En principe, les séparations L, R et C, S sont infinies alors que les séparations adjacentes sont de 30 dB. Dans ce procédé, les différences de phases entre canaux frontaux, issues de la prise de son ou du mixage (utilisation de réverbération), peuvent engendrer des artefacts, par exemple des transitoires de sources frontales apparaissant dans les canaux arrière.

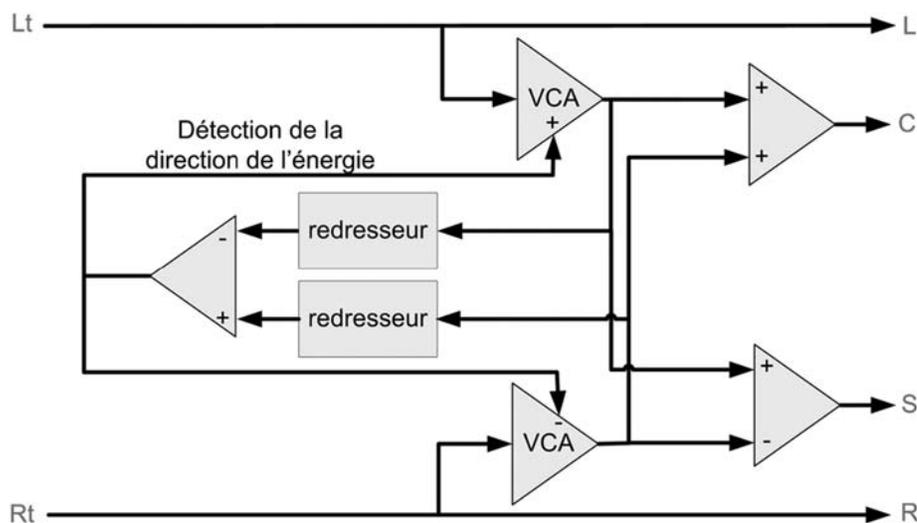


Figure 7.17 – Décodeur Dolby Surround Pro Logic avec fonction de steering,
© J.-L. Ohl d'après Dolby.

Dolby Pro Logic 2

Le Dolby Pro Logic 2 est une évolution du Dolby Surround reposant sur un algorithme de codage et de décodage adaptatif, avec des canaux surround stéréo de 100 Hz à 18 kHz. Les programmes codés en Dolby Pro Logic 2 sont compatibles avec les récepteurs/décodeurs Dolby Surround Pro Logic.

L'encodage est une matrice 6-5-2-5, le canal LFE peut être intégré dans L et R, avant la fabrication des canaux Lt Rt. Puis au décodage, les cinq canaux L C R Ls Rs sont recréés et le canal LFE est traité par Bass Management.

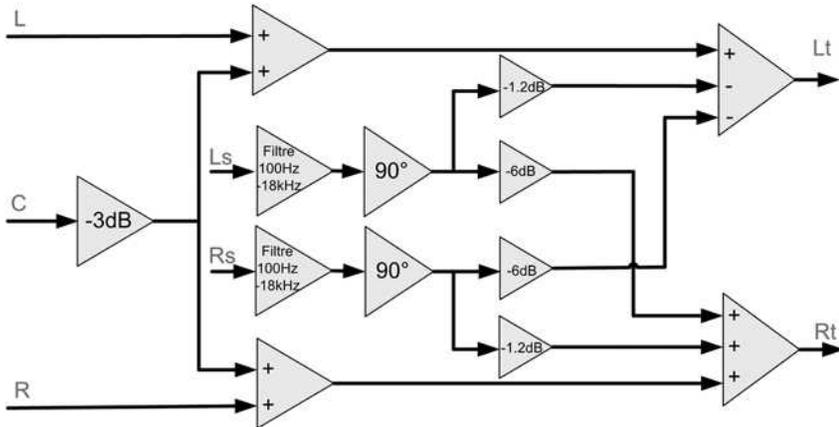


Figure 7.18 – Encodeur Dolby Surround Pro Logic 2, © J.-L. Ohl d'après Dolby.

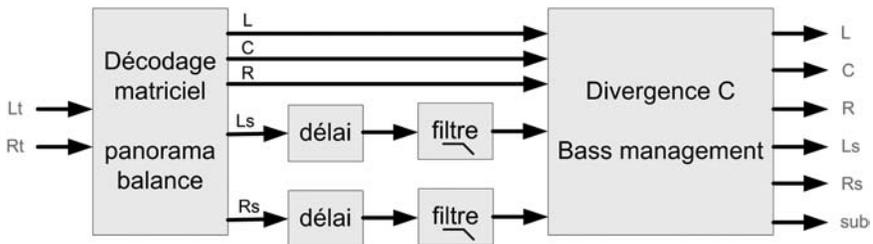


Figure 7.19 – Décodeur Dolby Surround Pro Logic 2, © J.-L. Ohl d'après Dolby

Décodage adaptatif :

Il existe deux modes de décodage : le mode « movie » garde les informations du canal central dans le haut-parleur du centre et ne permet pas de réglage de postproduction et le mode « music » qui propose des réglages pour la postproduction : divergence du centre « Center Width » pour répartir entre centre physique et centre fantôme, équilibre frontal/arrière par le réglage « Dimension » et divergence vers les autres canaux par le réglage « Panorama ».

Le décodeur Dolby Surround Pro Logic 2 offre une séparation infinie sur L, R et C, S, et une meilleure séparation sur les canaux adjacents de 40 dB contre 30 dB pour le Dolby Pro Logic 1.

Les appareils professionnels DP570 et le DP564 et de nombreux amplificateurs AV et téléviseurs intègrent les décodeurs Pro Logic 2.

Complément Web 7.10

Décodage adaptatif du Dolby Pro Logic 2, mode movie, mode music.

◆ Steering

Les technologies de matricage reposant principalement sur les signaux communs et différences, les performances de séparation entre canaux sont généralement assez faibles. Cet aspect est amélioré par une technologie appelée « steering » employée pour le Dolby Surround et ProLogic : la différence de niveau entre canaux est amplifiée dans le canal le plus fort et atténuée dans les autres canaux, ce qui accentue la séparation en conservant une énergie globale équivalente.

◆ Dolby Digital

Principe

Le Dolby Digital, de nom d'origine AC3, est un format de codage audio multicanal discret (sans matricage), utilisé pour la diffusion et devenu courant depuis l'arrivée du DVD. Cette technologie, apparue en 1992 avec la sortie du film « Batman Returns », a été prévue pour remplacer les matricages Dolby Stéréo/Dolby Surround. Utilisé dans la diffusion DVB, DVD, au cinéma, dans les jeux vidéo (PS3, PS2, Xbox), la téléphonie mobile, les tablettes numériques et la radio, le Dolby Digital permet la diffusion du son multicanal 5.1 et reste compatible mono, stéréo avec décodeurs Pro Logic des amplificateurs AV. Il transporte la metadata de loudness qui permet d'uniformiser le niveau de restitution des programmes, la metadata downmix qui optimise la compatibilité stéréo et mono, et la metadata DRC (Dynamic Range Control) qui permet d'adapter la dynamique aux conditions d'écoute.

Caractéristiques

Le Dolby Digital comporte de 1 à 6 canaux et le « channel mode » précise le nombre de canaux transportés avec le premier chiffre pour les canaux frontaux et le second pour les canaux arrière et la lettre L désigne la présence du canal LFE : mode 3/2L pour le 5.1, mode 2/0 pour la stéréo et mode 1/0 en mono. Le canal LFE, filtré entre 3 Hz et 120 Hz, est échantillonné à 240 Hz. La résolution est de 16, 20 ou 24 bits et les débits possibles vont de 32 kbit/s (mono) jusqu'à 640 kbit/s (5.1) :

- Cinéma 320 kbit/s
- DVB 384 kbit/s
- DVD 64 à 448 kbit/s
- Blu-ray 64 à 640 kbit/s

Dolby Digital Cinéma

Le Dolby Digital était présent au cinéma entre les perforations de la pellicule 35 mm, sous formes de pavés codés, conjointement à la version stéréo analogique codée LtRt avec réduction de bruit SR. Le format combiné s'appelle alors **Dolby SR-D**. Chaque bloc de données contient une matrice de 76×76 pixels, une image contenant 4 blocs en 24 images par seconde. Le débit maximal est de 554,5 kbit/s ($76 \times 76 \times 4 \times 24$) dont 320kbit/s pour l'audio, le reste en correction d'erreur. La version analogique sert de secours en cas de problème de lecture de la piste numérique ou pour les salles non équipées en Dolby Digital.

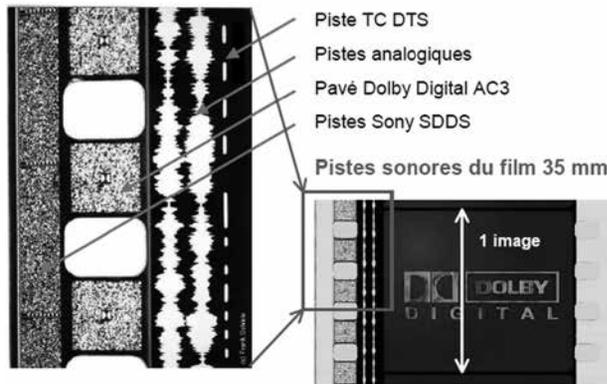


Figure 7.20 – Dolby Digital sur pellicule Film 35 mm, © Dolby.

Complément Web 7.11

Tableau résumé des formats Dolby.

DOLBY DIGITAL

- Codec utilisé pour le cinéma, le DVB, les supports DVD et Blu-ray.
- Formats du 1.0 jusqu'au 5.1.
- Intègre les metadata Dialnorm, DRC, Downmix.

Dolby Digital Plus

Le Dolby Digital Plus (DD + ou E-AC3) est une évolution du codec Dolby Digital, jusqu'au format 13.1 avec un débit jusqu'à 6 Mbps. Le DD + à 224 kbit/s remplace progressivement le Dolby Digital à 384 kbit/s dans la diffusion du son des chaînes TV HD. Un transcodage est possible de DD + vers Dolby Digital. Prévu pour une liaison

en HDMI 1.3, il n'est pas transportable dans une connexion numérique S/PDIF, sauf transcodage préalable.

◆ Dolby E

Principe et caractéristiques

Le Dolby E est un flux professionnel de contribution, conçu pour transporter jusqu'à huit canaux audio 48 kHz/20 bits, des métadonnées et du time code dans une trame PCM stéréo. Plusieurs répartitions de programmes sont possibles : 8×1 , 4×2 , $5.1 + 2$. La version la plus utilisée, $5.1 + 2$, transporte un mixage 5.1 et une version stéréo LtRt ou LoRo du programme.

Dolby annonce une intégrité de l'information jusqu'à 10 générations de codages. Les interfaces audio numériques standards (AES/EBU, S/PDIF, SDI etc..) permettent de transporter ces huit canaux, mais il est nécessaire de décoder le signal pour tout traitement audio ou écoute. Les magnétoscopes numériques peuvent enregistrer du Dolby E sur deux canaux en 20 bits. Si l'enregistrement est en format 16 bits, seuls 6 canaux peuvent être codés.

Le Dolby E est synchrone à l'image, ce qui simplifie son traitement et permet le montage et la commutation : chaque trame du Dolby E correspond à la trame vidéo en respectant un « guard band » qui garantit la protection contre la corruption des données audio.

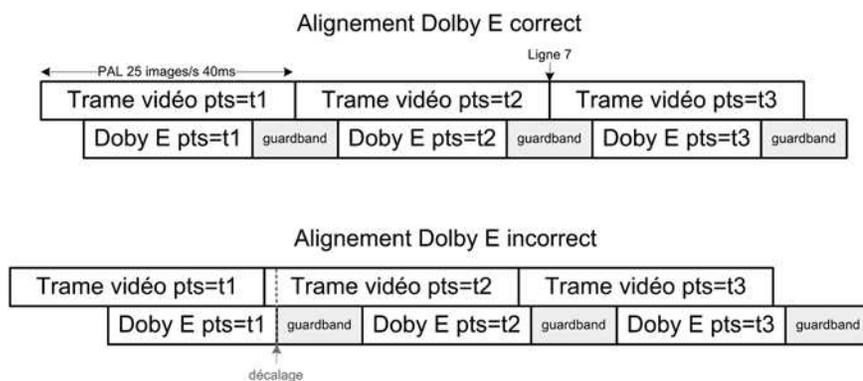


Figure 7.21 - Guard Band signal Dolby E, © J.-L. Ohl d'après Dolby.

La commutation du signal doit se faire sur ces « guard bands », pour assurer la validité des données. De plus, le décodeur Dolby E crée un crossfade d'une durée de 5 ms dans la zone de recouvrement.

Afin d'obtenir un bon alignement du Dolby E sur la trame vidéo, voici les valeurs des numéros de lignes du signal vidéo à respecter :

GUARD BAND : ALIGNEMENT SUR LA TRAME VIDÉO

- Signal vidéo SD 25 ips : de la ligne 604 à la ligne 11.
- Signal vidéo HD 25 ips : de la ligne 1087 à la ligne 20.

Complément Web 7.12

Structure de la trame Dolby E.

Le Dolby E accepte différentes configurations de canaux, mono, stéréo, 5.1, un groupe de canaux formant un programme. Plusieurs types de programmes sont configurables :

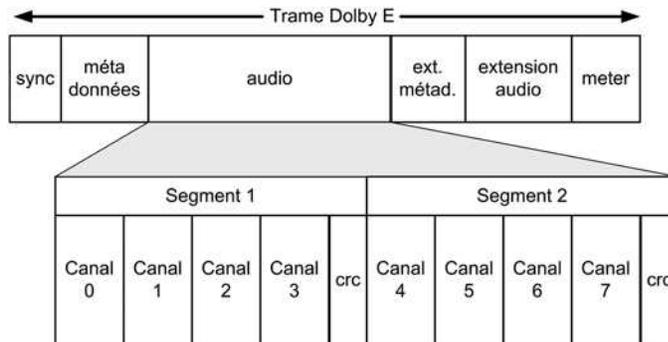


Figure 7.22 – Programmes Dolby E, © J.-L. Ohl d'après Dolby.

Fichiers Dolby E : MBWF, MXF

Le standard Dolby E n'exige pas de fichier spécifique et utilise un fichier .wav standard ou le format RF64-MBWF spécifié par l'EBU. Basé sur un broadcast wav multicanal, il intègre les huit canaux audio PCM du programme (5.1+2) ainsi que les métadonnées. Ce format d'échange peut s'exporter depuis les stations de travail vers des applications d'encodage Dolby comme le DP600.

Traitement du Dolby E

Réglage du niveau audio

Le seul traitement possible sur un signal encodé en Dolby E est l'ajustement du niveau, tous les autres traitements audio nécessitant un décodage en PCM puis un réencodage en Dolby E. Les mots de gain de canal sont transportés comme « Professional Metadata », sont modifiables dans le flux et sont utilisés pendant le décodage.

Montage des canaux

Chaque canal étant codé séparément, un canal peut être remplacé sans conséquence sur les autres canaux.

Temps de latence

L'encodage et le décodage Dolby E décalent le signal audio d'une image exactement (sauf en lecture d'un programme) à chaque traitement. Ce retard fixe simplifie l'alignement temporel avec la vidéo.

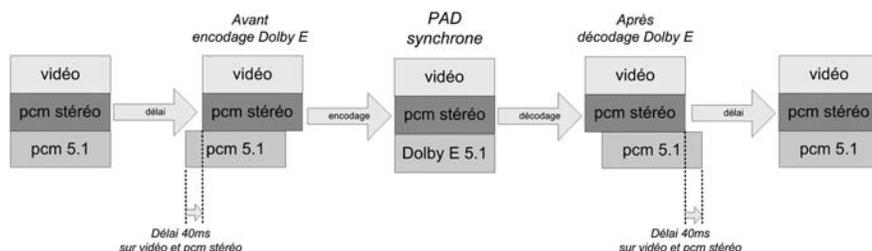


Figure 7.23 – Synchronisation des signaux vidéo/audio en Dolby E, © J.-L. Ohl.

DOLBY E

- Flux professionnel de contribution intégrant jusqu'à 8 canaux sur des infrastructures PCM stéréo.
- Plusieurs configurations de programmes, la configuration 5.1 + 2 la plus usuelle intègre le programme multicanal et la version stéréo LtRt ou LoRo.
- Transport des metadata et du time code.
- Débit 1920 kbps pour 8 canaux en 48 kHz.
- Synchrone à l'image : la trame du Dolby correspond à la trame vidéo.
- Temps de latence : l'encodage ou le décodage Dolby E prend une image.
- Nécessité du décodage pour tout traitement de signal ou écoute.

◆ Dolby Pulse

Suite au rachat en 2007 de la société Audio Coding Technologies par Dolby, le format Dolby Pulse a été proposé pour des applications à bas débit. Ce codage compatible avec les metadata Dolby fonctionne à partir des codecs AAC, HeAAC et HeAACv2.

7.5.3 Famille DTS

Le Digital Theater System a démarré en présentant un système de lecture optique sur CD synchronisé avec un projecteur cinéma, utilisant un codec apt-X avec audio 5.1 en 44,1 kHz et 16 bits.

Ce codage est basé sur un filtrage polyphase en 32 sous-bandes puis requantification ADPCM (Adaptative Differential Pulse Code Modulation) et prédiction linéaire (LPC). La technique du Joint Frequency Coding est utilisée au-dessus de 2,5 kHz mais seulement pour les débits faibles.

Dans la norme définissant le DVD, le format DTS, aussi appelé Coherent Acoustic, est optionnel (il ne peut donc pas être présent seul) et accepte une structure multicanal jusqu'au 7.1 en 48 kHz et 24 bits.

Sur le Blu-ray, certains formats DTS peuvent être utilisés seuls, à condition d'avoir un décodage possible de type DTS-Core (DTS Legacy), c'est-à-dire une extraction du flux de base compatible avec tous les lecteurs.

Complément Web 7.13

Famille DTS.

◆ Liste des codages de la famille DTS

- DTS Digital Surround : le format DTS utilisé dans les DVD, de 255 à 1509 kbs ;
- DTS Express : format compressé bas débit, de 64 kbs à 512 kbs, jusqu'au 7.1 ;
- DTS-HD High Resolution jusqu'à 7.1 canaux en 96 kHz et 24 bits, de 2 à 6 Mbs, compatible DTS Legacy ;
- DTS-HD Master Audio, codage sans pertes jusqu'à 7.1 à 96 kHz ou 5.1 à 192 kHz, compatible DTS Legacy ;

◆ Codages DTS avec traitement sonore

DTS propose aussi différents formats comprenant des traitements audio liés au multicanal :

- DTS Neo 6 : upmix stéréo vers 5.1 ;
- DTS Neo X : upmix 5.1, 6.1 ou 7.1 vers 11.1 ;

Complément Web 7.14

- DTS Neo X, © DTS. ;
- DTS-ES matrix : ES = Extended Surround, 6.1 avec matricage du centre arrière dans un flux 5.1 ;
- DTS-ES discret : 6.1 avec canaux séparés, dont le centre arrière, dans un flux 5.1 ;
- DTS Surround Sensation : virtualisation pour deux enceintes ou casque ;

- DTS Neural Surround : downmix et upmix pour transport jusqu'au 7.1 dans un flux stéréo.

7.5.4 Sony ATRAC et SDDS

Le codage perceptif ATRAC développé par Sony a été utilisé dans les lecteurs-enregistreurs minidisc et dans le système SDDS Sony Dynamic Digital Sound proposé pour le cinéma et inscrit sur les bords de la copie 35 mm du film. Les huit canaux audio sont répartis en 5 canaux frontaux, 2 canaux surround et un canal LFE.

7.5.5 Auro-3D Octopus codec

Le codec utilisé par les formats Auro-3D ne repose pas sur des principes psychoacoustiques de réduction de débit mais sur l'utilisation des bits de faible poids de l'échantillon 24 bits pour des informations utiles de spatialisation, conjointement à un codage entropique. Un fichier audio quantifié sur 24 bits possède une dynamique théorique de 140 dB. Le reste du système électro-acoustique ne pouvant exploiter au mieux que 110 à 120 dB soit 18 à 20 bits, ces 4 à 6 bits inutiles sont utilisés pour diffuser les canaux supplémentaires et les métadonnées des formats évolués type Auro-3D 9.1 ou 11.1.

Complément Web 7.15

Quantification du signal et dynamique.

Complément Web 7.16

- Principe du codec Auro-3D octopus.
- Le codec peut multiplexer jusqu'à trois formats différents en un seul flux avec 100 % de séparation. Par exemple, un fichier PCM 5.1 codé en Auro-3D peut être exploité pour transporter un format Auro-3D 9.1 et un mixage 5.1 pour le Blu-ray.

Autres exemples :

- 11.1 ou 13.1 Auro-3D et mixage 5.1, transportés par un fichier PCM 5.1 codé en Auro-3D pour le DCP ;
- Mixage 5.1 et 2.0 stéréo, transportés par un fichier PCM stéréo codé en Auro-3D pour le broadcast ;
- Mixages 7.1 et 5.1, transportés par un fichier PCM stéréo codé en Auro-3D pour le Blu-ray et le DCP ;
- 9.1 Auro-3D et mixage 5.1, transportés par un fichier PCM stéréo codé en Auro-3D pour le broadcast ;

Le décodage Auro-3D implique une latence de trois échantillons, soit 60 μ s pour un signal à 48 kHz. Les principales métadonnées sont :

- information des flux audio multiplexés ;

- information sur la réduction de la dynamique ;
- information sur les downmix artistiques ;
- paramètres utiles pour l'algorithme d'upmix Auro-3D ;
- metadata Program Loudness, Loudness Range conforme à la norme EBU R128 ;
- information sur l'utilisation de courbes isoX pour l'écoute.

Le premier Pure Audio Blu-ray compatible Auro-3D 9.1 en 96 kHz est sorti à l'été 2012 chez 2L, c'est un enregistrement de l'orchestre de chambre norvégien TrondheimSolistene. Le disque comporte aussi les flux 7.1 (96 kHz/24 bits), 5.1 (192 kHz/24 bits) en DTS Master Audio ainsi qu'une stéréo PCM (192 kHz/24 bits), la source de l'enregistrement étant au format DXD, soit 352,8 kHz/24 bits.

7.5.6 Codages libres de droit

Les codages perceptifs précédemment décrits sont issus des développements de sociétés commerciales ou d'organismes de recherche et l'utilisation est régie par les législations sur la propriété intellectuelle et le droit commercial. Le développement de solutions ouvertes existe, comme celles proposées par l'organisation Xiph.org dans le projet Ogg avec deux codecs audio libres de droits, Vorbis et Opus. Leurs qualités sont comparables à celles des solutions propriétaires et étant « open source », les codes sources sont publiés.

7.5.7 Évolution du codage

Les principales avancées sont attendues au niveau des latences, des très faibles débits et des codages hybrides avec/sans pertes.

Concernant le délai de codage/décodage, l'évolution technologique amène à des valeurs inférieures à 10 ms tout en gardant une bonne qualité. Quant aux débits, on peut estimer qu'entre 1995 et 2015, ceux-ci ont été divisés par 5 à qualité sonore constante.

Le codage hybride avec/sans pertes permet de travailler avec un flux ou fichier à débit réduit et d'y associer un algorithme de correction pour retrouver le format complet sans pertes.

7.5.8 Droit et protections

◆ DRM

Le premier support numérique grand public, le CD audio fut conçu avant l'avènement de l'ordinateur grand public avec graveur : la nécessité d'une protection contre la copie ne faisait pas partie du cahier des charges. Pour la génération suivante, le DVD, différents systèmes de protection (zones, CSS Content Scrambling,...) ont été inclus dans les spécifications mais en sous-estimant les capacités techniques des contrefacteurs. La copie illégale de CD et DVD a imposé des protections et la gestion des droits dès la définition de nouveaux supports ou formats de diffusion. Les SACD et les disques Blu-ray, ont des protections anti-copies bien plus solides que les formats antérieurs.

Les équipements Blu-ray utilisent la norme HDCP qui établit un échange d'informations entre deux appareils connectés et qui impose de n'envoyer la meilleure qualité de flux audio ou vidéo qu'à un appareil s'annonçant aussi HDCP. Les sorties non protégées sont limitées à un débit réduit : une sortie audio numérique est « dégradée » par exemple à 48 Hz et 16 bits.

Le cinéma numérique propose une norme de protection de contenu DCP qui définit les formats des médias, des descripteurs et aussi les clefs de protection KDM sur 128 bits, qui sont nécessaires au décodage des contenus MXF cryptés.

◆ **Watermarking**

La technologie du tatouage, Watermarking, a pour but de signer un contenu audio afin de suivre sa diffusion, d'en authentifier son contenu ou d'identifier les ayants droit.

Ce marquage, sans altérer la nature du signal, doit être suffisamment robuste pour résister à des traitements du signal en cascade tels que les conversions numériques analogiques, les traitements dynamiques ou fréquentiels et, bien entendu, les codages. Ces impératifs sont évidemment contradictoires : un codage étant prévu pour enlever la partie inaudible du signal, comment le marquage pourrait-il résister ?

Différentes technologies sont susceptibles d'être utilisées et souvent étalent les informations ajoutées soit dans le domaine temporel, soit dans le domaine fréquentiel : les effets de masquage sont alors mis à profit pour cacher ces modifications. Parmi les diverses possibilités, on trouve :

- Codage des LSB : des bits de poids faible peuvent être réservés au watermarking indépendamment du signal audio. Cette technique manque de robustesse : elle est dégradée par la plupart des traitements audio ;
- Codage par modulation de la phase : la perception auditive étant peu sensible à la phase, le signal peut être modulé en phase soit temporellement soit par bande de fréquence ;
- « Echo hiding » : le signal d'origine est rajouté à lui-même, atténué et décalé. Ces « échos » sont auditivement masqués mais les délais sont analysables par calcul d'auto-corrélation et représentent l'information de tatouage ;
- la technique de « spread spectrum watermarking » qui est l'addition d'un signal de bruit à faible niveau mais réparti sur une bande de fréquence étendue. La répartition du tatouage est importante pour cacher ce marquage, réduire le risque d'audibilité et améliorer la robustesse aux traitements ultérieurs.

7.5.9 Vidéo et audio

◆ **Multiplexage**

On parle de multiplexage des données quand différents contenus, audio, vidéo, sous-titres, sont organisés par paquets dans un seul flux ou fichier pour être présentés de

façon synchrone. Chaque type de données occupe alternativement une place dans le train des données. Il est aussi possible, comme pour le DVD, le Blu-ray ou les transmissions satellites, de multiplexer plusieurs flux de type identique : plusieurs programmes vidéos par exemple associés aux flux audio correspondant.

Formats de multiplexage MPEG

- **Elementary stream** : (.mpv, .mp2, .mpg), ce format ne contient qu'un type de données soit audio soit vidéo ;
- le **Program Stream** (MPEG-PS) est un format de flux ou de fichier multiplexé vidéo + audio comprenant les informations de code temporel. Ce multiplexage qui peut comprendre plusieurs flux audio-vidéo est utilisé en tant que conteneur VOB dans les DVD. Les fichiers sont généralement pourvus des extensions .mpg, .mpeg ou .ps ;
- le **Transport Stream** (MPEG-TS) est aussi un format multiplexé vidéo + audio mais plus robuste que le Program Stream vis à vis des erreurs de transmission, donc particulièrement adapté à la diffusion. Ce format est aussi présent sur les disques Blu-ray. Quand le Transport Stream est présent en tant que fichier, les extensions généralement utilisées sont .ts, .m2ts ou .m2t.

Structure de flux MPEG transport

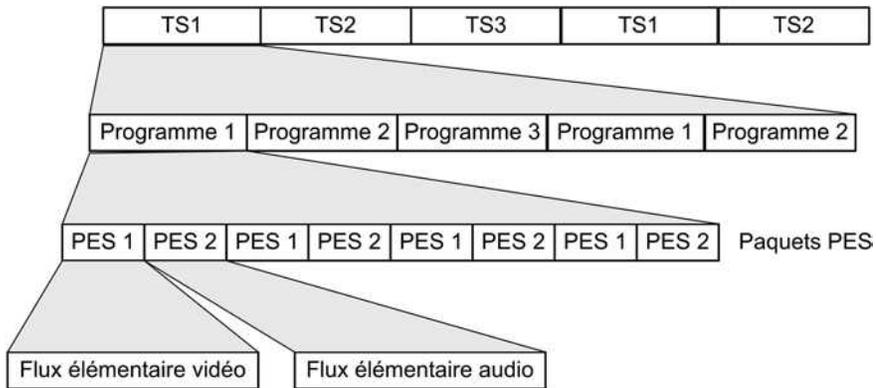


Figure 7.24 – Structure d'un flux TS, © J.-L. Ohi.

◆ Structures logiques

Indépendamment du format physique d'un support, les données sont organisées selon des structures logiques qui peuvent être totalement différentes. Ainsi, la normalisation du CD-Audio (Red Book) est antérieure à l'utilisation des ordinateurs personnels : un ordinateur ne comprend pas cette structure et ne peut pas présenter un CD audio comme une arborescence de répertoires et de fichiers. Le CD-ROM (Yellow Book), créé plus tard, a la même structure physique que le CD-Audio mais possède une structure logique compatible avec les systèmes informatiques. Les supports ultérieurs DVD,

Blu-ray, ont été prévus selon des normes ISO-UDF assurant la compatibilité avec les principaux systèmes informatiques ou plateformes de jeux.

7.5.10 Supports physiques

Nous ne nous intéresserons qu'aux principaux supports ayant eu une déclinaison dédiée au multicanal.

◆ **Disque 33t**

Différentes solutions ont été proposées mais sans grand succès : le CD4 de JVC codait quatre canaux discrets, deux par la gravure stéréophonique habituelle (sillon modulé horizontalement et verticalement) et deux canaux supplémentaires en ajoutant une modulation haute fréquence à flanc de sillon. Le SQ Quadraphonic de RCA et le QS de Sansui étaient des formats multiplexés avec de faibles performances de séparation entre canaux.

◆ **Laserdisc**

Le Laserdisc (LD) est un support vidéo analogique avec son numérique sur disque de 30 cm. Certains LD contiennent un flux audio multicanal codé en Dolby Surround, en Dolby Digital 5.1 ou en DTS.

◆ **CD DTS**

Le CD d'origine contient deux canaux audio en 16 bits à 44,1 kHz. La société Digital Theater System a proposé un format presque conforme au standard Red Book, mais avec le son codé en apt-X (débit de 1234 kbit/s pour un format 5.1) pour encoder un signal multicanal. Si le lecteur n'intègre pas de décodeur DTS, sa sortie analogique est inutilisable et il faut envoyer le signal numérique par une sortie optique ou coaxiale vers le décodeur.

◆ **DVD**

Le DVD, premier standard grand public avec format multicanal, a connu un succès considérable amenant à l'image la qualité et la facilité d'emploi que le CD audio a apporté au son.

Deux propositions, le MMCD (Multimedia CD) de Sony et Philips et le SD (Super Density) d'un consortium concurrent comprenant Toshiba, se sont opposées avant d'opter pour un format unique en 1995, le DVD (Digital Versatile Disc), « versatile » pour signifier que dès la définition du format, différentes déclinaisons étaient prévues.

Un ensemble d'évolutions marquantes a permis ce format :

- l'amélioration technologique des supports optiques, la densité d'information est multipliée par 5 par rapport au CD ;

- le format vidéo MPEG2 conserve une très bonne qualité d'image en divisant le débit par 50 par compression spatiale intra-image et compression temporelle supprimant les redondances entre images ;
- les codages audio MPEG, Dolby Digital ou DTS autorisent une réduction de débit jusqu'à un facteur 10 ;
- la norme DVD arrivant alors que le parc informatique est déjà très large, sa structure a été choisie pour être compatible avec la plupart des systèmes d'exploitation.

Possibilités du format

Les caractéristiques du DVD permettent jusqu'à 2 heures de vidéo, une possibilité de vidéo multi angles (jusqu'à 9 angles), et 8 flux audio. Un flux se définit par son format (LPCM, MPEG, Dolby Digital, DTS...) et par le nombre de canaux (de 1 à 8).

Codes et protections

Les protections sont optionnelles : il s'agit d'un choix d'éditeur et non d'un choix technique.

Code régional : Cette notion de zone permet une mise sur le marché des titres respectant la chronologie des médias propre à chaque pays, 8 zones géographiques sont définies, indépendamment du format vidéo PAL ou NTSC.

Protections contre la copie : la protection contre la copie est présente dans la norme par trois systèmes dont le principal est le CSS Content Scrambling. Les données audio/vidéo sont cryptées et une clef en 40 bits présente dans le lecteur permet le décryptage.

L'audio du DVD Vidéo

Un flux 5.1 en PCM à 48 kHz et 20 bits occupe 5.76 Mbps soit autant qu'un flux vidéo de débit moyen. Pour réduire le débit audio et ne pas trop perdre en durée de programme sur le DVD, différentes solutions ont été prévues :

- codage MPEG1 layer 2 ou MPEG2 : de la monophonie jusqu'au format 7.1, très peu utilisé parce que les encodeurs n'ont jamais été disponibles pour les studios d'authoring ;
- Dolby Digital avec des possibilités allant du 1/0 (mono) jusqu'au 3/2 + 1 (5.1 surround) et un débit entre 92 kbit/s et 448 kbit/s en 48 kHz avec métadonnées renseignées à l'encodage et permettant la gestion de la dynamique ;
- DTS : format en 48 kHz ou 96 kHz, 24 bits avec débit 754 ou 1509 kbps, incluant aussi la gestion de métadonnées. Le DTS étant optionnel, un autre format doit, en principe, être présent (MPEG ou PCM avec une vidéo PAL, Dolby ou PCM avec une vidéo NTSC) ;
- Le format PCM est possible jusqu'à 8 canaux en 48 ou 96 kHz, 16, 20 ou 24 bits. Mais le débit maximal pour l'audio étant défini par la norme DVD à 6.144 Mbs, la limite est de 8 canaux à 48 kHz, 20 bits ou 2 canaux à 96 kHz, 24 bits.

Remarque : l'échantillonnage à 96 kHz fait partie de la norme et donc tous les lecteurs DVD lisent ce format mais sans obligation à ce que la sortie numérique délivre effectivement du 96 kHz.

DVD-Audio

Le DVD-Vidéo favorisant plutôt l'image, un format plus orienté vers le son a été défini : le DVD-Audio. Le DVD Vidéo limitant le débit audio à 6.144 Mbps, le DVD-Audio permet un débit de 9.6 Mbps. Les fréquences d'échantillonnages possibles sont 44.1/48/88.2/96/176.4/192 kHz en 16/20/24 bits. Une réduction de débit sans pertes Packet PCM de type MLP (Meridian Lossless Packing) est possible pour 8 canaux MLP en 96 kHz, 24 bits par exemple. Les canaux peuvent être séparés en deux groupes et utiliser des fréquences et quantification différentes pour chaque groupe : par exemple, gauche, droite et centre à 192 kHz, 24 bits et le second groupe (surround et subwoofer) en 48 kHz, 24 bits.

Une fonction SMART permet de prédéfinir à l'authoring une table des coefficients de réduction en stéréo (downmix). Autre différence par rapport au DVD-Vidéo, la possibilité de « diaporama » (nommé ASV pour Audio Still Video) : en lisant continuellement le flux audio, on peut naviguer entre des images fixes (textes, biographie...) de façon séquentielle ou aléatoire.

Malgré sa compatibilité technique avec les systèmes DVD-vidéo existants de stations de montage, lignes de mastering et pressage, ce format n'a pas eu de succès commercial.

◆ Super Audio CD

Les disques SA-CD utilisent un format One-bit sigma-delta à 2,8224 MHz dénommé DSD (Direct Stream Digital). Ce procédé permet une bande passante très élevée et la possibilité d'utiliser un filtre passe-bas minimal en lecture. Par contre, le rapport signal/bruit est moins bon qu'en PCM mais par l'utilisation de noise-shaping rejetant le bruit de quantification dans les hautes fréquences, une dynamique de l'ordre de 120 dB peut être obtenue. Cette technologie impose l'utilisation de matériels spécifiques dans l'ensemble de la chaîne depuis la prise de son jusqu'au pressage. En pratique, un passage en PCM est fréquent en mixage ou montage.

Jusqu'à six canaux (format 5.1) en format DSD sont permis sur ce support. Comme il n'y a pas de matrice de downmix, les pistes stéréo sont présentes indépendamment des pistes multicanal.

La protection anti-copie est présente sous plusieurs aspects :

- ces disques ne sont pas lisibles par un ordinateur ;
- il n'existe pas de graveur pour flux DSD ;
- un encryptage des données avec la clef inscrite sur le disque et lisible seulement par un lecteur agréé ;

- par construction, un lecteur Super Audio CD n'offre pas de sortie numérique pleine résolution.

La fabrication est généralement en double couche hybride collant une couche SA-CD et une couche CD. La couche CD est lue à travers la couche SA-CD semi-transparente.

◆ Blu-ray Disc

Comme souvent, une bataille de format a précédé l'adoption du format : Blu-ray (Sony) contre HD-DVD (Toshiba et Microsoft).

Grâce à un laser bleu-violet et une gravure bien plus fine que pour le DVD, la densité d'information est 10 fois plus importante et le débit maximal 5 fois supérieur (54 Mbps au lieu de 10 Mbps).

Résumons l'apport de cette technologie par rapport au DVD :

- Vidéo HD en définition 1920x1080 aux formats MPEG2, MPEG4-AVC ou VC1 ;
- augmentation de la capacité de stockage jusqu'à 50Go ;
- nouvelles fonctionnalités d'interactivité (diaporama, animation des menus, programmation BD-java, liens vers internet avec BD-live...);
- Protection anti-copie Advanced Access Content System (AACS) et BD +, cryptographie dynamique (physique) ;
- watermarking ROM-Mark évitant la copie vers un disque BD enregistrable ;
- Protection HDCP ;
- Un choix important de formats audio multicanal : LPCM, Packed PCM (« lossless ») et codages avec pertes.

Dolby et DTS ayant proposé différents formats de codage avec/sans pertes afin de satisfaire les besoins de qualité ou de durée, voici les formats autorisés pour le maximum de canaux :

Format	Version Dolby	Version DTS
LPCM 8 canaux		
PPCM 8 canaux	Dolby True HD	DTS-HD Master audio
Avec pertes 5.1	Dolby Digital	DTS Digital Surround
Avec pertes 7.1	Dolby Digital +	DTS-HD High resolution audio

Figure 7.25 – Formats audio multicanal sur Blu-ray, © J.-L. Ohi.

Des versions enregistrables sont disponibles en BD-R (enregistrable une seule fois) et BD-RE (ré-enregistrable). Si la finalité est l'archivage, il faut rester prudent quand aux durées de vie, en particulier pour les versions ré-enregistrables.

◆ Production de supports optiques

Pour tous les supports optiques, le mastering est une étape entre postproduction et le mastering, décomposée en deux phases : l'encodage correspond à la transformation des sources en fichiers compatibles avec le format final et l'autoring permet d'assembler les flux et de créer la navigation et l'interactivité.

Bit Budgeting

En fonction de la durée des programmes, des langues, des flux audio (format et nombre de canaux), des sous-titres, des menus... le calcul des débits de la vidéo et de l'audio tient compte de la capacité du disque et du débit instantané admissible. Souvent en fonction de ces impératifs se fait le choix du format audio : sans pertes ou avec réduction de débit.

Encodage et compression

Selon le budget des données, il est possible de choisir des qualités différentes pour les éléments : par exemple privilégier la qualité du film et de la bande-son et réduire les débits des bonus.

Pour les formats multicanal, certains points doivent être vérifiés : l'ordre des pistes, la correspondance des codes temporels avec la vidéo, la validité des métadonnées. Trop fréquemment, des informations manquent et le studio encode avec les valeurs par défaut du logiciel ! Dans ce cas, la qualité du mixage du film n'est pas reproduite de façon optimale chez le spectateur.

◆ Supports à bandes

Les bandes ne sont plus présentes ni en grand public ni en professionnel : les formats multicanal ADAT et TDIF (8 pistes Tascam) ne sont plus utilisés. En vidéo, les cassettes Digital Betacam, 4 canaux audionumériques en 48 kHz et 20 bits, compatibles avec le format multicanal Dolby E sont remplacées par des fichiers.

◆ Connexions

Vous trouverez ci-après une liste non exhaustive de standards grand public et professionnels utilisés en multicanal.

S/PDIF et Toslink

Le nom du format S/PDIF est l'abréviation de **Sony-Philips Digital Interface**, il s'agit d'une norme de liaison numérique grand public asymétrique sur câble 75 Ohms, avec un niveau nominal à 0.5V et définie par la norme IEC60958.

Sa version optique est la liaison Toslink (Toshiba Link) qui utilise une diode électroluminescente et une fibre plastique. L'atténuation importante de ce type de fibre limite la distance à environ 15 m.

Les formats Dolby Digital et DTS, transportables dans un format PCM stéréo, sont compatibles avec les liaisons S/PDIF ou Toslink à condition que le récepteur sache décoder le flux.

AES

Le format professionnel AES3, défini aussi par la norme IEC60958, est similaire au S/PDIF mais avec tension nominale de 5V par liaison symétrique 110 Ohms ou par liaison asymétrique sur connecteur BNC en 75 Ohms.

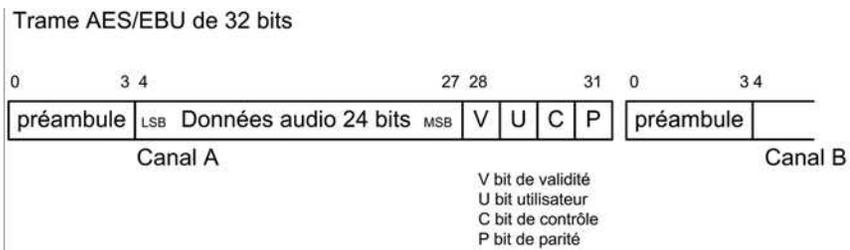


Figure 7.26 – Structure de la trame AES, © J.-L. Ohl.

ADAT

Le format **Alesis Digital Audio Tape** a été utilisé pour des enregistreurs et permet de véhiculer 8 canaux en 48 kHz/24 bits sur liaison optique ADAT Lightpipe identique physiquement à celle utilisée pour le Toslink.

DTRS

Le format **DTRS Digital Tape Recording System** de Tascam permet d'enregistrer 8 canaux sur bande Hi-8 jusqu'à 24 bits/48 kHz. De plus hautes fréquences d'échantillonnage sont possibles avec moins de pistes. Ce support était utilisé par les stations d'authoring DVD. La liaison bidirectionnelle TDIF (Tascam Digital Interface Format), utilise un câble multiconducteurs pour transmettre 8 canaux d'entrée et 8 de sortie.

MADI

Le **Multichannel Audio Digital Interface** est un format standardisé par l'AES10-2003 pour transmettre 56 ou 64 canaux, en 44.1, 48 kHz ou 96 kHz (limitée alors à 32 canaux). La connexion peut s'effectuer en liaison cuivre coaxiale BNC ou en fibre optique.

SDI

L'audio multicanal est souvent transmis comme un signal AES dans les VANC (Vertical Ancillaries Data) d'un signal vidéo SDI **Serial Digital Interface**, SD, HD ou 3G. On parle de signal audio « embeddé » qui signifie « intégré » qui peut transmettre jusqu'à 16 canaux audio 48 kHz, 24 bits. Ces signaux sont présentés par 4 groupes de 2 AES

chacun (4 canaux). Dans un signal HD, les groupes peuvent être différents en fréquence et en quantification.

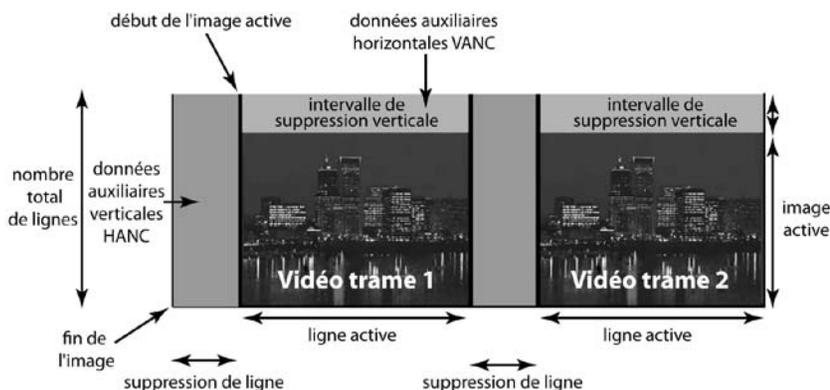


Figure 7.27 – Structure d'une image vidéo, © J.-L. Ohi.

Les signaux Dolby E, étant compatibles avec les trames AES, peuvent être contenus dans un ou plusieurs groupes, chaque groupe pouvant contenir à la fois des flux AES comme des flux Dolby E.

HDMI High Definition Multimedia Interface

Cette norme d'interfaçage entre appareils grand public permet de transporter vidéo et audio numériques multiplexés et cryptés ainsi que des informations de commande. La protection HDCP (High-bandwidth Digital Content Protection) agit en empêchant l'émetteur de transmettre en pleine définition si le récepteur n'est pas conforme à cette norme. Nombreux sont les formats audio pouvant être transmis en HDMI.

Complément Web 7.17

Tableau des contenus audio selon version de la norme HDMI.

7.5.11 Diffusion

◆ Fichiers audio

Sur support informatique, les données audio, dans les différents formats physiques et logiques, sans ou avec codage perceptif, sont présent en tant que fichiers.

Un fichier LPCM peut être sous format brut, .pcm ou .raw mais plus souvent en .wav ou .aif, formats qui ont l'avantage de contenir aussi les informations de fréquence d'échantillonnage, quantification, nombre de canaux, paramètres qui facilitent la lecture. Plus précisément, le format WAV est un conteneur (voir ci-dessous) de type RIFF.

L'EBU a normalisé le standard BWF par le document Tech-3285, qui constitue une extension du format WAV avec ajout des métadonnées et d'un code temporel pour faciliter la liaison avec la vidéo dans les stations de montage. La version RF64 permet de surmonter la limitation à 4 Go du format BWF de base. À noter qu'un fichier BWF peut aussi contenir de l'audio codé en MPEG.

◆ **Formats Conteneurs**

Un format conteneur (« wrapper » ou « container » en anglais) décrit et contient un ensemble de données organisées. Les conteneurs sont utilisés pour des données multimédias à l'aide de codecs normalisés accompagnés de métadonnées auxiliaires comme le sous-titrage, le chapitrage, le time code, etc. Les principaux formats conteneurs actuels sont wav, avi, Quicktime, MXF, Ogg, VOB, MKV. Il faut bien comprendre que le conteneur n'est pas un codec et ne contient pas le codec mais donne l'information sur le codec à utiliser pour la lecture.

En postproduction ainsi que dans les normes du cinéma numérique, le conteneur MXF (Material eXchange Format) est généralement utilisé. Les flux contenus sont appelés "essences". Ce conteneur inclut les informations de code temporel ainsi que des métadonnées. Le format MXF est complexe et a connu des problèmes de compatibilité et d'échange qui, depuis la redéfinition de la norme en 2009, ont été majoritairement résolus.

◆ **Audio sur réseau**

La diffusion audio sur réseau informatique a connu de nombreuses variantes surtout au niveau professionnel. Ces standards se différencient d'abord par la couche OSI (Open Systems Interconnection, prescription issue des organismes ISO et ITU) qui définit la couche physique ou logicielle des protocoles réseau.

MODÈLE OSI OPEN SYSTEMS INTERCONNECTION

- Couche 7 : Application.
- Couche 6 : Présentation et cryptage.
- Couche 5 : Session : gestion des sessions entre applications.
- Couche 4 : Transport : transfert des données, contrôle de flux.
- Couche 3 : Réseau : adressage, routage.
- Couche 2 : Liaison : syntaxe et erreurs.
- Couche 1 : Couche physique, électrique et fonctionnelle.

- En couche 1 OSI : A-Net, SuperMac, Rocknet, l'audio utilise le câblage Ethernet mais pas son routage ;
- En couche 2 OSI : Cobranet, Ethersound, AES51, AVB... l'audio est encapsulée dans des trames Ethernet ce qui permet souvent d'utiliser une partie de l'infrastructure réseau (commutateurs par exemple).

- En couche 3 OSI : Dante, Ravenna, Q-LAN, Livewire, l'audio est présentée dans des paquets IP, généralement en UDP, ce qui assure une bonne compatibilité avec une structure locale mais empêche un routage internet.

On peut donc globalement distinguer l'audio sur Ethernet (AoE) et l'audio sur IP (AoIP) qui occupent deux couches OSI différentes.

Les protocoles se différencient par de nombreux paramètres :

- La couche OSI décrite précédemment ;
- La compatibilité avec un routage local ou global ;
- La topologie du réseau : anneau, étoile ;
- la diffusion unicast ou multicast ;
- Le nombre et type (fréquence et échantillonnage) des canaux transportés, liés à la bande passante utile du réseau ;
- La latence ;
- La synchronicité et le transport d'horloge ;
- La robustesse.

L'incompatibilité des propositions a freiné l'adoption de ces technologies mais la norme d'interopérabilité AES67 pourrait changer cette situation.

◆ Téléchargement et streaming

Les réseaux informatiques privés ou publics (LAN, internet) permettent deux types d'échanges de données audio : le téléchargement et le streaming.

Le téléchargement est une transmission asynchrone de données qui n'est pas soumise aux contraintes de l'écoute en temps réel et permet d'échanger, plus ou moins rapidement n'importe quel format compressé ou non.

Par contre, le streaming, une diffusion en flux, est limitée par la bande passante disponible, d'où des contraintes sur le débit et donc sur le format des données.

◆ DVB : Télévision numérique

Alors que la télévision hertzienne analogique limitait les formats audio aux codages matricés en stéréophonie comme le Dolby Surround, les principaux standards de diffusion sont issus d'une norme DVB ouverte et évolutive. Cette norme s'appuie sur des formats MPEG-2 et MPEG-4 qui permettent d'inclure un signal multicanal dans les codage audio prévus en MPEG-2 et les variantes de l'AAC.

L'ensemble des normes DVB se décline selon le type de diffusion :

- DVB-T diffusion TNT terrestre ;
- DVB-S diffusion satellite ;
- DVB-C diffusion par câble ;

- DVB-H diffusion pour récepteurs portables ;
- DVB-IPTV diffusion sur réseaux IP.

En pratique, les télévisions émettant dans des formats HD se limitent au multicanal 5.1 avec une partie des programmes obtenus par traitement d'upmix de programmes stéréo.

◆ **DAB : radio numérique**

Dans la plupart des pays, la radio numérique Digital Audio Broadcasting est basée sur une norme DAB + qui contient un format audio MPEG Audio Layer 2 ou HE-AAC V2. La France a aussi opté pour DAB + tout en autorisant ce qui fut un premier choix très contesté, le format T-DMB qui impose la présence d'un flux vidéo en MPEG4.

◆ **Cinéma**

Le cinéma est à l'origine des premières diffusions en multicanal. Fantasia, dessin animé de Disney, fut en 1940 un précurseur en employant 3 canaux, le Fantasia-Sound.

Le cinéma avec diffusion par serveur informatique a remplacé le 35 mm. Ce cinéma numérique est basé sur la recommandation DCI (Digital Cinema Initiative) qui régit l'ensemble de la chaîne de production, tournage, postproduction et projection numérique.

Trois étapes sont présentes :

- Le DSM (Digital Source Master) est le master numérique issu de la postproduction ;
- Le DCDM (Digital Cinema Distribution Master) est créé pour réaliser les DCP en rajoutant les informations nécessaires pour la projection : images, sons, sous-titrage ;
- Le DCP (Digital Cinema Package) est destiné à l'exploitation en salle. Les images vidéo sont compressées en format JPEG2000, multiplexées en MXF puis cryptées. Cet ensemble de fichiers est transmis aux salles pour être stocké dans le serveur D-Cinema de la cabine de projection.

Pour attirer le public, le cinéma doit conserver une avance technologique par rapport aux installations domestiques. Les configurations multicanal 5.1 étant devenues courantes chez le particulier, le cinéma a besoin d'un nouveau saut qualitatif.

Le format DCP prévoit jusqu'à 16 canaux sonores. Au-delà du format 7.1, diverses évolutions sont proposées, toutes ayant vocation d'augmenter la sensation d'immersion, en rajoutant en particulier une "dimension" de hauteur.

On peut regrouper les technologies en trois concepts :

- « Canal » : la notion habituelle dite « core mix » ou « beds » où un canal de stockage/transmission est associé à un canal de diffusion, une piste correspond à un haut-parleur ;

- « Objets sonores » : le signal d'une piste peut être répartie dynamiquement par le processeur audio dans les canaux de diffusion selon des métadonnées inscrites dans le flux et selon la configuration, nombre et emplacements des haut-parleurs ;
- « Scènes » : le concept High Order Ambisonics HOA simule précisément le champ acoustique par un ensemble réduit de vecteurs de pression.

Les propositions en partie basées sur ces concepts, sont :

- Barco Auro-3D : du 9.1 au 13.1, ce codage propose une dimension verticale en rajoutant en hauteur des canaux frontaux et surround ainsi que des canaux de plafond, avec empaquetage et compatibilité dans un flux LPCM de 6 canaux et solution d'up-mix ;
- Dolby Atmos avec comme canaux supplémentaires : Lc et Rc en canaux d'écran centraux, des canaux Side surround, deux rangées de canaux au plafond et un bass-management des surrounds avec 2 SUB arrière. Le format est variable selon la taille de la salle, avec un rendu sur 64 enceintes maximum. Dolby a racheté le procédé Imm Sound qui permettait un déploiement de formats 14.1 à 23.1 avec au moins 3 canaux de plafond, les canaux supplémentaires permettant un placement plus fin des sources sonores ;
- Iosono 3D : basé sur la technologie WFS Wave Field Synthesis, avec rendu de 64 sources sur 128 canaux au maximum ;
- DTS MDA : un format ouvert avec audio « lossless » et la possibilité d'intégrer pistes, objets et scènes.

Complément Web 7.18

MDA (Multi-Dimensional Audio, un format de contenus ouverts pour le son immersif).

Les dispositions correspondantes des enceintes sont précisées aux chapitres 1 et 2.

Processeurs de cinéma

La chaîne de traitement comprend un serveur de lecture comme le Doremi DCP-2K4 ou le Dolby DSS200, qui décryptent et décodent les fichiers vidéo, audio, sous-titrage pour les envoyer vers le projecteur d-cinéma et le processeur audio.

Les fonctions du processeur audio sont les suivantes :

- gestion des entrées et sorties aux différents formats ;
- décodage du flux multicanal entrant ;
- traitement spatial ;
- filtrage et égalisation des enceintes ;

- possibilité d'insérer un autre traitement de décodage externe pour les formats non pris en charge nativement ;
- gestion des niveaux ;
- commandes externes (pilotage, GPIO...).

Quelques exemples de processeurs audio répandus :

- Dolby CP850, le successeur du CP750, propose un décodage Atmos ;
- Datasat AP20 : avec correction par processeur Dirac-Live, travaillant en convolution basée sur des mesures audio relevées en plusieurs points de la salle ;
- le processeur Trinnox Ovation permet une égalisation fine par convolution du signal audio après mesures acoustiques mono ou multipoints.

◆ Jeux vidéo

Depuis les flippers, les machines à sous puis les premiers jeux électroniques, le son a toujours eu une place importante dans les sensations du jeu. En réalité virtuelle, l'audio est un élément fondamental de l'**immersion** et du **réalisme**. De plus, la musique et les effets dans les jeux vidéo sont dynamiques : le rythme, les timbres, la localisation sont liés à l'évolution de la partie et appuient l'interactivité. Les sources virtuelles et l'environnement sonore participent incontestablement à la qualité perçue et une bande multicanal représente un atout incontestable pour la réussite d'un jeu.

Certaines **consoles de jeu** sont conçues pour supporter nativement des formats multicanal. Par exemple, la console Wii admet le Dolby Pro Logic II, la Xbox 360 supporte le Dolby Digital et la PlayStation 3 est prévue pour le Dolby TrueHD et le DTS-HD, jusqu'au 7.1 et 192 kHz. La console portable PSP dispose d'un traitement audio 3D. Certains **téléphones portables** sont suffisamment puissants en traitement DSP pour intégrer la lecture de flux multicanal et les diffuser sur casque ou haut-parleurs intégrés, avec éventuellement un traitement transaural.

La plupart des jeux récents **sur ordinateur** sont créés en multicanal, mais les qualités d'immersion et de réalisme dépendent des possibilités des cartes son et du nombre d'enceintes acoustiques. Dans ce domaine, les traitements de rematriçage peuvent s'avérer intéressants et le joueur étant à une place fixe devant l'écran, les technologies de spatialisation transaurale proposent une excellente immersion auditive avec seulement deux haut-parleurs.

Codage multicanal	Codage	Format	Echantillonnage et débit max.	Pistes	Affectation standard en 7 canaux max.
MPEG Multichannel (MPEG1 layer 2 extended)	Perceptif	5.1	48kHz, 24bits 1130kbps	L R C LFE Ls Rs	L C R Ls Rs LFE
MPEG2 AAC-LC MPEG4 HE-AAC	Perceptif	7.1	48kHz, 24bits 1152kbps	C L R Ls Rs LFE	L C R Ls Rs LFE
Dolby Surround ProLogic	Matricé	4.0		Lt Rt	L C R S
Dolby ProLogic II	Matricé	5.0		Lt Rt	L C R Ls Rs
Dolby Digital	Perceptif	5.1	48kHz, 24bits 640kbps	L C R Ls Rs LFE	L C R Ls Rs LFE
Dolby EX	Perceptif, matricé	7.1	48kHz, 24bits 640kbps	L C R Ls Rs LFE	L C R Ls Cs Rs LFE
Dolby Digital Plus	Perceptif	7.1 (13.1)	48kHz, 24bits 6Mbps	L C R Ls Rs LFE	L C R Ls Cs Rs LFE
Dolby True HD	Sans pertes	7.1 (14.0)	192kHz, 24bits 18Mbps	L R C LFE Ls Rs	L C R Ls Cs Rs LFE
DTS	Perceptif	5.1	48kHz, 24bits 1509kbps	C L R Ls Rs LFE	L C R Ls Rs LFE
DTS2496	Perceptif	5.1	96kHz, 24bits 1509kbps	L R C LFE Ls Rs	L C R Ls Rs LFE
DTS-ES	Perceptif	6.1	48kHz, 24bits 1509kbps	L R C LFE Ls Rs Cs	L C R Ls Cs Rs LFE
DTS Express	Perceptif	7.1	48kHz, 24bits 512kbps	L R C LFE Ls Rs Lb Rb	L C R Ls Rb Lb Rs LFE
DTS-HD High Resolution	Perceptif	7.1	192kHz, 24bits 6Mbps	L R C LFE Ls Rs Lb Rb	L C R Ls Rb Lb Rs LFE
DTS-HD Master Audio	Sans pertes	7.1	192kHz, 24bits 24.5Mbps	L R C LFE Ls Rs Lb Rb	L C R Ls Rb Lb Rs LFE
Sony SDDS	Perceptif	7.1	44.1kHz, 20bits 2.2Mbps	L Lc C Rc R LFE Ls Rs	L Lc C Rc R LFE Ls Rs

Figure 7.28 – Tableau des principaux codages multicanal, © J.-L. Ohl.

Chapitre 8

Exploitation Son Broadcast, supports, nouveaux médias

POINTS ESSENTIELS

- Chaîne d'exploitation broadcast.
- Dolby Surround.
- Metadata Dolby : loudness, downmix, DRC.
- Metadata AAC, DTS.
- Mesure du loudness.
- Normes PAD TVHD, mesure des niveaux, recommandations CST, EBU, R128.
- Exploitation des outils de mesure et de traitement broadcast 5.1 : traitement loudness, gestion des metadata, upmix de diffusion.
- Les formats DTS pour le support, DTS Neural.
- Diffusion 5.1 et binaurale pour les nouveaux médias.

La phase de mixage multicanal s'accompagne de métadonnées (metadata) qui facilitent la diffusion et permet une meilleure maîtrise de celle-ci chez le particulier. Ces metadata, associées principalement au flux Dolby Digital, se retrouvent sur le support home-cinéma DVD, Blu-ray et sont notamment présentes dans la diffusion

broadcast mais ne touchent pas le secteur du cinéma. Ces métadonnées transportent la mesure loudness du programme afin d’harmoniser les niveaux entre différents programmes ou différentes chaînes. Elles permettent l’adaptation au lieu d’écoute domestique à travers des profils dynamiques que l’on peut attribuer au programme ainsi qu’elles traitent de la compatibilité stéréo et mono par des coefficients de downmix.

8.1 La chaîne d’exploitation Broadcast

8.1.1 Synoptique du signal depuis la captation à la diffusion

Le schéma ci-dessous représente le cheminement du signal audio depuis la captation d’un événement sportif dans un car HD, jusqu’à la diffusion chez le particulier. Il illustre le transport des programmes 1 et 2 en Dolby E (5.1 + 2), le décodage Dolby E en PCM pour la postproduction avec l’ajout éventuel de commentaires, l’encodage Dolby E pour le stockage sur cassette HDCAM ou sur serveur, le décodage Dolby E en PCM pour la vérification PAD et l’encodage Dolby Digital suivi du multiplexage avec la vidéo pour la diffusion.

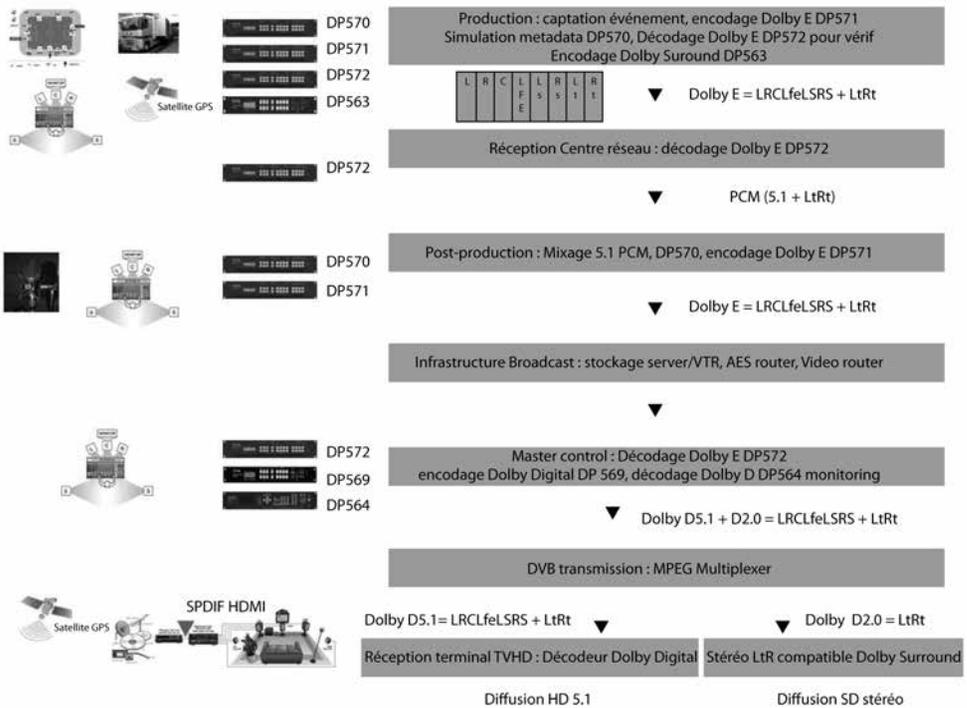


Figure 8.1 – De la captation à la diffusion, © B. Périaux. En couleur sur dunod.com

8.1.2 *Dolby Surround*

Les formats Dolby Surround, l'encodage et le décodage Pro Logic 1 et 2 ont été décrits dans le chapitre 7. L'encodage Dolby Surround offre la compatibilité multicanal sur deux canaux, mais pose néanmoins des contraintes lors du mixage. Certains artefacts peuvent survenir lorsque le mixage est très spatialisé, avec notamment l'utilisation de réverbérations. Ces artefacts se produisent aussi lorsque trop d'information de phase est générée dans l'espace frontal. Ainsi, des prises de son avec différences de phase (systèmes de microphones distants, type AB), des séquences avec une réverbération importante, risquent d'être dénaturées par l'encodage/décodage car certains éléments peuvent se retrouver à l'arrière, notamment des transitoires. En effet, le décodeur analyse la différence de phase entre Lt et Rt, et spatialise des informations frontales pertinentes vers l'arrière, en plus des canaux Ls et Rs ayant subi le décalage de phase. L'encodage Dolby Surround (Pro Logic 1) réduit la spatialisation arrière puisque les canaux arrière sont en mono, l'équilibre spectral arrière est modifié (bande passante restreinte de 100 Hz à 7 kHz). L'encodage Dolby Pro Logic 2 améliore le rendu de l'image sonore et respecte l'équilibre spectral de la zone arrière. En revanche, cet algorithme a toujours tendance à réduire l'impression spatiale et à produire une focalisation vers le canal central. Les décodeurs Dolby Surround et Dolby Surround Pro Logic 2 retardent les canaux Ls et Rs de 15 ms ce qui, grâce à l'effet de précedence, permet de conserver les canaux frontaux comme étant le premier front d'onde et limite ainsi les artefacts de localisation. Ce réglage de 15 ms est renseigné par défaut dans le DP570, le DP564 et les décodeurs dans les amplificateurs AV et les téléviseurs.

Dans la spatialisation multicanal Dolby Surround, il faut éviter de positionner des sources dans le dernier quart de la zone arrière, parce que le downmix mono somme les canaux Lt et Rt, qui eux-mêmes contiennent Ls + Rs en opposition de phase, lesquels s'annulent lors de cette addition.

DOLBY SURROUND

- Signal stéréo LtRt compatible 5.1 : Dolby pro logic1, Dolby pro logic2.
- Canaux arrière en opposition de phase.
- Éviter la spatialisation de sources dans le dernier quart arrière pour la compatibilité mono.
- Le recours à une grande spatialisation de phase peut induire des artefacts lors du décodage.
- Après décodage, l'image sonore est focalisée vers le canal central, avec impression spatiale réduite (diaphonie entre les canaux).

8.1.3 *Les metadata Dolby*

Les metadata contenues dans le flux Dolby Digital permettent de restituer une écoute de qualité chez le particulier, en fonction des caractéristiques acoustiques du lieu

d'écoute. Les metadata sont transportées dans chacune des trames, à travers les interfaces AES ou SPDIF. Elles jouent un rôle de description du contenu audio et servent aussi à contrôler le signal audio lors du décodage. Ces metadata sont renseignées par le mixeur du programme et sont utilisées au décodage, combinées avec le réglage de l'écoute de l'utilisateur sur l'amplificateur AV home-cinéma (Bass management, ajustement optionnel des metadata loudness et DRC). La première metadata, le Dialnorm, porte la valeur de loudness du programme, la deuxième, les DRC (Dynamic Range Control), intègre des profils de compression dynamique et la troisième, le Downmix, permet la compatibilité stéréo et mono du signal.

◆ Dialnorm

Le Dialnorm est la metadata qui porte la valeur de loudness du programme, c'est-à-dire le niveau moyen mesuré sur toute la durée du programme (voir mesure du Loudness). Le plus souvent, c'est le niveau de la parole qui sert de référence, d'où le nom de Dialog level. On distingue donc, le Dialog level qui mesure le niveau moyen des dialogues d'un programme de type film, du loudness, qui mesure le niveau moyen d'un programme sans dialogue sur toute sa durée. La metadata Dialnorm sera un Dialog level pour la mesure d'un film, d'un documentaire, elle sera un loudness pour la mesure d'un programme musical ou une publicité.

Nous aborderons les metadata Dolby par la mesure loudness telle qu'elle a été pensée par Dolby à l'origine, c'est-à-dire le Dialog level. Nous verrons ultérieurement que la metadata Dialnorm en Europe se mesure depuis 2011-2012 suivant la norme de loudness EBU R128.

Voici ci-dessous la représentation de niveaux de loudness (mesure du Dialog level Dolby) de programmes courants :

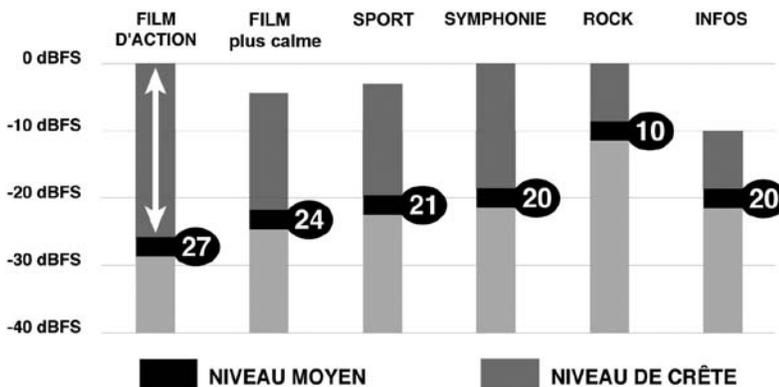


Figure 8.2 – Niveaux de loudness de programmes courants, © Dolby.

La diffusion analogique TV ou radio FM est un vecteur de transmission avec un niveau de bruit et une réserve dynamique réduite. Le niveau de référence est +4 dBu (0 VU) ou

– 18 dBFS, le niveau crête max est à + 13 dBu ou – 9 dBFS ou 0 crête, l'écart référence/crête est donc de + 9 dB, c'est une valeur très réduite.

De l'analyse des films les plus délicats, Dolby préconise une valeur moyenne de loudness de – 31 dBFS (31 dB au-dessous du niveau de loudness maximal 0 dBFS) mesuré avec la pondération K soit –31 LKFS (voir paragraphe mesure du loudness). Cette valeur permet d'exprimer pleinement toute la dynamique du programme, notamment sur les passages d'action (forts transitoires, explosions).

Lorsqu'on mesure le Dialnorm d'un programme, prenons – 24 LKFS par exemple, le décodeur Dolby Digital appliquera alors à la diffusion une atténuation de 7 dB pour restituer le programme à la valeur moyenne de – 31 LKFS. Pour un programme de – 21 LKFS, le décodeur appliquera une atténuation de 10 dB, ainsi de suite. Les programmes à – 31 LKFS ne seront pas atténués et seront diffusés tels quels.

On définit donc le Dialnorm comme suit.

31 + Dialnorm = atténuation appliquée

La diffusion Dolby Digital des programmes courants sera la suivante, tous les programmes sont ramenés à – 31 LKFS avec leur dynamique originale préservée :

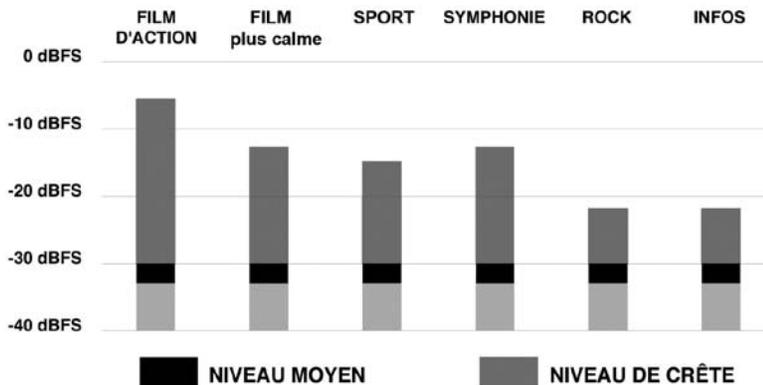


Figure 8.3 – Diffusion Dolby Digital des programmes courants, © Dolby.

En résumé, la metadata Dialnorm, par la mesure systématique du loudness de chaque programme, homogénéise le niveau moyen de restitution à la diffusion, sans que l'utilisateur ait besoin de changer le niveau d'écoute, et préserve la dynamique originale des programmes sans faire appel à des traitements dynamiques d'antennes (limiteurs, compresseurs).

Les différentes interfaces de réception

Les signaux de réception à la diffusion TV utilisent différentes interfaces en fonction de la nature analogique ou numérique du signal, le nombre de canaux mono, stéréo,

5.1. Le décodeur alimente une sortie remodulée RF analogique en mono, une sortie analogique stéréo ou LtRt compatible Dolby Surround sur Péritel, une sortie numérique stéréo PCM sur coaxial, optique ou HDMI (stéréo et 5.1) et enfin une sortie 5.1 numérique Dolby Digital sur coaxial ou optique ou HDMI.

◆ DRC

Les divers environnements d'écoute domestiques introduisent de nombreuses contraintes pour la dynamique de reproduction. La dynamique acceptable dans un salon est limitée par sa taille, son isolation, les bruits environnants. D'autre part, suivant la nature du programme, l'heure, les conditions d'écoute ou simplement ses goûts personnels, un auditeur ne désirera pas bénéficier de la même dynamique. Au lieu de compresser le signal audio pour qu'il se comporte au mieux sur les plus mauvais systèmes d'écoute, une metadata appelée DRC est véhiculée dans le Dolby Digital jusqu'au décodeur, ce dernier applique alors le profil de compression renseigné par le mixeur afin de réduire la dynamique du signal.

Les profils de traitement dynamique associés à la metadata DRC, compressent à des taux variables les hauts niveaux et relèvent les bas niveaux. Le seuil de compression est la valeur renseignée dans le Dialnorm, d'où l'importance de bien l'indiquer pour éviter une détérioration du signal.

La plupart des décodeurs Dolby Digital offrent la possibilité d'activer ou de désactiver les DRC, mais certains ne le permettent pas. Tous les décodeurs avec sorties 5.1 discrètes le proposent généralement. Les décodeurs avec sorties stéréo, mono, ou RF remodulée sur les lecteurs DVD ou set-top boxes activent automatiquement la metadata DRC en sortie analogique. On peut cependant récupérer le flux Dolby Digital numérique et le décoder sur un ampli home-cinéma plus complet.

Le flux Dolby Digital transporte deux profils de compression suivant les deux modes **Line mode** et **RF mode**. Ces modes sont identifiés par le type de sortie que le décodeur alimente. Généralement, le Line Mode gère les sorties de niveaux lignes sur les décodeurs en 5.1 ou stéréo, le RF mode est associé aux décodeurs qui alimentent une entrée mono RF/antenne d'un téléviseur.

Les décodeurs avancés proposent d'activer ou désactiver les DRC et de sélectionner le Line mode ou le RF mode. Généralement, les options proposées sont « off », « light compression » et « heavy compression » pour désigner pas de compression (DRC désactivés), Line mode, et RF mode. Sur les décodeurs complets l'utilisateur pourra même régler par lui-même le taux de compression des hauts et bas niveaux (mode « Custom DRC »). Le RF mode est généralement associé à un profil de compression importante pour alimenter les écoutes avec des enceintes de petite dimension, pour les ordinateurs portables, ou les téléphones portables.

Les sorties lignes analogiques sur connecteurs RCA des lecteurs DVD, des set-top boxes, des télévisions, des amplificateurs AV sont associées au Line mode.

Lors du décodage du flux Dolby Digital et quel que soit le mode choisi, la normalisation du dialogue se fait à chaque fois et en premier. Lorsque le programme est ajusté au bon niveau la compression s'applique. Normalement (sauf si erreur dans le renseignement des métadonnées) la « Null Band » du profil *DRC* doit être centrée sur la valeur de « Dialogue Level » (- 31 LKFS dans tous les décodeurs Dolby Digital).

DRC : DYNRRNG ET COMPR

La compression DRC fonctionne par blocs audio, dont le niveau est contrôlé par des variables *dynrng* et *compr*. La variable *dynrng* correspond à un fonctionnement du décodeur en Line Mode. Ainsi, à chaque bloc audio (5,3 ms) une valeur de *dynrng* indique le gain à appliquer au bloc concerné. Ce gain peut varier entre +/- 24 dB avec une résolution de 0,25 dB. La variable *compr* correspond à un fonctionnement du décodeur en RF Mode. Le mot *compr* est envoyé non pas à chaque bloc, mais à chaque trame (32 ms). Il commande un gain de +/- 48 dB avec une résolution de 0,5 dB.

En mode 5.1, l'utilisateur choisit de désactiver les DRC et profite pleinement de la dynamique du programme, ou de les activer avec le Night mode (light ou heavy, soit Line Mode ou RF Mode). En revanche, **en sortie stéréo downmix, les DRC sont automatiquement activés, avec le profil RF mode.**

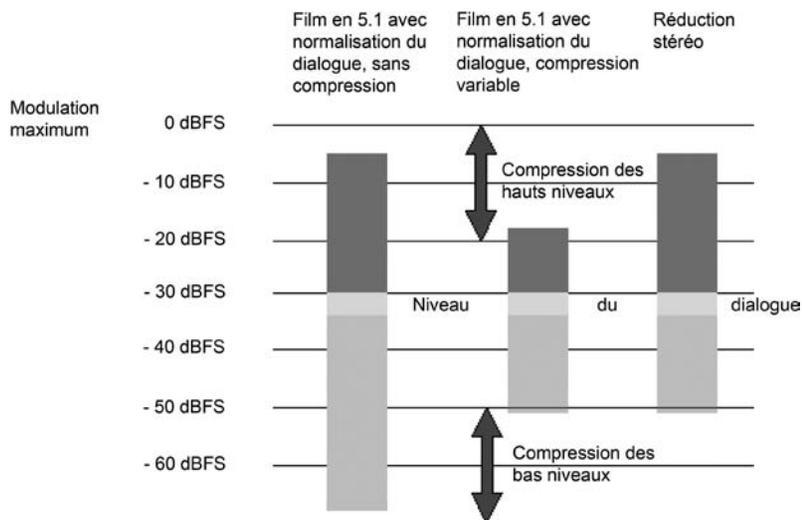


Figure 8.4 - Compression Line Mode, © A. Libolt, Dolby

Le RF mode associe une compression permanente des hauts et bas niveaux, il limite les crêtes, avec un gain de + 11 dB pour être conforme aux niveaux transportés en télévision analogique RF. Le dialogue, réglé suivant la valeur de Dialnorm - 31 LKFS et combiné avec le changement de gain de +11 dB, est reproduit au niveau constant de - 20 LKFS.

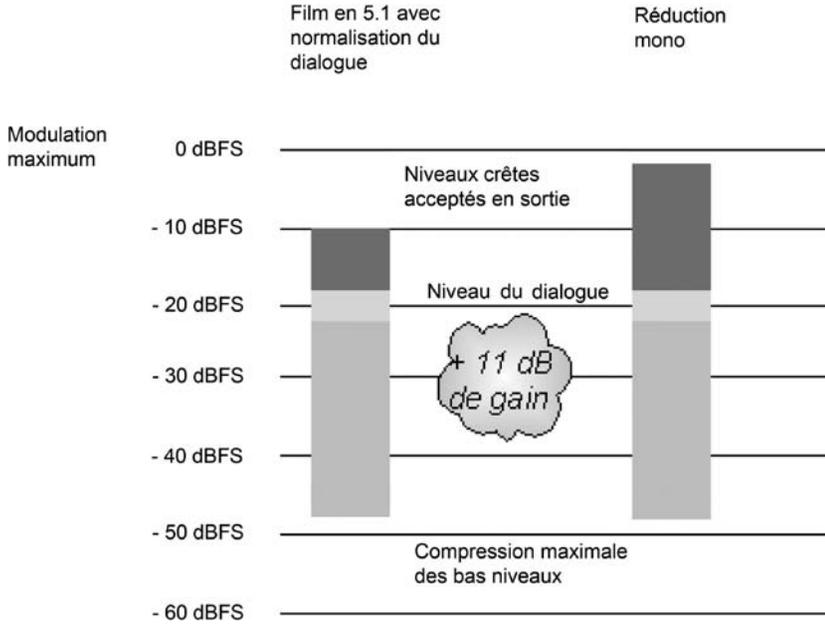
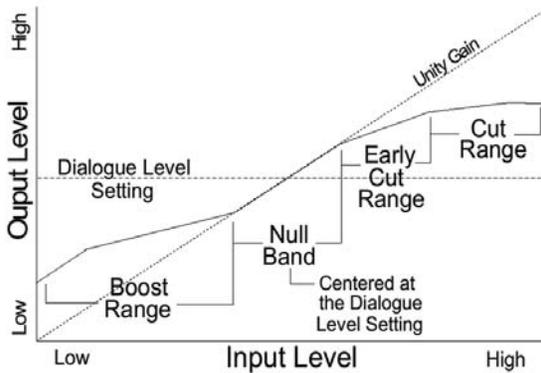


Figure 8.5 – Compression RF Mode, © A. Libolt, Dolby.

Profils DRC

Il existe six profils DRC : film standard, film light, music standard, music light, speech et none.



- Film Light**
 Max Boost: 6 dB (sous -53 dB)
 Boost Range: -53 dB à -41 dB (2:1 ratio)
 Null Band Width: 20 dB (-41 dB à -21 dB)
 Early Cut Range: -26 dB à -11 dB (2:1 ratio)
 Cut Range: -11 dB à +4 dB (20:1 ratio)
- Film Standard**
 Max Boost: 6 dB (sous -43 dB)
 Boost Range: -43 dB à -31 dB (2:1 ratio)
 Null Band Width: 5 dB (-31 dB à -26 dB)
 Early Cut Range: -26 dB à -16 dB (2:1 ratio)
 Cut Range: -16 dB à +4 dB (20:1 ratio)
- Music Light** (aucun early cut range)
 Max Boost: 12 dB (sous -65 dB)
 Boost Range: -65 dB à -41 dB (2:1 ratio)
 Null Band Width: 20 dB (-41 dB à -21 dB)
 Cut Range: -21 dB à +9 dB (2:1 ratio)
- Music Standard**
 Max Boost: 12 dB (sous -55 dB)
 Boost Range: -55 dB à -31 dB (2:1 ratio)
 Null Band Width: 5 dB (-31 dB à -26 dB)
 Early Cut Range: -26 dB à -16 dB (2:1 ratio)
 Cut Range: -16 dB à +4 dB (20:1 ratio)
- Speech**
 Max Boost: 15 dB (sous -50 dB)
 Boost Range: -50 dB à -31 dB (5:1 ratio)
 Null Band Width: 5 dB (-31 dB à -26 dB)
 Early Cut Range: -26 dB à -16 dB (2:1 ratio)
 Cut Range: -16 dB à +4 dB (20:1 ratio)
- None**

Figure 8.6 – Profils de compression DRC, © Dolby.

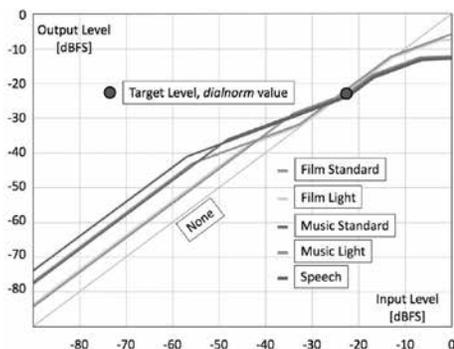


Figure 8.7 – Profils de compression DRC, © EBU Tech 3343.

Comme indiqué sur la **figure 8.7**, les profils comportent tous une zone linéaire appelée « la Null Band » où le gain est unitaire, il n'y a pas de compression. Cette zone, centrée sur la valeur du Dialog level, est plus ou moins large suivant le réglage choisi. Ainsi, si la valeur du Dialog level est correctement renseignée par l'opérateur, la compression du signal sera cohérente. De part et d'autre de la « Null Band », une compression est appliquée afin de réduire les forts niveaux et de remonter les bas niveaux. Les paramètres pour ces deux opérations dépendent du profil choisi mais sont définis sur le même modèle : un réglage pour l'amplification des bas niveaux (« Boost Range ») et deux réglages pour l'atténuation des forts niveaux (« Early Cut Range » avec un rapport 2 : 1 et « Cut Range » avec un rapport 20 : 1). Le sixième choix est « None », soit aucun profil DRC sélectionné. Cependant cela n'empêche pas de réduire la dynamique pour éviter les saturations lors des downmix.

◆ Downmix

La metadata Downmix permet la diffusion à travers le flux Dolby Digital vers tous les formats d'écoute : stéréo, mono et LtRt compatible Dolby Surround. Le flux Dolby Digital est transmis en 5.1, il assure une compatibilité stéréo par une opération de « Downmix » qui consiste à sommer les canaux arrière en phase dans le mode LoRo ou hors phase dans le mode LtRt. Le LtRt peut être alors écouté tel quel, en stéréo, ou redirigé vers un décodeur Dolby Prologic pour une écoute multicanal matricée. Le choix du mode de Downmix est fait par le mixeur, en conformité avec le cahier des charges du diffuseur. Le mixeur active ou non le mode « Phase shift », qui décale la phase des canaux arrière de 90° dès l'encodage. À la diffusion, l'information hors phase entre les canaux Lt et Rt est analysée et redistribuée vers les canaux arrière Ls et Rs.

À la diffusion chez le particulier, le flux Dolby Digital est décodé en 5.1 ou s'adapte au mode de diffusion grâce à la metadata Downmix :

- les modes LoRo ou LtRt pour les formats stéréo ;
- la sommation Lo + Ro ou Lt + Rt pour le format mono ;
- le décodage Dolby Pro Logic du LtRt pour le mode Dolby Surround.

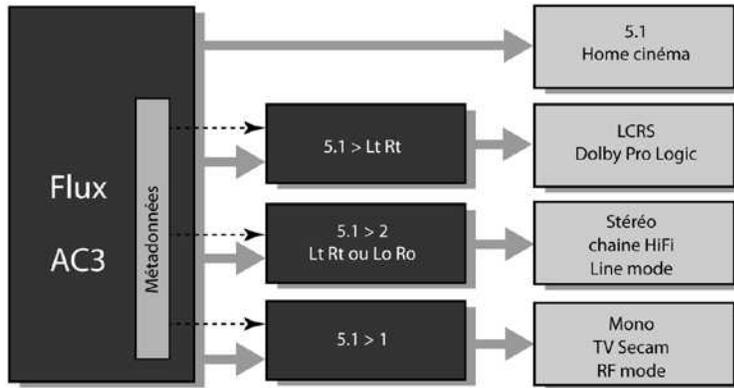


Figure 8.8 – Les différents downmix issus du Dolby Digital, © B. Périaux.

Le mode de downmix est renseigné dans la metadata « preferred downmix », il utilise les coefficients de réduction pour le canal central et les canaux arrière, associés au mode choisi. On trouve ainsi les coefficients Center et Surround pour les modes LoRo et LtRt. Par défaut, ces coefficients sont -3 dB mais ils peuvent être modifiés si les downmix stéréo ou mono ne fonctionnent pas de façon optimale. Le changement se fait par pas de $1,5$ dB, ce qui donne des coefficients de $-\infty$, -6 dB, $-4,5$ dB, -3 dB, $-1,5$ dB, 0 dB, $+1,5$ dB, $+3$ dB. Dans la pratique, il arrive fréquemment d’atténuer le coefficient du canal central à $-4,5$ dB, et d’atténuer les canaux arrière avec un coefficient de $-4,5$ dB ou -6 dB lorsque les ambiances arrière gênent l’intelligibilité du signal. Le canal Lfe n’est pas réintégré dans les downmix du Dolby Digital.

Voici les équations de downmix :

ÉQUATIONS DE DOWNMIX DOLBY

- $Lo = L + (Cmix)C + (Smix)Ls$.
- $Ro = R + (Cmix)C + (Smix)Rs$.
- $Lt = L + (Cmix)C - (Smix) (Ls + Rs)$.
- $Rt = R + (Cmix)C + (Smix) (Ls + Rs)$.

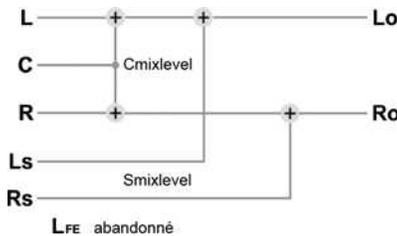


Figure 8.9 – Le downmix stéréo LoRo, © B. Périaux.

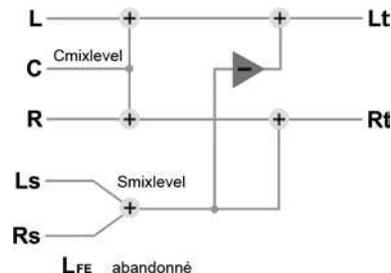


Figure 8.10 – Le downmix stéréo LtRt, © B. Périaux.

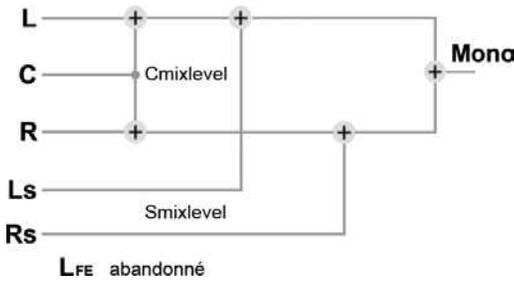


Figure 8.11 – Le downmix mono, © B. Périaux.

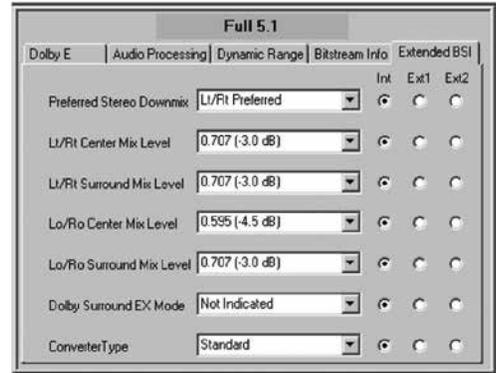


Figure 8.12 – Metadata Downmix dans l'interface Dolby DP570 Remote, © Dolby.

Le mixage 5.1 pour un programme TV doit par conséquent être compatible stéréo et mono, le mixeur doit assurer un compromis entre l'image sonore multicanal et l'image stéréo. Les difficultés portent essentiellement sur les critères d'homogénéité, d'intelligibilité et d'impression spatiale. Une image multicanal très enveloppante peut donner une image stéréo downmixée manquant d'homogénéité. Si les downmix ne donnent pas de résultats satisfaisants, il convient de modifier d'abord l'équilibre des sources dans le mixage 5.1 puis, en dernier recours, d'ajuster les coefficients de downmix.

METADATA DOLBY

- Dialnorm : renseigner le loudness du programme mesuré en fonction de la norme de loudness, le Dialog level Dolby (films, documentaires), le loudness Dolby (musique, publicités), ou le loudness R128 (mesure universelle).
- DRC : renseigner les profils de compression associés au Line Mode et au RF Mode.
- Downmix : renseigner le mode de downmix, LoRo ou LtRt (surround phase shift activé) ainsi que les coefficients de réduction associés.

8.1.4 Les metadata AAC

On trouve aussi des metadata dans la famille de codecs AAC, elles sont équivalentes à celles de Dolby, avec quelques modifications. Les tableaux ci-dessous font la synthèse des différences concernant la gestion du loudness, les DRC et le downmix. La metadata loudness s'appelle le « program reference level », comparable au « Dialnorm », ses valeurs varient par pas de 0,25 dB contre 1 dB pour le Dialnorm. Le niveau de sortie du décodeur n'est pas fixe contrairement aux modes line et RF imposant – 31 ou – 20 dB. La gestion des DRC est comparable dans les deux familles. En revanche, le downmix propose deux modes en AAC, le mode « matrix-mixdown » imposant – 3 dB pour le

centre et $(-\infty -6 -3 0)$ dB pour les canaux surround, ou bien un mode plus souple, mode auxiliaire, avec la possibilité de renseigner les coefficients $(-\infty -9 -7,5 -6 -4,5 -3 -1,5 0)$ dB pour le centre et l'arrière.

Comparaison des metadata Dolby et AAC, cf. Fraunhofer.

	AC3	(HE)-AAC
Loudness Normalization	Dialnorm	Program Reference Level
Dinamic Range Control : light compression	Line Mode	Dynamic Range Control
Dinamic Range Control : heavy compression	RF Mode	compression value
Downmix	Downmix	matrix-mixdown Downmixing levels

Comparaison du loudness Dolby et AAC, cf. Fraunhofer.

	AC3	(HE)-AAC
Loudness Normalization	Dialnorm	Program Reference Level
Bitstream field name	dialnorm	prog_ref_level
Range	-1 ... -31 dB	0 ... -31,75 dB
Granularity	1 dB	0,25 dB
Bits per Value	5 bits	7 bits
Repetition rate	1 value per Frame (1536 samples)	1 value per Frame (1 024/2048 samples)
Decoder Output Level	-31 dB Line Mode -20 dB RF Mode	0 ... -31,75 dB

Comparaison des DRC Dolby et AAC, cf. Fraunhofer.

	AC3	(HE)-AAC
Light Compression	Line Mode	MPEG dynamic Range Control
Bitstream field name	dynrng	dyn_rng_ctl, dyn_rng_sgn
Range	-24 ... +24 dB	-31,75 ... +31,75 dB
Granularity	0,25 dB	0,25 dB
Repetition rate	6 values per Frame (1536 samples)	1 value per Frame (1 024/2048 samples) + interpolation scheme

	AC3	(HE)-AAC
Heavy Compression	RF Mode	DVB compression_value
Bitstream field name	compr	compression_value
Range	-48 ... +48 dB	-48 ... +48 dB
Granularity	0,5 dB	0,5 dB
Repetition rate	1 value per Frame (1536 samples)	1 value per Frame (1 024/2048 samples)

Comparaison du Downmix Dolby et AAC, cf. Fraunhofer.

	AC3	(HE)-AAC
Coefficients		
Center	(-6 -4,5 -3 -1,5 0 +1,5 +3) dB	matrix-mixdown -3 dB
Surround	(-∞ -6 -4,5 -3 -1,5) dB	matrix-mixdown (-∞ -6 -3 0) dB
Center		Ancillary Data (-∞ -9 -7,5 -6 -4,5 -3 -1,5 0) dB
Surround		Ancillary Data (-∞ -9 -7,5 -6 -4,5 -3 -1,5 0) dB

Le flux HeAAC peut être utilisé pour transporter le son sur une chaîne TVHD, le CSA a communiqué sur cette option en 2009. Il est préconisé de pouvoir alors le transcoder en Dolby Digital ou en DTS avec conversion des metadata. Aujourd'hui, peu d'amplificateurs AV semblent intégrer des DSP de décodage compatibles AAC.

8.1.5 Metadata DTS

Le DTS intègre depuis peu des metadata comparables à celles du Dolby Digital, ou de l'AAC. Ainsi, lorsque l'on encode un programme en DTS pour les supports CD ou DVD Vidéo, on doit renseigner un Loudness. Le Loudness est renseigné comme pour le Dialnorm de Dolby, en revanche le Downmix n'est compatible que LoRo et **les coefficients de downmix sont fixes** (-3 dB, -3 dB). Pour les formats évolués DTS HD Master que l'on trouve sur le Blu-ray, l'utilisateur peut fabriquer une matrice de Downmix par l'utilisation de faders, par pas de 0,1 dB. Ainsi, les formats 7.1 d'un Blu-ray en DTS HD ou DTS HD Master sont compatibles 5.1, et les formats 5.1 sont compatibles stéréo, par le biais de cette matrice. La gestion de la dynamique du programme en DTS diffère des DRC Dolby ou AAC, on n'attribue pas au programme de profils dynamiques types, mais une courbe de niveau plus précise comparable à une automation de niveau. Cette courbe de niveau est exportée dans un fichier AAF depuis un Pro Tools et s'importe

dans le flux secondaire lors de l'encodage DTS, dans la suite DTS HD Master Audio Suite (voir exploitation des formats DTS).

8.1.6 Intégration des encodeurs Dolby E

Le DP570 est le processeur qui simule les metadata, il pilote les encodeurs Dolby Surround DP563, et Dolby E DP571. Il permet de renseigner les metadata du flux Dolby E et de simuler l'écoute des différents flux. La fabrication du signal Dolby E est faite à partir de ces trois boîtiers dans les cars HD ou dans les régies de postproduction 5.1.

Le DP570 comporte une section monitoring 5.1 + 2 analogique ou numérique pour écouter les différents flux : le full 5.1, le Dolby Surround simulé en 5.1, les downmix stéréo LtRt ou LoRo, et le downmix mono. Son intégration en régie de mixage peut ou non associer le signal LtRt aux metadata et au flux Dolby E. Le DP570 se câble en sortie de console de mixage ou en sortie d'une station audionumérique, sur le signal de mixage 5.1 et sur la stéréo LtRt. Généralement, le 5.1 fera l'objet du programme 1 et la stéréo LtRt sera le programme 2. Le programme 5.1 + 2 est le programme le plus souvent utilisé, mais il est possible d'utiliser d'autres configurations de programmes, il en existe 24 au total.

Les différentes configurations de programmes	
5.1 + 2	4 + 2
5.1 + 2 × 1	4 + 2 × 1
4 + 4	3 × 2
4 + 2 × 2	2 × 2 + 2 × 1
4 + 2 + 2 × 1	2 + 4 × 1
4 + 4 × 1	6 × 1
4 × 2	4
3 × 2 + 2 × 1	2 + 2
2 × 2 + 4 × 1	2 + 2 × 1
2 + 6 × 1	4 × 1
8 × 1	7.1
5.1	7.1 Scrn

Figure 8.13 – Configurations des programmes du DP570, © Dolby.

Les boîtiers Dolby se raccordent en BNC 75 ohms, et se synchronisent sur une référence vidéo, il convient donc d'utiliser des boîtiers transformateurs d'impédance 110 ohms-75 ohms. Le DP570 comporte une section « router » qui permet de redistribuer les canaux de mixage 5.1 qui alimentent les entrées, comme le proposent tous les boîtiers. La redistribution peut alimenter l'encodeur Dolby Surround DP563 qui distribue le mixage 5.1 vers l'encodeur Dolby E DP571. Dans le mode où le signal LtRt est intégré au flux Dolby E, la sortie principale Main Out LtRt du DP563 est redistribuée à l'entrée 7/8 du DP570, elle alimente alors le programme 2 du flux Dolby E et on utilise la sortie router 7/8 du DP570 pour alimenter l'entrée 7/8 de l'encodeur Dolby E DP571. Enfin, la sortie metadata du DP570 permet de véhiculer les metadata à l'encodeur DP571, sur une interface série RS485.

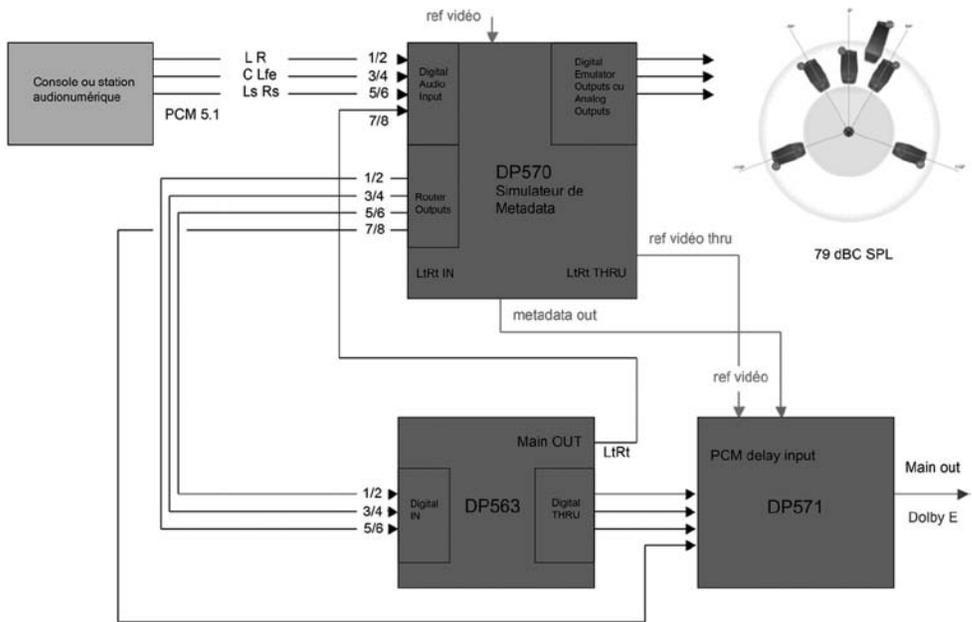


Figure 8.14 – Chaîne d'encodage Dolby E associant le LtRt généralement utilisée, © B. Périaux.

L'autre mode qui n'associe pas le signal LtRt au flux Dolby E, permet de sortir deux signaux, le Dolby E et la sortie LtRt synchronisés pour un report sur bande. La sortie Main Out du DP563 est câblée sur l'entrée LtRt in du DP570 pour être monitorée et alimente l'entrée PCM delay du DP571, via la sortie LtRt thru. L'entrée PCM delay permet de garder la synchronisation des deux signaux, compte tenu du retard d'une image induit par l'encodage Dolby E. Le DP572 alimente le PCM delay, le signal stéréo LtRt, vers l'entrée 1/2 du VTR, et la sortie principale Main Out, le Dolby E 5.1, vers l'entrée 3/4 du VTR.

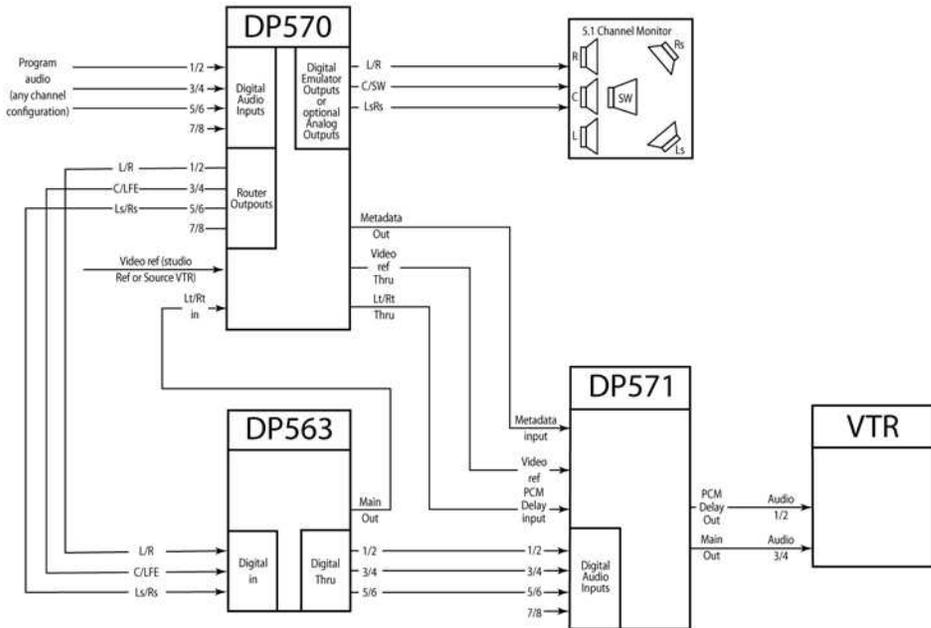


Figure 8.15 – Chaîne d'encodage Dolby E n'associant pas le LtRt, © Dolby.

8.1.7 DM100, vérification de la bonne synchronisation des trames Dolby E et vidéo

Le DM100 Dolby Bitstream Analyser est un outil tout particulièrement adapté au test de bout en bout de la chaîne d'exploitation Dolby. Il se synchronise sur la référence vidéo d'une installation, et grâce à son générateur de signaux tests interne, il permet de contrôler rapidement l'intégrité et la composition des signaux Dolby Digital, Dolby E et PCM acheminés par des équipements de production. Son utilisation valide particulièrement la bonne synchronisation des trames Dolby E et vidéo.

8.1.8 Utilisation du DP570, metadata Dolby

Le DP570 est contrôlé par une application Windows DP570 Dolby Remote via le port RS-232 sur le devant, ou via le port RS-485 à l'arrière de l'appareil. Cette application comporte une section monitoring qui permet d'écouter les différents flux : le mixage 5.1 en position Full mix, la simulation 5.1 du Dolby Surround Pro Logic 1 et 2 (PL et PL2), la stéréo LtRt en mode Pro Logic 1 ou 2 ou le LoRo, et la mono issue du LtRt ou du LoRo. On trouve ensuite les deux modes associés aux DRC, le Line mode et le RF mode. Cette section monitoring comporte des fonctions Mute et Solo ainsi que la gestion du niveau d'écoute valable pour la sortie analogique uniquement, avec une position de référence. Le DP570 comporte une page de réglage de monitoring permettant d'ajuster les niveaux et les délais pour chaque canal, avec la possibilité de générer un bruit rose pour vérifier l'écoute. La partie monitoring Dolby Surround permet d'ajuster les délais des canaux arrière, ils sont réglés à 15 ms par défaut.

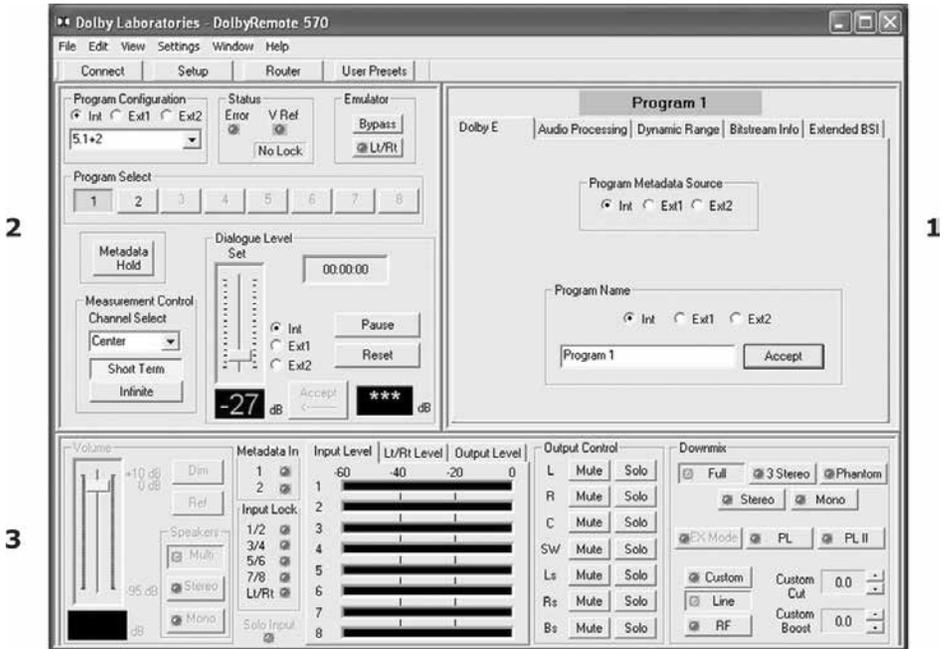


Figure 8.16 – DP570 Dolby Remote, © Dolby.

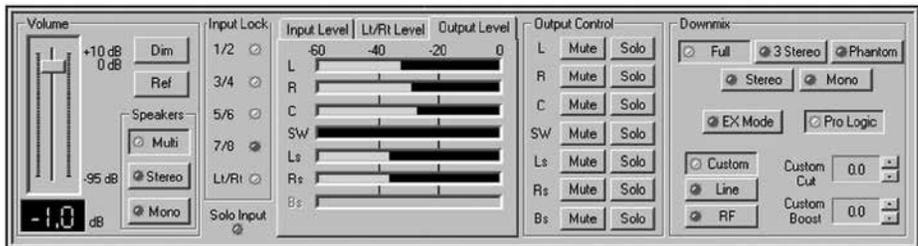


Figure 8.17 – Section Monitoring, Metering, © Dolby.

Le DP570 permet le réglage du Bass Management avec différentes possibilités de renvois vers les Sub ainsi que trois fréquences de coupure 80, 100 ou 120 Hz.

L'utilisateur doit renseigner dans la page principale en haut à gauche le type de programme du signal Dolby E (5.1 + 2 par exemple) puis, juste au-dessous, il accède à la sélection du programme (« program select »), ici 1 et 2.

Lorsqu'on renseigne les metadata, le mode interne est activé, cela signifie que les metadata sont librement renseignées par l'utilisateur, les autres modes Ext1 et Ext2 permettent de lire les metadata d'un flux externe câblé en RS-485 Ext1 ou Ext2, et de les verrouiller sur ces entrées. Il est possible d'activer ces modes pour chacune des metadata et de rectifier une metadata spécifique, que l'on passerait alors en interne.

◆ Loudness

Comme représenté sur la **figure 8.16**, la page principale du DP570 gère la mesure de la loudness : on y trouve les deux modes de mesure « short term » ou « infinite » ainsi que le mode de sélection des canaux qui alimentent la mesure (tous les canaux, ou simplement le centre, le canal LFE étant abandonné). La fonction « accept » permet d'associer la valeur de Loudness mesurée à la metadata Dialnorm, l'utilisateur peut aussi régler manuellement la valeur. Comme dans tout outil de mesure de loudness, on trouve les fonctions « start », « pause » et « reset ».

◆ Audio Processing

Sur le même **schéma 8.16**, la page de droite gère les autres metadata, tout d'abord le nom du programme que l'on retrouvera lors du décodage du signal Dolby E, puis un premier groupe de metadata appelées les metadata Audio Processing. Dans ce premier groupe, on renseigne « le channel mode », le mode du format multicanal (3/2 ou 2/0 généralement), l'activation du canal LFE ou non, le mode de fabrication du downmix LoRo ou LtRt intégrant la fonction « surround phase shift », l'atténuation -3 dB des canaux arrière (valable pour l'encodage d'un mixage film afin d'assurer la compatibilité avec l'écoute ITU TVHD). Attention, au lancement de l'application Dolby Remote, l'atténuation de 3 dB des canaux arrière est validée. Le mode « LtRt preferred stereo Downmix » (mode « surround phase shift ») garantit la compatibilité LtRt du Dolby Digital.

Les autres metadata sont généralement activées : le filtre passe haut « DC High pass filter » bloque la composante continue d'un signal audio à 3 Hz sur les entrées, le filtre passe bas « Lowpass filter » coupe en entrée les hautes fréquences à 20 kHz évitant l'aliasing au décodage, le filtre passe bas « Lfe filter » coupe à 120 Hz le canal LFE, c'est un filtre avec une pente très raide.

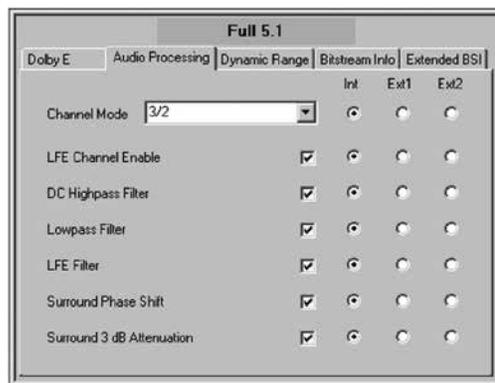


Figure 8.18 – Metadata Audio processing, © Dolby.

◆ DRC

La page DRC du DP570 permet d'affecter le profil de compression pour chacun des modes Line mode et RF mode, elle intègre des indicateurs de niveau qui représentent

l'activité des profils, lors de la compression ou de l'expansion. Le paramètre « RF overmodulation protection » déclenche une protection contre la saturation si le flux Dolby Digital décodé est raccordé à un appareil RF/antenne. Ce paramètre doit toujours être désactivé.

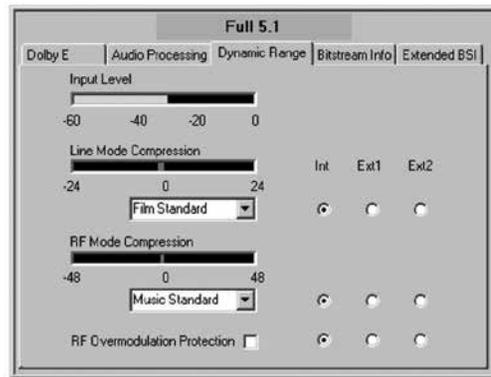


Figure 8.19 – Metadata DRC, © Dolby.

◆ Bitstream info

Le « bitstream mode » décrit le type de service transporté par le flux Dolby Digital. Un programme audio complet peut comprendre un service audio principal (un mixage complet du programme audio) et un service associé ou un service principal avec un service associé. Généralement, la plupart des programmes utilisent le réglage par défaut, « complete main », qui correspond au programme audio complet (de la mono au 5.1). Le programme constitué d'un service audio principal et d'un service associé peut être utilisé pour transporter plusieurs langues, un voice over, des commentaires, une description sonore de l'image, une version karaoké, sur un débit total maximal de 512 kbps.

Sur cette page bitstream info figure aussi le « Dolby Surround mode », il convient de préciser ici si le signal est un LtRt Dolby Surround ou non. Dans le programme 5.1, on valide « not indicated » et sur le programme stéréo LtRt, on précise bien le mode « Dolby Surround encoded ». Ce mode force le décodage Dolby Pro Logic, après décodage Dolby Digital d'un signal Dolby 2.0 à la réception sur un ampli home-cinéma par exemple.

Ensuite figurent des metadata de description, « le copyright bit » indique si le flux Dolby Digital est protégé par un copyright, « l'original bitstream » indique si le flux Dolby Digital est une version master ou copie. « L'Audio Production Information » précise le niveau maximal SPL atteint dans la régie de mixage afin de contrôler le niveau d'écoute sur les équipements haut de gamme ainsi que le type d'acoustique de l'auditorium, le réglage « small » étant une courbe de réponse plate, le réglage « large » correspondant à une correction ISO-X type cinéma.

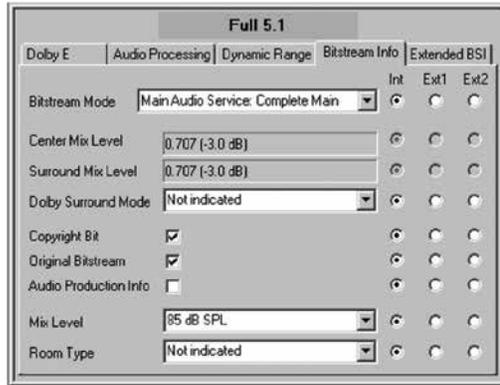


Figure 8.20 – Bitstream info, © Dolby.

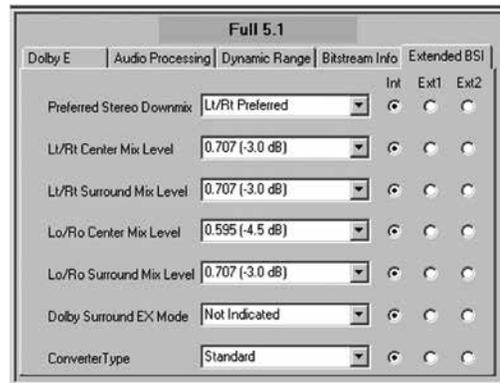


Figure 8.21 – Extended BSI, © Dolby.

◆ Extended BSI

« Les extended BSI » sont des metadata supplémentaires intégrées au flux Dolby Digital, elles agissent essentiellement sur le paramétrage des downmix. « Le preferred stereo downmix » indique si le flux Dolby Digital est downmixé en LtRt ou LoRo, la sortie stéréo du décodeur sera donc pilotée par cette metadata sans que l'auditeur ait à le régler. Les coefficients de downmix du LtRt sont renseignés dans les metadata « Lt/Rt Center Mix level » pour le canal central et « Lt/Rt Surround Mix level » pour les canaux surround. Les coefficients de downmix LoRo sont renseignés dans les metadata « Lo/Ro Center Mix Level » et « Lo/Ro Surround Mix Level ». Comme indiqué précédemment, les coefficients possibles sont : -6 dB, -4,5 dB, -3 dB, -1,5 dB, 0 dB, +1,5 dB, +3 dB pour le centre et les canaux arrière.

Le Dolby Digital est compatible Dolby Surround EX, il existe une metadata appelée « Dolby Surround EX », qui indique si le flux est encodé dans ce format. Le Dolby Surround EX est un format 6.1 qui encode un canal arrière central supplémentaire BS, matricé dans les canaux Ls et Rs, identique au Dolby Surround

Dans les extended BSI figure aussi la metadata « AD converter type » qui indique si le signal transporté a été converti par la technologie HDCD, cette technologie étant pratiquée sur certains master CD.

Ensemble des Metadata Dolby, cf. Dolby.

Paramètres Metadata	Production	Diffusion
Program Configuration	×	
Program Description Text	×	
Dialog Level		×
Channel Mode		×
Lfe Channel		×
Bitstream Mode		×
Line Mode Compression		×
RF Mode Compression		×
RF Overmodulation Protection		×
Center Downmix Level		×
Surround Downmix Level		×
Dolby Surround Mode		×
Audio Production Information		×
Mix Level		×
Room Type		×
Copyright Bit		×
Original Bitstream		×
Preferred Stereo Downmix		×
Lt/Rt Center Downmix Level		×
Lt/Rt Surround Downmix Level		×
Lo/Ro Center Downmix Level		×
Lo/Ro Surround Downmix Level		×
Dolby Surround EX Mode		×
À/D Converter Type		×
DC Filter	×	
Lowpass Filter	×	
LFE Lowpass Filter	×	
Surround 3 dB Attenuation	×	
Surround Phase Shift	×	

8.1.9 Encodage/décodage Dolby Digital, Dolby Digital Plus, Dolby Pulse

L'encodage Dolby Digital pour la diffusion se fait avec le boîtier Dolby DP569, piloté par les metadata du DP570. Il existe une application Dolby DP569 Remote qui permet de contrôler l'encodeur par un PC. Voici le schéma représentant la chaîne d'encodage Dolby Digital depuis un signal Dolby E transmettant les metadata, à l'issue, deux signaux Dolby Digital sont encodés, le programme 1 (5.1) et le programme 2 (stéréo LtRt ou LoRo) :

Dolby a récemment développé le DP568, un décodeur pour la diffusion, compatible Dolby Digital, Dolby Digital + et Dolby Pulse (codec He AAC v2 + metadata, voir chapitre 7). Ce décodeur compatible avec la mesure loudness EBU R128, pourrait à terme remplacer le DP564.

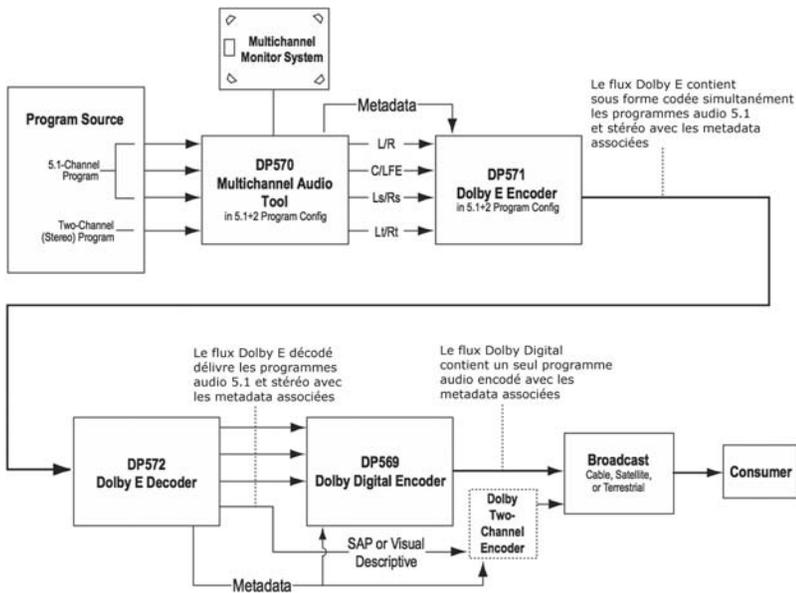


Figure 8.22 – Chaîne d'encodage Dolby Digital, © Dolby

8.1.10 Solutions Logiciels

Complément
Web 8.1

8.2 Mesure du loudness

8.2.1 Principe, mesure Dolby

Le loudness est une nouvelle mesure apparue avec les formats Dolby, elle exprime le niveau d'un signal audio réellement perçu par l'oreille et repose sur un niveau

équivalent Leq, moyenné sur des temps variables. Cette mesure permet tout simplement d'enchaîner des programmes lors de la diffusion, à la même valeur, de telle sorte qu'il n'y ait plus de sauts de niveaux. En effet, dans la chaîne de fabrication audiovisuelle, les produits ont tous des niveaux de loudness différents. Par exemple, le film a son niveau de loudness assez bas, avec une grande dynamique, c'est le cas aussi de la musique classique. En revanche, la publicité a des niveaux plus forts, quant au CD de musique actuelle, il a considérablement repoussé les niveaux vers le haut et réduit la dynamique. Le particulier est d'ailleurs contraint de baisser le niveau sonore pendant la page de publicité et de remonter le niveau à la reprise du programme. A partir de 2013 en France, la norme loudness permet un enchaînement fluide en niveau des différents programmes et des différentes chaînes de télévision.

Le loudness s'exprime par un Leq, moyenné au départ sur une fenêtre glissante de 10 secondes et dont l'échelle s'exprimait en dBA (courbe de pondération A). La fenêtre glissante de 10 secondes définit le mode de mesure du loudness « short term » **chez Dolby**. C'est une valeur qui évolue sur toute la durée du programme. Si l'on souhaite mesurer un programme entier, il faut procéder à la mesure « long term », c'est la moyenne du loudness sur toute la durée.

La méthode de mesure **Dolby** diffère suivant la nature du programme, si c'est un film, un documentaire, on fait la mesure sur les parties avec dialogues, la mesure s'appelle alors le « dialog level », ou niveau de dialogue. Le dialogue est l'élément primordial de ce type de programmes, il doit être intelligible sur toute la longueur du film et garder un niveau constant. Pour tout autre programme sans dialogues notamment les programmes musicaux, on procède à la mesure de toute la durée du programme, cela s'appelle « la mesure loudness ». Afin de simplifier la méthode de mesure du Dialog level, Dolby a développé un algorithme de mesure « Dialog Intelligence » que l'on trouve sur la plupart des outils Dolby (LM100, Dolby Media Meter...). Il permet d'isoler les parties avec dialogues, et ne prend pas en compte les passages intermédiaires. Autrement, sans « Dialog Intelligence », il convient d'isoler quatre passages représentatifs du programme, et d'en moyenniser les quatre valeurs obtenues, en combinant la mesure des passages en continu.

8.2.2 *Recommandation ITU BS 1770-1*

La mesure de loudness a ensuite évolué, utilisant une courbe de pondération différente, la courbe RLB, donnant des valeurs plus proches de la réalité. La courbe RLB est une extension de la courbe B dans les basses fréquences (Revised Low Frequency B curve). Puis, depuis début 2009, la mesure intègre la courbe R2LB ou courbe K, qui diffère quelque peu de la courbe RLB avec une préaccentuation dans l'aigu. La courbe K est aujourd'hui la courbe retenue pour la mesure du loudness, elle donne alors l'échelle logarithmique de mesure en LKFS.

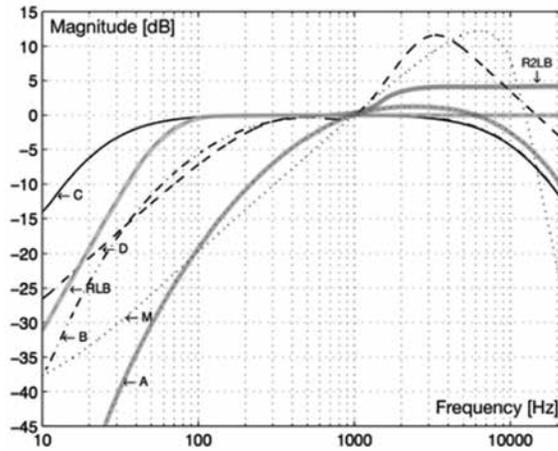


Figure 8.23 – Courbes de pondération A, B, RLB, R2LB (K), © TC electronic.

La méthode de mesure fait l'objet d'une norme ITU-R BS1770-1 dont la **figure 8.24** décrit l'algorithme.

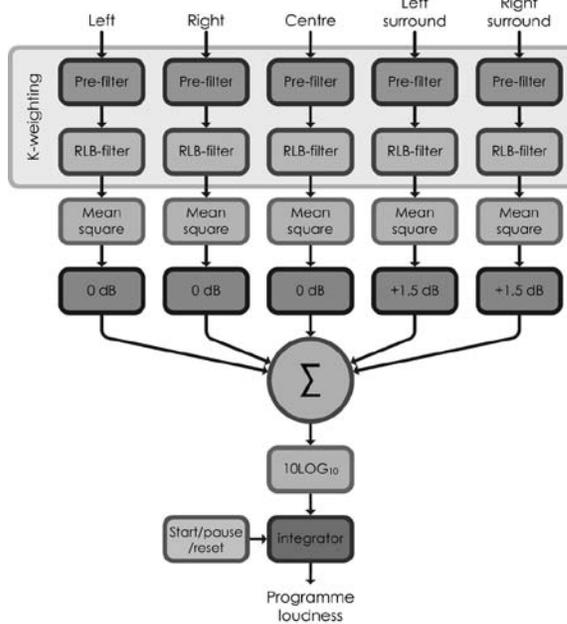


Figure 8.24 – Norme ITU-R BS1770-1, © ITU.

Cette norme intègre la courbe de pondération K sur l'analyse des niveaux des cinq canaux du multicanal, le canal LFE n'étant pas pris en compte. Les canaux arrière ont un coefficient plus élevé, + 1,5 dB, car ils ont un pouvoir attentionnel plus important. En effet, comme précédemment évoqué, tout son provenant de la zone arrière est synonyme de danger et attire l'attention.

Dialog Range

Le Dialog Range permet simplement de suivre l'excursion du Dialog level, il donne les valeurs minimum et maximum du Dialog level sur la durée du programme mesuré. On le trouve sur les outils Dolby, il est corrélé au Dialog Intelligence. Ces valeurs permettent de vérifier si le niveau des dialogues reste constant ou bien s'il évolue beaucoup.

MESURE LOUDNESS DOLBY

- Mesure short term : Leq sur 10 secondes.
- Mesure long term ou infinite all : moyenne du Leq sur toute la durée du programme.
- Short term Dialog : Leq sur 10 secondes avec Dialog Intelligence.
- Infinite Dialog : moyenne du short term Dialog sur tous les passages de dialogues.

8.2.3 Loudness EBU R128 Europe, recommandation ITU BS 1770-3

En 2008, l'EBU crée un groupe de travail nommé le PLOUD dont la finalité est de mettre en place une norme européenne de loudness. Aujourd'hui, il regroupe l'ensemble des chaînes de télévisions et radios européennes. Cette norme intègre des notions nouvelles dans la mesure du loudness notamment l'utilisation d'un outil universel permettant de mesurer tous les programmes, on ne fait plus alors de différence entre un film et un programme sans dialogues. Pour cela, cet outil intègre un gate avec un seuil relatif. C'est donc tout d'abord une simplification dans la mesure des programmes. Par ailleurs, l'ancienne norme Dolby ne prenait pas en compte les passages intermédiaires, entre séquences avec dialogues, si bien qu'elle ne mesurait pas réellement l'ensemble du programme. De plus, il s'est avéré que les programmes de grande dynamique, ramenés à une valeur de loudness cible, sonnaient plus forts que ceux à dynamique réduite, d'où la nécessité d'intégrer le gate. Le gate permet donc d'écarter les passages dont le niveau de loudness décroît subitement, ce sont des passages dont le niveau passe au second plan, derrière des passages plus représentatifs du programme. Par exemple, un match de tennis comporte ce type de changements dans les niveaux de loudness, l'oreille s'appuie sur les passages forts, marqués par les bruits de foule, les commentaires, pour estimer le niveau perçu, les autres passages silencieux passent au second plan. L'algorithme de mesure avec gate, permet justement d'isoler les passages forts, utiles à l'oreille et d'écarter les passages faibles.

La norme R128 définit trois mesures qui permettent de mieux suivre l'évolution du loudness :

- **Momentary** : mesure instantanée du loudness, moyennée sur 400 ms ;
- **Short Term** : mesure de courte durée, moyennée sur 3 secondes ;

- **Program loudness ou Integrated** : mesure de toute la durée du programme, avec utilisation du gate.

Le gate est défini par un seuil variable, égal à **-10 dB** en dessous du loudness momentané (momentary 400 ms).

Le canal LFE n'est toujours pas intégré dans la mesure loudness bien que des interrogations au sein du PLOUD sur cette option soient d'actualité aujourd'hui.

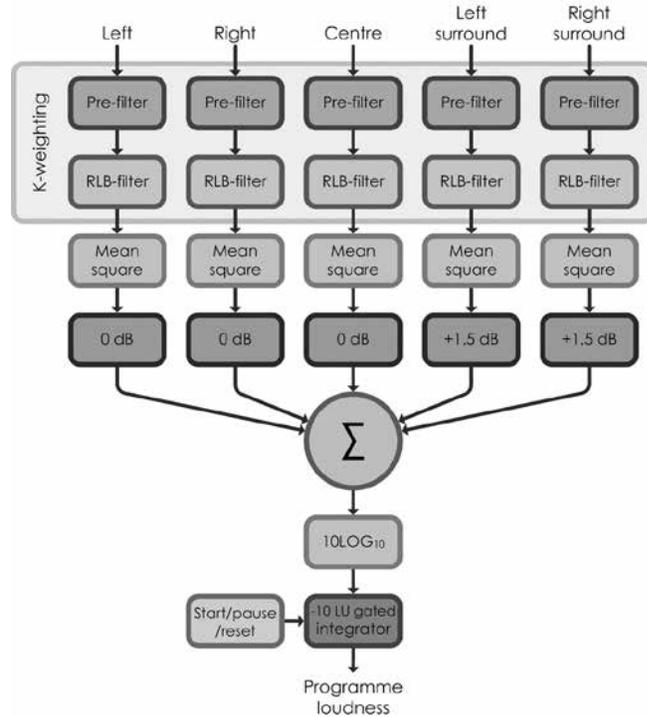


Figure 8.25 – Recommandation ITU-R BS 1770-2 intégrant le gate dans la mesure, © EBU PLOUD.

Une nouvelle échelle exprimée en LUFS (Loudness Unit referenced to Full Scale) remplace l'ancienne, exprimée en LKFS. L'échelle LUFS est donnée pour les valeurs absolues, le LU est utilisé pour les valeurs relatives.

Nous pouvons exprimer les valeurs en LUFS ou bien en LU, avec la correspondance suivante, conformément à la recommandation ITU-R BS 1771 :

$$0 \text{ LU} = -23.0 \text{ LUFS}$$

Deux échelles sont possibles, suivant les valeurs mesurées : l'échelle +18 LUFS utilisée pour les programmes à grande dynamique, et l'échelle +9 LUFS pour les programmes à dynamique plus réduite.

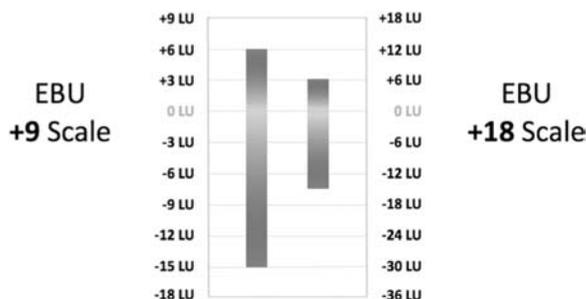


Figure 8.26 – Échelles LUFS, +18 LUFS et +9 LUFS, © EBU Tech 33430.

La valeur du program loudness pour chacun des programmes est normée à :

Program loudness = – 23.0 LUFS

Pour les programmes en postproduction, la valeur du loudness doit être de – 23.0 LUFS, pour les programmes Live une tolérance de ± 1 LU autour de la valeur cible est admise.

LOUDNESS EBU R128

- Mesure sur toute la durée du programme.
- N'intègre qu'un seul mode de mesure, compatible avec tout programme comportant des dialogues ou non.
- Compatible avec les programmes de grande dynamique.
- 3 valeurs de loudness.
 - Momentary : mesure instantanée 400 ms.
 - Short term : mesure courte durée, moyennée sur 3 secondes.
 - Integrated ou program loudness : mesure du programme avec gate – 10 LU, conformément à la norme ITU-R BS1770-3.
- Program Loudness = – 23.0 LUFS ± 1 LU pour les programmes Live.
- 0 LU = – 23 LUFS.

8.3 Normes françaises PAD TVHD

Le concept de normalisation des niveaux crêtes en référence à un niveau maximum permis (à –9 dBFS) a mené à des niveaux de loudness très variés. La variation du loudness dépend du niveau de compression des programmes. La normalisation du loudness des programmes, intégrée dans les nouvelles recommandations, permet d'obtenir un niveau moyen constant du loudness, avec des crêtes variant en fonction du contenu des programmes. Le spectateur perçoit alors un niveau moyen de loudness égal sur tous les

programmes, il n'a plus besoin de changer le niveau de monitoring. L'écoute se fait sans gêne à la condition que la variation du loudness du programme soit adaptée au lieu.

8.3.1 Mesure des niveaux

◆ Rappel sur la mesure crête DIN 45406 QPPM

Le contrôle des niveaux des programmes stéréo est assuré par un crête mètre quasi-instantané de 10 ms de temps de montée suivant la norme DIN 45406. On appelle cette mesure de crête un QPPM ce qui signifie quasi peak program meter (quasi crête mètre). En diffusion broadcast analogique stéréo, le niveau des crêtes du signal ne doit pas dépasser 0 QPPM sur le crête mètre DIN 45406 équivalent à -9 dBFS. Pour rappel, le niveau d'alignement 0 VU (+4 dBu) en analogique doit correspondre à -9 QPPM sur le crête mètre, ce qui doit correspondre sur l'échelle numérique à -18 dBFS en PPM. Le PPM signifie que la mesure du niveau crête est instantanée, elle se fait pendant la durée d'un échantillon.

◆ Mesure crête instantanée PPM

Le crête mètre numérique contrôle généralement un niveau de crête instantané 0 ms que l'on appelle le niveau PPM. Sa mesure est faite sur chacun des échantillons, et généralement, le 0 dBFS se déclenche lorsqu'au minimum trois échantillons successifs ont une valeur de 0 dBFS. Mais il existe d'autres crête-mètres qui déclenchent le 0 dBFS au premier échantillon dont la valeur est 0 dBFS. Dans la recommandation CST RT17-TV v2 éditée en 2009 (voir le paragraphe Normes PAD TVHD), le niveau du signal multi-canal ne doit pas dépasser -3 dBFS PPM instantané.

Voici un schéma représentant l'alignement des niveaux exprimés en VU, crête-mètre QPPM et crête-mètre numérique avec leur correspondance :

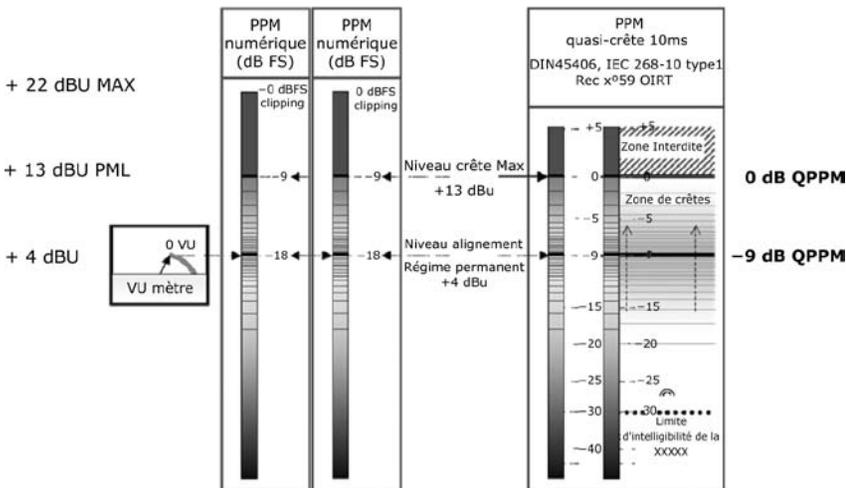


Figure 8.27 – Alignement des niveaux VU, QPPM, PPM numérique et dBu, © CST.

◆ Mesure True Peak « dBTP »

De nombreuses études dans le contrôle du signal à la conversion numérique-analogique ont révélé le manque de précision du crête mètre instantané. Le 0 dBFS mesuré en PPM peut engendrer la saturation du signal analogique en sortie de convertisseur. Les études ont évalué une valeur maximale proche de 3 dB d'incertitude dans la mesure, ce qui a remis en cause la mesure crête instantanée. Elle est remplacée par une nouvelle mesure crête appelée True Peak (crêtes vraies). Cette mesure est réalisée entre chacun des échantillons, ce qui revient à suréchantillonner à quatre fois le signal original afin « d'insérer » des mesures supplémentaires et mieux suivre l'évolution du signal. Le schéma ci-dessous représente l'algorithme de la mesure True Peak :

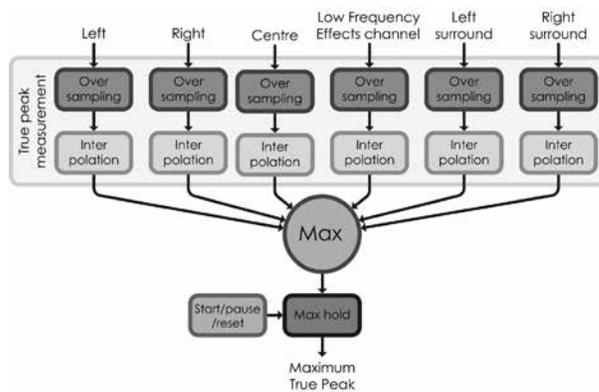


Figure 8.28 – Algorithme de mesure True Peak, © EBU PLOUD.

Les valeurs sont données en dBTP et l'échelle du crête mètre True Peak s'étend alors de -60 dBTP à $+3$ dBTP.

8.3.2 Anciennes normes françaises PAD CST RT16, CST RT17 v2, CST RT19

Complément Web 8.2

Anciennes normes françaises PAD CST RT16, RT17, RT19.

8.3.3 Norme européenne actuelle EBU R128

La norme EBU R128 est une norme européenne en cours de validation dans les pays européens. Elle est l'aboutissement du groupe de travail PLOUD au sein de l'EBU, actif sur ce sujet depuis 2010. Des études portant sur l'enchaînement de programmes de grande dynamique et de dynamique réduite ont révélé que les écarts de niveaux étaient moins importants lorsque la normalisation de loudness se fait avec une mesure universelle de type EBU R128, qu'avec une mesure dialogue Leq (K) ou une mesure

dialogue Leq (A). Le manque de précision dans la mesure du dialogue s'explique en partie par des esthétiques de mixages très différentes observées sur les programmes. Un seul paramètre, le loudness, n'est pas suffisant pour caractériser différents genres de programmes et différents styles de mixage. Trois caractéristiques doivent être mesurées :

- le niveau des crêtes du signal ne doit pas dépasser un certain seuil afin d'éviter la distorsion du signal tout le long de la chaîne (processeurs, codecs, convertisseurs), dans le domaine numérique on utilise la mesure True Peak ;
- le niveau de Loudness du programme mesuré doit être conforme à la valeur cible retenue par le diffuseur. Cette valeur doit être suffisamment basse afin de laisser vivre la dynamique du signal sans atteindre la saturation. La mesure est faite par le « Program Loudness » ou « Integrated Loudness », avec l'utilisation d'un gate relatif ;
- l'enchaînement des programmes doit se faire sans saut de niveaux, et la variation de loudness dans un programme doit être adaptée aux capacités en dynamique du système d'écoute de l'auditeur. Le Loudness Range permet de quantifier cette excursion.

◆ Loudness, LRA

La mesure du loudness abordée dans le paragraphe Loudness met en évidence les points suivants :

- la mesure est faite sur toute la durée du programme, elle n'intègre qu'un seul mode de mesure, compatible avec tout programme comportant des dialogues ou non ;
- elle est compatible avec les programmes de grande dynamique ;
- elle repose sur 3 valeurs de loudness :

Momentary (M) : la mesure instantanée 400 ms, la mesure se fait sans gate

Short term (S) : la mesure courte durée, moyennée sur 3 secondes, la mesure se fait sans gate

Integrated ou Program Loudness (I) : la mesure du programme avec gate à -10 LU, conformément à la norme ITU-R BS 1770-2, utilisation du start et du stop pour démarrer et stopper la mesure. La valeur cible du Program Loudness est -23.0 LUFS et 23.0 LUFS ± 1 LU pour les programmes Live. L'échelle LU s'exprime avec la correspondance suivante, conformément à la norme ITU-R BS 1771 : 0 LU = -23 LUFS

Cette norme utilise donc une méthode de mesure de loudness intégrant deux gate en cascade, la gate absolue de -70 LUFS et la gate relative à -10 LU relatif au niveau de loudness utilisant le gate absolu.

PRINCIPE DU SEUIL DE MESURE RELATIF

- On effectue à intervalles réguliers une mesure glissante de l'énergie moyenne de chaque canal du programme sur une fenêtre de 400 ms. La recommandation précise que le chevauchement entre deux fenêtres de mesure successives doit être égal à 75 %. Un chevauchement de 75 % implique de faire une mesure toutes les 100 ms.
- On peut donc calculer le loudness de chaque fenêtre (ou bloc). Les blocs dont le loudness est inférieur à -70 LUFS (absolute gating threshold) sont purement et simplement ignorés. Pour les autres, on mémorise la mesure d'énergie moyenne par canal dans une table.
- On peut ainsi calculer le loudness que produirait un signal composé uniquement des blocs mémorisés dans la table. Le seuil relatif de mesure est placé 10 dB en dessous de cette mesure.
- On peut maintenant déterminer le relative gated loudness : c'est celui que produirait un signal composé uniquement des blocs mémorisés dans la table et dont le loudness est supérieur au seuil relatif.
- L'objectif du seuil relatif est de produire un résultat plus fiable en ignorant les passages faibles relativement au loudness moyen du programme.
- Dans la figure suivante, les zones bleues sont celles qui ne seront pas prise en compte, après application du seuil de mesure.

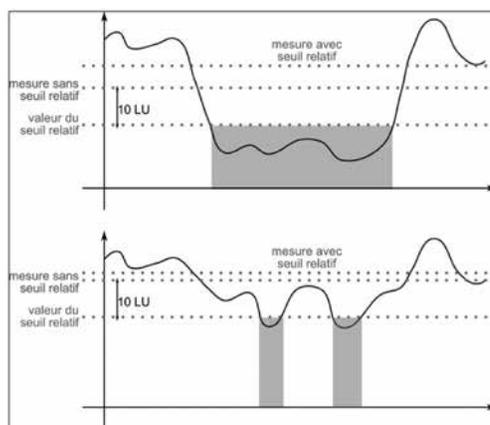


Figure 8.29 – Principe du seuil de mesure relatif, © France Télévisions.
Guide de loudness France Télévisions, Manuel Naudin et Claire Mérienne.

Cette méthode de mesure intégrée dans la norme ITU-R BS 1770-2 permet d'écarter les moments de silence, les passages de bas niveau qui sont perçus en second plan par rapport à un premier plan constitué des dialogues par exemple. La distribution des niveaux de loudness mesurés avec l'utilisation des deux gate est alors quantifiée suivant une échelle statistique exprimée en pourcentage. Le **Loudness Range (LRA)** est alors défini par l'écart de Loudness exprimé en LU sur la distribution entre 10 % et 95 %,

les valeurs de loudness extrêmes étant écartées. La valeur 10 % permet par exemple d'écarter les fades out de fin d'une musique et la valeur 95 % permet de ne pas tenir compte des coups de pistolets dans un film. Le LRA traduit l'excursion du loudness sur toute la durée du programme, les valeurs de loudness les plus basses représentent alors les niveaux les plus faibles du contenu pertinent du programme, les passages de silence et les ambiances de faible niveau étant écartés.

Le Loudness Range permet de mesurer la dynamique d'un programme et de s'assurer qu'elle est conforme à la dynamique maximale acceptée suivant le lieu de diffusion. La **figure 8.30** représente les différentes valeurs de LRA retenues pour ces différents lieux, le cinéma, le home-cinéma, la voiture etc. La norme R128 n'impose pas une valeur de LRA maximale, elle encourage l'utilisation du LRA pour suivre l'évolution du loudness du programme, les valeurs types étant laissées à la discrétion des diffuseurs, en tenant compte du genre de programme, du type d'audience et du mode de distribution.

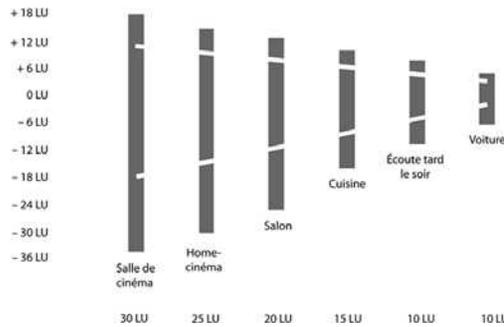


Figure 8.30 – Exemples de valeurs de LRA maximales acceptées suivant les lieux de diffusion, © EBU. Cf. EBU Tech 3343.

◆ Niveaux crête

Les niveaux crêtes sont mesurés suivant un crête-mètre **True Peak**, et un maximum de -1 dBTP ne doit pas être dépassé, ceci est valable pour les niveaux de mixage en production. La méthode de mesure True Peak est définie par la norme ITU-R BS 1770-3. En France les niveaux crête maximums pour des signaux encodés sont plus bas, environ -3 dBTP (cf. EBU guide de distribution 3344).

◆ Programmes courts

Sont considérés comme programmes courts, les publicités ou bandes annonces d'une durée inférieure à 30 secondes. La valeur du Program Loudness doit être normalisée à -23.0 LUFS ± 0.5 LU. Pour ce type de programmes, on introduit les valeurs maximales de loudness Momentary et loudness Short Term, qui permettent d'éviter d'avoir des publicités trop fortes. Il convient alors de ne pas dépasser les valeurs suivantes :

- Maximum Short Term Loudness.
MAX SL = + 5 LU = - 18 LUFS.
- Ou de façon alternative Maximum Momentary Loudness
MAX ML = + 8 LU = - 15 LUFS.

L'une ou l'autre des deux valeurs limites maximales doit être respectée, pas les deux en même temps. Les valeurs loudness Max Short Term et loudness Max Momentary sont respectées généralement par le travail des niveaux des différentes sources sur les passages concernés. Dans les cas les plus difficiles, le traitement dynamique s'impose.

SYNTHÈSE POUR LES PROGRAMMES COURTS NORMALISÉS R128

- Program Loudness -23.0 LUFS ± 0.5 LU
 - Max True Peak -1 dBTP
 - Max Short Term Loudness -18.0 LUFS
 - Le LRA n'est pas applicable
- Ou de façon alternative :
- Program Loudness -23.0 LUFS ± 0.5 LU
 - Max True Peak -1 dBTP
 - Max Momentary Loudness -15.0 LUFS
 - Le LRA n'est pas applicable

◆ **Méthode de travail et traitements dynamiques**

L'approche recommandée par le PLOUD de l'EBU, consiste à intégrer directement les nouveaux outils de mesure pendant la phase de mixage. Elle permet d'obtenir une dynamique plus importante des programmes et, après un certain temps d'expérimentation, de mixer à l'oreille, à un niveau constant de - 23 LUFS.

La plupart des programmes s'accommodent d'une légère atténuation sur le master de la console afin d'obtenir le niveau cible de - 23.0 LUFS. Pour certains programmes à grande dynamique, il convient de limiter les crêtes avec des outils compatibles True Peak et de compresser le signal afin d'obtenir un LRA adéquat.

Suivant la norme R128, le traitement dynamique pour la compression du LRA recommande un seuil bas, inférieur à - 40 dBFS, un rapport de compression bas, entre 1 :1,2 et 1 :1,5, et l'utilisation du gain de rattrapage afin d'obtenir la cible - 23 LUFS.

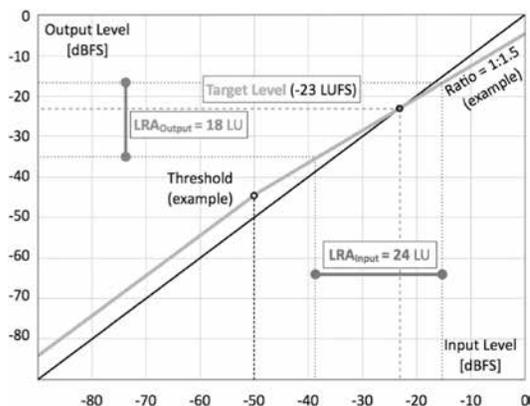


Figure 8.31 – Traitement du LRA, © EBU Tech 3343.

EBU R128

- 3 paramètres caractérisent le signal.
 - le Program Loudness.
 - le Loudness Range (LRA).
 - les niveaux crêtes max True Peak.
- Le Program Loudness doit être normalisé à $-23.0 \text{ LUFS} \pm 0,5 \text{ LU}$.
- Une tolérance de $\pm 1 \text{ LU}$ pour les programmes pour lesquels il est difficile d'obtenir une normalisation exacte comme le Live.
- L'outil de mesure doit être conforme à la norme ITU-R BS 1770-3 et EBU Tech Doc 3341.
- La mesure du LRA permet de savoir si une compression du signal est nécessaire suivant le genre du programme, le type d'audience et le mode de distribution.
- Le niveau crête True Peak maximal est de -1 dBTP en production.
- La metadata loudness doit être renseignée à -23.0 LUFS , si pour une raison particulière il n'est pas possible de normaliser à -23.0 LUFS , la metadata loudness doit toujours indiquer la valeur mesurée.

8.3.4 Normes françaises PAD R128 2011 CST RT-17-TV version V3

Voici les principaux éléments à retenir sur les normes PAD actuelles. Attention, la norme française reprend la norme EBU R128 mais elle ajoute des spécificités comme par exemple la prise en compte des dialogues.

◆ Installations 5.1

Le niveau d'écoute de référence pour une installation TVHD Son est autour de 79 dBc SPL par canal, pour un bruit rose généré à -18 dBFS RMS , le canal LFE est réglé à 83 dBc SPL (sonomètre pleine bande).

NIVEAU RÉFÉRENCE D'ÉCOUTE ET NORME EBU R128

- En conséquence de la normalisation des niveaux de loudness des mixages à -23 LUFS, le niveau référence d'écoute est autour de 79 dBC SPL par canal. On garde une certaine souplesse dans le réglage du niveau d'écoute, autour de 79 dBC SPL, suivant la taille de la régie et la distance des enceintes. L'EBU travaille actuellement sur la mise à jour du réglage du niveau de référence (Tech 3343) :
- Lref = 73 dBA SPL par canal pour un bruit rose filtré entre 500 Hz et 2 kHz dont le program loudness est de -23.0 LUFS (ce fichier est fourni dans la session Pro Tools test au chapitre 2).

◆ Niveau d'alignement

Le niveau d'alignement lu sur un crête-mètre numérique Full Scale est de -18 dBFS pour un signal sinusoïdal à la fréquence de $1\,000$ Hz.

Un signal d'alignement $1\,000$ Hz à -18 dBFS crête doit afficher un niveau de loudness de -18 LUFS sur un équipement de mesure EBU, si le signal est présent sur les canaux frontaux gauche et droite d'un programme stéréo ou d'un programme 5.1.

Si le signal n'est présent que sur un seul canal frontal, le niveau affiché sera de -21 LUFS.

◆ Niveaux crête

Le niveau des crêtes du signal audio, mesuré en True Peak, ne doit pas dépasser -3 dBTP (compatible production PCM et diffusion Dolby).

◆ Niveau de Loudness

La valeur cible du loudness intégré est :

- 23.0 LUFS ± 1 LU pour les programmes de durée supérieure à 2 minutes ;
- 23.0 LUFS précisément pour les programmes de durée inférieure à 2 minutes.

◆ Dynamique

Pour les programmes de durée supérieure à 2 minutes, le LRA doit être inférieur ou égal à 20 LU, et au minimum égal à 5 LU. Les dialogues doivent moduler autour de la valeur cible de loudness sur une dynamique de ± 7 LU, mesuré en short term (3 secondes), soit entre -30 et -16 LUFS.

Pour les programmes de durée inférieure à 2 minutes, le niveau maximal Short Term ne doit pas dépasser $+3$ LU, soit -20 LUFS, pour l'ensemble du programme. On ne mesure pas le LRA de ce type de programmes.

Synthèse CST RT17 V3.

	programmes durée ≤ 2 minutes	programmes durée > 2 minutes
Max True Peak	- 3 dBTP	- 3 dB TP
Program Loudness	≤ - 23.0 LUFS	- 23.0 LUFS ± 1 LU
LRA	non spécifié	entre 5 LU et 20 LU
Max Short Term Loudness	≤ - 20 LUFS	non spécifié
Dynamique loudness des paroles	non spécifié	± 7 LU autour de - 23 LUFS entre - 30 et - 16 LUFS

La synthèse des metadata Dolby est donnée par le schéma ci-dessous, reprenant les données de la CST RT17 V2 avec les mesures de loudness R128 :

Paramètres	Programme 5.1	Programme 5.0	Programme Stéréo (2.0)	Programme mono/dual	Signal de référence
"Dialog level"	- 23 LUFS ± 1 (A mesurer)	- 31 LUFS			
Program Name	Choix Editeur**	Choix Editeur**	Choix Editeur**	Choix Editeur**	Test_Check
Channel Mode	3/2	3/2	2/0	2/0	3/2 ou 2/0
LFE Channel	Enable	Disable	N/A	N/A	5.1: Enable 5.0: Disable 2.0: N/A
DC Filter	Enable	Enable	Enable	Enable	Enable
Low pass Filter	Enable	Enable	Enable	Enable	Enable
LFE Low pass Filter	Enable	Disable	N/A	N/A	5.1: Enable 5.0: Disable 2.0: N/A
Srnd Phase Shift*	Enable	Enable	N/A	N/A	3/2: Enable 2.0: N/A
Srnd 3 dB Attenuation	Disable	Disable	N/A	N/A	3/2: Disable 2.0: N/A
Line Mode Pro	Film LIGHT	Film LIGHT	Film LIGHT	Film LIGHT	None
RF Mode Pro	Film Standard	Film Standard	Film Standard	Film Standard	None
RF Overmod Protect	Disable	Disable	Disable	Disable	Disable
Bitstream Mode	Main Complete				
Center Downmix Level	0.707 (-3 dB)				
Srnd Downmix Level	0.707 (-3 dB)				
Dolby Srnd Mode	N/A	N/A	Dolby Surround enabled	Dolby Surround enabled	N/A
Copyright bit	Yes	Yes	Yes	Yes	Yes
Original Bitstream	Yes	Yes	Yes	Yes	Yes
Audio Production Info	Disable	Disable	Disable	Disable	Disable
Mix Level	N/A	N/A	N/A	N/A	N/A
Room type	N/A	N/A	N/A	N/A	N/A
Preferred Stereo Downmix*	L/Rt	L/Rt	N/A	N/A	3/2: L/Rt 2.0: N/A/N/A
L/Rt Center Downmix Level	0.707 (-3 dB)	0.707 (-3 dB)	N/A	N/A	N/A
L/Rt Surround Downmix Level	0.707 (-3 dB)	0.707 (-3 dB)	N/A	N/A	N/A
Lo/Ro Center Downmix Level	0.707 (-3 dB)	0.707 (-3 dB)	N/A	N/A	N/A
Lo/Ro Surround Downmix Level	0.707 (-3 dB)	0.707 (-3 dB)	N/A	N/A	N/A
Dolby Surround EX Mode	Not Surround EX	Not Surround EX	N/A	N/A	N/A
A/D Converter Type	Standard	Standard	Standard	Standard	Standard

Figure 8.32 – Tableau de synthèse des metadata Dolby E, © CST RT17 V3.

◆ Autres données

Synchronisation vidéo et audio

Les signaux vidéo et audio sont synchrones entre eux dans la tolérance suivante :

Soit une avance maximale du son sur l'image de 20 ms ;

Soit un retard maximum du son sur l'image de 40 ms.

Format audionumérique

Le format audionumérique est 48 kHz 16 bits pour le PCM, 48 kHz 20 bits obligatoire pour le Dolby E. Le signal Dolby E doit être impérativement exempté d'erreur CRC (Cyclic Redundance Code).

Fiches de Mixage et d'encodage

Chaque livraison d'un mixage doit être accompagnée d'une « **Fiche d'information de mixage** » standardisée.

Chaque support PAD contenant une à deux pistes Dolby E doit être accompagné d'une « **Fiche d'encodage Dolby E** » standardisée.

Dolby Surround

Comme le signale la recommandation, « les réductions Surround (LCRS) stéréo Lt/Rt doivent conserver la cohérence spatiale de l'image sonore et ne doivent pas altérer l'intelligibilité ni le timbre du message sonore. Les produits dont l'origine est un signal Lt/Rt Surround ne doivent en aucun cas être décodés en LCRS pour alimenter les canaux du Dolby E en configuration 5.1. De plus, ils ne doivent pas subir une génération de décodage et de re-matçage Lt/Rt. ».

Synchronisation des trames Dolby E et de la vidéo

Le positionnement relatif du début de la trame Dolby E et de la vidéo doit être conforme aux recommandations du tableau ci-dessous.

Synchronisation des trames Dolby E et de la vidéo.

Format Vidéo	Début de la trame Dolby E	tolérance
SD	Ligne 12	± 1 ligne
HD	Ligne 20	± 3 lignes

Affectation des canaux

Le Dolby E prévoit 4 ou 8 canaux suivant la configuration mono/stéréo ou 5.1.

En stéréo, la configuration du programme est 2 + 2, les programmes 1 et 2 étant en « channel mode » 2/0, les canaux 1 et 2 reçoivent la double mono ou la stéréo, et les

spécifications du contenu des pistes 7 et 8 du Dolby E sont laissées au choix des diffuseurs.

En multicanal, la configuration du programme est 5.1 + 2, le programme 1 étant en « channel mode » 3/2 L, le programme 2 étant en 2/0, la répartition des canaux se faisant comme indiqué dans la figure suivante, généralement la stéréo LtRt ou LoRo est affectée aux canaux 7 et 8, on peut trouver pour certains diffuseurs l'audio description :

Affectation des canaux pour les signaux Dolby E 5.1 + 2.

Paire AES d'origine	Piste dans le Dolby E	Contenu
AES 1	Piste 1	Piste Gauche
	Piste 2	Piste Droite
AES 2	Piste 3	Piste Centre
	Piste 4	Piste LFE
AES 3	Piste 5	Piste Arrière Gauche
	Piste 6	Piste Arrière Droite
AES 4	Piste 7	Voie optionnelle A contenu à préciser par le diffuseur
	Piste 8	Voie optionnelle B contenu à préciser par le diffuseur

Compatibilité des downmix stéréo

La recommandation précise que les downmix stéréo doivent conserver :

- la cohérence spatiale de l'image sonore ;
- le niveau et l'intelligibilité de la voix ;
- l'équilibre entre les éléments du mixage ;
- l'équilibre fréquentiel du mixage.

Metadata Dolby E

La recommandation distingue deux familles de programmes, les programmes musicaux qui sont susceptibles de poser problème pour leur downmix stéréo, on les encode généralement en LoRo, et les autres programmes. Elle précise les points suivants pour le renseignement du Dialnorm.

La mesure s'effectuera selon la recommandation EBU R128

La valeur exacte mesurée devra être renseignée dans les métadonnées

Dans le cas des programmes courts dont la valeur de Loudness est inférieure à - 23 LUFs, la valeur réelle mesurée sera affichée.

Organisation du contenu du programme HD

L'organisation du contenu du programme HD se fait comme présenté dans la **figure 8.33**.

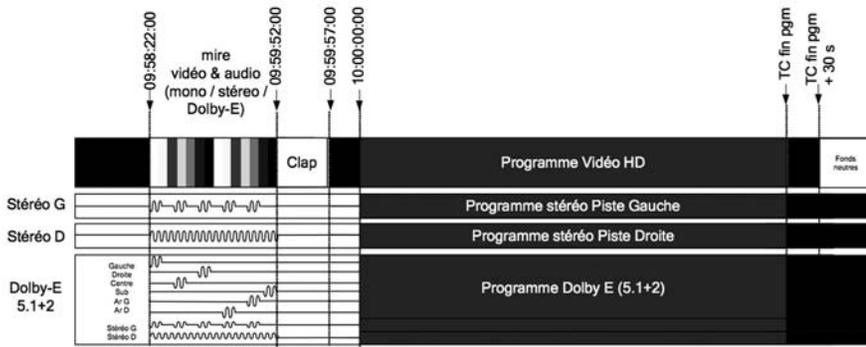


Figure 8.33 – Organisation du contenu d'un programme HD, © CST.

De 09 :58 :22 :00 (inclus) à 09 :59 :51 :24 (inclus)

Signaux d'identification

Ces signaux d'identification en début de bande sont les suivants :

Pour un programme mono : fréquence 1 000 Hz continu au niveau de référence - 18 dBFS, les tonalités des 2 pistes doivent être cohérentes (même source) et en phase.

Pour un programme multicanal et stéréo : les signaux audio d'identification et d'alignement multicanal et stéréo doivent correspondre à la recommandation **EBU Tech 3304** « Multichannel Audio line-up Tone », en voici la représentation pour un signal multicanal :

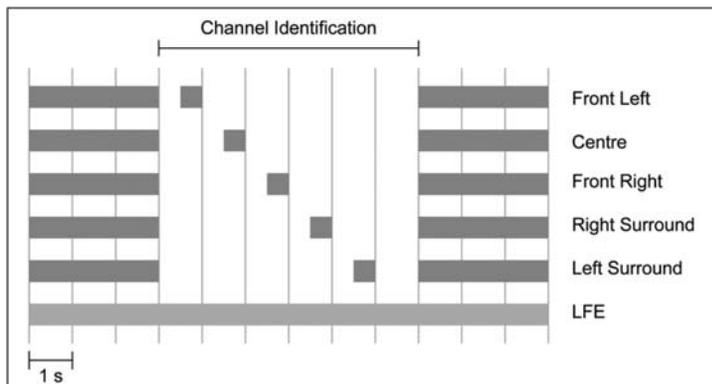


Figure 8.34 – Signal d'identification EBU, © EBU Tech 3304.

L'identification des canaux se fait avec l'ordre **L, C, R, Rs, Ls**, pour un signal de fréquence pure 1 kHz. Le canal LFE est identifié par un signal de fréquence pure de 80 Hz.

Les metadata du Dolby E, sur la durée des signaux de tests doivent être paramétrées comme suit :

Dialog level = -31 LKFS

Line mode = none

RF mode = none

Surround 3 dB attenuation = disable

De 09:59:52:00 (inclus) à 09:59:59:24 (inclus)

Format PCM : Silence d'une durée de huit secondes (8'')

Format Dolby E : Silence d'une durée de huit secondes (8'') codé Dolby E avec les metadata du programme utile

À partir de 10 :00 :00 :00 (inclus)

Début du programme utile.

Du TC de fin de programme

Du TC de fin de programme utile, et pendant une durée de trente secondes (30'') : noir codé et muet.

Lors de la vérification d'un PAD, l'écoute se fait en stéréo et en multicanal (Dolby Surround, Dolby-E), la compatibilité du Downmix Stéréo et Mono est vérifiée et doit être validée en **RF Mode Film Standard**. Les valeurs des metadata Dolby sont vérifiées.

Voici des exemples de PAD sous forme de cassette HDCAM et HDCAM-SR indiquant la répartition des pistes audio, avec une ou plusieurs versions 5.1 (VF et VO), dans tous les cas, la première paire AES de la cassette sera obligatoirement utilisée pour une version stéréo PCM.

Exemple d'un PAD HDCAM avec version stéréo et 5.1

	AES 1	AES 2
Pistes	1-2	3-4
Format	Stéréo	Dolby E 5.1
Contenu	LR	5.1 + (LtRt)

Exemple d'un PAD HDCAM SR avec versions stéréo et 5.1 VF et VO

	AES 1	AES 2	AES 3	AES 4
Pistes	1-2	3-4	5-6	7-8
Format	Stéréo VF	Stéréo VO	Dolby E 5.1 VF	Dolby E 5.1 VO
Contenu	LR	LR	5.1 + (LtRt)	5.1 + (LtRt)

Le BLITS est un autre type de signal d'identification du 5.1 souvent utilisé en broadcast, il est fourni dans la session Pro Tools test (chapitre 2).

SIGNAL D'IDENTIFICATION BLITS

- Le signal d'identification BLITS permet d'identifier les canaux surround dans un signal 5.1. Il permet aussi d'identifier la présence ou l'absence d'un canal pour un downmix stéréo. Ce signal est souvent utilisé en début de programme pour identifier les canaux dans les cars HD.
- Les fréquences utilisées sont indiquées sur le schéma ci-dessous, elles diffèrent suivant les canaux, 880 Hz pour L et R, 1 320 Hz pour C, 660 Hz pour Ls et Rs et 82,5 Hz pour le LFE. Elles sont ordonnées dans l'ordre de rotation sens horaire pour un meilleur suivi.

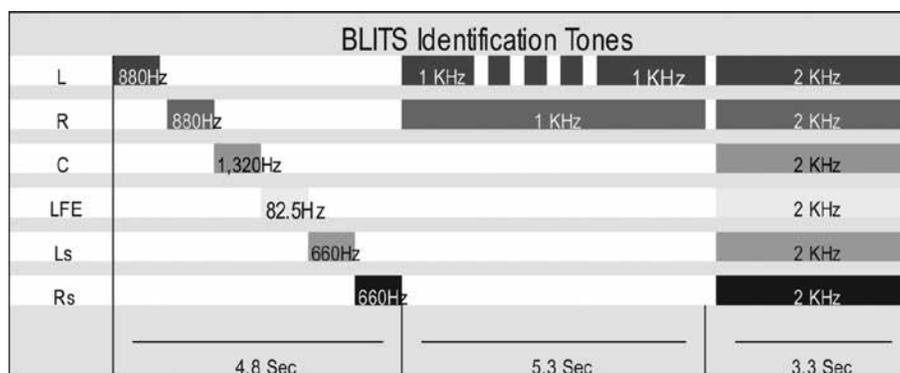


Figure 8.35 – Signal d'identification BLITS, © DK technologies.

8.3.5 Gestion du loudness des différents flux à la diffusion

Le signal 5.1 issu du DVB peut être diffusé en Dolby Digital ou Dolby Digital +, ou bien en HeAAC. Dans le cas d'une diffusion HeAAC, le signal est transcodé en Dolby Digital ou en DTS (optionnel). Le niveau de loudness associé au signal 5.1 présent sur une sortie HDMI codé ou PCM, coaxiale SPIDF (codé Dolby D ou DTS) et analogique est dans tous les cas – **31 LUFS**. Le niveau de loudness d'un signal stéréo dans les autres cas est de – **23 LUFS**. C'est pourquoi il est prévu un système d'atténuations pour respecter ces niveaux. Le RF mode d'un signal Dolby Digital associé à un downmix impose

un niveau de -20 LUFS, d'où la nécessité de prévoir une atténuation de 3 dB pour être raccord avec le niveau -23 LUFS. Certains programmes en Dolby Digital étaient auparavant renseignés avec un niveau de loudness de -27 LUFS, une atténuation de 4 dB les feront passer au niveau -31 LUFS.

Les schémas 8.36 et 8.37 issus du document Tech 3344 de l'EBU synthétisent la diffusion DVB pour la partie qui nous intéresse en Dolby D dans le premier cas et en HeAAC dans le second.

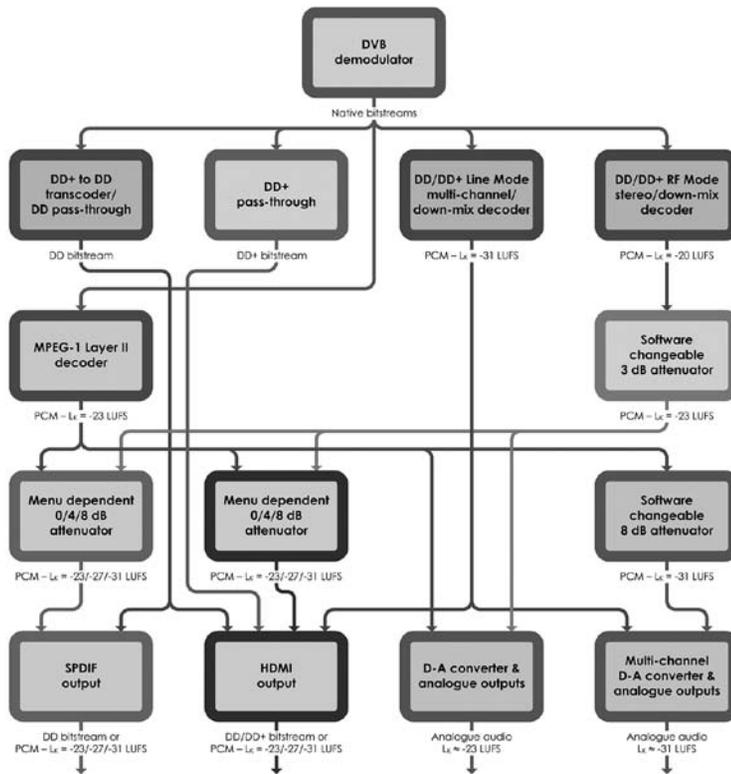


Figure 8.36 – Gestion des niveaux loudness à la diffusion, Dolby D, © EBU Tech 3344.

DOLBY VOLUME, DTS NEURAL LOUDNESS CONTROL

Dolby et DTS ont développé des algorithmes de mesure de loudness appelés Dolby Volume et DTS Neural Loudness Control que l'on trouve progressivement sur les amplificateurs home-cinéma. Ils permettent d'harmoniser en fin de chaîne les niveaux de loudness de toutes les sources audio présentes en entrée.

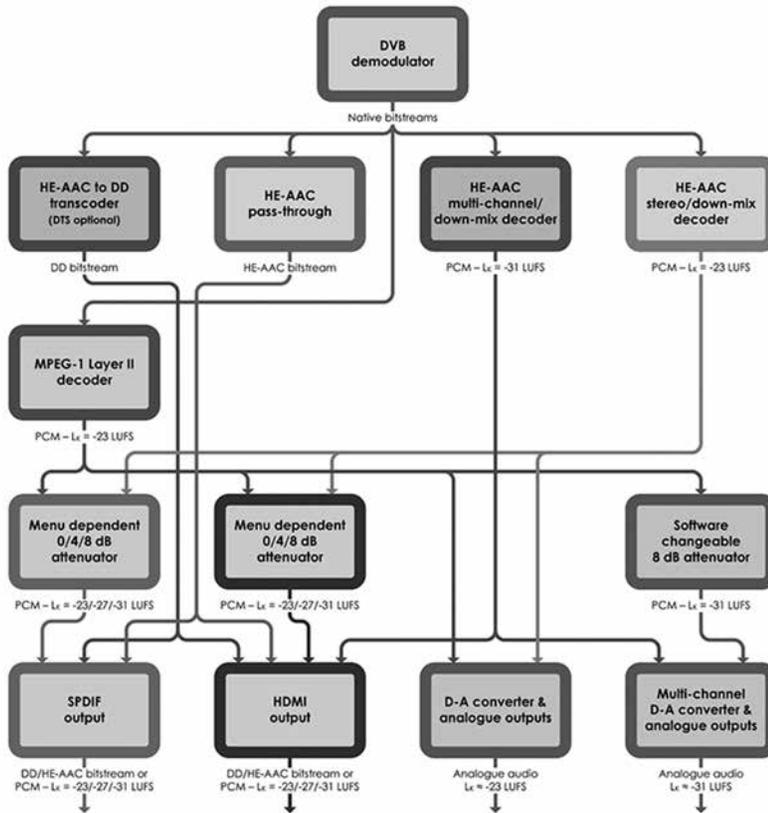


Figure 8.37 – Gestion des niveaux loudness à la diffusion, HeAAC, © EBU Tech 3344.

8.4 Exploitation des outils de traitement broadcast 5.1

On utilise des processeurs de diffusion dans les cars HD et en régie de diffusion, ceux-ci permettent de traiter automatiquement le loudness du programme et la dynamique du signal 5.1 (compression multibandes). Ces processeurs de diffusion limitent les crêtes du signal en dB True Peak, intègrent un module d'upmix pour enchaîner des programmes stéréo upmixés en 5.1, ils comportent un module de downmix pour vérifier la compatibilité stéréo du programme 5.1, et, pour certains, ils encodent le signal en Dolby E, Dolby Digital, Dolby Pulse. On trouve en fin de chaîne dans le processeur un module de mesure Loudness R128 pour valider la diffusion à la norme.

8.4.1 Traitement du Loudness et de la dynamique 5.1

Ces processeurs broadcast permettent de traiter le loudness, et limitent les crêtes du signal pour faciliter la mesure en direct dans les cars HD ou en régie de diffusion. Ces

traitements sont compatibles 5.1, ils intègrent en général un module de correction de loudness automatique « ALC » (automatic loudness correction) qui sort automatiquement un signal à -23.0 LUFS, ils comportent un module de traitement dynamique 5.1 multi bandes, un limiteur true peak réglé à -3 dBTP ou -1 dBTP, et un module de mesure de loudness. Parmi les processeurs compatibles 5.1, on trouve généralement le TC DB8, DB6 ou DB4, et le Jünger TAP. Les traitements dynamiques multi bandes 5.1 ont été traités au chapitre 6, le principe de fonctionnement est le même.

Exposons les bases du principe de fonctionnement du module de traitement ALC du DB8.

L'utilisateur renseigne la valeur de loudness cible, -23.0 LUFS, et définit des valeurs maximales et minimales de gain qui limitent le traitement de loudness (les valeurs « max reduction », « max boost »). Il doit préciser le temps avec lequel le changement de loudness va se faire, le rate (dB/s) et le paramètre « slow window » (dB). Ce dernier détermine une échelle en dB autour de la valeur cible, à l'intérieur de laquelle les variations de loudness seront plus lentes. Cela permet de stabiliser le procédé vers la valeur voulue -23.0 LUFS. Enfin, un seuil supplémentaire appelé « freeze level », définit un domaine en niveaux au-dessous duquel rien ne se passe, les bas niveaux restent intacts. Afin d'obtenir une continuité du signal, le passage par le niveau de freeze se fait avec un temps de maintien appelé « freeze hold », le processeur revient au gain 0 dB au bout de ce temps.

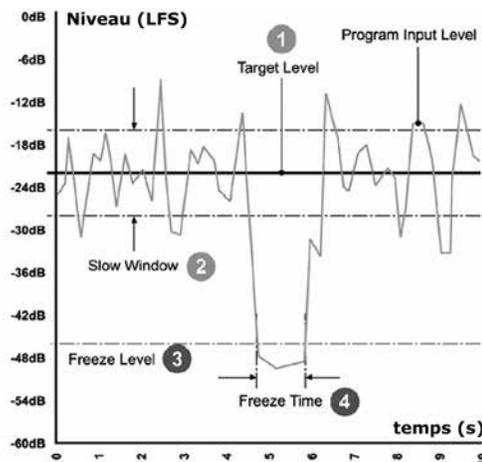


Figure 8.38 – Mode de fonctionnement de l'ALC DB8, © TC electronic.

Le correcteur automatique de loudness comporte un limiteur true peak placé en sortie du module de loudness.

8.4.2 « Upmixeurs »

Il existe de nombreux processeurs d'upmix pour la diffusion, lesquels permettent de fabriquer un flux 5.1 d'une chaîne HD à partir de programmes stéréo. Ces procédés sont utilisés de façon provisoire, avant la mise en place de nouvelles régies de diffusion

compatibles 5.1. Ces processeurs servent aussi à spatialiser des programmes disponibles qu'en stéréo. On peut penser que ces processeurs serviront toujours en secours ou dans le cas d'archives.

Dans les régies de diffusion TVHD ou dans les cars HD, on trouve ces traitements d'upmix. En France beaucoup de cars sont équipés de processeurs TC DB8, France 2 et NRJ12 utilisent des processeurs Isostem, Arte exploite un processeur Linear Acoustic Aeromax. La chaîne anglaise SKY utilise par exemple des processeurs UPM1 Soundfield.

Certains des processeurs fabriquent un upmix compatible 100 % avec le downmix stéréo LoRo, c'est à dire que la version stéréo downmixée en LoRo du flux 5.1 obtenue est 100 % identique à la stéréo d'origine (voir procédé Isostem). D'autres, fabriquent un programme 5.1 sans obtenir un downmix 100 % compatible.

Comme évoqué au chapitre 6, ces procédés reposent sur l'analyse de la phase entre les canaux gauche et droite de la stéréo, tel un décodeur Dolby Pro logic. Certains algorithmes rajoutent de la réverbération, d'autres n'en rajoutent pas et permettent de séparer le champ acoustique direct du champ réverbéré et de les spatialiser comme souhaité.

Il existe de nombreux processeurs d'upmix, citons les principaux : Isostem, Soundfield UPM1, TC Upcon/Unwrap (DB8, DB6, DB4), Junger TAP, Penteo, Aeromax (Linear Acoustic).

L'Isostem a la particularité innovante de séparer, dans un mixage stéréo, le champ direct du champ réverbéré. Il donne des résultats intéressants, compatibles 100 % avec la stéréo d'origine par downmix LoRo.

◆ Isostem

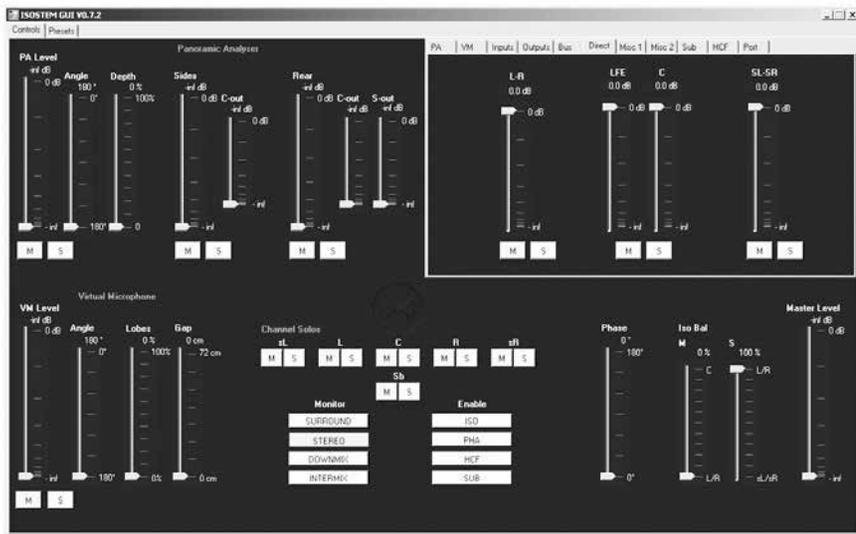


Figure 8.39 – Interface de contrôle d'Isostem, © Isostem. En couleur sur dunod.com

Isostem est un processeur d'upmix de diffusion qui repose sur un principe acoustique avec un downmix "ISO" égal à la stéréophonie. La technologie d'upmix est fondée sur l'analyse des échanges d'énergie acoustique au point d'écoute. Le point de départ du réglage repose sur l'analyseur panoramique PA, qui analyse la largeur ou la décorrélation en temps réel de la stéréo d'origine et qui permet d'isoler en quelque sorte la source de son environnement.

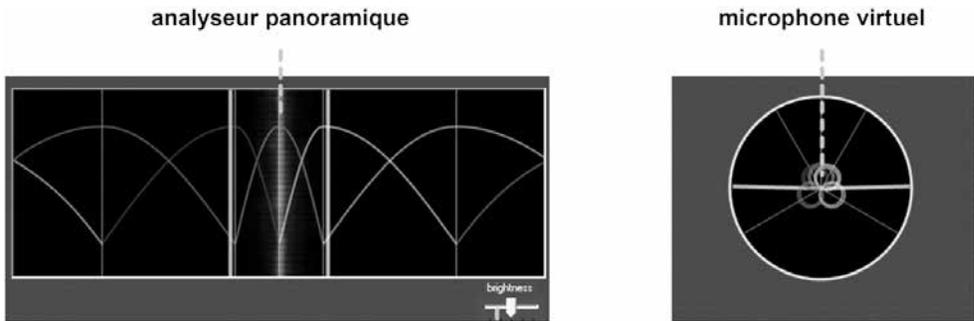


Figure 8.40 – Analyseur panoramique et microphone virtuel, © Isostem.

Le travail sur les paramètres de l'analyseur panoramique nous permet d'isoler l'information très corrélée et de l'extraire de l'image fantôme LR du multicanal pour en créer un centre. Nous procédons de la même façon pour l'information des canaux surround avec la gestion du niveau d'extraction du canal C et des canaux L, R. Ce réglage est dépendant de la profondeur de la loi panoramique et de l'angle d'ouverture de la stéréo. L'optimisation du réglage peut se faire en monitorant les différents éléments de l'analyseur grâce aux solo et aux mute.

Le module suivant est le microphone virtuel VM, une sorte de simulation d'un système de prise de son multicanal, qui permet de décorréler l'image sonore grâce à l'intégration de délais et des directivités des microphones. L'approche est intéressante, elle permet d'ajuster la décorrélation spatiale mais les délais peuvent induire quelques problèmes de filtrage en peigne.

Le module ISO nous permet d'optimiser le procédé d'upmix de telle sorte que le downmix soit identique à la stéréo d'origine, en réintégrant l'intermix, la différence entre la stéréo et le downmix.

Le module de sortie intègre les fonctions de monitoring (les différents canaux du multicanal, la stéréo d'origine, le downmix avec les mêmes coefficients de Dolby et l'intermix) avec les différentes options ISO (optimisation du downmix), HCF (atténuation des aigus à l'arrière), SUB (création du SUB par Bass Management), PHA (correction de phase).

Mis à part la qualité d'upmix, l'Isostem gère aussi de façon indépendante plusieurs flux stéréo pouvant recevoir des stems différents (musique, ambiances, VO, VI...) jusqu'à

4 entrées stéréo ce qui permet d'optimiser le rendu sonore de l'upmix en fonction des différentes versions d'un programme ou des stems disponibles.

◆ Upcon

L'Upcon est un processeur automatique d'upmix dans le DB8 qui gère les transitions de programmes 5.1 aux programmes stéréo upmixés. Le processeur repose sur l'algorithme Unwrap que l'on trouve aussi dans la M6000 (voir chapitre 6). Le programme analyse en temps réel le format audio en entrée et, si le signal d'entrée passe en stéréo, il enchaîne avec « crossfade » sans interruption ou artefact sur un programme stéréo upmixé. Le processeur assure ainsi la diffusion d'un signal 5.1 ininterrompue.

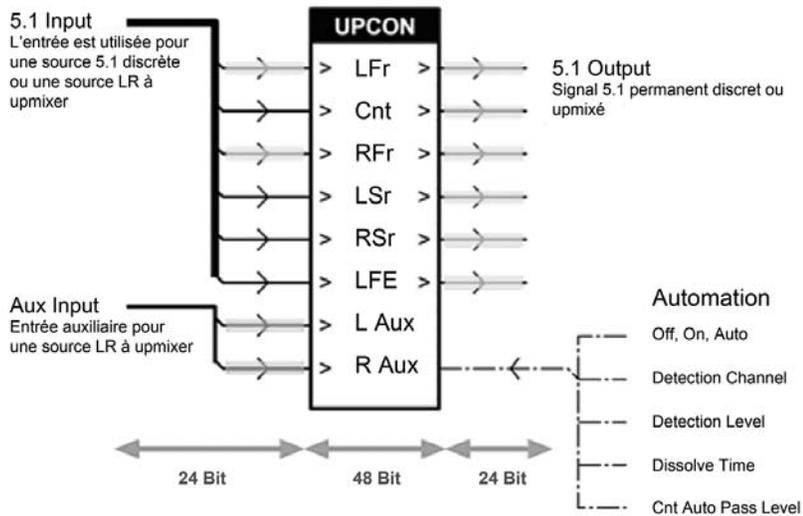


Figure 8.41 – Mode de fonctionnement de l'Upcon DB8, © TC electronic.

8.4.3 Encodage Dolby D, Dolby D+, Dolby pulse et gestion des metadata

Parmi les solutions hardware exposées ci-dessus, certaines proposent l'encodage Dolby E, Dolby Digital+, et Dolby Pulse avec la gestion des metadata : le DP600, le Jünger TAP, et l'Optimod le proposent. Ces processeurs gèrent aussi les metadata Dolby des programmes et forment ainsi des solutions d'Asset Management de fichiers : le DP600, ou le Minnetonka Audio Tools Server le proposent. Ces solutions fonctionnent à partir de tâches programmées, elles peuvent analyser et corriger le loudness de fichiers de façon automatique, encoder ou décoder un signal Dolby, modifier une metadata particulière, traiter les fichiers MXF, « upmixer » en 5.1 ou 7.1, « downmixer » etc. Elles équipent progressivement les chaînes de télévision et facilitent la diffusion. L'asset management des fichiers est contrôlé chez Arte par le Minnetonka Audio Tools Server.

8.5 Exploitation des formats DTS

8.5.1 *DTS Neural*

Le format DTS Neural existe depuis le rachat de Neural Audio par DTS en 2009, c'est un format compatible 5.1 ou 7.1, matricé en deux canaux, de type LtRt. Neural Audio développait auparavant des processeurs d'upmix 5.1 pour la diffusion. Le DTS Neural a été conçu pour la diffusion broadcast, il est assez répandu aux Etats Unis notamment pour la fabrication du multicanal, la diffusion se faisant en Dolby E. L'encodeur appelé le DTS Neural Downmix, produit un flux stéréo de type LtRt PCM qui peut être « monitoré » tel quel en stéréo, ou bien décodé en DTS Neural. On trouve des décodeurs DTS Neural sur certains amplificateurs home-cinéma récents (Onkyo par exemple). Le décodeur professionnel, le DTS Neural Upmix, restitue les différents canaux du multicanal avec paramétrage de l'image sonore comme le réglage de l'équilibre frontal/arrière, le réglage de la largeur frontale. Le DTS Neural upmix peut aussi servir à « upmixer » des programmes stéréo.

On trouve aussi depuis 2012 des solutions DTS Express utilisant des débits appropriés pour la diffusion broadcast.

8.5.2 *Exploitation des formats DTS pour le support*

Le DTS est un codage très répandu pour le support, on le trouve dans le film, la musique, la radio. C'est un codec de très bonne qualité, son débit comme mentionné au chapitre 7 est de 1,5 Mbps. DTS propose des encodeurs hardware, les CAE-4 DTS et CAD-4 DTS, ainsi que des solutions software avec notamment la suite DTS HD Master Audio pour encoder un DVD ou un Blu-ray.

Il existe aussi le plug-in Soundcode DTS Neyrinck pour encoder depuis Pro Tools, ou l'application Minnetonka Surcode DTS. Le plug-in permet d'encoder la metadata de loudness (« Dialog norm »), le downmix étant de type LoRo et utilisant les coefficients de downmix – 3 dB pour le centre et les canaux arrière.

8.6 Diffusion 5.1 et binaural pour les nouveaux médias

À l'heure où le broadcast et Internet convergent vers des solutions multimédia, se pose la question de diffuser des fichiers vidéo HD son 5.1 sur Internet, sur mobile ou sur tablette numérique. L'évolution des codecs permet aujourd'hui de diffuser un son 5.1 de plutôt bonne qualité avec un débit de 192 kbps. L'écoute binaurale accompagne progressivement ces solutions de diffusion du son en 5.1, grâce à des systèmes de codage utilisant des fonctions de transfert HRTF. On peut citer le mp3surround parmi les premières solutions qui sont apparues ou le MPEG Surround. Ces technologies proposées par Fraunhofer sont compatibles 5.1, stéréo et binaural.

L'étude de ces procédés met en évidence la diversité des codecs, la compatibilité avec les lecteurs, la compatibilité avec les metadata Loudness, Downmix, DRC afin d'optimiser

un downmix stéréo par exemple, et la possibilité d'activer un mode d'écoute spatialisée pour le casque.

8.6.1 Quelques recommandations

Voici tout d'abord quelques recommandations utiles avant d'encoder un flux audio 5.1 pour Internet. Rappelons l'ordre de grandeur du débit que demanderait la lecture d'un fichier 5.1 PCM en 48 kHz, 24 bits.

$$D = 6 \times 48000 \times 24 = 6,912 \text{ Mbps}$$

La lecture du flux 5.1 suppose l'utilisation d'une carte son 6 canaux connectée à un ordinateur ou bien la présence d'un port HDMI que l'on connecte sur un ampli home-cinéma.

Le fichier, avant l'encodage doit être entrelacé, mais il n'existe aujourd'hui aucune norme pour Internet, ce qui ne facilite pas du tout la lecture. L'ordre des canaux, par exemple, est variable entre le Dolby, le DTS, l'AAC, le fichier QuickTime... d'où l'intérêt d'avoir un lecteur qui permette de choisir l'ordre à la lecture, bien qu'il soit possible de le changer dans le paramétrage de la carte son. Si aucune contrainte d'ordre n'est imposée par le codec, il est peut être préférable de choisir l'ordre que l'on trouve en TVHD pour assurer la compatibilité entre les deux modes de diffusion TV et Internet, l'ordre ITU SMPTE étant LRCLfeLsRs. La lecture de fichiers 4.0, LRLsRs, ou 5.0 LRCLsRs, est possible si l'encodeur renseigne le mode 4.0 ou 5.0 comme dans le flux Dolby Digital, le lecteur reconnaîtra alors le « channel mode ». Mais pour les autres codecs, l'information de « channel mode » n'existe pas forcément, il convient alors de convertir tous les flux en 5.1 avec des pistes silencieuses en respectant l'ordre des canaux. Notons que les fichiers QuickTime largement utilisés, imposent l'ordre CLRLsRs Lfe.

8.6.2 Les codecs

Il existe aujourd'hui une dizaine de codecs compatibles 5.1 que l'on peut utiliser à un débit variant entre 160 et 600 kbps. Voici les applications des principaux codecs que l'on utilise pour la diffusion des contenus sur ces nouveaux médias (Internet, mobiles, tablettes numériques).

◆ **Mp3surround, mp3D**

Le mp3surround est l'évolution du mp3 en multicanal, développé par Fraunhofer. Le débit utilisé est normalisé à 192 kbps, les fichiers ont des extensions .mp3 comme pour la stéréo. Fraunhofer met à disposition pour l'encodage l'application mp3 encoder et le plug-in Sonnox Pro Codec. Les lecteurs compatibles sont l'application Fraunhofer mplayer, Winamp avec le plug-in mp3surround, ainsi que les applications Internet Explorer, Firefox, et Safari avec le plug-in mp3surround. Il existe par exemple deux webradios allemandes qui proposent des flux mp3surround : Rock Antenne et Antenne Bayern. Radio Classique a proposé un temps sur son site Internet un flux Internet

Radio Classique HD, compatible mp3surround. Le mp3D est un codec compatible mp3 et mp3surround qui permet la restitution du son en binaural. Il comporte un algorithme d'upmix pour les flux stéréo et propose différents réglages HRTF.

◆ **Dolby Surround**

Le Dolby Surround est une solution très peu exploitée sur Internet, car peu d'applications sont compatibles à ce jour, VLC Player affiche l'être.

◆ **Dolby Digital, Dolby Digital +, Dolby Headphone**

Les codecs Dolby Digital et Dolby Digital Plus permettent de diffuser un son 5.1 de qualité comparable au DVD, à un débit de 384 kbps ou 256 kbps pour le Dolby Digital Plus. Les fichiers Dolby Digital sont des .ac3, compatibles 5.1 et les fichiers Dolby Digital Plus sont des .ec3 compatibles 5.1 et 7.1. Ces codecs sont compatibles avec VLC player. L'encodage peut se faire avec des applications ou plug-ins tels que Dolby Media Encoder, Neyrinck Soundcode et Minnetonka Surcode. Les metadata Dolby ne sont en revanche pas exploitées par VLC Player à ce jour, VLC Player crée par exemple son propre downmix LoRo lorsque l'application voit en sortie une configuration stéréo. On trouve aussi ces codecs dans certains mobiles et tablettes numériques. Le mode binaural peut être activé dans un flux Dolby Digital avec le procédé Dolby Headphone, certains mobiles l'intègrent (Dolby mobile utilisant le Dolby Digital Plus).

◆ **DTS Neural, DTS express, DTS Surround sensation headphones**

Le DTS Neural est un format PCM matricé de type LtRt compatible 5.1 et 7.1, il existe des encodeurs logiciels développés par DTS mais aucun lecteur n'est vraiment compatible à ce jour, seul l'amplificateur compatible DTS Neural complète la chaîne. Quelques radios HD diffusent aux États-Unis un son DTS Neural, encodé en mp3 sur Internet. À défaut de décodeur DTS Neural, un décodeur Dolby Pro logic 1 ou 2 décodera le signal compatible LtRt. D'autre part DTS communique depuis 2012 sur le DTS Express adapté aux applications bas débit, mais peu d'expérimentations ont pu être menées à ce jour. DTS propose aussi une solution de diffusion binaurale apparue en 2012 appelée le DTS Surround sensation headphones qui malheureusement n'est pas encore très accessible à ce jour.

◆ **AAC (AAC LC), HeAAC, HeAACv2, HD AAC**

Le codec AAC LC, fichier .aac, est un format assez répandu sur Internet, il est exploité par Apple pour le son des bandes annonces de certains Films HD 5.1. Ce codec est directement compatible avec les lecteurs QuickTime Player, VLC Player et iTunes. Les applications QuickTime Pro, Adobe Media Encoder ou les plug-ins Sonnox Pro Codec/Codec Toolbox permettent d'encoder en AAC LC. Il n'existe pas encore de solution accessible pour intégrer des metadata dans le flux. Le downmix stéréo est directement fabriqué

par le lecteur, généralement avec les coefficients par défaut pour le centre et les arrières (-3 dB, -3 dB).

Le codec HeAACv2 est aujourd'hui un des codecs les plus performants, présentant une grande efficacité. La qualité est étonnante pour un taux de compression de rapport 40 correspondant à un débit de 192 kbps. Les fichiers sont des .mp4 et .m4a. Il n'existe pas encore de solution accessible pour intégrer des metadata dans le flux. Il existe aujourd'hui quelques applications accessibles capables d'encoder en HeAACv2, parmi elles on citera QuickTime Pro, Adobe Media Encoder, VLC Player, Orban, les applications Sonnox Pro Codec/Codec Toolbox. Les fichiers encodés sont compatibles avec VLC Player, QuickTime Player et iTunes. Il existe aussi un codec AAC sans perte, le HD AAC qui utilise les applications Sonnox Pro Codec/Codec Toolbox.

◆ **MPEG Surround**

Le MPEG Surround, développé par Fraunhofer en collaboration avec LSI Corp., Dolby Laboratories et Philips Applied Technologies, fonctionne à partir de n'importe quel codec, généralement l'AAC, le HeAAC et le MPEG layer 2, à des débits comparables à ceux de la stéréo. L'information de l'espace est généralement codée par un flux de 32 kbps. Les applications Sonnox Pro codec/Codec Toolbox permettent d'encoder en MPEG Surround utilisant l'AAC-LC et l'HeAAC. Aujourd'hui peu d'applications sont compatibles MPEG Surround, seul Winamp Media Player supporte le MPEG Surround en lecture, ce n'est pas encore le cas des lecteurs Apple (iTunes, QuickTime). D'autre part, Fraunhofer met à disposition une license pour des applications sur mobile ou tablette numérique.

◆ **MPEG-H 3D**

Le MPEG-H 3D en cours de développement apportera sans doute une solution de diffusion complète, compatible 3D, binaural à des débits variables.

◆ **Dolby Pulse**

Le Dolby Pulse utilise les codec HeAACv1 et HeAACv2 et intègre les metadata Dolby. L'encodage se fait avec le DP600 ou l'application AudioTools Server Minnetonka. Il est compatible avec les lecteurs QuickTime Player et iTunes.

◆ **WMA**

Le WMA (Windows Media Audio, fichier .wma) est un codec Microsoft compatible jusqu'au 7.1. L'encodage se fait avec l'application Windows Media Encoder. Ce codec intègre une metadata downmix qui permet de contrôler la version stéréo downmixée LoRo si le fichier est lu avec Windows Media Player. Le WMA supporte aussi la metadata DRC. Le fichier .wma peut être lu avec VLC Player.

8.6.3 Les lecteurs compatibles 5.1

Aujourd'hui le 5.1 n'est pas toujours compatible avec les navigateurs Internet, un downmix stéréo est généralement fabriqué par le lecteur. Il existe en revanche des plug-ins pour la compatibilité mp3surround. L'application Flash player, présente sous forme de plug-in dans les navigateurs, n'est pas à ce jour compatible 5.1, elle fabrique un downmix stéréo. D'autre part, le langage HTML 4 n'est compatible qu'avec le codec mp3. Le langage HTML 5 représente une grande évolution pour la diffusion 5.1 car il est directement compatible avec les codecs AAC et avec les fichiers .wav 5.1 ou 7.1. Aujourd'hui, les navigateurs ne sont pas tous compatibles, à ce jour Internet Explorer 9, Safari et Chrome le sont. Le codec HeAAC est nativement compatible dans Android depuis la version 4.1, le mobile et la tablette numérique peuvent donc facilement délivrer un flux 5.1 en HeAAC. Côté Apple, il semblerait que la sortie 5.1 en iOS8 reste encore limitée au flux Dolby Digital.

CODAGE FICHIER	COMPRESSION	TYPE	FORMAT	DECODAGE possible	DEBIT approximatif utilisé	EXTENSION habituelles	METADATA
Dolby Surround	perte	matricé	matricé 2.0, 5.1	amplis AV Dolby PL1, Dolby PL2		.wav	Loudness/ Downmix/ DRC
Dolby Surround 2X	perte	matricé	matricé 2.0, 7.1	amplis AV Dolby PL2X		.wav	Loudness/ Downmix/ DRC
Dolby surround 2Z	perte	matricé	matricé 2.0, 9.1	amplis AV Dolby PL2Z		.wav	Loudness/ Downmix/ DRC
Dolby D	perte	discret	1.0 à 5.1, Dolby EX 6.1	VLC	384 kbps 640 kbps max	.ac3	Loudness/ Downmix/ DRC
Dolby D +	perte	discret	1.0 à 13.1	VLC	256 kbps 3Mbps max	.ec3	Loudness/ Downmix/ DRC
Dolby D TrueHD	sans perte	discret	jusqu'à 16 canaux	VLC	jusqu'à 18 Mbps		Loudness/ Downmix/ DRC
Dolby Headphone	perte	binaural associé au Dolby Surround, Dolby Digital, DD+	binaural	amplis AV, PC, mobiles			
DTS	perte	discret	5.1	VLC	1509 kbps	.cpt	Loudness/ Downmix
DTS express	perte	discret	jusqu'au 5.1		64 à 768 kbps		
DTS HD	perte	discret	jusqu'au 7.1	VLC	2 à 6 Mbps	.dtshd	Loudness/ Downmix
DTS Master	sans perte	discret	jusqu'au 7.1	VLC	jusqu'à 24 Mbps	.dtshd	Loudness/ Downmix
DTS Neural	perte	matricé	matricé 2.0	Dolby PL2 ou DTS Neural amplis AV		.wav	non
DTS Surround sensation headphones	perte	binaural associé au DTS (5.1 ou upmix)	binaural	amplis AV, PC, mobiles			
AAC LC	perte	discret	jusqu'au 7.1	QT, iTunes, VLC	32 à 800 bbps	.mp4a	non
HeAAC	perte	discret	jusqu'au 5.1	QT, iTunes, VLC	192 kbps 640 kbps max	.mp4	metadata Fraunhofer

CODAGE FICHER	COMPRESSION	TYPE	FORMAT	DECODAGE possible	DEBIT approximatif utilisé	EXTENSION habituelles	METADATA
HEAAC V2	perte	discret	jusqu'au 7.1	QT, iTunes, VLC	192 kbps 640 kbps max	.mp4	metadata Fraunhofer
Dolby Pulse	perte	discret	1.0 2.0 5.1	QT, iTunes	192 kbps 640 kbps max	.mp4	Loudness/ Downmix/ DRC
HDAAC	sans perte	discret	1.0 2.0 5.1		la moitié du débit PCM	.mp4	non
Windows Media Audio	perte	discret	jusqu'au 7.1	WMP, VLC	192 kbps 768 kbps max	.wma	Downmix/ DRC
mp3surround	perte	discret, binaural associé	5.1	Mplayer, Winamp, plug-in Fraunhofer	192 kbps 320 kbps max	.mp3	non
FLAC	sans perte	discret	jusqu'au 7.1	VLC	la moitié du débit PCM	.flac	non
mp3D	perte	binaural (5.1 ou upmix)	binaural	Mplayer			non
MPEG Surround (AAC LC ou HeAAC)	perte	discret, binaural associé	jusqu'au 7.1	Winamp, Orban	128 kbps 320 kbps max	.m4a	compatible stéréo

Figure 8.42 – Synthèse des principaux codecs 5.1 pour la diffusion des nouveaux médias.

8.6.4 Les lecteurs compatibles 5.1, 3D et binaural

Il existe des solutions de lecteurs multimédia compatibles 5.1 et binaural, comme l'application Fraunhofer Mplayer. Le lecteur propose trois flux possibles : le 5.1 pour une écoute sur home-cinéma, la stéréo downmixée ou stéréo originale, et la version binaurale dont l'encodage peut être réalisé avec trois réglages HRTF « dry room », « living room » et « cinéma ». Le codec exploité par l'application Mplayer est le mp3surround, mais il existe des solutions analogues avec le MPEG Surround.

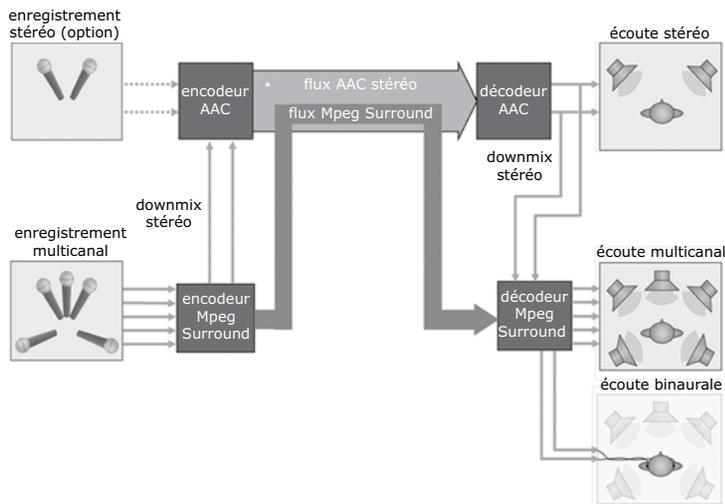


Figure 8.43 – Encodage décodage Fraunhofer MPEG Surround, © Fraunhofer.

Le schéma ci-dessous représente les différents modes d'exploitation d'un même fichier MPEG Surround en 5.1 diffusé depuis un mobile/iPod dans une voiture, sur un home-cinéma depuis le même lecteur mobile ou depuis un pc, et spatialisé en binaural.

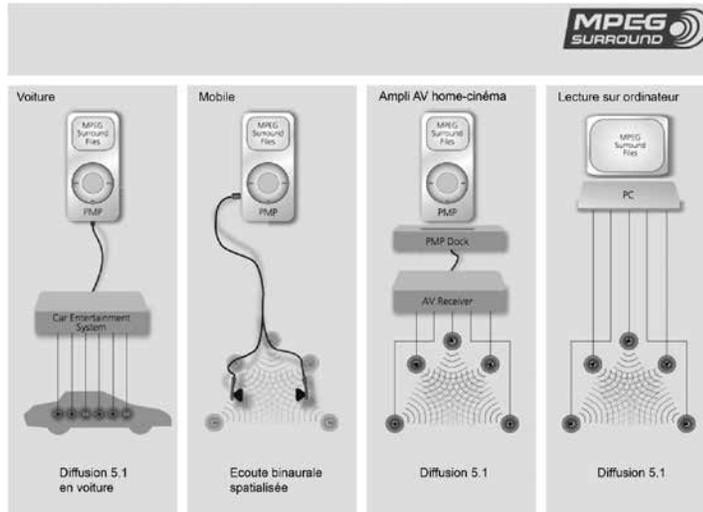


Figure 8.44 – Exploitation d'un fichier MPEG Surround en 5.1 compatible binaural, © Fraunhofer.

La diffusion sur Nouvoson utilise la même approche, les fichiers sont encodés en HeAACv2, l'auditeur peut alors choisir entre les 3 formats 5.1, stéréo ou binaural. La synthèse binaurale est opérée par la web API qui gère aussi la norme de diffusion R128. Une application comme Orange Radio complète le procédé pour l'écoute binaurale sur mobile ou tablette numérique, il suffit de pointer vers les flux de diffusion des sites concernés. Une diffusion 5.1 via HDMI est aussi envisagée pour une écoute sur home cinéma.

À terme, le codec MPEG-H 3D aura des possibilités élargies pour une diffusion 3D channel ou objets, avec compatibilité HOA et downmix binaural possible.

Conclusion

La spatialisation du son met en scène des nouveaux concepts de fabrication par la prise de son, le mixage et la diffusion. Ces notions en constante évolution sont au cœur des technologies numériques que l'audiovisuel connaît aujourd'hui. Le multicanal est depuis une vingtaine d'années abordé dans les séminaires de recherche sur l'audio, il s'enrichit chaque année de nouvelles publications. Nous avons derrière nous vingt ans d'expérimentations sur la stéréophonie et sommes à l'aube du développement du son 3D pour le cinéma, accompagné de représentations par objets sonores qui simplifient la diffusion. D'autres procédés reposant sur l'analyse complexe du champ sonore entrent en scène et sont appliqués dans la diffusion WFS par exemple, ou dans des systèmes de captation à haute résolution spatiale tels que le HOA. Les récents développements du binaural convergent vers cette prise en compte de l'espace et modifient, du fait de leur accessibilité pour tous, les modes d'écoute au casque. Ces technologies sur le son spatialisé constituent très certainement les modes de diffusion de demain.

De nombreuses réflexions et expérimentations ont abouti sur la recherche de nouvelles formes d'écriture pour différents champs d'application : le film, le documentaire, la musique, la radio, le spectacle, la scénographie, l'interactivité. Le son 3D ouvre vers d'autres formes d'écritures encore plus sophistiquées. C'est pourquoi la réalisation sonore de l'espace est une étape complexe, elle nécessite une écriture avant la production.

Le dispositif d'écoute s'homogénéise progressivement, le cinéma tend vers une écoute 3D plus décorrélée, plus universelle, offrant une meilleure compatibilité avec d'autres contenus comme l'événementiel diffusé en direct. Le cinéma explore ces nouvelles possibilités avec une longueur d'avance sur le broadcast, lui-même à la recherche de nouveaux formats son 3D accompagnant l'image HD 4K ou 8K.

Par ailleurs, l'écoute 5.1 domestique pose toujours la contrainte de la place qu'elle occupe chez le particulier, c'est pourquoi la miniaturisation des enceintes et l'utilisation d'un subwoofer facilitent les installations à la maison. La simplification de la diffusion est aussi indispensable, la présence du port HDMI sur les lecteurs Blu-ray, sur les ordinateurs portables, les tablettes numériques, le mobile ou les set-top box, permet la liaison vers l'ampli home-cinéma.

Si nous regardons derrière nous, sur quoi repose l'échec de la quadriphonie dans les années 1970 ? Une technologie pas encore mûre, non compatible avec le film, très coûteuse et surtout le manque d'accessibilité aux contenus (supports, diffusion). Aujourd'hui, la norme 5.1 ITU existe et le numérique simplifie la diffusion grâce à la réduction de débit. D'autre part, la diffusion sur Internet semble progresser avec l'apparition de webradios et de webtv.

Le son 5.1 est souvent accompagné d'une image. La convergence de l'image et du son est de plus en plus forte, notamment grâce à l'évolution des lecteurs audiovisuels mobiles comme la tablette numérique et le téléphone portable. La présentation d'informations sur l'écran, d'images fixes ou en mouvement est une demande croissante, c'est aussi ce que l'on voit sur les plates-formes comme YouTube ou Dailymotion. La radio numérique s'adapte à cette évolution vers les nouveaux médias en proposant de l'image.

Le cheminement vers l'écoute spatialisée est porté essentiellement par la reproduction la plus exacte possible d'une scène sonore et par l'écriture de nouvelles formes immersives interactives. Les images diffusées sont aussi de plus en plus grandes et il est cohérent de faire correspondre l'angle de vision avec l'angle auditif.

Finalement, c'est l'apport d'émotions nouvelles, plus fortes qui motive.

Bien sûr, la spatialisation répond à une écoute attentive du spectateur, lui-même en attente de découvertes et de divertissement. Mais lorsque l'écriture est pensée pour l'espace, il est difficile de revenir vers une écoute en stéréo, l'auditeur perd alors beaucoup d'informations et le rendu sonore est bien pauvre. La comparaison d'une ambiance multicanal et de son format en stéréo nous laisse toujours rêveur ! L'espace est confortable et la nature a bien fait les choses, nous l'entendons !

Bibliographie

Le lecteur est invité à consulter la bibliographie complète disponible gratuitement sur le site www.dunod.com.

BECH S., ZACHAROV N., *Perceptual Audio Evaluation: Theory, Method and Application*, Wiley, 2006

BLAUERT Jens, *Spatial Hearing*, MIT Press, 1996

BOSI M., GOLDBERG R.E., *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, 2002

CAMERER F., *Storytelling with Surround-Sound*, Audio Design and Esthetics, ORF, UDT, 2005

DANIEL J., *High Order Ambisonics, une approche complète de la spatialisation sonore*, Orange Labs, 2009

FUCHS J. & THEVENOT P., *De la monophonie à la synthèse de front d'onde WFS*, Présentation à l'Institut du Monde Arabe, 2012

GUILLOIN P., *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*, thèse Université du Maine, 2009

HAMASAKI K., *Three dimensional sound recording and reproduction of natural sound field*, NHK, VDT, 2005

HOLMANN Tomlinson, *Surround Sound : Up and Running*, Focal Press, 2008

HUGONNET C. & WALDER P., *Prise de Son, stéréophonie et multicanal*, Eyrolles, 2012

LAGNEL B., *Méthodes de prise de son et de mixage développées en multicanal*, présentation AES section française, 2013

LIBOLT A., *La création des métadonnées du Dolby E en prévision d'une diffusion multicanal en Dolby Digital*, Mémoire de fin d'étude, ENSLL, 2006

MOORE B. C. J., *An Introduction to the Psychology of Hearing*, Emerald, 2008

- MOREAU N., *L'Orientation réflexe de l'attention dans le cinéma multicanal*, Mémoire de fin d'étude, ENSLL, 1999
- NICOL R., *Binaural Technologies*, AES monograph 2011, Perception sonore spatialisée, Orange Labs
- Proceedings of ICSA, International Conference on Spatial Audio VDT, 2011 & 2014
- RUMSEY F., *Spatial Audio*, Focal Press, 2001
- SCHOEPS, *Surround Recording Techniques*, 2006
- STEINKE G., *High Definition Surround Sound with Accompanying HD Picture*, VDT, 2005
- THEILE G., *Multichannel Natural Music Recording Based on Psychoacoustic Principles*, IRT, 2001
- TRINNOV AUDIO, *La prise de son 5.0 en haute résolution spatiale*
- WILLIAMS M., *Microphone Arrays for Stereo and Multichannel Sound Recording*, Volume 1 & 2, Il Rostro, Milan, voir www.mmad.info
- WITTEK H., *Systems and Technics for 3D recording*, Schoeps, 2011

Index

5.1 cinéma, 27
5.1 ITU, 26
6.1 cinéma, 29
7.1 Blu-ray, 31, 32
7.1 cinéma, 30
8.1, 35
10.2, 38
22.2 NHK, 39, 102

A

AAC, 330, 408
ABX, 324
AC3, 335
Acousmatique, 162
Acousmonium, 9, 162
ADAT, 350
AES, 350
AES67, 353
Ambiances, 254
Ambisonie/Ambisonics, 18, 211, 268
Angle minimum audible, 126
Arbre Decca, 201
Arbre Fukada, 204
Artefacts, 326
ASPEC, 328
ATRAC, 341
Attention auditive, 139
Auro-3D, 36, 37, 341, 355
Autopan, 289

B

Bande Critique, 314
Barks, 314

Bass management, 59, 77, 252
Bed, 301
BiLi, 46
Binaural, 19, 406
BLITS, 399
Blu-ray, 7, 348
BRIR, 133
Bruitages, 254
Bruit rose, 51
BSI, 138
BWF, 352

C

C20, 66
Cadres sonores, 143
Carré Hamasaki, 224
Cascade, 326
Casque multicanal, 84
CBR, 321
Channel mode, 335
Channel oriented, 17
Cinéma, 2, 354
Codage, 309
Codage entropique, 310, 318
Codage M/S, 320
Codage paramétrique, 321
Codage sinusoïdal, 320
Codage spatial, 320
Code régional, 346
Coherent Acoustic, 340
Compensation, 91
Compresseur, 284
Cône de confusion, 117
Conteneur, 352

Correction, 91
 Courbes ISO X, 82
 Création sonore, 303
 Critères, 48, 146
 Croix IRT, 222
 CSI, 138
 CSS, 342
 CST RT-17-TV version V3, 393
 CST RT17 v2, 387
 CST RT19, 387

D

DAB, 354
 Dante, 353
 DBAP, 268
 DCI, 354
 DCP, 354
 Débit, 311
 Décorrélation, 261
 Décorrélation interaurale, 138
 Dialnorm, 362
 Dialog Intelligence, 381
 Dialog level, 362, 381
 Dialog Range, 383
 Diaphonie acoustique, 111, 113
 Digital Theater System, 339, 345
 Directs, 254
 Dislocalisation, 141, 153
 Dispositif 5.1, 113
 Distance, 127
 Distraction, 142, 153
 Dither, 312
 Divergence, 261, 271
 Divergence centrale, 261, 269, 276
 Divergence vers les autres canaux, 271, 277
 Dolby Atmos, 42, 101, 301, 355
 Dolby Digital, 335, 363, 380
 Dolby Digital Plus, 336
 Dolby Digital Surround EX, 4
 Dolby E, 337, 360
 Dolby Pro Logic, 333
 Dolby Pro Logic 2, 333
 Dolby Pulse, 339, 409
 Dolby SR, 3
 Dolby SRD, 3
 Dolby Stéréo, 3, 28
 Dolby Surround, 26, 332, 361
 Dolby Surround 7.1, 5, 30
 Dolby Surround EX, 378
 Dolby Surround IIX, 35
 Double MS, 207
 Double ORTF, 219
 Downmix, 289, 360, 362, 367
 DPA 5100, 227
 DRC, 335, 362, 364, 376
 DRM, 342
 DSD, 328, 347
 DTRS, 350
 DTS, 32, 340, 406
 DTS-Core, 340
 DTS ES 7.1, 4
 DTS express, 408
 DTS HD, 371
 DTS HD Master, 371
 DTS Neo X, 39
 DTS Neural, 406, 408
 DTS Surround sensation headphones, 408
 DVB, 353
 DVD, 345
 DVD-Audio, 347
 DXD, 328

E

EBU R128, 383, 387
 Échantillonnage, 311
 Ecouter, 2
 Edison, 46
 Effet cocktail party, 139
 Effet de fusion, 63
 Effet Haas, 66, 130
 Égalisation physiotemporelle, 65
 Elementary stream, 344
 Élévation, 118
 Enveloppement, 136, 138, 151
 Équilibre CD/CR, 153
 Équilibre frontal/arrière, 152
 Équilibre spectral, 153
 ERB, 314
 ESI, 138
 Esthétique, 147
 Esthétique équirépartie, 147
 Esthétique frontale, 147
 Externalisation, 131
 Extradiégétique, 142

F

Fantasound, 2
 FEC, 134
 Filtres FIR, 92
 Filtres IIR, 92
 Focus. Voir *divergence centrale*
 Formants, 318
 Format A, 212
 Format B, 212
 Formats multicanal, 20
 Fusion frontale/arrière, 140, 144
 FX, 254

G

GRM, 9, 163
 Guard bands, 337

H

Haute résolution spatiale, 229
 HDCP, 343
 HDMI, 351
 HE-AAC, 328
 Head tracking, 133
 High Order Ambisonics (HOA), 134, 235, 355, 412
 Holophone, 226
 Home-cinéma, 7
 HPTF, 133
 HRIR, 119, 133
 HRTF, 118, 133, 237, 268
 HRTF personnalisées, 135

I

IACC, 138
 ILD, 115, 321
 Image fantôme, 110
 Image réelle, 110, 147
 Image virtuelle, 111, 147, 148, 149
 IMAX, 3, 29
 Immersion, 110
 Imm Sound, 40, 100
 Impression spatiale, 136, 150
 INA5, 204
 Incertitude de localisation, 125
 Indices de localisation dynamiques, 118

Indices spectraux, 118, 119, 120
 Intensité acoustique, 95
 Interactivité, 303
 Internalisation, 110, 131, 149
 Iosono, 299, 355
 IPD, 115
 Ircam, 10
 ISO-UDF, 345
 ITD, 115, 321
 ITU BS 1770-1, 381
 ITU BS 1770-2, 383
 ITU-R BS.775, 76
 ITU-R BS.1116, 324
 ITU-R BS.1283-1, 323
 ITU-R BS.1387, 325
 ITU-R BS 2051, 41

J

Jeu vidéo, 15, 165, 166, 167, 168, 169, 170

K

K, 381
 KFM 360, 217
 KU 100, 237

L

Largeur apparente, 136
 Latence, 311, 353
 LEDE, 69
 LF, 138
 Lfe, 3, 252, 264, 265, 335
 Limiteur, 288
 Line mode, 364
 LKFS, 381
 Loi de pan, 230, 267
 Long term, 381
 L'oreille ipsilatérale, 115
 LoRo, 289, 367, 372
 Lossless, 328
 Loudness, 360, 362, 376, 380, 388, 401
 Loudness Range, 389
 LPCM, 327
 LRA, 388, 391
 LtRt, 290, 332, 367, 372
 LUFs, 384

M

MADI, 350
 Manager, 261
 Masquage fréquentiel, 313
 Masquage temporel, 313
 Matricage, 327
 Mesure, 258
 Mesure loudness, 381
 Metadata, 16, 17, 359
 Metadata AAC, 369
 Metadata Dolby, 374
 Metadata DTS, 371
 Métadonnées. Voir *metadata*
 MLS, 51
 MMAD, 183
 Modèle psychoacoustique, 315
 Momentary loudness, 383, 388
 Montage son multicanal, 253
 Mp3D, 407
 Mp3surround, 407
 MPEG, 328
 MPEG2, 329, 346
 MPEG4, 330
 MPEG-H 3D, 409, 412
 MPEG Surround, 321, 328, 332, 409, 412
 Multichannel Stereo, 17
 Multiplexage, 343
 MUSHRA, 323
 MUSICAM, 328
 Musique, 256
 MXF, 352

N

Night mode, 365
 Niveaux, 329
 Noise-shaping, 312
 Normes cinéma, 80
 NouvOson, 412

O

Object oriented, 19
 Objet, 19, 300, 301
 OCT Surround, 197
 ODG, 325
 Ogg, 342
 Opus, 342

Orange Radio, 412
 Ordre, 229, 236
 Ordre des canaux, 253
 Oreille controlatérale, 115
 ORTF Surround, 220
 OSI, 352

P

PAD CST RT16, 387
 Pan-Pot, 269, 267
 PDR, 134
 Phase shift, 367
 PLOUD, 383
 Post-synchro, 254
 PPCM, 327
 Prédiction linéaire, 318
 Profils DRC, 366
 Program loudness, 384, 388
 Program Stream, 344
 Pro Logic, 361
 Puissance acoustique, 95
 Pure Audio Blu-ray, 7

Q

Quadriphonie, 5, 25
 Quantification, 311

R

R2LB, 381
 Réaction d'orientation réflexe
 de l'attention, 139, 140
 Réalisation, 171
 Realiser A8, 133
 Rendement, 57
 Réservoir, 321
 Réverbérance, 136
 Réverbération, 280
 RF mode, 364
 RLB, 381
 Rotation, 271

S

Saillance cognitive physique, 140
 Scene oriented, 18
 SDDS, 30, 341

SDI, 350
Sensibilité, 57
Short term loudness, 381, 383, 388
SMART, 347
SMPTE, 80
SOFA, 133
Sonde intensimétrique, 94
Son hors champ, 142
Son in, 142
Son off, 142
Soundfield, 211
Sonde physio-angulaire, 94
S/PDIF, 349
SPS 200, 216
SQAM, 325
ST350/ST450, 214
Standard stéréo, 110
Steering, 335
Stems, 254, 292
Streaming, 353
Super Audio CD, 328, 347
Super Hi-Vision, 14
Supra High Vision, 12
Sweet spot, 111
Synthèse binaurale, 294

T

Taille apparente, 136
Tatouage, 343
TDS, 51
Tête artificielle, 237
TG, 58, 59
THX, 89

TOSlink, 349
Transaural, 105, 275
Transformées de Fourier, 313
Transport Stream, 344
Trinnov SRP, 230
True peak, 288, 387, 390
TVHD, 7, 14, 44

U

UHDTV, 14
Upcon, 405
Upmix, 290, 402

V

VBAP, 267
VBR, 321
VI, 292
Vision, 131
Vorbis, 342

W

Waterfall, 56
Watermarking, 343
Wavelet, 56
WFS, 18, 43, 103, 297
WMA, 409

Z

Zone d'écoute, 129, 152