

Conception d'un Capteur Sonore pour la Localisation de Source en Robotique Mobile.

THÈSE

présentée et soutenue publiquement le 8 Décembre 2006

pour l'obtention du

Doctorat de l'Université Paul Sabatier
(spécialité systèmes automatiques)

par

Sylvain Argentieri

Composition du jury

<i>Rapporteurs :</i>	Frédéric BIMBOT Hiroshi G. OKUNO	Chargé de Recherche HdR, IRISA-INRIA Rennes Professeur, Université de Kyoto
<i>Examineurs :</i>	Yannick DEVILLE Michel DEVY Patrick PÉREZ Jean-Luc SCHWARTZ	Professeur, Université Paul Sabatier Directeur de Recherche, LAAS-CNRS Directeur de Recherche, IRISA-INRIA Rennes Directeur de Recherche, ICP Grenoble
<i>Directeurs de thèse :</i>	M. Patrick DANÈS M. Philippe SOUÈRES	Maître de Conférence, Université Paul Sabatier Chargé de Recherche HdR, LAAS-CNRS

Remerciements

Voici donc venu le temps des remerciements. C'est étonnant de voir avec quelle impatience j'ai attendu ce moment, mais aussi de constater avec quelle angoisse je commence à écrire ces lignes. Impatience, parce que ces remerciements mettent enfin un terme à un peu plus de trois années de recherche et à l'écriture parfois difficile de ce manuscrit ; mais angoisse également, de peur d'oublier quelqu'un, quelque chose. Aussi, je m'excuse d'avance de mes oublis éventuels . . .

Pour commencer, merci à Raja Chatila, d'abord responsable du groupe RIA dans lequel j'ai débuté cette thèse et maintenant directeur du LAAS, de m'avoir accueilli ici à Toulouse. Un très grand merci également à Michel Devy, responsable du groupe RAP dans lequel j'ai achevé mes travaux, pour avoir accepté d'être président de mon jury de soutenance. Je souhaite également remercier chaleureusement les autres membres de mon jury, à savoir Patrick Pérez, Jean-Luc Schwartz et notre voisin Yannick Deville, mais également mes rapporteurs Frédéric Bimbot et Hiroshi G. Okuno, ce dernier étant venu spécialement du Japon pour assister à ma soutenance.

Que serait une thèse sans un directeur ? Sans *des* directeurs devrais-je même dire. J'ai en effet eu cette chance d'avoir été encadré par deux formidables chercheurs auxquels je souhaiterais exprimer ma plus profonde gratitude. Philippe Souères tout d'abord : merci pour tes innombrables conseils avisés, ton charisme scientifique à tout épreuve, ta patience inébranlable, et surtout pour m'avoir laissé mener mes recherches et mes enseignements en toute liberté. Patrick Danès ensuite, dont la modestie n'a d'égale que la compétence et la remarquable rigueur scientifique. J'ai énormément appris à ton contact, bien plus qu'avec quiconque auparavant. Un énorme merci pour ta présence et tes nombreux conseils lors de nos discussions, qu'elles aient été scientifiques ou non. Je ne sais au final comment mieux remercier ces deux personnes autrement qu'en leur promettant d'agir comme eux avec des étudiants dans la situation qui a été la mienne, si un jour l'occasion m'en est donnée.

Que serait une thèse sans ces contributions quotidiennes venant des collègues permanents, post-doc, doctorants, ingénieurs, stagiaires ? Ils ont tous été présents lors des moments de doute et de remise en cause, et c'est avec eux également que j'ai partagé le meilleur. Petite dédicace spéciale aux membres des bureaux 180 et 181, machine à café *Senseo* comprise, ayant dû –entre autre– me supporter lors de la phase de rédaction de ce manuscrit. Un grand merci donc à "Chef" Aurélie, Jean-Michel "l'aviateur", Thierry "le chevelu", Claudia et Gustavo. Un petit coucou également aux anciens ayant déjà quitté le navire : Vince et Ludo devenus depuis les patrons de leur *petite entreprise*, mais également Efrain "piltrafa" et Gabi. Certains sont partis, mais d'autres sont également arrivés : je pense en particulier à "Master Geek" Sylvain et à Martial, récemment promu *permanent parisien*. Je ne voudrais surtout pas oublier de remercier le "petit bonhomme" Jérôme (et ses blagues inoubliables) d'avoir bien voulu partager avec moi ses grandes compétences d'électronicien pour la réalisation du capteur acoustique objet d'une partie de ma thèse. Je voudrais enfin remercier Viviane Cadenat pour sa gentillesse, Michel Taïx pour avoir accepté d'être mon "tuteur" de monitorat, et Marc Renaud pour sa relecture critique, efficace et pertinente de ce manuscrit.

Enfin, un immense merci à Elodie, la femme de ma vie. Jamais je ne serais arrivé à faire tout ça sans toi, sans ta patience et tes encouragements.

*J'ai toujours aimé le désert.
On s'assoit sur une dune de sable.
On ne voit rien.
On n'entend rien.
Et cependant quelque chose rayonne en silence ...*

*Antoine de Saint-Exupéry
Le petit prince*

Table des matières

Table des figures

xi

Avant-propos

Chapitre 1

De la propagation du son à sa perception chez l'homme

1.1	Propagation d'une onde sonore	5
1.1.1	Démonstration de l'équation d'onde	5
1.1.1.1	Équation d'état	6
1.1.1.2	Equation d'Euler	7
1.1.1.3	Equation de continuité	8
1.1.1.4	Equation d'onde et relations fondamentales	8
1.1.2	Résolution de l'équation d'onde homogène	9
1.1.2.1	Propagation des ondes sonores selon un système de coordonnées cartésiennes	10
1.1.2.2	Propagation des ondes sonores dans un système de coordonnées sphériques	12
1.1.3	Equation d'onde non homogène	14
1.2	Le système auditif humain	15
1.2.1	Aspects physiologiques	15
1.2.1.1	L'oreille externe	15
1.2.1.2	L'oreille moyenne	16
1.2.1.3	L'oreille interne	17
1.2.2	La localisation du son chez l'homme	19
1.2.2.1	Localisation horizontale	19
1.2.2.2	Localisation verticale	21
1.2.2.3	Perception de la distance et influence de l'environnement	22
1.2.2.4	L'effet de précedence	23
1.3	Conclusion	24

Chapitre 2

Les méthodes de localisation de source sonore en robotique : un état des lieux

2.1	Les méthodes biomimétiques	25
2.1.1	Les indices interauraux	26
2.1.1.1	Modèles pour la détermination des indices interauraux	28
	Un modèle élémentaire	28
	Un second modèle simple prenant en compte la présence de la tête	29
	Un troisième modèle plus développé basé sur la diffusion acoustique	30
	Autres approches	31
2.1.1.2	Stratégies de localisation	32
2.1.2	Les indices spectraux	34
2.1.3	L'effet de précedence	36
2.1.4	Autres approches biomimétiques	37
2.1.5	Conclusion	39
2.2	Les méthodes d'antennerie	39
2.2.1	Notations	40
2.2.2	Une méthode haute résolution : MUSIC	43
2.2.2.1	Description de l'approche	43
2.2.2.2	Exploitation en robotique	45
2.2.3	Localisation par corrélation	46
2.2.3.1	Détermination des TDOAs	47
	Généralités	47
	Une première amélioration	48
	Généralisation	48
	Exploitation en robotique	49
2.2.3.2	Utilisation des TDOAs pour la localisation	50
2.2.4	Formation de voie	52
2.2.4.1	Principe de la formation de voie	52
	Etude de l'influence des paramètres de l'antenne	54
	Conséquences pour la robotique	55
2.2.4.2	Exploitation des techniques de formation de voie pour la localisation de source en robotique	56
	Performances des formations de voie	56
	Résultats de localisation	58
	Autres utilisations de la formation de voie en robotique	59
2.3	Conclusion	61

Chapitre 3

Synthèse d'une formation de voie large bande

3.1	Généralités et notations	63
3.1.1	Formation de voie faible bande	64
3.1.2	Formation de voie large bande	66
3.1.2.1	Implémentation dans le domaine fréquentiel	66
3.1.2.2	Implémentation dans le domaine temporel	67
3.1.3	Définition du problème de synthèse	69
3.1.3.1	Remarques générales	69
3.1.3.2	Définition du comportement angulaire de référence	70
3.2	Synthèse naïve large bande par optimisation convexe	71
3.2.1	Optimisation et programmation semi-définie	71
3.2.2	Méthode faible bande	72
3.2.2.1	Présentation du problème d'optimisation	73
3.2.2.2	Résultats d'optimisation	74
	Optimisation sans limite sur G_B	74
	Optimisation sous la contrainte $G_B \leq \delta$	74
3.2.3	Méthode large bande, dans le domaine fréquentiel	75
3.2.3.1	Présentation du problème d'optimisation	75
3.2.3.2	Résultats d'optimisation	76
	Optimisation sans contrainte sur G_B	76
	Optimisation sous la contrainte $G_B \leq \delta$	77
3.2.4	Méthode large bande, dans le domaine temporel	78
3.2.4.1	Formulation du problème d'optimisation	78
3.2.4.2	Résultats d'optimisation	79
3.2.4.3	Synthèse large bande en gain seulement	81
3.2.4.4	Etude de l'influence des paramètres de synthèse	82
3.2.5	Conclusion	84
3.3	Influence de la distance : du champ lointain au champ proche	85
3.4	Synthèse modale large bande par optimisation convexe	88
3.4.1	Eléments d'analyse modale	88
3.4.1.1	Généralités	88
3.4.1.2	Expression modale d'une formation de voie	89
3.4.1.3	Relation de Parseval	91
3.4.2	Calculs des coefficients modaux	92
3.4.2.1	Dans le cas général	92

3.4.2.2	Pour une antenne linéaire	93
3.4.3	Synthèse modale large bande dans le domaine fréquentiel	94
3.4.3.1	Présentation du problème de synthèse	95
3.4.3.2	Calcul des coefficients modaux du comportement de référence	95
3.4.3.3	Résultats d'optimisation	98
	Optimisation sans contraintes sur G_B	98
	Optimisation sous contraintes sur G_B	99
3.4.4	Synthèse modale large bande dans le domaine temporel	99
3.4.4.1	Présentation du problème de synthèse	100
3.4.4.2	Résultats d'optimisation	101
3.5	Conclusion	104

Chapitre 4

Exploitation de formations de voie large bande pour la localisation

4.1	Conception de la chaîne d'acquisition et traitement	105
4.1.1	Choix des microphones	106
4.1.2	Dispositif d'acquisition	108
4.1.2.1	Description d'une chaîne d'acquisition classique	108
4.1.2.2	Choix des composants	109
	Filtre passe-haut	109
	Gain Réglable	110
	Filtre anti-repliement	110
	Conversion analogique-numérique	112
4.1.2.3	Modélisation de la chaîne d'acquisition	112
4.1.3	Traitement numérique embarqué	114
4.1.3.1	Le DSP	114
4.1.3.2	Le FPGA	114
4.1.4	Aspects logiciels	115
4.1.5	Conclusion	116
4.2	Formation de voie large bande optimisée et cartes de puissance	116
4.2.1	Chaîne d'acquisition et formation de voie	117
4.2.2	Résultats de localisation	120
4.2.2.1	Localisation de signaux sinusoïdaux	121
4.2.2.2	Influence de la distance	122
4.2.2.3	Influence du bruit	125
4.2.3	Conclusion	126
4.3	Formation de voie large bande optimisée et méthodes haute résolution	127

4.3.1	Rappels sur MUSIC et notations	127
4.3.2	Extensions de MUSIC au cas de sources large bande	128
4.3.2.1	Alignement et matrices de focalisation	128
4.3.2.2	MUSIC large bande dans l'espace des sorties des formations de voie	129
	Généralités	129
	Invariance selon la distance	131
	Matrice de covariance focalisée	131
	Utilisation pour la localisation	131
4.3.3	Synthèse modale des harmoniques sphériques	133
4.3.4	Résultats de localisation	133
	Estimation de l'azimut	135
	Estimation de la distance	135
4.3.5	Conclusion	136
4.4	Conclusion	137

Conclusion

Bibliographie

143

Table des figures

1.1	Tranche d'air de section S constante	7
1.2	Système de coordonnées cartésiennes	10
1.3	Système de coordonnées sphériques	12
1.4	Aperçu général de l'oreille humaine	16
1.5	Détails de l'oreille externe et moyenne	17
1.6	Détails de l'oreille interne et de la cochlée	18
1.7	Représentation tonotopique des fréquences dans la cochlée	18
1.8	Différences interaurales en amplitude pour différentes directions d'incidence . . .	20
1.9	Fonction de transfert de la tête, pour différentes fréquences et positions de la tête	22
1.10	Illustration de l'effet de précedence	23
2.1	Humanoïde SIG (université de Tokyo et Japan Science and Technolgy Corp.) . .	26
2.2	Géométrie épipolaire auditive	28
2.3	Influence du rôle de la tête dans les mesures d'IPD	29
2.4	Mesures de l'IPD dans différentes conditions acoustique	30
2.5	Plateforme expérimentale pour l'étude de l'influence de la tête	31
2.6	Mesures de l'IPD sur le robot SIG	32
2.7	Active Direction Pass Filter	33
2.8	Résultats de localisation obtenus sur le robot SIG	34
2.9	Plate-forme expérimentale pour l'évaluer du filtrage opéré par l'oreille externe . .	35
2.10	Modèle du pavillon externe	35
2.11	Détections de zones libres d'écho, et localisation en résultant	36
2.12	Illustration des caractéristiques temporelles et fréquentielles des filtres gammatone	38
2.13	Organisation d'une antenne de microphone	40
2.14	Représentation d'une antenne linéaire de microphones	42
2.15	Utilisation de l'algorithme haute résolution MUSIC en robotique	46
2.16	Comparaison des résultats de corrélation obtenus avec PHAT et RWPHAT . . .	49
2.17	Robot guide et antenne murale de microphones	50
2.18	Principe de la formation de voie	52
2.19	Réponse normalisée de l'antenne pour différentes valeurs de N et L	54
2.20	Etude de l'influence de la fréquence, pour N et L fixés.	56
2.21	Localisation de source à partir d'une antenne de 128 microphones	57
2.22	Robot EvBoyII et son antenne de 8 microphones	58
2.23	Antenne circulaire proposée et résultats de localisation résultants	58
2.24	Grille de discrétisation de l'espace et résultats de localisation	59
2.25	Estimation de la position du robot et de la direction d'émission du son.	60
2.26	Formation de voie pour l'émission sonore directive	60

3.1	Principe de la formation de voie faible bande	64
3.2	Formation de voie large bande dans le domaine fréquentiel	66
3.3	Formation de voie large bande dans le domaine temporel	68
3.4	Définition du comportement de référence	70
3.5	Optimisation faible, sans limite sur G_B	74
3.6	Résultats d'optimisation, pour $N = 25$ et différentes contraintes sur G_B	75
3.7	Formations de voie faibles bandes pour $f \in [300 : 100 : 3000]$	76
3.8	Evolution de $G_B = \mathbf{W}^H \mathbf{W}$ en fonction de la fréquence	77
3.9	Formations de voie faibles bandes pour $f \in [300 : 100 : 3000]$, avec $\mathbf{W}^H \mathbf{W} \leq 100$	77
3.10	Formations de voie faible bande.	80
3.11	Formations de voie faibles bandes pour $f \in [400; 3000]\text{Hz}$, avec $\mathbf{W}_{\text{LB}}^H \mathbf{W}_{\text{LB}} = 6.10^3$	80
3.12	Diagramme de Bode des 8 filtres réalisant la formation de voie	81
3.13	Synthèse large bande en gain seulement	83
3.14	Etude de l'influence de l'ordre des filtres	83
3.15	Influence de la limite δ	84
3.16	Comparaison des formations de voie conventionnelle et optimisée	85
3.17	Réponse d'antenne synthétisée en l'infini, à une distance de 1 m	86
3.18	Influence de la distance sur la composante 400 Hz	87
3.19	Tracés de l'amplitude des harmoniques sphériques $Y_{mn}(\theta)$	90
3.20	Illustration du théorème de Parseval	92
3.21	Décomposition modale de la réponse d'antenne de référence.	97
3.22	Formation de voie faible bande pour $r_0 = 1$ m	99
3.23	Evolution de $G_B = \mathbf{W}^H \mathbf{W}$ en fonction de la fréquence	100
3.24	Formations de voie faibles bandes pour $f \in [400 : 100 : 3000]$, avec $G_B = 100$	101
3.25	Formations de voie large bande pour $f \in [400 : 50 : 3000]$	102
3.26	Formations de voie large bandes pour $f \in [400 : 50 : 3000]$, avec $G_B = 1000$	103
3.27	Diagramme de Bode des 8 filtres réalisant la formation de voie	103
4.1	Photo du type de microphone retenu et principales caractéristiques techniques.	108
4.2	Photo du conditionneur	109
4.3	Schéma électronique de la carte d'acquisition	111
4.4	Diagramme de Bode de la chaîne d'acquisition	113
4.5	Photos des cartes d'acquisition et traitement	115
4.6	Architecture logicielle implémentée dans le FPGA	117
4.7	Phénomène d'explosion numérique en dehors du domaine d'optimisation.	118
4.8	Réalisation pratique d'une formation de voie dans une direction de l'espace.	119
4.9	Comparaison des diagrammes d'une formation de voie conventionnelle et optimisée	121
4.10	Cartes de puissance pour des formations de voie conventionnelle et optimisée	122
4.11	Représentation temporelle des signaux des sources	122
4.12	Carte de puissance pour des sources en champ lointain	123
4.13	Réponses d'antenne en champ proche	123
4.14	Influence de la distance sur la carte de puissance	124
4.15	Formation de voie en champ proche et carte de puissance	125
4.16	Influence du bruit blanc spatial sur la carte de puissance	125
4.17	Carte de puissance en présence de bruit blanc spatial	126
4.18	Synthèse modale des harmoniques sphériques pour $f = 400$ Hz	134
4.19	Synthèse modale des harmoniques sphériques pour f compris entre 400 Hz et 3 kHz	135
4.20	Détermination de l'azimut par MUSIC large bande	136

4.21 Estimation de la distance par MUSIC large bande	137
--	-----

Avant-propos

Le système auditif de l'homme fournit de nombreuses informations sur son environnement et s'avère être un complément idéal à ses autres modalités sensorielles. Ainsi, contrairement à la vision, notre perception du son est quasi-omnidirectionnelle et indépendante de la présence ou non de lumière dans la scène. De la même façon, nous sommes en mesure de traiter des informations auditives issues d'une pièce voisine sans la moindre information visuelle sur l'origine de celles-ci. Mais nos capacités ne se limitent pas seulement à la *localisation* des sons. Nous pouvons aussi *extraire*, au milieu de plusieurs locuteurs parlant simultanément, le discours de la personne sur laquelle nous souhaitons porter notre attention. Connues sous le nom de *Cocktail Party Effect*, ces capacités de séparation nous permettent tous les jours de traiter de manière efficace et sélective l'ensemble des données sonores de l'environnement. Sensibles à la moindre variation de ton et d'amplitude d'un message sonore, nous savons en *reconnaître* l'origine (sonnerie de téléphone, voix de notre collègue, ...) pour en *interpréter* l'information extraite. Toutes ces propriétés de localisation, extraction, reconnaissance et interprétation nous permettent d'évoluer aujourd'hui dans un univers dynamique, dans lequel il apparaît extrêmement difficile de se passer des informations auditives. Plus généralement encore, l'audition joue un rôle fondamental dans l'apprentissage culturel, en particulier pour tout ce qui a trait au langage, et donc à la communication entre les hommes.

Si nous sommes capables d'autant de tâches différentes à partir de seulement deux oreilles, c'est aussi que l'intégration des informations auditives s'opère en grande partie dans le cerveau. C'est le cas également pour la vision, pour laquelle l'emplacement d'un objet se trouve initialement codé sur chacune de nos deux rétines. Des cartes de l'environnement visuel peuvent alors être mises en avant dans la structure même des aires corticales, de sorte que la topographie de notre rétine soit directement codée au niveau du cerveau. Au contraire, dans le cas de l'audition, seuls des indices acoustiques sont disponibles et permettent au cerveau d'intégrer les informations sonores. Ces indices sont extraits du champ acoustique par nos deux oreilles, véritables transducteurs miniatures entassés dans le volume d'un petit pois. Les performances de ce capteur, aussi petit soit-il, sont impressionnantes : au seuil de perception et dans un environnement totalement silencieux, celui-ci est capable de détecter un déplacement des molécules d'air de l'ordre de la taille d'un atome d'or (c'est à dire quelques pico-mètres) tout en localisant l'origine d'un son avec une précision d'environ deux degrés.

L'ensemble des capacités auditives chez l'homme est extrêmement motivant pour chercher à étendre les capacités d'interaction des robots mobiles. Pour autant, les recherches dans le domaine de l'Acoustique *appliquée* à la robotique sont restées très longtemps anecdotiques. Il suffit pour s'en convaincre de compter le nombre d'articles de conférence sur le sujet, et de le comparer à la quantité de travaux déjà proposés en vision artificielle. Les progrès récents de l'électronique permettent d'ailleurs aujourd'hui de concevoir des systèmes capables de traiter un flux vidéo en temps réel pour l'extraction d'informations complexes sur l'environnement. Cependant, comme

nous avons tenté de l'illustrer précédemment, l'audition peut constituer un complément idéal à la vision, et l'émergence de travaux sur l'interaction entre l'homme et la machine ont mis en avant cette nouvelle modalité. L'audition peut ainsi être utilisée pour *localiser* un locuteur se situant en dehors du champ de vision du robot et le *reconnaître*, pour ensuite en *interpréter* les demandes : nous retrouvons là l'ensemble des capacités du système auditif humain que les roboticiens essaient aujourd'hui de reproduire.

Bien que l'utilisation du son en robotique soit très récente, l'exploitation et le traitement de signaux acoustiques ne sont pas des domaines de recherche nouveaux. Les communautés du Traitement du Signal et de l'Acoustique se sont emparées depuis longtemps des problématiques de localisation et de séparation de sources sonores et ont proposé de nombreuses solutions à ces problèmes. Cependant, les contraintes liées au contexte de la robotique mobile sont nombreuses et inédites, trop rarement prises en compte dans le formalisme des méthodes déjà existantes. Listons ici un certain nombre d'entre elles :

- **Contrainte géométrique** : Si l'objectif est d'aboutir à un système d'audition artificiel capable des mêmes performances que le système auditif humain, il n'est pas forcément nécessaire de limiter l'approche à un système biomimétique constitué de seulement deux microphones. Compenser la puissance de traitement de notre cerveau par l'utilisation de multiples transducteurs permet en effet d'augmenter la redondance d'informations en vue d'améliorer la résolution de la localisation, ou la robustesse vis-à-vis du bruit. Ainsi, il n'est pas rare en Acoustique de prélever l'information sonore à partir de capteurs – d'antennes – de grande dimension, souvent constitués de nombreux récepteurs. L'exploitation de tels dispositifs sur une plateforme mobile de taille modeste est cependant difficilement envisageable. Nous définissons là une contrainte d'*embarquabilité* venant limiter la dimension totale du capteur sonore. L'objectif est donc d'obtenir les meilleures performances à partir d'un nombre minimal de récepteurs.
- **Contrainte temporelle** : Beaucoup des méthodes de traitement déjà existantes nécessitent de nombreux calculs en ligne, et sont donc rarement compatibles avec un objectif de traitement en temps réel des données audio. S'il n'est pas gênant pour un acousticien de patienter 10 s pour obtenir une carte sonore de l'environnement, la rapidité de traitement des informations acoustiques est un critère fondamental en robotique. Que ce soit pour l'interaction ou l'inclusion dans une boucle d'asservissement haut niveau, la réactivité est un objectif fondamental. De ce fait, de nombreux algorithmes très performants sont à priori inexploitable et nous devons veiller à conserver une complexité algorithmique raisonnable, compatible avec une utilisation en robotique.
- **Contrainte fréquentielle** : La plupart des signaux sonores utiles dans le cadre d'une utilisation en robotique sont dits "large bande". Que ce soit la voix ou l'immense majorité des sons que nous percevons, tous ces signaux acoustiques sont constitués de multiples fréquences dont il faudra tenir compte lors de leur exploitation. Cela conduit tout naturellement à ne considérer que des algorithmes capables de traiter ce type de signal. De ce fait, les méthodes de la littérature s'avèrent souvent plus complexes que dans le cas de signaux mono-fréquentiels, et nécessitent une puissance de calcul plus importante.
- **Contrainte environnementale** : Les robots mobiles sont amenés à se déplacer dans un grand nombre d'environnements différents. Chacun possède ses propres caractéristiques

acoustiques, que ce soit en terme de réverbération ou de bruit ambiant. A l'origine de perturbations des signaux sonores, ces caractéristiques dépendent largement de la structure de la pièce dans laquelle le robot sera amené à évoluer : dimensions, présence de murs, nature des matériaux ... chacun de ces éléments étant susceptible de venir perturber le processus de traitement. Il est donc important que le système auditif embarqué soit capable d'adaptation à son environnement, ou du moins qu'il soit robuste vis à vis de ces perturbations acoustiques.

▷ Objectif de la thèse

L'étude que nous présentons constitue le premier travail effectué au sein du groupe Robotique Action Perception du LAAS en rapport avec l'audition artificielle. Notre objectif à terme est de doter les plateformes mobiles du laboratoire de capacités auditives pour localiser, reconnaître et interpréter les demandes de l'homme à la machine par la parole. Cependant, ce manuscrit se focalise uniquement sur l'aspect localisation de source sonore. Dans ce but, nous pouvons distinguer principalement trois objectifs à ce travail de thèse :

- Faire le point sur les méthodes de localisation déjà existantes, et évaluer leur pertinence pour une utilisation en robotique ;
- Proposer une solution de localisation satisfaisant les contraintes énoncées précédemment ;
- Réaliser le capteur sonore proprement dit, en vue d'une intégration sur un des robots mobiles du laboratoire.

▷ Organisation du document

Ce manuscrit est structuré en quatre chapitres, de la manière suivante :

- Dans le premier chapitre, le lecteur est initié aux bases de la propagation sonore et au fonctionnement de l'oreille humaine. Ainsi, nous étudions dans un premier temps les modes de propagation d'une onde acoustique dans l'air, en nous attachant à distinguer le cas des ondes planes ou sphériques. Puis nous présentons les mécanismes de l'audition à travers l'étude des différentes étapes de transduction opérées à l'intérieur de nos oreilles, pour enfin détailler les indices acoustiques nous permettant de localiser un son dans l'espace.
- Sur ces bases, le second chapitre est consacré à un état de l'art sur la localisation de source sonore en robotique mobile. Ce travail constitue la première contribution de ce manuscrit, car à notre connaissance une telle étude bibliographique n'a jamais été encore proposée. Après avoir détaillé les différentes approches dites biomimétiques, nous passons en revue les méthodes de localisation issues du Traitement du Signal et exploitées dans le cadre de la robotique.
- Le chapitre 3, quant à lui, est consacré à la synthèse de formations de voie large bande. Après avoir montré les limites d'une extension large bande "naïve" de méthodes déjà existantes, nous proposons une nouvelle technique de synthèse, valide en champ proche ou en champ lointain, et basée sur la résolution d'un problème d'optimisation convexe dans l'espace des harmoniques sphériques. Cette deuxième contribution est illustrée par de nom-

breux résultats obtenus sous Matlab, et nous permet de discuter de l'influence de la distance sur les performances des formations de voie.

- Le chapitre 4 fait le lien avec le précédent en proposant d'exploiter les formations de voie synthétisées pour le tracé de cartes de puissance acoustique de l'environnement, d'une part, et le calcul de pseudo-spectres MUSIC d'autre part. Cette dernière contribution sera également pour nous aussi l'occasion de présenter le travail de conception et de développement du capteur sonore, du choix des microphones à la réalisation des cartes d'acquisition.
- Une conclusion termine le document. Nous y résumons les principaux résultats obtenus et évoquons quelques perspectives à plus ou moins court terme de ce travail.

Chapitre 1

De la propagation du son à sa perception chez l'homme

Nous proposons dans ce premier chapitre d'étudier la propagation d'un son depuis une source sonore jusqu'à sa perception par le système auditif humain. Ces généralités physiques et physiologiques sont reprises par la suite afin d'appréhender les méthodes de localisation de source utilisées en robotique. Dans une première partie, nous redémontrons tout d'abord l'équation d'onde, équation de la physique modélisant la propagation du son dans un fluide. Nous rappelons ensuite ses deux solutions classiques conduisant à considérer des fronts d'ondes plans ou sphériques. Puis, dans une seconde partie, nous nous attachons à présenter le fonctionnement de l'oreille humaine. Après une étude physiologique rapide, nous présentons quels sont les indices auditifs utilisés par notre cerveau pour la localisation de source sonore.

1.1 Propagation d'une onde sonore

Les ondes sonores dans l'air ou l'eau sont des ondes *longitudinales*, i.e. les molécules du fluide ne se déplacent que dans leur direction de propagation, produisant des zones adjacentes de compression et de dilatation. D'une manière générale, les fluides possèdent deux caractéristiques mécaniques principales : leur élasticité et leur masse volumique. L'élasticité d'un fluide est à l'origine de sa résistance à la compression ou à la dilatation, et lui permet le retour à l'équilibre, tandis que la masse volumique produit quant à elle un effet d'inertie au cours du mouvement.

Supposons qu'à l'équilibre un élément infinitésimal d'un fluide possède une masse volumique ρ_0 (kg.m⁻³) sous une pression p_0 (Pa, ou N.m⁻²) à la température T_0 (°C, ou K). Dans la suite, nous désignons par δp la variation de pression autour de p_0 induite par une source sonore, et appelée *pression acoustique*. Si l'amplitude de cette composante variable est grande, alors δp n'est plus négligeable devant p_0 : des effets non linéaires doivent être pris en compte et conduisent à l'utilisation des modèles non linéaires de l'acoustique. Un tel cas n'est pas envisagé ici, et nous posons donc l'hypothèse de linéarité dans toute la suite de ce manuscrit : nous travaillons donc dans le domaine de l'acoustique linéaire. Ainsi, en présence d'une source sonore, l'ensemble des grandeurs à l'équilibre varient de manière proportionnelle à l'amplitude de l'excitation sonore.

1.1.1 Démonstration de l'équation d'onde

L'objectif de cette partie est de démontrer l'équation d'onde. De façon à simplifier l'exposé sans perdre pour autant en généralité, nous considérons dans un premier temps une propagation

mono-dimensionnelle en espace libre. Nous étendons ensuite la solution obtenue au cas à trois dimensions dans la partie suivante.

1.1.1.1 Équation d'état

Considérons une kilomole de gaz parfait placé dans les conditions normales de pression et de température pour lesquelles nous avons

$$T_0 = 273 \text{ K}, \quad p_0 = 1,013.10^5 \text{ Pa.} \quad (1.1)$$

Ce gaz vérifie l'équation d'état des gaz parfaits

$$\frac{p}{\rho} = \frac{R}{M}T, \quad (1.2)$$

où R désigne la constante des gaz parfaits ($R = 8,314 \text{ J.mole}^{-1}.\text{K}^{-1}$), p la pression, ρ la masse volumique et T la température. M est la masse molaire du gaz parfait, avec $M = 29 \text{ kg.kmole}^{-1}$ pour l'air. Dans toute la suite, nous admettrons qu'en l'absence de perturbation acoustique, le gaz au repos est parfaitement *homogène* et *isotrope*. De plus, ses caractéristiques physiques au repos (pression p_0 , masse volumique $\rho_0 = 1,293 \text{ kg.m}^{-3}$ et température T_0) seront considérées comme *constantes* et *uniformes* dans l'espace et le temps. Enfin, nous faisons l'hypothèse que le gaz est *sans viscosité*, que ses états d'équilibres sont *réversibles* et que les processus considérés sont *adiabatiques*, c'est à dire sans échange de chaleur avec l'extérieur. Sous ces conditions, l'entropie du gaz reste constante, et nous pouvons écrire l'équation adiabatique de changement d'état

$$pV^\gamma = \text{cte}, \quad (1.3)$$

où γ est le rapport des capacités molaires calorifiques et thermiques du fluide. La différentielle logarithmique de l'équation (1.3) conduit à l'équation de Laplace

$$\frac{\delta p}{p} + \gamma \frac{\delta V}{V} = 0. \quad (1.4)$$

Dans un fluide de masse volumique ρ_0 occupant un volume V_0 , une perturbation acoustique provoque une déformation. Le principe de conservation de la masse s'écrit alors

$$\rho_0 V_0 = (\rho_0 + \delta\rho)(V_0 + \delta V) = \rho_0 V_0 + V_0 \delta\rho + \rho_0 \delta V + \delta V \delta\rho. \quad (1.5)$$

Sous l'hypothèse d'acoustique linéaire, les termes du second ordre peuvent être négligés. D'où

$$\frac{\delta V}{V_0} = -\frac{\delta\rho}{\rho_0}. \quad (1.6)$$

Ainsi, l'équation de Laplace (1.4) peut s'écrire indifféremment

$$\boxed{\delta p = -\gamma p_0 \frac{\delta V}{V_0} = \gamma p_0 \frac{\delta\rho}{\rho_0}}, \quad (1.7)$$

ce qui conduit à constater que

$$\frac{\delta p}{\delta\rho} = \text{cte}. \quad (1.8)$$

L'équation précédente traduit donc l'existence d'une relation linéaire entre les variations de pression et de masse volumique dans le cadre de l'acoustique linéaire.

1.1.1.2 Equation d'Euler

Nous venons de rappeler les hypothèses sur le fluide nécessaires à l'obtention de l'équation d'onde en acoustique linéaire, ainsi que quelques-unes de ses propriétés. Nous allons maintenant étudier l'influence d'une perturbation acoustique sur son mouvement. Pour cela, considérons le cas simple d'une onde unidimensionnelle, se propageant dans une tranche d'air de section constante S , et délimitée par deux faces situées respectivement en les abscisses x et $x + dx$, voir figure 1.1. Au repos, chacune de ces faces est soumise à une pression p_0 . Suite à une excitation en $x = 0$, il existe une perturbation de pression $\delta p(x, t)$ autour de p_0 sur la première de ces faces. La perturbation de pression en $x + dx$ s'exprime par

$$\delta p(x + dx, t) = \delta p(x, t) + \frac{\partial(\delta p)}{\partial x} dx. \quad (1.9)$$

Sous l'action de cette même excitation, la tranche d'air située en x se déplace pour atteindre la position $x + s(x, t)$ à l'instant t , où $s(\cdot)$ désigne l'*élongation*. A l'instant $t + dt$, cette même tranche se situe en $x + s(x, t + dt) = x + s(x, t) + \frac{\partial s(x, t)}{\partial t} dt$. Ainsi, pendant dt , la tranche en x se sera déplacée à une vitesse $v(x, t)$ vérifiant

$$v(x, t) = \frac{\partial s(x, t)}{\partial t}. \quad (1.10)$$

Remarquons que v désigne la vitesse de la matière, aussi appelée *vitesse particulière*. Elle n'est pas à confondre avec la vitesse de déformation de l'air.

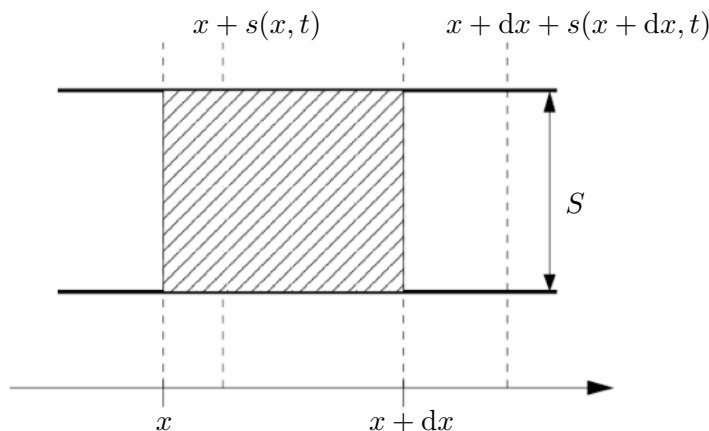


FIG. 1.1 – Déformation d'une masse élémentaire de fluide contenue dans un cylindre de section constante, sous l'action d'une perturbation acoustique.

Calculons maintenant l'accélération d'une particule située au temps t en x et animée d'une vitesse particulière $v(x, t)$ définie selon (1.10). Cette même particule sera en $x + dx'$ au temps $t + dt$. La variation de vitesse $dv = dx'/dt$ entre ces deux instants est égale à

$$dv = \frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial x} dx', \quad (1.11)$$

et l'accélération Γ est donnée par

$$\Gamma = \frac{dv}{dt} = \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x}. \quad (1.12)$$

L'application du théorème de la résultante dynamique conduit alors à

$$(\rho_0 S dx) \Gamma = -S \frac{\partial(\delta p)}{\partial x} dx, \quad (1.13)$$

qui peut être ré-écrite à l'aide de (1.12) sous la forme

$$\boxed{\rho_0 \frac{\partial v}{\partial t} + \rho_0 v \frac{\partial v}{\partial x} + \frac{\partial(\delta p)}{\partial x} = 0.} \quad (1.14)$$

1.1.1.3 Equation de continuité

Sur la tranche en mouvement, la différence entre le flux de matière entrant et sortant est égal à l'accroissement, par unité de temps, de matière dans cette même tranche. Cette propriété de conservation de la masse peut s'écrire

$$(\rho S v)_{x+s} - (\rho S v)_{x+dx+s+ds} = \frac{\partial}{\partial t} [\rho S(dx + ds)]. \quad (1.15)$$

Comme $\rho = \rho_0 + \delta\rho$, l'expression précédente peut être simplifiée par application de la formule des accroissements finis pour donner

$$\boxed{\frac{\partial(\delta\rho)}{\partial t} + v \frac{\partial(\delta\rho)}{\partial x} + \rho_0 \frac{\partial v}{\partial x} = 0.} \quad (1.16)$$

1.1.1.4 Equation d'onde et relations fondamentales

Nous disposons maintenant de toutes les équations nécessaires à l'obtention de l'équation d'onde. Sous l'hypothèse de l'acoustique linéaire, les équations (1.14) et (1.16) peuvent être simplifiées pour obtenir

$$\rho_0 \frac{\partial v}{\partial t} + \frac{\partial(\delta p)}{\partial x} = 0, \quad (1.17)$$

$$\frac{\partial(\delta\rho)}{\partial t} + \rho \frac{\partial v}{\partial x} = 0. \quad (1.18)$$

En utilisant l'équation de Laplace (1.7), et par dérivation du système précédent, nous obtenons

$$\rho_0 \frac{\partial^2 v}{\partial t^2} = \gamma p_0 \frac{\partial^2 v}{\partial x^2}. \quad (1.19)$$

Au final, comme $v = \partial s / \partial t$, et en notant c la grandeur définie par

$$c = \sqrt{\frac{\gamma p_0}{\rho_0}} = \sqrt{\frac{\gamma R T}{M}}, \quad (1.20)$$

l'équation (1.19) devient l'équation d'onde à une dimension, et est donnée par

$$\boxed{\frac{\partial^2 s}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} = 0.} \quad (1.21)$$

La grandeur c définit la vitesse de propagation de la déformation selon x . Elle est appelée *célérité de l'onde*, ou simplement *vitesse du son*, et a pour valeur $c \approx 340 \text{ m.s}^{-1}$ pour $T = 20^\circ \text{C}$. Notons qu'elle dépend de la température selon la formule

$$c \approx 20\sqrt{T}, \quad (1.22)$$

où T s'exprime en degrés Kelvin. Pour simplifier l'étude et la résolution de l'équation d'onde, les acousticiens ont pris l'habitude d'introduire le *potentiel des vitesses*, noté Φ , et vérifiant

$$v = -\frac{\partial \Phi}{\partial x}. \quad (1.23)$$

L'introduction de ce potentiel permet d'exprimer l'ensemble des grandeurs de l'acoustique, puisque nous avons, à partir de (1.17), la relation

$$\delta p = \rho_0 \frac{\partial \Phi}{\partial t}. \quad (1.24)$$

Les propriétés que nous venons d'énoncer sont également valables dans le cas 3D. Ainsi, tout comme en électricité où le champ électrique $\vec{E}$ dérive d'un potentiel V , ou en mécanique où le champ de force $\vec{F}$ dérive d'une énergie potentielle U , le potentiel des vitesses Φ vérifie de manière générale

$$\mathbf{v} = -\overrightarrow{\text{grad}}\Phi. \quad (1.25)$$

Finalement, et nous ne le démontrerons pas ici, l'équation d'onde *homogène* (1.21), dans le cas à trois dimensions, peut se mettre sous la forme générale

$$\boxed{\nabla^2 \Phi(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \Phi(\mathbf{r}, t) = 0}, \quad (1.26)$$

où $\Phi(\mathbf{r}, t)$ désigne le potentiel des vitesses à l'instant t pour un vecteur de position $\mathbf{r}$, et ∇^2 est le *Laplacien* dont l'expression, tout comme celle de $\mathbf{r}$, s'exprime différemment selon le système de coordonnées retenu. Nous allons donc maintenant résoudre cette équation dans deux systèmes de coordonnées différents : le cas cartésien tout d'abord, puis le cas sphérique.

1.1.2 Résolution de l'équation d'onde homogène

Pour résoudre l'équation (1.26) selon différents systèmes de coordonnées, il est classiquement supposé que le potentiel des vitesses Φ possède une dépendance temporelle harmonique telle que pour une fréquence f exprimée en Hz,

$$\Phi_f(\mathbf{r}, t) = \Phi(\mathbf{r}, f) e^{j2\pi f t}, \quad (1.27)$$

où $\Phi(\mathbf{r}, f)$ décrit la dépendance du potentiel des vitesses en la variable d'espace $\mathbf{r}$ et la fréquence f , et $j \triangleq \sqrt{-1}$. Cette hypothèse revient donc à considérer un champ acoustique mono-fréquentiel. Dans le cas général, rares sont les champs acoustiques de ce type. Néanmoins, leur étude est fondamentale, dans la mesure où sous l'hypothèse d'additivité des potentiels des vitesses, un potentiel multichromatique $\Phi(\mathbf{r}, t)$ s'écrit

$$\Phi(\mathbf{r}, t) = \int_{\mathbb{R}} \Phi(\mathbf{r}, f) e^{j2\pi f t} df. \quad (1.28)$$

Ainsi, $\Phi(\mathbf{r}, t)$ et $\Phi(\mathbf{r}, f)$ sont tous deux reliés par une *transformation de Fourier*. L'expression (1.27), combinée avec l'équation d'onde (1.26) conduit alors à

$$\boxed{\nabla^2 \Phi(\mathbf{r}, f) + \left(\frac{2\pi f}{c}\right)^2 \Phi(\mathbf{r}, f) = 0,} \quad (1.29)$$

qui est l'équation d'onde indépendante du temps, ou *équation d'Helmholtz*. Nous pouvons faire apparaître dans 1.29 la grandeur k , appelée *nombre d'onde*, et définie par

$$k = \frac{2\pi f}{c} = \frac{2\pi}{\lambda}, \quad (1.30)$$

avec λ la longueur d'onde. Nous rappelons ici que nous faisons l'hypothèse que la vitesse du son c est constante ; de ce fait, le nombre d'onde k est directement proportionnel à la fréquence f . Ainsi, dans la suite de ce manuscrit, nous désignerons aussi par *fréquence* le nombre d'onde k . Finalement, la résolution de l'équation d'onde (1.26) s'effectue à partir de l'équation d'Helmholtz (1.29).

1.1.2.1 Propagation des ondes sonores selon un système de coordonnées cartésiennes

Nous considérons dans cette partie un système de coordonnées cartésiennes. Ainsi, le vecteur position $\mathbf{r}$ est constitué des trois coordonnées $\mathbf{r} = (x, y, z)$, cf figure 1.2. La solution de l'équation

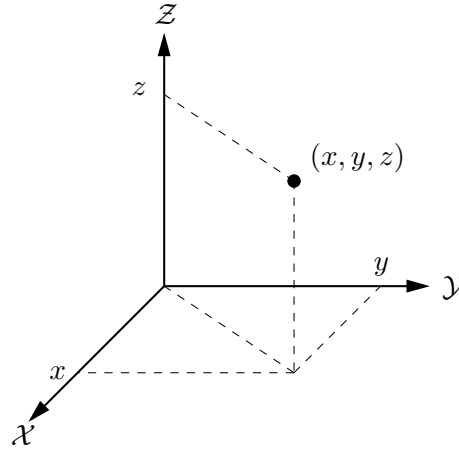


FIG. 1.2 – Système de coordonnées cartésiennes

d'Helmholtz (1.29) est obtenue en utilisant la méthode de séparation des variables, qui consiste à poser

$$\Phi(\mathbf{r}, f) = \Phi(x, y, z) = X(x)Y(y)Z(z), \quad (1.31)$$

où la dépendance des fonctions Φ , X , Y et Z à la fréquence f est sous-entendue. En introduisant (1.31) dans (1.29), avec, par définition du Laplacien dans un repère cartésien

$$\nabla^2 \triangleq \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \quad (1.32)$$

nous aboutissons à

$$\frac{1}{X(x)} \frac{d^2}{dx^2} X(x) + k^2 = -\frac{1}{Y(y)} \frac{d^2}{dy^2} Y(y) - \frac{1}{Z(z)} \frac{d^2}{dz^2} Z(z). \quad (1.33)$$

Remarquons que la partie gauche de l'équation précédente est une fonction de la variable x , tandis que la partie droite est une fonction des variables y et z . Comme nous souhaitons que (1.33) tienne pour tous les triplets (x, y, z) , il faut nécessairement que chacun de ses termes soit égal à une constante, que nous noterons ν_1^2 . Nous avons donc, en posant $k_x^2 = k^2 - \nu_1^2$,

$$\frac{d^2}{dx^2} X(x) + k_x^2 X(x) = 0, \quad (1.34)$$

qui admet pour solution exacte

$$X(x) = \mathcal{A}_x e^{-jk_x x} + \mathcal{B}_x e^{jk_x x}, \quad (1.35)$$

où $\mathcal{A}_x$ et $\mathcal{B}_x$ désignent deux constantes d'intégration. Compte tenu de l'équation (1.27), le premier terme de (1.35) est relatif à la propagation d'une onde sonore se déplaçant dans le sens des x croissants, tandis que son second terme décrit une onde se déplaçant en sens contraire. En procédant de la même façon pour la détermination des fonctions $Y(y)$ et $Z(z)$, nous pouvons montrer alors que la solution $\Phi(x, y, z)$ à l'équation (1.29) est donnée par

$$\Phi(x, y, z) = (\mathcal{A}_x e^{-jk_x x} + \mathcal{B}_x e^{jk_x x}) (\mathcal{A}_y e^{-jk_y y} + \mathcal{B}_y e^{jk_y y}) (\mathcal{A}_z e^{-jk_z z} + \mathcal{B}_z e^{jk_z z}), \quad (1.36)$$

où $\mathcal{A}_y, \mathcal{B}_y, \mathcal{A}_z$ et $\mathcal{B}_z$ désignent les constantes d'intégrations des fonctions $Y(y)$ et $Z(z)$. L'expression précédente peut être simplifiée en introduisant le vecteur de propagation $\mathbf{k} = (k_x, k_y, k_z)^T$, ou *vecteur d'onde*, dont la direction est identique à celle de la propagation, et dont la norme vérifie

$$\|\mathbf{k}\| = k_x^2 + k_y^2 + k_z^2 = k^2. \quad (1.37)$$

Si de plus nous supposons que la propagation s'effectue sans obstacles, c'est à dire en *espace libre*, alors l'onde se propage uniquement dans le sens pointé par $\mathbf{k}$, de sorte que nous pouvons poser $\mathcal{B}_x = \mathcal{B}_y = \mathcal{B}_z = 0$. Dans ce cas, l'équation 1.27 s'écrit simplement

$$\Phi_f(\mathbf{r}, t) = \mathcal{A} e^{-j\mathbf{k} \cdot \mathbf{r}} e^{j2\pi f t}, \quad (1.38)$$

où $\cdot$ désigne le produit scalaire et $\mathcal{A}$ une constante d'intégration dépendant de f . Finalement, $\Phi_f(\mathbf{r}, t)$ représente une onde harmonique se propageant selon la direction du vecteur de propagation $\mathbf{k}$, et pour laquelle la surface définie par

$$\mathbf{k} \cdot \mathbf{r} = \text{cte} \quad (1.39)$$

représente l'ensemble des points de l'espace possédant la même valeur de phase à un instant t . Une telle surface est appelée *front d'onde*. Elle est orthogonale à la direction de propagation $\mathbf{k}$, et définit ici un plan : on parle alors d'*onde plane*.

1.1.2.2 Propagation des ondes sonores dans un système de coordonnées sphériques

Nous venons de détailler l'obtention de la solution de l'équation d'Helmholtz dans un repère cartésien. Nous nous proposons ici de résoudre cette même équation dans un système de coordonnées sphériques. Pour cela, une position $\mathbf{r}$ dans l'espace est cette fois représentée par ses trois coordonnées sphériques r , θ et ψ , comme indiqué figure 1.3. Dans ce cas, le Laplacien est défini par

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \psi^2} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) \quad (1.40)$$

Tout comme précédemment, la résolution est obtenue par application de la méthode de séparation des variables, qui consiste à poser ici

$$\Phi(\mathbf{r}, f) = \Phi(r, \theta, \psi) = R(r)\Theta(\theta)\Psi(\psi), \quad (1.41)$$

où la dépendance des fonctions Φ , R , Θ et Ψ à la fréquence f est sous entendue.

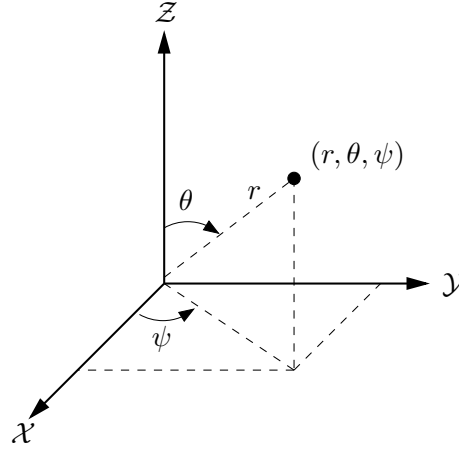


FIG. 1.3 – Système de coordonnées sphériques

De cette façon, partant de la définition du Laplacien (1.40), et en introduisant (1.41) dans (1.29), il vient

$$r^2 \sin^2 \theta \left(\frac{1}{R(r)} \frac{d^2 R(r)}{dr^2} + \frac{2}{r R(r)} \frac{dR(r)}{dr} + k^2 \right) + \frac{\sin^2 \theta}{\Theta(\theta)} \left(\frac{d^2 \Theta(\theta)}{d\theta^2} + \frac{1}{\tan \theta} \frac{d\Theta(\theta)}{d\theta} \right) = -\frac{1}{\Psi(\psi)} \frac{d^2 \Psi(\psi)}{d\psi^2}. \quad (1.42)$$

Ainsi, la partie droite de (1.42) est une fonction de ψ seulement, tandis que la partie gauche est quant à elle une fonction de r et θ . Comme l'équation précédente est valide pour tous les triplets (r, θ, ψ) , il faut que chacun de ses termes soit égal à une constante que nous noterons n^2 , appelée constante de séparation. Il vient alors immédiatement

$$\frac{d^2 \Psi(\psi)}{d\psi^2} + n^2 \Psi(\psi) = 0, \quad (1.43)$$

qui admet pour solution

$$\Psi(\psi) = \mathcal{E}_n \cos(n\psi) + \mathcal{F}_n \sin(n\psi), \quad (1.44)$$

où $\mathcal{E}_n$ et $\mathcal{F}_n$ représentent deux constantes d'intégration. Il reste donc à traiter la partie gauche de l'équation (1.42), qui se ré-écrit en

$$\frac{r^2}{R(r)} \frac{d^2 R(r)}{dr^2} + \frac{2r}{R(r)} \frac{dR(r)}{dr} + k^2 r^2 = \frac{n^2}{\sin^2 \theta} - \frac{1}{\Theta(\theta)} \left(\frac{d^2 \Theta(\theta)}{d\theta^2} + \frac{1}{\tan \theta} \frac{d\Theta(\theta)}{d\theta} \right). \quad (1.45)$$

Cette fois encore, l'équation précédente doit être vraie pour tout couple (r, θ) , de sorte que ses deux parties ne peuvent être égales qu'à une constante de séparation M . On admet qu'on peut poser $M^2 = m(m+1)$, avec $m \in \mathbb{N}$. Avec ces notations, le terme de droite de (1.45) peut s'écrire

$$\frac{d^2 \Theta(\theta)}{d\theta^2} + \frac{1}{\tan \theta} \frac{d\Theta(\theta)}{d\theta} + \left(m(m+1) - \frac{n^2}{\sin^2 \theta} \right) \Theta(\theta) = 0, \text{ avec } (m, n) \in \mathbb{N}^2.$$

Cette équation est connue sous le nom d'*équation de Legendre associée*, et admet pour solution

$$\Theta(\theta) = \mathcal{C}_{mn} P_m^n(\cos \theta) + \mathcal{D}_{mn} Q_m^n(\cos \theta), \quad (1.46)$$

où $P_m^n(\cdot)$ et $Q_m^n(\cdot)$ désignent respectivement les *fonctions de Legendre associées de degré m et d'ordre n de première et seconde espèces*, et $\mathcal{C}_{mn}$, $\mathcal{D}_{mn}$ deux constantes d'intégration. Enfin, la dernière équation différentielle de la variable r issue de (1.45) est de la forme

$$\frac{d^2 R(r)}{dr^2} + \frac{2}{r} \frac{dR(r)}{dr} + \left(k^2 - \frac{m(m+1)}{r^2} \right) R(r) = 0,$$

équation similaire à l'*équation différentielle de Bessel sphérique*, dont la solution est

$$R(r) = \mathcal{A}_m h_m^{(1)}(kr) + \mathcal{B}_m h_m^{(2)}(kr), \text{ avec } m \in \mathbb{N},$$

où $h_m^{(1)}(\cdot)$ et $h_m^{(2)}(\cdot)$ désignent respectivement les *fonctions sphériques de Hankel d'ordre m de première et seconde espèces*. Ainsi, la solution générale de l'équation de Helmholtz (1.29) peut s'exprimer comme une combinaison linéaire de modes ϕ_{mn} , chacun de la forme

$$\Phi_{mn}(r, \theta, \psi) = \left(\mathcal{A}_m h_m^{(1)}(kr) + \mathcal{B}_m h_m^{(2)}(kr) \right) \left(\mathcal{C}_{mn} P_m^n(\cos \theta) + \mathcal{D}_{mn} Q_m^n(\cos \theta) \right) \left(\mathcal{E}_n \cos(n\psi) + \mathcal{F}_n \sin(n\psi) \right).$$

Supposons maintenant que la propagation s'effectue en espace libre. Dans ce cas, celle-ci ne s'effectue que dans un seul et unique sens, que nous prendrons comme étant celui des r croissants. Sous une telle convention, il devient nécessaire de poser $\mathcal{A}_m = 0$ puisque la fonction $h_m^{(1)}$ décrit, nous l'admettrons ici, une propagation en sens contraire. Ensuite, selon les propriétés des fonctions de Legendre associées de seconde espèce, nous avons $|Q_m^n(\cos \theta)| \rightarrow \infty$ pour $\theta \rightarrow 0$ ou π . Physiquement, il n'y a a priori aucune raison pour que le potentiel des vitesses tende vers l'infini pour de telles valeurs de θ . En conséquence, nous posons $\mathcal{D}_{mn} = 0$. Finalement, la solution de l'équation de Helmholtz homogène dans un repère sphérique peut donc s'écrire

$$\boxed{\Phi(r, \theta, \psi, f) = \sum_{m=0}^{\infty} \sum_{n=0}^m h_m^{(2)}(kr) P_m^n(\cos \theta) (\mathcal{A}_{mn} \cos(n\psi) + \mathcal{B}_{mn} \sin(n\psi))}, \quad (1.47)$$

avec $\mathcal{A}_{mn} = \mathcal{B}_m \mathcal{C}_{mn} \mathcal{E}_n$ et $\mathcal{B}_{mn} = \mathcal{B}_m \mathcal{C}_{mn} \mathcal{F}_n$. De plus, nous pouvons remarquer que pour une valeur de m fixée, la condition $kr = \text{cte}$ définit une surface possédant une phase constante. Comme r est la distance à l'origine, les fronts d'onde considérés ici sont donc sphériques : on parle alors d'*onde sphérique*.

1.1.3 Equation d'onde non homogène

Nous venons de rappeler la résolution de l'équation d'onde homogène. Les solutions obtenues nous seront nécessaires dans la suite de ce manuscrit afin de décrire convenablement les effets de la propagation et éventuellement des réflexions acoustiques en présence d'obstacles sur le trajet de l'onde. Néanmoins, il est fondamental de traiter le cas de l'équation d'onde non homogène, qui suppose la présence d'une source acoustique dans l'environnement et permet de décrire les variations du potentiel des vitesses en fonction de l'excitation de la source. Dans ce cas, l'équation d'onde (1.26) est modifiée de façon à faire apparaître le signal sonore $s(\mathbf{r}, t)$ de la source et s'écrit maintenant [Ziomek 1995]

$$\nabla^2 \Phi(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \Phi(\mathbf{r}, t) = s(\mathbf{r}, t). \quad (1.48)$$

De façon à respecter les hypothèses de l'acoustique linéaire, nous supposons que la source produit un signal acoustique de faible amplitude. La solution à (1.48) peut être obtenue en traitant le milieu de propagation de manière analogue à un $\Phi(\mathbf{r}, t)$. Ainsi, il est possible d'écrire que

$$\Phi(\mathbf{r}, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\mathbf{r}_0, t_0) h(\mathbf{r}, t; \mathbf{r}_0, t_0) dt_0 d\mathbf{r}_0, \quad (1.49)$$

où $h(\mathbf{r}, t; \mathbf{r}_0, t_0)$ est la *réponse impulsionnelle* du milieu de propagation. Nous pouvons montrer via l'application des transformées de Fourier temporelle puis spatiale aux deux membres de l'équation (1.48), que

$$h(\mathbf{r}, t; \mathbf{r}_0, t_0) = -\frac{1}{4\pi \|\mathbf{r} - \mathbf{r}_0\|} \delta\left(t - \left(t_0 + \frac{\|\mathbf{r} - \mathbf{r}_0\|}{c}\right)\right), \quad (1.50)$$

ce qui permet d'écrire que

$$\Phi(\mathbf{r}, t) = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{1}{\|\mathbf{r} - \mathbf{r}_0\|} s(\mathbf{r}_0, t - \frac{\|\mathbf{r} - \mathbf{r}_0\|}{c}) d\mathbf{r}_0. \quad (1.51)$$

Considérons maintenant une source ponctuelle et omnidirectionnelle, située en $\mathbf{r} = \mathbf{r}_s$, possédant une variation temporelle $b(t)$. Alors,

$$s(\mathbf{r}, t) = b(t) \delta(\mathbf{r} - \mathbf{r}_s),$$

et nous pouvons déduire de (1.51) et (1.50) que

$$\Phi(\mathbf{r}, t) = -\frac{1}{4\pi \|\mathbf{r} - \mathbf{r}_s\|} b(t - \tau), \text{ avec } \tau = \frac{\|\mathbf{r} - \mathbf{r}_s\|}{c}, \quad (1.52)$$

où τ désigne le retard lié à la propagation. L'équation (1.52) indique que l'amplitude de l'onde sonore décroît de manière inversement proportionnelle à la distance entre la source et le point considéré. Cette propriété sera utilisée dans la suite afin d'exprimer le transfert entre une source

et un microphone. Dans le cas particulier d'une source monochromatique, avec $b(t) = e^{jkt}$, on obtient

$$\Phi(\mathbf{r}, t) = -\frac{1}{4\pi\|\mathbf{r} - \mathbf{r}_s\|} e^{-jk\|\mathbf{r} - \mathbf{r}_s\|} e^{jkt}, \quad (1.53)$$

correspondant à l'expression d'une onde harmonique sphérique.

1.2 Le système auditif humain

Nous venons de rappeler les bases théoriques de l'acoustique linéaire. Ces premières équations fondamentales nous sont nécessaires afin de modéliser le phénomène de propagation entre une source et un récepteur. Ce récepteur, en robotique, est un microphone dont le rôle est de transformer une variation de pression acoustique en un signal électrique. Ce signal peut être mis en forme, converti numériquement, et traité par le système de localisation artificiel embarqué sur la plateforme mobile. Comme de nombreuses méthodes en robotique s'inspirent du fonctionnement de l'oreille humaine, il semble intéressant de s'intéresser un instant au système auditif humain. C'est l'étude que nous proposons dans la section suivante. Pour cela, nous débutons cette section par une description physiologique de l'anatomie de l'oreille. Puis, nous proposons dans une seconde partie une étude des indices auditifs utilisés par le cerveau nous permettant, dans la vie de tous les jours, de localiser un son dans notre environnement. Disposant ainsi des connaissances de base en acoustique et sur le fonctionnement de l'oreille humaine, nous pourrions alors, dans le chapitre 2, étudier en détail les méthodes de localisation de source sonore utilisées en robotique.

1.2.1 Aspects physiologiques

Le système auditif a pour but de transformer une onde sonore complexe en une suite de configurations d'activités nerveuses transmises sous forme de potentiels au cerveau. Celui-ci se charge ensuite de les intégrer aux autres données sensorielles pour guider notre comportement ou nos réactions. L'oreille est donc un transducteur mécanique, et ce sont les différentes étapes de cette conversion physique que nous proposons d'étudier dans cette partie. L'oreille humaine peut être scindée en trois parties, chacune ayant un rôle distinct lors de la conversion de l'information de pression en une information électrique. On distingue ainsi *l'oreille externe*, *l'oreille moyenne* et *l'oreille interne*. Une coupe de l'oreille proposée sur la figure 1.4 indique la position de chacun de ces éléments. Selon les individus, et selon leur âge (les enfants, en particulier, sont sensibles à des sons de plus hautes fréquences que les adultes), la gamme des fréquences audibles par notre oreille s'étend d'environ 20 Hz à 20 kHz. Extrêmement sensible, notre oreille nous rend capables de distinguer une variation de pression infime causée par une source sonore de 1W à 3 kHz située à une distance d'environ 300km ... mais dans un monde idéal complètement silencieux [Purves 2004]. Ce sont ces caractéristiques incroyables pour un récepteur de si petite taille que nous allons tenter d'expliquer en détaillant dans les 3 parties suivantes le fonctionnement des oreilles externe, moyenne et interne.

1.2.1.1 L'oreille externe

La partie principale et visible de l'oreille externe est le *pavillon*, formant un angle moyen de 30° avec la tête. Sa partie centrale, appelée la *conque*, se prolonge par le *conduit auditif* externe, lequel admet une longueur de 2 à 3 cm pour un diamètre moyen de 0.7 cm (voir figure 1.5(a)). Le

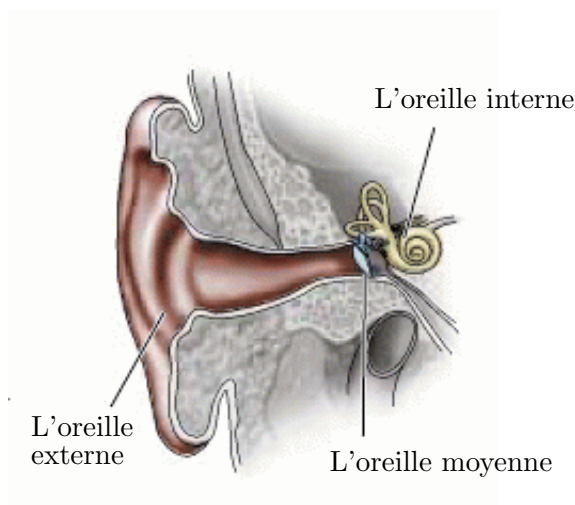


FIG. 1.4 – [Purves 2004] - Aperçu général de l'oreille humaine

rôle de l'oreille externe est de prélever l'information sonore de pression et de la conduire jusqu'à l'oreille moyenne, plus précisément jusqu'au *tympan*. Il est possible de montrer que la forme particulière de l'oreille externe joue le rôle d'amplificateur sélectif des variations de pression sonore. En effet, la conque et le conduit auditif externe peuvent être assimilés à une cavité résonante, dont la fréquence de résonance se situe aux environs de 3 kHz. Ainsi, les fréquences du signal acoustique situées autour de cette fréquence peuvent se trouver amplifiées d'un facteur pouvant aller jusqu'à 20 dB. Un deuxième rôle important du pavillon et de la conque est le filtrage spatial de l'information de pression, filtrage qui nous permet d'obtenir de précieuses informations d'élévation sur la position d'une source sonore. En effet, les formes alambiquées du pavillon sont à l'origine de combinaisons constructives ou destructives des certaines fréquences selon leur provenance spatiale. En pratique, il apparaît que l'oreille externe transmet davantage de hautes fréquences pour une source émettant en hauteur qu'une source au niveau de l'oreille. Notons par ailleurs que chez de nombreux mammifères la focalisation des sons de hautes fréquences est facilitée par la mobilité de leur oreille externe. Pour suppléer à ce manque de mobilité, nous effectuons des rotations inconscientes de la tête dans le sens le plus favorable à la pénétration des ondes sonores.

1.2.1.2 L'oreille moyenne

Une fois filtrée et amplifiée par l'oreille externe, l'onde de pression atteint le *tympan*. Cette membrane, de 10 mm de diamètre pour 0.1 mm d'épaisseur, a pour rôle de transmettre les fluctuations de pression à l'oreille interne via un système mécanique, appelé *chaîne des osselets*. On y distingue le *marteau*, fixé par son manche sur le versant supérieur du tympan, l'*enclume*, étroitement emboîtée avec le marteau, et l'*étrier*, dont l'extrémité est en contact avec l'oreille interne au niveau de la *fenêtre ovale* (voir figure 1.5(b)). L'étrier est le plus petit os de l'organisme, et pèse environ 3 mg pour une taille de 3 mm. Le rôle principal de l'oreille moyenne est d'adapter l'impédance mécanique relativement basse du milieu sonore aérien avec l'impédance plus élevée du liquide contenu dans l'oreille interne, le *pérylymphe*. En effet, sans cette adaptation d'impédance, presque toute l'énergie acoustique serait réfléchi au lieu d'être transmise à l'oreille interne. Ainsi, ce système mécanique permet de récupérer la variation de pression faisant vibrer

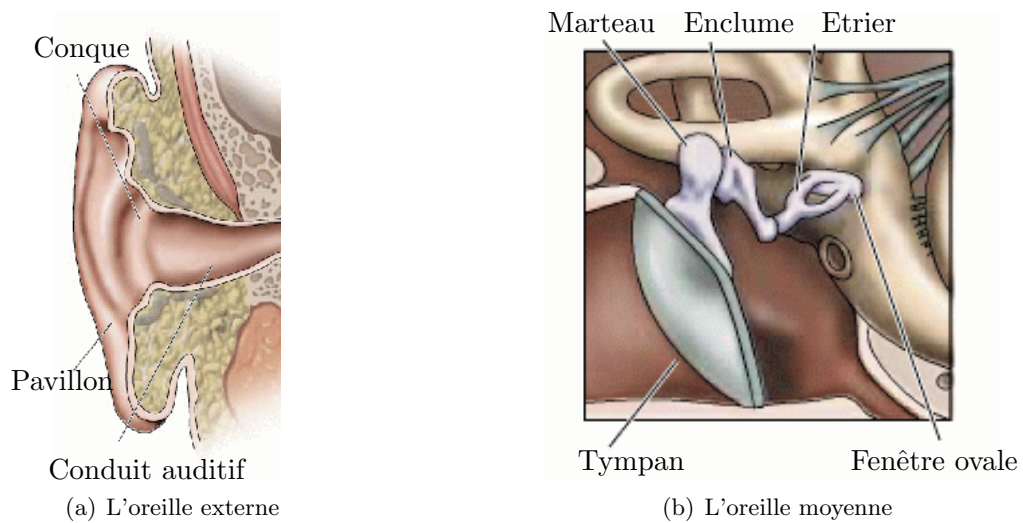


FIG. 1.5 – [Purves 2004] - Détails de l'oreille externe et moyenne

le tympan en l'amplifiant à un niveau près de 200 fois supérieur. Pour information, au seuil auditif et pour une fréquence de 1 kHz, l'amplitude du mouvement du marteau atteint seulement environ $2 \cdot 10^{-10}$ m. L'amplification est obtenue par la combinaison de deux phénomènes : la chaîne des osselets est à l'origine d'un effet de levier, ensuite combiné à une amplification des pressions, due au fait que la pression appliquée à la surface relativement étendue de la membrane tympanique est reportée sur la surface beaucoup plus petite de la fenêtre ovale, située sur l'oreille interne. Evidemment, le couplage présenté ici est dans la réalité fonction de la fréquence, si bien qu'on peut modéliser l'oreille moyenne par un filtre fournissant une amplification maximale autour de 1 kHz.

1.2.1.3 L'oreille interne

Si les deux premiers traitements effectués par l'oreille peuvent se résumer à une amplification sélective des fréquences suivie d'une adaptation d'impédance, l'oreille interne a un rôle beaucoup plus central dans le traitement des signaux acoustiques que nous percevons. En effet, c'est ici que la véritable transduction mécanique s'effectue, permettant de transformer une onde de pression en potentiels transmis par le *nerf auditif* jusqu'au cerveau. Cette conversion des grandeurs physiques s'effectue au sein de la *cochlée*, aussi appelée *limaçon*, sorte de tube tronconique de 35 mm de long enroulé autour d'un axe creux sur près de 3 tours (voir figure 1.6). La cochlée est reliée à l'oreille moyenne au niveau de la *fenêtre ovale*, qui communique son mouvement induit par les osselets à un liquide appelé *périmylphe* contenu dans la *rampe vestibulaire*, séparée de la *rampe tympanique* par le *canal cochléaire*. La membrane située entre le canal cochléaire et la rampe tympanique, appelée *membrane basilaire*, joue un rôle fondamental dans le traitement du son effectué par l'oreille interne, puisque c'est sur elle que repose l'*organe de Corti* dont les *cellules ciliées* internes assurent la transduction de la stimulation mécanique en phénomène bio-électrique. La membrane basilaire possède une structure beaucoup plus rigide et étroite à sa base qu'à son extrémité. De ce fait, des études ont montré qu'elle vibre à des endroits distincts en réponse à des sons de fréquences différentes. En particulier, les points de la membrane qui répondent aux fréquences élevées sont à sa base, tandis que ceux qui répondent aux basses

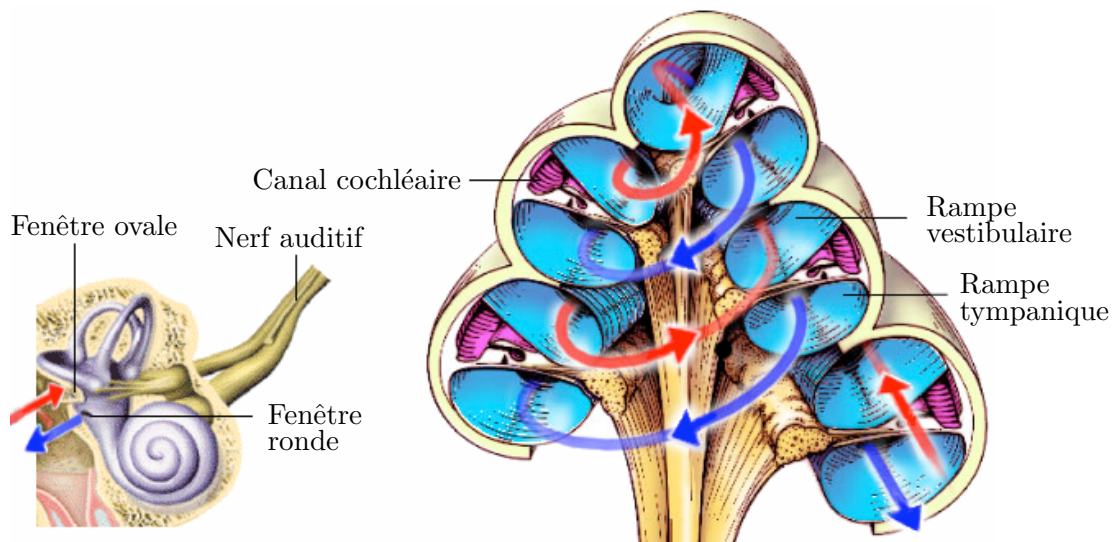


FIG. 1.6 – [Purves 2004] - L'oreille interne (gauche). Détails de la cochlée (droite) ; la flèche rouge provient de la fenêtrée ovale, la flèche bleue aboutit à la fenêtrée ronde

fréquences se situent à son extrémité. Ainsi, l'information fréquentielle est codée de manière géométrique à l'intérieur de la cochlée, ce qui constitue une *représentation tonotopique* des fréquences (voir figure 1.7). Cette caractéristique fondamentale prévoit ainsi qu'un son complexe détermine un profil de vibration de la membrane équivalent à la superposition des vibrations causées par les fréquences individuelles dont il est composé. La membrane basilaire joue donc le rôle de "séparateur fréquentiel" de l'onde sonore.

La propagation de l'onde déclenche donc la transduction sensorielle en faisant bouger les cellules ciliées de l'organe de Corti disposé sur la membrane basilaire. Ce mouvement courbe de minuscules prolongements des cellules ciliées, appelés *stéréocils*, et entraîne la modification du potentiel de membrane de celles-ci par échanges ioniques. L'échelle à laquelle travaillent les cellules ciliées est étonnante et leur capacité d'adaptation à un stimulus persistant nous rend capable d'extraire des signaux acoustiques du bruit ambiant. Pour autant, aussi rapides soient-elles, les cellules ciliées ne sont pas capables d'osciller à des fréquences supérieures à

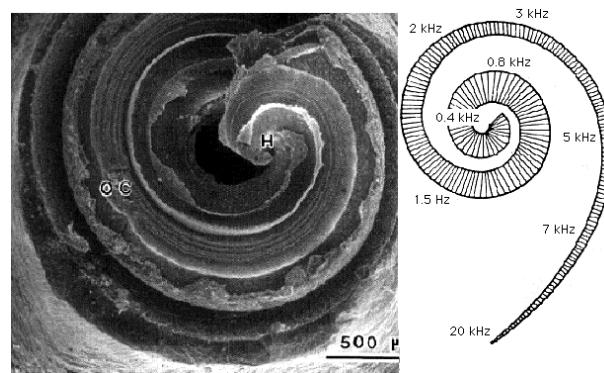


FIG. 1.7 – [Pujol 2006] - Représentation tonotopique des fréquences dans la cochlée

environ 3 kHz. Pour des fréquences supérieures, c'est le codage tonotopique de l'information fréquentielle qui permet le traitement par un mécanisme de "lignes dédiées". Les fréquences inférieures produisent quant à elles un potentiel impulsionnel important, résultat d'une décharge de neurotransmetteur ayant lieu uniquement durant les phases positives des sons. Ce phénomène, valable seulement pour les basses fréquences, peut être vu comme une sorte de *verrouillage de phase* fournissant des informations *temporelles*, et non plus fréquentielles. La comparaison de ces informations temporelles relatives aux deux oreilles permet la localisation ultérieure d'un signal sonore.

1.2.2 La localisation du son chez l'homme

Une onde sonore générée par une source externe est modifiée par la présence de notre tête préalablement à son interaction avec notre pavillon. Les diffractions résultantes viennent alors modifier les caractéristiques de l'onde acoustique perçue. Ces modifications sont à l'origine de nos capacités à localiser un son et fournissent des *indices acoustiques* qui sont extraits par notre oreille interne, puis intégrés par notre cerveau. L'objectif de cette partie est de caractériser ces indices acoustiques et de présenter leur rôle dans le processus de localisation. Cette étude est nécessaire car comme nous le verrons dans le chapitre suivant, de nombreuses méthodes de localisation de source sonore utilisées en robotique cherchent à utiliser le même type d'indices acoustiques. La localisation du son chez l'homme a fait l'objet de très nombreuses études dans le domaine des neurosciences. En particulier, les physiologistes et anatomistes ont accompli de grands progrès durant les années 1980 à 1990 dans la compréhension des mécanismes utilisés par le système nerveux pour l'intégration des indices sonores [Middlebrooks 1991]. Ce sont les résultats de leurs travaux que nous résumons ici.

Le système auditif humain possède une précision étonnante. Nous sommes capables de localiser une source sonore avec une précision d'environ 3° en azimuth comme en élévation. Evidemment, cet ordre de grandeur reste une moyenne et varie selon les individus, mais une constatation reste valable pour tous : nous localisons de manière très précise les sons en provenance de l'avant, et l'erreur de localisation augmente sensiblement dans les directions périphériques. Ainsi, notre précision devient très relative pour des sons provenant de l'arrière de notre tête, puisque ceux-ci peuvent être détectés avec une erreur de 20° . D'ailleurs, il semble que nos capacités de discrimination entre deux sources situées en face et/ou derrière nous soient assez limitées, en particulier pour des sons de faibles largeurs de bande. Heureusement, ce type de sons est très rare dans la nature où la quasi-totalité des stimuli sonores est large bande.

1.2.2.1 Localisation horizontale

Différence interaurale en amplitude - Historiquement, la découverte du premier indice acoustique utilisé pour la localisation horizontale est attribuée à *Lord Rayleigh*, qui comprit il y a un peu plus d'un siècle qu'un son émis depuis la droite d'un auditeur atteint son oreille gauche en étant modifié par la présence de sa tête. Il s'en suit que le signal acoustique perçu par l'oreille droite est plus intense que celui relevé par l'oreille gauche. Aujourd'hui, la différence des intensités des signaux, exprimés en dB, et perçue par les oreilles droite et gauche est appelée *différence interaurale en amplitude* (Interaural Level Difference ou ILD en anglais). La figure 1.8 montre que l'ILD est une fonction de la fréquence sur la totalité du spectre audible. Nous constatons aussi la présence de deux domaines particuliers, pour lesquels l'ILD est soit négligeable, soit au contraire très importante. En effet, pour des longueurs d'onde supérieures au diamètre de la tête, la diffraction de celle-ci devient négligeable et rend l'ILD quasi-nulle : nos oreilles

perçoivent alors le même niveau acoustique. Au contraire, si la source émet avec une fréquence supérieure à environ 750 Hz, ces mêmes diffractions viennent modifier sensiblement les niveaux acoustiques relevés par les deux oreilles. La différence en amplitude peut alors atteindre plus de 30 dB. Evidemment, le modèle utilisé pour représenter les variations de l'ILD sur la figure

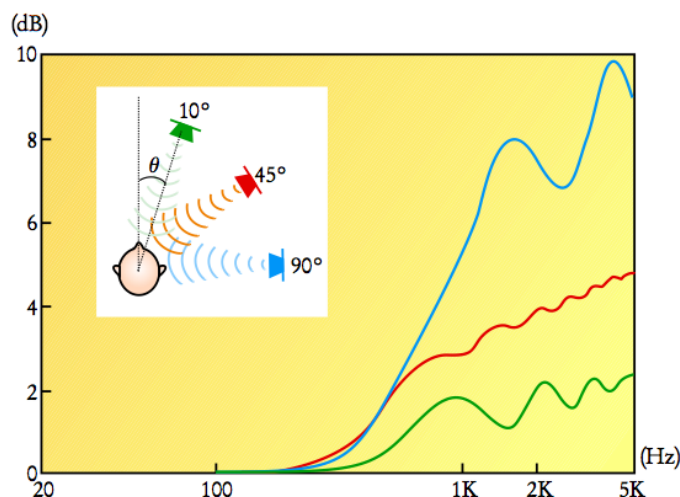


FIG. 1.8 – [Hartmann 1999] - Différences interaurales en amplitude, calculées pour une source dans le plan azimutal défini par les deux oreilles, pour des directions d'incidence égales à 10° (courbe verte), 45° (rouge) ou 90° (bleue), en supposant que la tête se réduit à une sphère rigide.

1.8 est relativement simpliste, puisque la tête est assimilée à une simple sphère rigide munie de deux oreilles sur le plan azimutal. Ainsi, pour les hautes fréquences, n'importe quel détail de petite taille du visage, non pris en compte jusqu'ici, fait office de "perturbateur" de l'onde acoustique et vient modifier significativement l'allure de cette courbe simplifiée. Nous verrons d'ailleurs par la suite comment ces variations sont utilisées pour la localisation verticale du son. Cependant, il faut alors nécessairement que ces modifications de l'ILD puissent être détectées par le système nerveux central. En fait, il apparaît que la plus petite modification détectable des rapports d'amplitude est de l'ordre de 1 dB [Mills 1960], et ce quelle que soit la fréquence. Ainsi, l'ILD est un indice potentiel de localisation pour toutes les fréquences. Cependant, nous avons vu que l'ILD admet de très faibles valeurs, inférieures au seuil détectable, pour les plus faibles d'entre elles. Pourtant, nous sommes tout à fait capables de localiser un son composé de grandes longueurs d'onde. Ceci implique donc qu'un autre indice sonore est utilisé pour traiter le cas des basses fréquences.

Différence interaurale en phase - C'est encore Lord Rayleigh qui conclut en 1907 que notre système auditif est aussi capable de détecter une différence en phase entre nos deux oreilles. Ce nouvel indice acoustique est appelé de nos jours *différence interaurale en phase* (Interaural Phase Difference ou IPD en anglais), ou parfois *différence interaurale temporelle* (Interaural Time Difference ou ITD) lorsqu'on considère un signal pluri-chromatique. Son existence est très simplement justifiée par la différence de chemin que doit parcourir l'onde pour parvenir à nos deux oreilles. En pratique, il apparaît que nous sommes sensibles à un large plage de valeurs d'ITD. Cependant, la valeur maximale effectivement utilisée pour la localisation est de l'ordre de 700 μ s, ce qui correspond à la durée de la période d'un son de 1400 Hz, tandis que nous sommes

capables de discerner une différence d'environ 2° en azimuth correspondant à un ITD d'environ $25\mu s$. Malgré la présence de délais synaptiques importants (environ 1ms), nous sommes donc capables de traiter des informations sonores bien plus rapides : le parallélisme du traitement effectué dans le système nerveux central, réalisant des opérations analogues à des lignes à retards et des opérations de corrélations, permet de traiter ce problème [Hartmann 1999]. La limite d'utilisation de l'ITD, située donc aux environs de 1400 Hz, est à mettre en parallèle avec le fonctionnement des cellules ciliées de l'organe de Corti, jouant le rôle de "boucle à verrouillage de phase" pour les faibles fréquences. Notons que physiologiquement, c'est bien l'IPD et non l'ITD qui intervient dans la mesure des décalages liés à la propagation entre les deux oreilles puisque la membrane basilaire de la cochlée joue le rôle de "séparateur de fréquences".

En résumé - Il existe donc deux domaines de fréquences pour lesquelles deux types d'indices acoustiques différents sont utilisés. Les fréquences inférieures à environ 1 kHz sont localisées à l'aide de l'IPD tandis que les fréquences supérieures à environ 3 kHz exploitent l'ILD. Entre ces deux limites, des tests acoustiques ont effectivement montré une perte de précision dans la localisation, aucun indice sonore n'étant réellement disponible. Une autre constatation a été effectuée sur des personnes équipées de casques, par la modification indépendante de chacun des indices sonores : l'ILD et l'IPD, s'ils permettent effectivement de décider de la présence d'une source à droite ou à gauche, ne produisent pas d'"externalisation" de la perception ; les auditeurs perçoivent la source dans la bonne direction, mais à l'intérieur de leur tête. Enfin, il est simple de constater qu'une source se situant dans le plan de symétrie verticale de la tête n'est à l'origine d'aucune différence interaurale. Pourtant, nous sommes bien capables de réaliser la localisation dans ces conditions ; c'est donc qu'un autre processus permet de déterminer l'élévation d'une source sonore.

1.2.2.2 Localisation verticale

Les ondes sonores qui proviennent de différentes positions dans l'espace sont modifiées par la présence d'obstacles dans une scène, y compris par notre propre corps, nos épaules, notre tête, notre oreille externe, etc., qui jouent tous le rôle de réflecteurs venant modifier leurs caractéristiques fréquentielles. Cet effet est tout à fait assimilable à une opération de filtrage, et est modélisé par la *fonction de transfert de la tête* (Head-Related Transfer Function ou HRTF) ou *fonction de transfert anatomique*. La HRTF joue un rôle central dans notre capacité à localiser en élévation une source sonore. En effet, la somme combinée de toutes les réflexions ayant lieu autour de la tête conduit à la création de zones amplifiées ou coupées dans le spectre de la source, comme cela est illustré sur la figure 1.9 pour trois positions différentes d'une même source sonore dans l'espace. L'indice acoustique utilisé pour la localisation verticale est donc un indice spectral, aussi appelé *indice monaural* puisqu'il est obtenu sans comparaison entre les signaux reçus par les deux oreilles. L'examen attentif de fonctions de transfert mesurées de l'oreille externe montre des particularités spectrales (zones coupées ou amplifiées) qui varient systématiquement avec l'emplacement de la source. En particulier, il semble que l'élévation de celle-ci vienne influencer significativement la position de coupures pointues (notches) dans le spectre du signal reçu [Humanski 1988]. Ce type de caractéristique est très probablement utilisé par le cerveau pour la détermination de l'élévation. Mais dès lors, comment décider si un indice spectral est issu d'une modification induite par la HRTF ou si au contraire, cette caractéristique était déjà présente dans la source ? Il semble que nous ne soyons pas capables d'effectuer cette discrimination. Des tests d'écoute ont prouvé que nous sommes plus sensibles au contenu fréquentiel émis par la source qu'à la position depuis laquelle elle émet [Butler 1983]. Si nous demandons ainsi à un

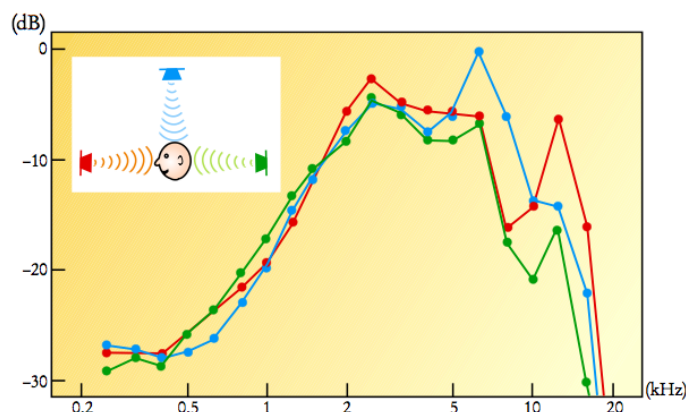


FIG. 1.9 – [Hartmann 1999] - Fonction de transfert de la tête, représentée en fonction de la fréquence pour différentes positions de la source dans le plan vertical, mesurée sur le mannequin KEMAR de Knowles Electronics.

sujet de décider d'où peut provenir les sons qu'il perçoit, sa réponse dépendra uniquement de la fréquence d'émission, et non de la position depuis laquelle ils sont effectivement émis. Enfin, si les indices spectraux nous permettent de déterminer l'élévation d'une source sonore, ils nous permettent aussi d'externaliser nos perceptions. Des essais ont en effet montré que simuler ces indices spectraux est *suffisant* pour ne plus entendre une source sonore comme émettant depuis l'intérieur de notre tête.

1.2.2.3 Perception de la distance et influence de l'environnement

L'ensemble des indices auditifs que nous venons de présenter permet de déterminer très précisément l'azimut et l'élévation d'un son, mais en aucun cas la distance entre nos oreilles et la source sonore. Relativement peu de recherches existent encore sur le sujet, mais il semble que notre perception de la distance ne soit pas très bonne car principalement basée sur le niveau sonore émis par la source. Ceci est confirmé par des tests effectués en chambre anéchoïque, pour des distances supérieures à 3 m [Strybel 1984]. Pour des sources proches, les variations de volumes nécessaires pour obtenir une discrimination de la distance sont encore plus importantes. D'autres observations laissent cependant penser que notre estimation de la distance est meilleure dans un espace non anéchoïque. La différence d'amplitudes entre l'onde sonore reçue directement et celle ayant subi des réflexions peut constituer un bon indicateur sur la distance à la source. Évidemment, de telles constatations sont très dépendantes de l'environnement et il est aujourd'hui difficile de généraliser cette propriété à notre perception en environnement réel : aucune étude n'a encore prouvé l'influence du rapport entre les amplitudes des ondes directes et réfléchies sur notre jugement de la distance. Plus généralement encore, l'ensemble des réflexions que subit une onde sonore dans un environnement réel vient perturber les autres indices acoustiques IPD/ITD et ILD. Ces modifications dépendent totalement de l'environnement et il est donc, de fait, très difficile de les prévoir. Il s'avère que cette remarque est particulièrement vérifiée pour l'ITD, très sensible aux conditions environnementales puisque basée sur la cohérence des signaux perçus par les deux oreilles. Dans de grandes pièces, l'importance des ondes réfléchies est telle que l'ITD devient même très difficilement mesurable. Au contraire, la mesure de l'ILD reste possible dans de telles conditions, notamment du fait que l'absorption acoustique d'une grande majorité de matériaux réfléchissants augmente avec la fréquence. Comme l'ILD est précisément l'indice

acoustique adapté aux hautes fréquences, la réflexion moindre des ondes rend la localisation de source envisageable. Des essais acoustiques ont d'ailleurs montré qu'il était probable, dans des environnements fortement réverbérants, que notre localisation soit uniquement basée sur l'ILD, plus robuste à de telles situations. Evidemment, cette stratégie reste inconsciente et témoigne de nos capacités importantes d'adaptation à notre environnement sonore.

1.2.2.4 L'effet de précedence

Dans un environnement réel, un son atteint nos oreilles en parcourant plusieurs chemins. Pour autant, malgré la présence de multiples réflexions, nous sommes tout à fait capables sous certaines conditions de localiser l'origine d'une source sonore. La première onde arrivée à nos oreilles joue donc un rôle fondamental dans notre perception, expliqué par l'*effet de précedence* ou *effet de Haas*. Par exemple, il est extrêmement délicat de localiser la source d'un son mono-fréquentiel établi dans une pièce. Par contre, il suffit de stopper ce son et de le ré-émettre pour pouvoir préciser son origine. Des tests effectués en chambre sourde avec deux haut-parleurs, le premier simulant le chemin direct et le second un écho, ont permis de mettre en évidence que tout se passe comme si un écho parvenant à nos oreilles dans un délais compris entre 1ms et 5ms n'était pas perçu. Pour des délais supérieurs, les deux sons sont distingués comme admettant deux origines distinctes, tandis que pour les délais inférieurs à 1ms, les deux sources sont fusionnées en une seule située entre les deux haut-parleurs (figure 1.10). L'origine de cet effet

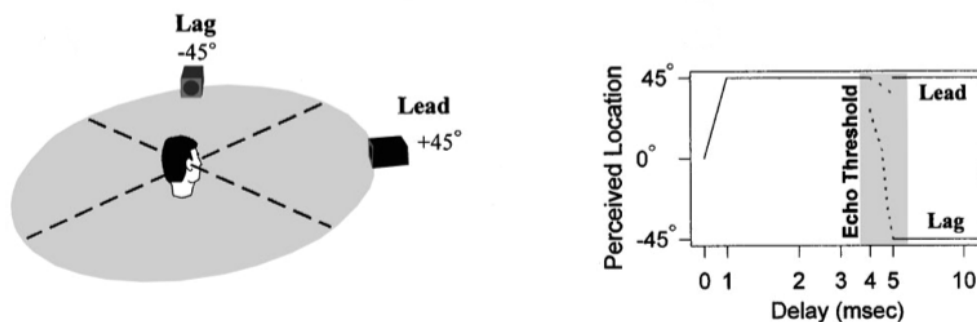


FIG. 1.10 – [Litovsky 1999] - Illustration de l'effet de précedence. Deux haut-parleurs situés en $\pm 45^\circ$ émettent un son court. Le signal émis par le haut-parleur "Lag" est retardé d'un délai variable par rapport au second haut-parleur. La figure de droite indique les directions perçues en fonction du décalage entre ces deux sources.

peut s'expliquer par la présence de sortes de "barrières" neuronales, dont le rôle est de stocker les informations utiles pour la localisation pendant une durée d'environ 1ms. Les indices auditifs parvenant à l'oreille dans un délai supérieur ne sont alors pas pris en compte. Un autre modèle propose d'expliquer ce phénomène par une pondération forte des indices auditifs permettant de mettre en avant les sons arrivant en premier à l'oreille. De cette façon, même les échos perçus tardivement peuvent être intégrés, mais avec moins d'importance. Ceci expliquerait d'ailleurs certaines observations montrant que l'ensemble du signal est malgré tout pris en compte dans le processus de localisation.

1.3 Conclusion

Ce chapitre d'introduction a été pour nous l'occasion d'introduire l'ensemble des notations et notions nécessaires pour la suite de ce manuscrit. Nous y avons tout d'abord caractérisé la propagation d'une onde sonore dans le cadre de l'acoustique linéaire. Ces mises en équation aboutissent à considérer deux types de fronts d'ondes, plans ou sphériques. Puis nous nous sommes attachés d'une part à décrire le fonctionnement mécanique de l'oreille, et d'autre part à présenter les indices acoustiques exploités par le cerveau pour la localisation des sons. Nous proposons maintenant d'étudier comment la communauté robotique a su s'inspirer de l'ensemble de ces phénomènes pour implémenter certaines fonctionnalités auditives aux plateformes mobiles.

Chapitre 2

Les méthodes de localisation de source sonore en robotique : un état des lieux

Nous venons de voir dans le chapitre précédent comment le cerveau humain est capable d'intégrer les indices sonores extraits des signaux reçus par nos deux oreilles pour déterminer l'origine spatiale d'une stimulation sonore. Si l'on souhaite reproduire artificiellement cette fonction pour une utilisation en robotique, la prise en compte *nécessaire* des diffractions et modifications spectrales du signal sonore pose des problèmes particulièrement difficiles à résoudre lors de l'exploitation de seulement deux microphones. Dans une telle configuration, la localisation ne peut souvent s'effectuer qu'en azimut, avec l'impossibilité de distinguer si une source émet de l'avant ou de l'arrière du capteur. Toutefois, le fait de pouvoir limiter le système de perception à deux microphones est d'un intérêt certain en robotique ; le capteur est alors de petite taille et ne nécessite pas de chaîne d'acquisition lourde. C'est probablement pour cette raison qu'une part importante des chercheurs de la communauté s'est attachée à développer des algorithmes de localisation basés sur l'utilisation de deux microphones, et/ou fortement inspirés du fonctionnement de notre oreille. Nous présentons l'ensemble de ces techniques, dites *bio-inspirées*, dans une première partie de ce chapitre.

Le fait que l'homme et de nombreux animaux soient capables de percevoir leur environnement sonore avec seulement deux oreilles a motivé le développement de capteurs à deux microphones dans le contexte de la robotique. Il est toutefois évident que l'utilisation de plusieurs microphones est susceptible de fournir des informations redondantes pouvant permettre une amélioration de la précision de la localisation ou de la robustesse des algorithmes. C'est ce principe qui a été largement retenu en Acoustique et Traitement du Signal, dans le domaine de l'*antennerie*. Si les méthodes biomimétiques ont longtemps pris le pas sur les techniques d'antennerie, l'utilisation de ces dernières et leur adaptation au contexte de la robotique est particulièrement d'actualité. Ainsi, la seconde partie de ce chapitre est dédiée à la présentation de méthodes de localisation en robotique basées sur des antennes de microphones.

2.1 Les méthodes biomimétiques

Avant d'entrer dans le détail de ces méthodes, il convient de préciser le sens du mot "biomimétique" en robotique. On retrouve ainsi les techniques basées sur l'utilisation de seulement deux microphones : le biomimétisme est ici "géométrique", puisque le capteur est volontairement

limité de façon à reproduire extérieurement l'apparence du système auditif humain. D'autres méthodes proposent de travailler avec les mêmes indices sonores que ceux manipulés par notre cerveau pour la localisation. Ici, le biomimétisme devient "algorithmique", le nombre de microphones employés pouvant être supérieur à deux. En robotique, très peu de méthodes à deux microphones sont basées sur l'extraction des indices auditifs habituels. En pratique, le traitement extrêmement complexe effectué par le cerveau sur les informations de pression issues de nos deux oreilles s'avère très difficile à reproduire, si bien que la majorité des méthodes que nous allons présenter exploite plus de deux microphones. De façon à mettre en avant le parallèle existant entre le fonctionnement de notre appareil auditif et les méthodes biomimétiques proposées, nous proposons dans cette section d'étudier successivement comment chaque indice auditif a déjà été exploité en robotique.

2.1.1 Les indices interauraux

L'exploitation des indices interauraux pour la localisation en azimuth a été étudiée en détail par l'équipe du professeur Okuno de l'université de Tokyo. De nombreux résultats expérimentaux ont été proposés depuis 1999, obtenus sur le robot humanoïde SIG et SIG2 dont la structure est présentée figure 2.1. Celui-ci fait office de banc d'essai pour l'intégration multi-sensorielle. Il est basé sur une mécanique fournissant quatre degrés de liberté contrôlés par quatre moteurs à courant continu. Une paire de caméras CCD, chacune pourvue de trois degrés de liberté (pan, tilt et zoom), est embarquée sur la plate-forme. SIG est également équipé de deux paires de

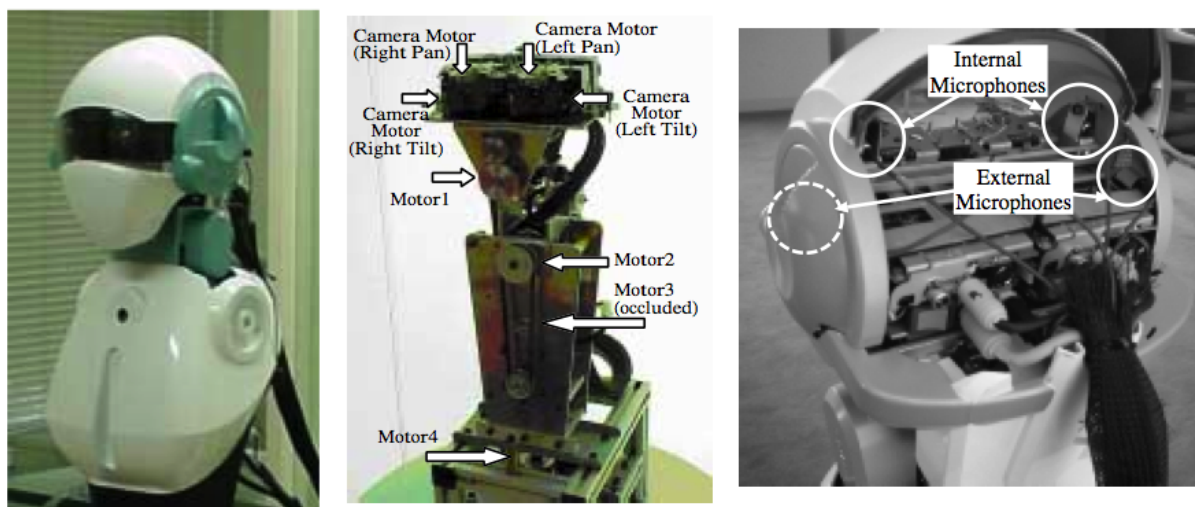


FIG. 2.1 – [Nakadai 2000a] - A gauche : l'humanoïde SIG (université de Tokyo et Japan Science and Technology Corp.). Au milieu : sa structure mécanique interne. A droite : détails montrant les microphones intégrés à l'intérieur de la tête et le placement des caméras.

microphones omnidirectionnels : l'une est placée à l'intérieur de la tête, de façon à enregistrer les sons internes au robot provoqués par les moteurs, l'autre se situe à l'extérieur pour la perception extéroceptive. L'ensemble de la mécanique est caché par une carcasse externe jouant le rôle de tête et visant à isoler, dans une certaine mesure, les bruits des moteurs émis à l'intérieur du robot [Nakadai 2000b]. Évidemment, cette isolation phonique n'est pas parfaite ; un filtre adaptatif exploitant les sorties des deux microphones internes permet de supprimer le bruit des

moteurs dans les signaux perçus à l'extérieur du robot. Il devient donc envisageable d'écouter *pendant* un mouvement, y compris si celui-ci s'avère bruyant : on parle alors d'*audition active* [Nakadai 2000a]. Cette possibilité est originale, puisque la plupart des méthodes précédentes exploitant des données sonores étaient basées sur une approche du type *Arrêt-Ecoute-Action*. Dans ce cadre, [Huang 1997a] propose de comparer la puissance des signaux de sortie de trois microphones placés en triangle afin de déterminer l'azimut de la source. Toujours pendant des phases sans mouvement, le robot humanoïde de l'université de Waseda exploite la fonction de transfert de la tête (Head Related Transfer Function, ou HRTF) en environnement contrôlé pour localiser un unique locuteur [Matsusaka 1999], tandis que l'humanoïde COG du MIT est doté de deux micros embarqués dans un pavillon simplifié et utilise un apprentissage visuo/auditif pour la localisation de source sonore [Brooks 1999].

L'objectif de toutes ces méthodes est d'aboutir aux mêmes performances que le système auditif humain en terme d'*extraction*, *reconnaissance* et *localisation*. L'ensemble de ces trois fonctionnalités élémentaires nous permet en effet d'analyser les données sonores en provenance du monde extérieur. En robotique, *l'analyse de scène auditive* (Computational Auditory Scene Analysis, ou CASA) fournit un cadre de travail pour l'étude de ces phénomènes [Brown 1993]. Son objectif est la compréhension d'une mixture de sons (extraction) pouvant être constituée de voix, de bruit, de musique, etc. Une telle analyse nécessite non seulement l'interprétation de la signification d'une donnée sonore spécifique (reconnaissance), mais aussi la détermination des relations géométriques existant entre les sources (localisation). Dès lors, être capable d'écouter en mouvement peut permettre de compléter les autres modalités extéroceptives "habituelles" pour obtenir d'avantage d'informations sur la scène sonore.

De tous les indices sonores extraits à partir de deux microphones, les indices interauraux demeurent les plus exploités en robotique. Les mesures expérimentales des différences interaurales en phase (IPD) $\Delta\varphi_{exp}$ ou/et en amplitude (ILD) $\Delta\rho_{exp}$ permettent en effet d'aboutir, à partir de seulement deux microphones et de calculs simples, à une localisation en azimut. Notons $S_{P_g}(f)$ et $S_{P_d}(f)$ les transformées de Fourier des signaux reçus par les capteurs gauche et droite respectivement. Les IPD $\Delta\varphi_{exp}(f)$ et ILD $\Delta\rho_{exp}(f)$ expérimentaux sont alors simplement déterminés par les expressions :

$$\begin{aligned}\Delta\varphi_{exp}(f) &= \text{Arg}(S_{P_g}(f)) - \text{Arg}(S_{P_d}(f)) \\ \Delta\rho_{exp}(f) &= 20 \log_{10} \frac{|S_{P_g}(f)|}{|S_{P_d}(f)|}.\end{aligned}\tag{2.1}$$

L'enjeu est donc ensuite, à partir des mesures expérimentales de $\Delta\varphi_{exp}$ et $\Delta\rho_{exp}$, de déterminer l'azimut de la source sonore. Ce passage peut s'effectuer rigoureusement en identifiant préalablement la HRTF à partir de mesures impulsionnelles en chambre anéchoïque. Les signaux reçus par les microphones peuvent alors être exprimés théoriquement tout en tenant compte des réflexions et diffusions de l'onde acoustique provoquées par la présence éventuelle d'une tête ou de tout autre élément du robot. Cependant, en environnement réel, la HRTF doit inclure la réponse acoustique de la pièce dans laquelle se situe le robot. Son utilisation est donc de fait limitée à un environnement maîtrisé et identifié. Il est par conséquent nécessaire d'établir un autre moyen pour déterminer, à partir des indices interauraux, l'origine d'un son.

2.1.1.1 Modèles pour la détermination des indices interauraux

Afin d'établir une relation simple entre la position de la source sonore à localiser et les valeurs mesurées des indices interauraux (2.1), il est nécessaire de posséder un modèle de la propagation permettant de relier leur valeur théorique à l'azimut θ à déterminer. Différents modèles plus ou moins complexes ont été proposés. Ils sont examinés et comparés ci-dessous.

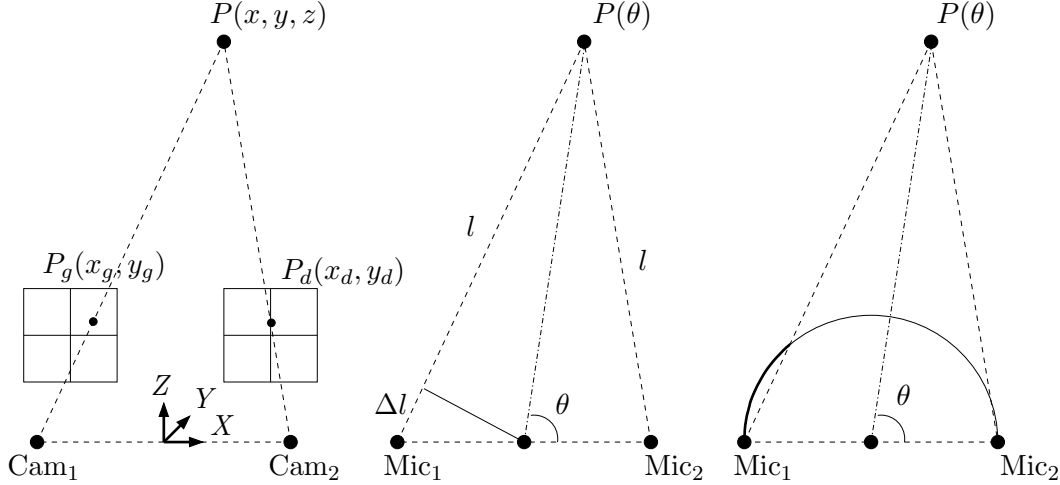


FIG. 2.2 – [Nakadai 2001] - Géométrie épipolaire auditive. A gauche : parallèle avec la géométrie épipolaire en vision. Au centre : géométrie épipolaire auditive. A droite : géométrie épipolaire auditive corrigée, prenant en compte la présence de la tête.

Un modèle élémentaire En vision par ordinateur, la contrainte épipolaire permet de simplifier la recherche de points homologues dans une paire d'images stéréoscopiques pour ensuite, par triangulation, aboutir à une localisation 3D. La *géométrie épipolaire auditive* (Auditory Epipolar Geometry, ou AEG) proposée par [Nakadai 2000b] peut être vue comme une extension naïve de la géométrie épipolaire utilisée en vision. Une représentation de ces deux géométries est représentée sur la figure 2.2, où θ représente l'azimut à déterminer de la source placée en P . L'AEG consiste à relier, pour chaque fréquence f , la différence interaurale en phase $\Delta\varphi_{AEG}$ à l'azimut θ par la formule élémentaire [Nakadai 2001]

$$\Delta\varphi_{AEG}(\theta, f) = \frac{2\pi f d}{c} \cos \theta, \quad (2.2)$$

où c désigne la vitesse du son et d l'espacement entre les deux microphones. Notons que l'équation (2.2) suppose implicitement que la source P se situe en champ lointain et que donc la distance l définie sur la figure 2.2 tend vers l'infini.

L'avantage principal de cette formulation réside dans l'absence de l'utilisation de la HRTF : aucune identification n'est nécessaire et la comparaison de l'IPD mesurée $\Delta\varphi_{exp}$ avec l'IPD théorique $\Delta\varphi_{AEG}$ par inversion de (2.2) permet d'approximer très simplement l'azimut θ . Cependant, si le modèle proposé est intéressant du point de vue de sa complexité calculatoire, il ne permet pas de décrire les modifications induites sur les mesures expérimentales par la tête du robot placée entre les deux microphones. Ceci est illustré sur les figures 2.3(a) et 2.3(b), où les mesures expérimentales $\Delta\varphi_{exp}$, en traits épais, effectuées en présence ou non de la tête, sont comparées aux valeurs, en traits fins, prédites par le modèle de l'AEG. Sans surprise, la

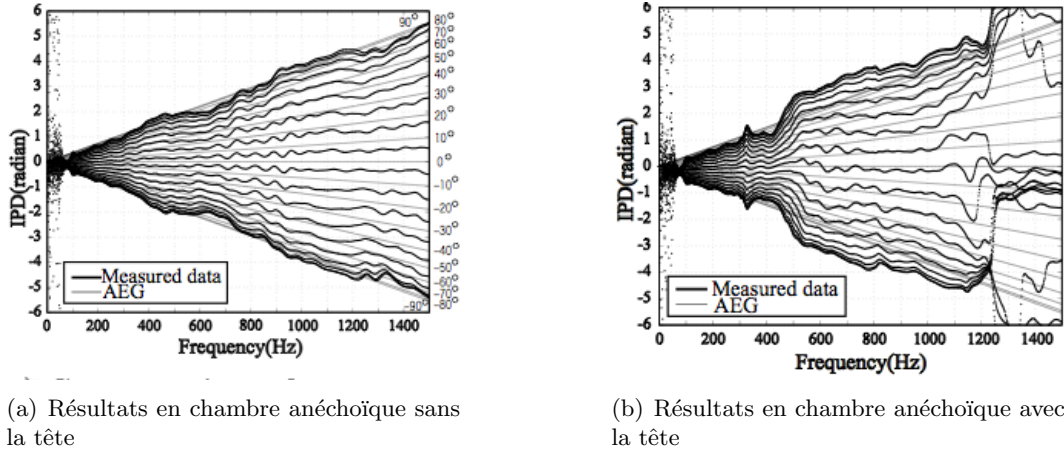


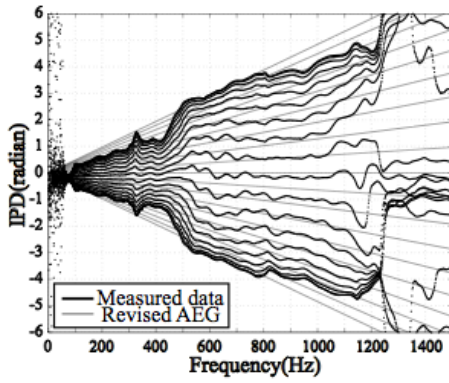
FIG. 2.3 – [Nakadai 2001] - Illustration de l'importance du rôle de la tête dans la modification des mesures expérimentales de l'IPD.

présence des phénomènes de réflexions et diffusions vient perturber sensiblement les valeurs des IPDs mesurés. Cependant, notons que les mesures obtenues sans la tête montrent une grande ressemblance avec les valeurs données par le modèle simple de la géométrie épipolaire auditive. Ce même modèle, pour les raisons évoquées précédemment, perd très nettement en précision pour des fréquences supérieures à 300Hz lorsque la tête est présente. Il est donc nécessaire de proposer un second modèle prenant cette fois explicitement en compte sa présence.

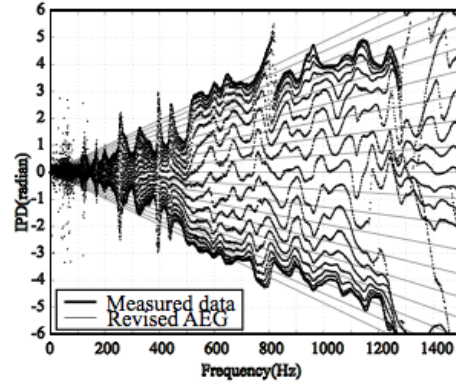
Un second modèle simple prenant en compte la présence de la tête Afin d'obtenir une plus grande concordance avec les données expérimentales, [Nakadai 2001] et [Nakadai 2002a] proposent une *géométrie épipolaire auditive modifiée* (Revised Auditory Epipolar Geometry, ou RAEG), prenant en compte, dans une certaine mesure, l'existence de la tête. Celle-ci est représentée par un demi-disque dans le plan azimutal comme indiqué sur la figure 2.2, et peut donc occulter la source. Dans ce cas, l'onde sonore ne suit pas nécessairement un chemin direct vers les deux oreilles, mais peut être amenée à se propager le long de la tête. Cet effet de "guide d'onde" conduit naïvement à l'écriture suivante de l'IPD $\Delta\varphi_{RAEG}(\theta, f)$:

$$\Delta\varphi_{RAEG}(\theta, f) = \frac{\pi f d}{c}(\theta + \sin \theta). \quad (2.3)$$

La figure 2.4(a) montre que ce nouveau modèle permet de mieux décrire les mesures expérimentales obtenues en chambre anéchoïque, en particulier dans le domaine de fréquence $500\text{Hz} - 800\text{Hz}$. Cependant, bien qu'il prenne en compte une partie de la diffusion induite par la présence de la tête, les réflexions causées par la présence d'éventuels murs ou obstacles ne sont pas capturées. Ainsi, si les mesures expérimentales $\Delta\varphi_{exp}$ de l'IPD sont effectuées dans une chambre acoustiquement préparée, mais non anéchoïque, les résultats obtenus restent très mauvais. La figure 2.4(b) montre les différences mesurées dans ces deux environnements acoustiques distincts. Cette fois, le modèle simple de la géométrie épipolaire révisée ne décrit pas les distorsions causées par l'acoustique de la pièce, et il est très difficile d'effectuer une correspondance entre l'IPD mesurée $\Delta\varphi_{exp}$ et l'IPD théorique $\Delta\varphi_{RAEG}$. Nous constatons également que les fréquences supérieures à 1200Hz ne sont plus du tout décrites correctement. A partir de cette



(a) Résultats comparatifs avec l'AEG modifié



(b) Mesures en environnement acoustique préparé

FIG. 2.4 – [Nakadai 2001] - Mesures de l'IPD dans différentes conditions acoustique, et comparaison avec le modèle donné par la géométrie épipolaire auditive (AEG).

limite, la valeur de l'IPD dépasse $\pm 2\pi$ et devient ambiguë, phénomène lié à l'espacement d entre les deux microphones (cf. chapitre précédent, page 20).

En conclusion, l'IPD ne peut donc pas être utilisée pour localiser les hautes fréquences. On retrouve là le même type de comportement que celui observé chez l'homme, où le cerveau base la localisation des hautes fréquences sur l'ILD seulement. [Nakadai 2001] propose le même type de distinction dans l'utilisation des indices interauraux, et intègre l'ILD pour les fréquences supérieures à environ 1500Hz . Cependant, l'obtention d'un modèle simple décrivant les variations théoriques de l'ILD s'avère beaucoup moins immédiat que pour l'IPD. [Huang 2002] propose d'en représenter les variations en faisant l'hypothèse que l'énergie de l'onde sonore décroît de manière inversement proportionnelle à la distance qu'elle parcourt à la surface d'une tête sphérique. Ce modèle de simulation simple ne peut décrire que les tendances des courbes expérimentales relevées en chambre anéchoïque, si bien que l'erreur de modélisation s'avère parfois très importante. De ce fait, [Nakadai 2001] propose d'utiliser l'ILD uniquement pour détecter si un son provient de la droite ou de la gauche, par une simple comparaison des puissances des signaux reçus, tandis que l'IPD produit une estimation permanente de l'azimut grâce à l'AEG. Il est donc opportun de proposer un troisième modèle permettant de décrire aussi fidèlement que possible les évolutions de l'ILD en fonction de l'azimut.

Un troisième modèle plus développé basé sur la diffusion acoustique [Handzel 2002] et [Nakadai 2003] exploitent tous deux la théorie de la diffusion acoustique (Scattering Theory, ou ST) pour exprimer le champ de pression capté par deux microphones placés diamétralement de part et d'autre d'une sphère supposée rigide. [Handzel 2003] propose de valider expérimentalement ce formalisme sur une sphère en PVC d'un diamètre d'environ 9.5cm , cf. figure 2.5. Une telle sphère rigide permet donc de modéliser plus finement les propriétés de diffusion acoustique de la tête. En particulier, quand une source se situe sur le côté, l'onde de pression parcourant l'arrière de la tête vient modifier le signal perçu par le microphone placé du côté opposé. La géométrie épipolaire ne prend pas en compte ce phénomène de diffusion et se révèle donc incomplète.

La théorie de la diffusion est utilisée ici pour exprimer la valeur du potentiel des vitesses $\phi(\mathbf{r})$, dont la définition a été introduite dans le chapitre précédent page 9, en fonction des ondes



FIG. 2.5 – [Handzel 2003] - Plateforme expérimentale permettant d'étudier l'influence de la présence d'une tête entre deux microphones. La tête est ici émulée par une sphère en plastique, d'un rayon d'environ 10 cm, dont les pôles sont plats.

acoustiques incidente et réfléchi à la surface de la sphère. Ainsi, en supposant qu'une source sonore émet depuis la position repérée par $(r_0, 0, 0)$ dans le système de coordonnées sphériques défini figure 1.3 page 12, le potentiel des vitesses $\phi(\mathbf{r}) = \phi(\theta, f)$ s'écrit [Ziomek 1995]

$$\phi(\theta, f) = \sum_{m=0}^{\infty} (-j)^m (2m+1) j'_m(k \frac{d}{2}) P_m(\cos \theta) \frac{h_m^{(2)}(kr_0)}{h_m^{(2)'}(k \frac{d}{2})}, \quad (2.4)$$

où $j_m(\cdot)$ désigne la fonction sphérique de Bessel de première espèce et d'ordre m , tandis que la notation $'$ représente l'opération de dérivation. Si les deux microphones gauche et droite sont placés en $M_g = (d/2, \pi/2, 0)$ et $M_d = (d/2, -\pi/2, 0)$, alors les indices interauraux sont donnés par [Nakadai 2003] :

$$\begin{aligned} \Delta\varphi_{ST}(\theta, f) &= \text{Arg}(\phi_g(\theta, f)) - \text{Arg}(\phi_d(\theta, f)) \\ \Delta\rho_{ST}(\theta, f) &= 20 \log_{10} \frac{|\phi_g(\theta, f)|}{|\phi_d(\theta, f)|}, \end{aligned} \quad (2.5)$$

$$\text{avec } \phi_g(\theta, f) = \phi(\frac{\pi}{2} - \theta, f) \text{ et } \phi_d(\theta, f) = \phi(-\frac{\pi}{2} - \theta, f) \quad (2.6)$$

Contrairement au modèle basé sur la géométrie épipolaire, la théorie de la diffusion permet donc l'obtention d'expressions théoriquement mieux étayées de l'IPD et ILD en fonction de l'azimut θ , et l'inversion de ces formules pourra être utilisée pour la localisation. La figure 2.6 propose une comparaison de ces deux modèles, et montre que la prise en compte de la diffusion acoustique permet d'approximer avec nettement moins d'erreur la valeur de l'IPD mesurée en chambre anéchoïque. Cependant, bien que ceci ne soit pas mentionné dans [Nakadai 2003], le modèle de la diffusion acoustique ne peut toujours pas prendre en compte l'acoustique de la pièce, et la comparaison entre les IPD/ILD mesurées et les IPD/ILD théoriques sera très probablement fortement perturbée en environnement réel.

Autres approches D'autres utilisations simples des indices interauraux ont été proposées dans la communauté robotique. Par exemple, [Berglund 2005] présente un système construit à partir d'un réseau de neurones et implémenté sur un robot Aibo de Sony. Les IPDs et ILDs sont déterminées, pour chaque fréquence, à partir de (2.1), tandis que l'ITD (mesure du *décalage temporel* entre deux signaux) provient d'un calcul de corrélation. Enfin, un quatrième indice

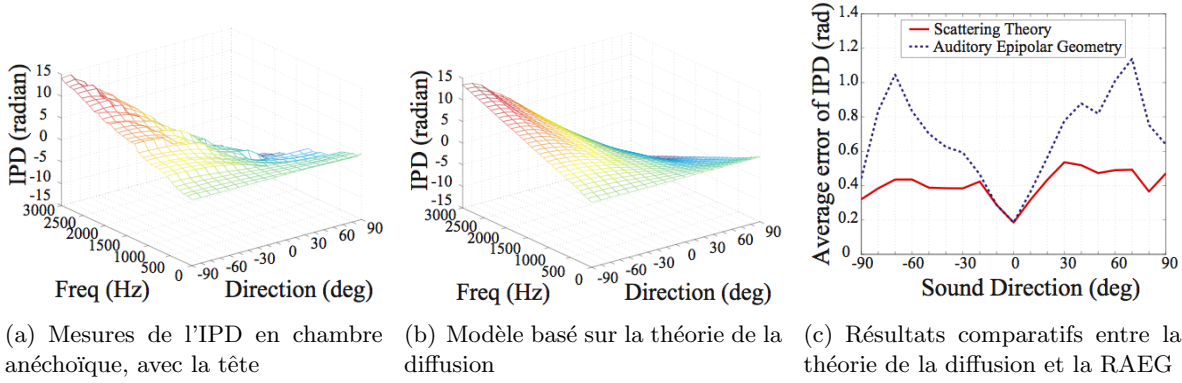


FIG. 2.6 – [Nakadai 2003] - Mesures de l'IPD sur le robot SIG, et comparaison entre le modèle donné par la géométrie épipolaire auditive (AEG) et la théorie de la diffusion.

interaural est introduit. Appelé *différence relative interaurale en amplitude* (Relative Interaural Level Difference, ou RILD), il représente une normalisation de l'ILD par rapport à la puissance du signal reçu à la fréquence considérée. L'ensemble de ces indices interauraux est ensuite regroupé en un unique vecteur d'entrée du réseau de neurones, puis est utilisé dans la procédure d'apprentissage.

Une autre approche, proposée dans [Cavaco 1999], est basée uniquement sur les ITDs. Les signaux issus des deux microphones sont tout d'abord traités par une banque de filtres de Bessel telle que le $i^{\text{ième}}$ filtre possède une bande passante $B_i = [f_1/2^{i-1}, f_2/2^{i-1}]$, les deux fréquences f_1 et f_2 vérifiant $f_1 > f_2 > f_1/2$. Ensuite, chacune des sorties des filtres est utilisée de façon à calculer l'ITD par une simple mesure du point de passage à zéro. Bien que cette méthode soit extrêmement simple, elle nécessite de fixer arbitrairement des seuils de détection et s'avère très sensible au bruit venant modifier les points de passage à zéro des signaux.

2.1.1.2 Stratégies de localisation

Nous avons détaillé dans la partie précédente la façon dont les indices interauraux peuvent être reliés théoriquement à l'azimut de la source. Nous allons maintenant parcourir les différentes méthodes permettant, à partir des différents modèles présentés précédemment, de localiser une source sonore. [Nakadai 2002a] propose un système complet, basé sur le modèle de la diffusion acoustique de la tête, pour déterminer l'origine géométrique d'un son et l'extraire de l'ensemble de la mixture sonore de l'environnement. Ce système, appelé "Active Direction Pass Filter" (ADPF), est représenté sur la figure 2.7. Les signaux reçus par les deux microphones sont d'abord décomposés dans le domaine fréquentiel par une FFT. Les pics du spectre obtenu sont ensuite utilisés pour la localisation comme suit. Pour des fréquences inférieures à $f_{th} = 1500\text{Hz}$, le système calcule, pour chaque valeur possible de θ , l'IPD théorique $\Delta\varphi_{ST}(\theta, f)$ prédit selon le modèle (2.4). Une première fonction de coût d_{IPD} est alors construite [Nakadai 2002a] :

$$d_{IPD}(\theta) = \frac{1}{n_{f \leq f_{th}}} \sum_{f=f_{min}}^{f_{th}} \frac{1}{f} (\Delta\varphi_{ST}(\theta, f) - \Delta\varphi(f))^2, \quad (2.7)$$

où $\Delta\varphi(f)$ représente la mesure expérimentale de l'IPD, et $n_{f \leq f_{th}}$ le nombre d'harmoniques dont la fréquence est inférieure à f_{th} . De la même façon, pour les fréquences supérieures à f_{th} , une

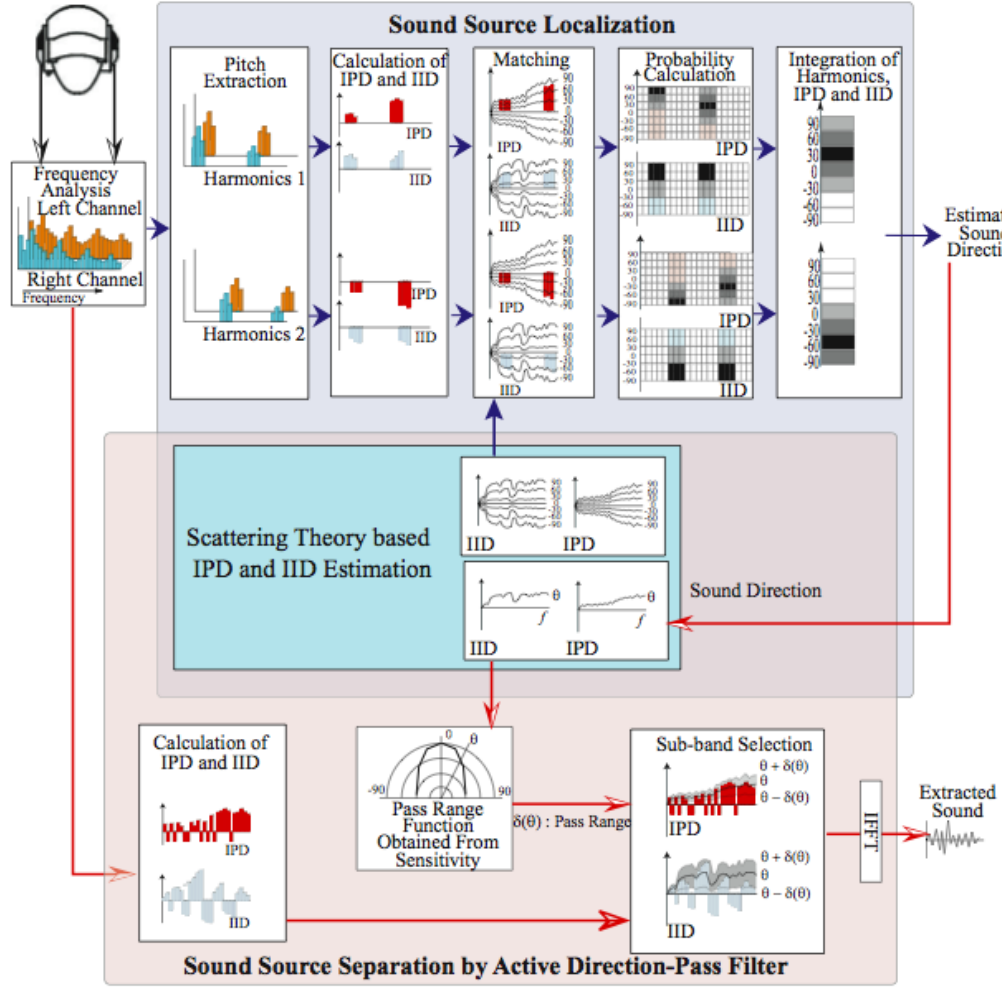


FIG. 2.7 – [Okuno 2004] - Architecture du système de localisation/extraction implémenté sur le robot SIG.

fonction de coût $d_{ILD}(\theta)$, basée sur les mesures de l'ILD, est proposée [Nakadai 2002a]

$$d_{ILD}(\theta) = \frac{1}{n_{f>f_{th}}} \sum_{f=f_{th}}^{f_{max}} f (\Delta\rho_{ST}(\theta, f) - \Delta\rho(f))^2, \quad (2.8)$$

où $\Delta\rho(f)$ désigne la mesure de l'ILD, et $n_{f>f_{th}}$ le nombre d'harmoniques dont la fréquence est supérieure à f_{th} . Ces deux fonctions de coût permettent ensuite de déterminer le θ optimal minimisant l'écart entre les modèles et les mesures. Pour cela, les deux distances d_{IPD} et d_{ILD} sont transformées en facteurs de croyance et fusionnées en un facteur $P_{IPD+ILD}(\theta)$ par la théorie de Dempster-Shafer. Au final, l'angle θ_s correspondant au maximum de $P_{IPD+ILD}$ est choisi comme azimut de la source sonore.

Des résultats de localisation selon cette méthode sont proposés sur la figure 2.8. Bien que l'utilisation du modèle basé sur la diffusion acoustique permette d'améliorer sensiblement la précision de la localisation, l'erreur moyenne pire cas obtenue demeure de l'ordre de 10° . Le

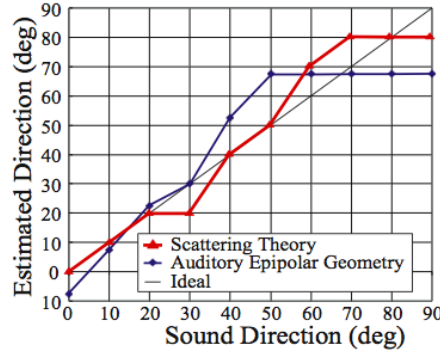


FIG. 2.8 – [Nakadai 2003] - Résultats de localisation obtenus sur le robot SIG.

résultat de localisation obtenu est ensuite utilisé pour la séparation par l'ADPF. Un intervalle de confiance est déterminé à partir de l'azimut estimé $\hat{\theta}_S$, à l'aide d'une "Pass Range Function" $\delta(\theta)$. Cette fonction correctrice empirique modélise la sensibilité amoindrie du système de localisation pour des sources se situant sur les côtés de la tête du robot, et aboutit à considérer l'ensemble des θ vérifiant $\theta_l < \theta < \theta_h$, avec $\theta_l = \hat{\theta}_S - \delta(\hat{\theta}_S)$ et $\theta_h = \hat{\theta}_S + \delta(\hat{\theta}_S)$. Cet ensemble correspond à un domaine dans lequel la source, détectée comme émettant depuis $\hat{\theta}_S$, a une très forte chance de se situer. De cette façon, les différentes fréquences contribuant à des valeurs d'ITD et d'ILD compatibles avec cet ensemble des θ sont collectées de façon à former un signal supposé provenir de la direction $\hat{\theta}_S$.

Les performances d'extraction sont évaluées dans [Nakadai 2002a] et comparées selon le modèle retenu pour exprimer le lien entre l'azimut et les indices interauraux. Sans surprise, le modèle le plus abouti de la diffusion produit les meilleures performances. Un tel système de localisation et d'extraction permet d'évaluer aussi l'amélioration produite par l'ADPF pour la reconnaissance vocale. Une telle étude est menée dans [Yamamoto 2004] et [Nakadai 2002b] où, à travers un scénario incluant deux locuteurs suffisamment proches de SIG et séparés l'un de l'autre, des informations visuelles et proprioceptives sont fusionnées de façon à améliorer l'interaction visuo-auditive avec le robot. L'importance de la fusion du son avec la vision est illustrée dans [Okuno 2004] : trois personnes parlent ici en même temps, et le robot est capable d'associer les mots prononcés à chaque personne. En cas d'échec de la reconnaissance, le robot tourne sa tête en direction du locuteur et lui demande de répéter, seul, le mot prononcé. Enfin, l'implémentation du système complet est abordée dans [Kuratoro 2005]. L'ADPF et la technique de localisation utilisée requièrent une puissance de calcul importante, si bien que le calcul est décentralisé et déporté à l'extérieur du robot. De façon à embarquer le traitement à l'intérieur de la plate-forme, l'utilisation d'un processeur dédié est proposée. Celui-ci est basé sur une architecture reconfigurable dynamiquement et propose des performances brutes assimilables à celles obtenues avec un ordinateur de bureau récent.

2.1.2 Les indices spectraux

Comme indiqué dans la partie 1.2.2.2 page 21, les indices spectraux ont pour origine les diverses réflexions acoustiques créées principalement par notre oreille externe et sont exploités chez l'humain pour la localisation verticale. En particulier, il semble que l'élévation de la source influe sur la position de fortes coupures dans le spectre du signal perçu par une oreille ("notches"). Le cerveau se charge ensuite d'intégrer ces indices monauraux (car déterminés sans comparaison

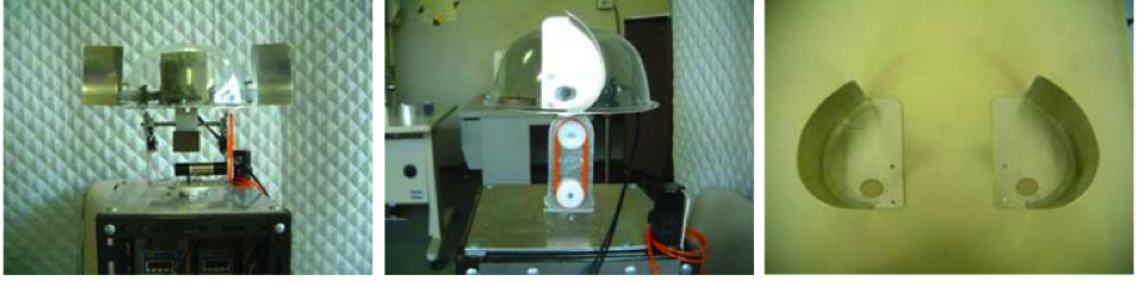


FIG. 2.9 – [Kumon 2005] - Plate-forme expérimentale permettant d'évaluer le filtrage effectué par l'oreille externe. A gauche et au milieu : le dispositif complet, avec les deux micros placés à l'intérieur de l'oreille artificielle ; une cloche recouvre les moteurs de la platine pour isoler le bruit. A droite : le détail des pavillons et de leur géométrie.

entre les signaux reçus par nos deux oreilles) avec les informations d'ITD et ILD pour localiser une source sonore dans l'espace.

L'utilisation des indices spectraux en robotique est très peu évoquée dans la littérature. Elle implique en effet la conception d'une oreille externe venant modifier le spectre des signaux perçus. Ainsi, [Kumon 2005] propose la conception d'un pavillon artificiel permettant de localiser verticalement un son. Celui-ci est représenté sur la figure 2.9, ainsi que l'ensemble du dispositif expérimental. Toute la difficulté réside dans le fait qu'aucun modèle théorique ne permet de définir la forme optimale de l'oreille externe conduisant à la présence de "notches" spécifiés pour différentes élévations. Cependant, une règle empirique doit être respectée : il est nécessaire que la forme du pavillon soit irrégulière ou asymétrique. La forme exponentielle proposée dans [Kumon 2005] est reportée sur la figure 2.10. Un modèle simplifié, inspiré de [Hebrank 1974] et basé sur la superposition de l'onde incidente avec une *unique* réflexion induite par le pavillon permet alors de prédire, à partir de la position des notches, l'élévation correspondante. La figure 2.10 présente également les spectres expérimentaux des signaux reçus, sur lesquels sont indiqués différents "notches" N_A , N_B et N_C [Kumon 2005]. Ces mesures montrent que plusieurs notches apparaissent ou disparaissent selon l'élévation, mais ne sont pas toujours facilement détectables. De plus, la présence d'obstacles dans l'environnement induit des réflexions multiples à l'origine de sommations destructives produisant de multiples coupures. Néanmoins, la position en fréquence des notches les plus importants varie bien en fonction de l'élévation selon la même tendance que celle prédite par le modèle simple proposé. La dispersion des valeurs obtenues est

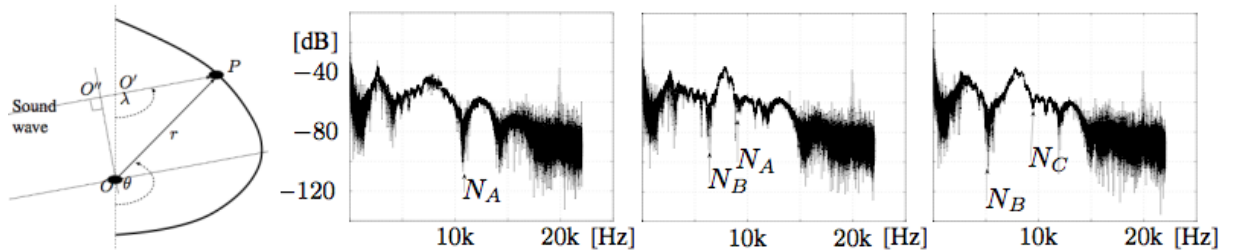
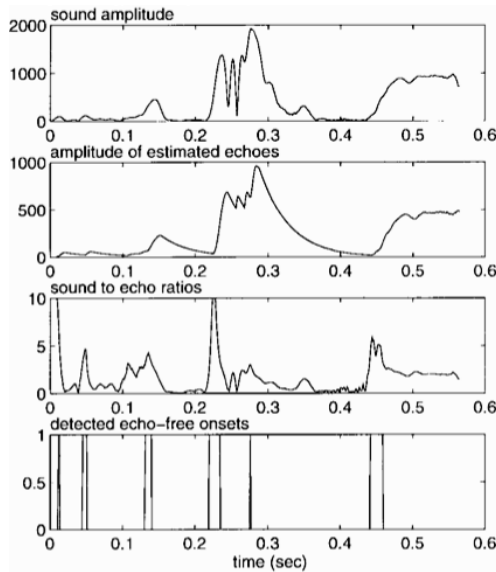


FIG. 2.10 – [Kumon 2005] - Modèle du pavillon externe proposé (gauche), et spectres obtenus en fonction de l'élévation (de gauche à droite : 45°, 75° et 105°).

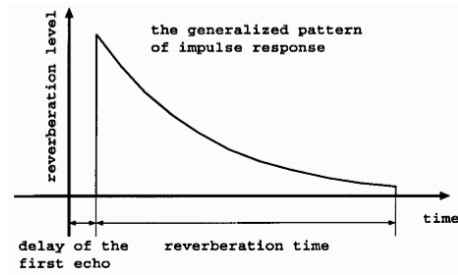
toutefois très importante et ne permet de déterminer l'élévation qu'à environ $\pm 25^\circ$ près. Une modélisation plus complète, disponible dans [Lopez-Poveda 1996], prend en compte les multiples réflexions et diffractions que subit l'onde sonore. Elle peut être utilisée pour modifier l'expression des intensités et des spectres des sons reçus par les microphones placés à l'intérieur de l'oreille artificielle. Cependant, cette modélisation reste difficilement exploitable pour la conception de pavillon.

2.1.3 L'effet de précedence

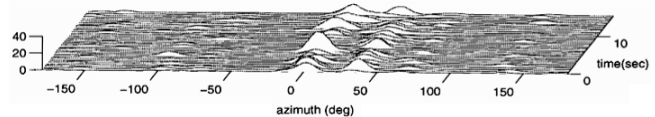
Nous avons vu dans le chapitre précédent que l'homme sait localiser précisément une source sonore, y compris dans un environnement fortement réverbérant. Dans de telles conditions, notre cerveau détecte dans les signaux perçus des zones dites "libres d'écho" qui sont ensuite utilisées pour la localisation. Ce phénomène est appelé *effet de précedence*, ou *effet de Haas*. La communauté robotique a tenté de transposer ce principe pour proposer de nouvelles stratégies de localisation de sources sonores. L'idée n'est donc plus d'effectuer la localisation par intégration temporelle des informations issues de microphones, mais plutôt de travailler sur un très court intervalle de temps pendant lequel le signal est supposé non entaché d'échos. [Huang 1995] en pro-



(a) Détection des zones libres d'écho.



(b) Modélisation de l'écho.



(c) Résultats de localisation en environnement réel, sur des intervalles de temps de 0.5s.

FIG. 2.11 – [Huang 1997b] - De la modélisation de l'écho aux résultats de localisation obtenus en travaillant sur les parties libres d'écho des signaux reçus.

pose une première implémentation, testée sur un capteur sonore constitué de trois microphones disposés en triangle. Les sorties de ces trois transducteurs sont tout d'abord comparées à un seuil permettant de décider de la présence ou non de signal. Si un signal est détecté, et s'il est précédé d'une durée de silence suffisamment longue, alors il est mémorisé pendant une courte durée T fonction des paramètres géométriques de la pièce et du capteur. Ensuite, la sortie de chaque microphone est décomposée en un ensemble de signaux à bande étroite par une banque de filtres passe-bande. Des délais temporels Δt_{ij} liés à la propagation entre chacune des paires de microphones sont alors déduits des points de passage à zéro de ces signaux bande étroite, et reportés dans des histogrammes accumulant leurs valeurs pendant une durée suffisamment

longue. Le passage d'un histogramme des durées à un histogramme moyen de l'azimut θ est basé sur une simple relation géométrique liée au placement des microphones. Cette méthode simple conduit à une précision de l'ordre de 10° , mais n'est opérationnelle qu'en présence de silences marqués dans le signal émis par la source. Ainsi, lorsque deux locuteurs parlent simultanément, seulement 9% de la durée des signaux est exploitable pour la localisation. Un autre inconvénient réside dans la nécessité d'accumuler des valeurs des décalages temporels pendant une durée non négligeable (environ 20s), ce qui ralentit significativement la localisation.

Une première amélioration est proposée dans [Huang 1997c] et [Huang 1999]. Elle est basée sur un modèle de l'écho, paramétré par la réponse impulsionnelle simplifiée $h(t)$ définie par la figure 2.11(b), avec

$$h(t) = \begin{cases} 1, & t = 0 & (\text{chemin direct}) \\ 0, & 0 < t < \tau_{fe} & (\text{zone sans écho}) \\ \alpha_{fe} e^{-(t-\tau_{fe})/\tau}, & t \geq \tau_{fe}, \end{cases} \quad (2.9)$$

où α_{fe} et τ_{fe} représentent respectivement l'amplitude et le retard du premier écho, et τ désigne une constante de temps. Après convolution du signal reçu avec le modèle (2.9), l'amplitude de l'écho est estimée et permet de calculer un "rapport son sur écho" (Sound to Echo Ratio, ou SER) au cours du temps. Les zones utiles pour la localisation sont alors isolées par simple seuillage. Si l'estimation des zones libres d'écho repose maintenant sur un modèle, la localisation est toujours basée sur l'approximation des délais Δt_{ij} à partir des points de passage à zéro des signaux perçus. L'azimut est ensuite estimé à partir de seulement deux des trois Δt_{ij} disponibles ; les décalages temporels les plus courts sont sélectionnés, car leur choix conduit à une sensibilité minimale à l'erreur d'approximation des Δt_{ij} [Huang 1999]. Cette utilisation du modèle (2.9) est illustrée sur la figure 2.11(a) où l'enveloppe du signal reçu sur un des trois microphones dans la bande 255 Hz est représentée.

Enfin, [Huang 1997b] propose de mixer les deux approches précédentes en utilisant les histogrammes des délais Δt_{ij} ainsi que le modèle d'écho de [Huang 1997c]. Cette fois, la détection des zones libres d'écho s'effectue avec le modèle donné par (2.9), mais en aval des filtres passe-bande réalisant la décomposition fréquentielle des signaux reçus. Ainsi, les zones utiles pour la localisation sont totalement indépendantes les unes des autres, et la fusion par simple moyennage des histogrammes n'est donc plus possible puisque ne correspondant plus nécessairement à la même fenêtre temporelle. Des heuristiques simples sont proposées de façon à filtrer les fausses détections et permettent à l'algorithme de fonctionner à une cadence plus élevée, de l'ordre de 2 Hz. Un résultat de localisation est reproduit figure 2.11(c), sur laquelle sont représentés les histogrammes de l'azimut obtenus en environnement réel. La précision mesurée est alors de l'ordre de 5° .

2.1.4 Autres approches biomimétiques

Parmi les méthodes que nous venons de présenter, un grand nombre s'appuie sur une décomposition fréquentielle des signaux perçus. La majorité effectue cette transformation par l'intermédiaire d'une transformation de Fourier, souvent implémentée sous sa forme "optimisée" FFT (Fast Fourier Transform). D'autres proposent l'utilisation de filtres passe-bande. Ainsi, [Cavaco 1999] choisit des filtres de Bessel d'ordre 8 de bande passante décroissante avec la fréquence centrale, tandis que [Huang 1997b] propose une approche voisine sans plus de précisions sur la nature des filtres utilisés.

Toutes ces méthodes, aussi efficaces soient elles, ne représentent pas véritablement la décomposition fréquentielle effectuée au sein de l'oreille interne. C'est pourquoi [Heckmann 2005]

suggère l'exploitation d'une banque de filtres passe-bande modélisant la vibration de la membrane basilaire située à l'intérieur de la cochlée. Ces *filtres gammatones* sont issus d'expériences physiologiques menées chez le chat dans [Johannesma 1972]. Ils émulent la séparation fréquentielle effectuée par la cochlée pour des niveaux modérés de stimuli sonores [Patterson 1995]. La réponse impulsionnelle $h(t)$ d'un filtre gammatone s'écrit [Van Immerseel 2003]

$$h(t) \sim t^{N-1} e^{-2\pi b \text{ERB}(f_c)t} \cos(2\pi f_c t + \phi), \quad (2.10)$$

où N , f_c , ϕ désignent respectivement son ordre, sa fréquence centrale et une phase constante, et b est un paramètre lié à son temps de réponse. La grandeur $\text{ERB}(f)$, dite "largeur de bande rectangulaire équivalente" (Equivalent Rectangular Bandwidth), représente, pour chaque filtre auditif, la largeur d'un filtre passe-bande idéal de même fréquence centrale qui, alimenté par un bruit blanc, produirait la même énergie en sortie. Traditionnellement, l'ERB vérifie la relation $\text{ERB}(f) = 24.7 + 0.108f$ pour des niveaux sonores modérés. Les filtres gammatones tirent leur nom du fait que leur réponse impulsionnelle fait apparaître l'expression de la fonction de répartition Gamma ainsi qu'un terme mono-fréquentiel de fréquence f_c . La figure 2.12(a) représente la réponse impulsionnelle d'un filtre gammatone de fréquence centrale $f_c = 1.5 \text{ kHz}$ et de largeur de bande $f_b = 50 \text{ Hz}$. Nous pouvons remarquer sur la figure 2.12 que la largeur de bande des

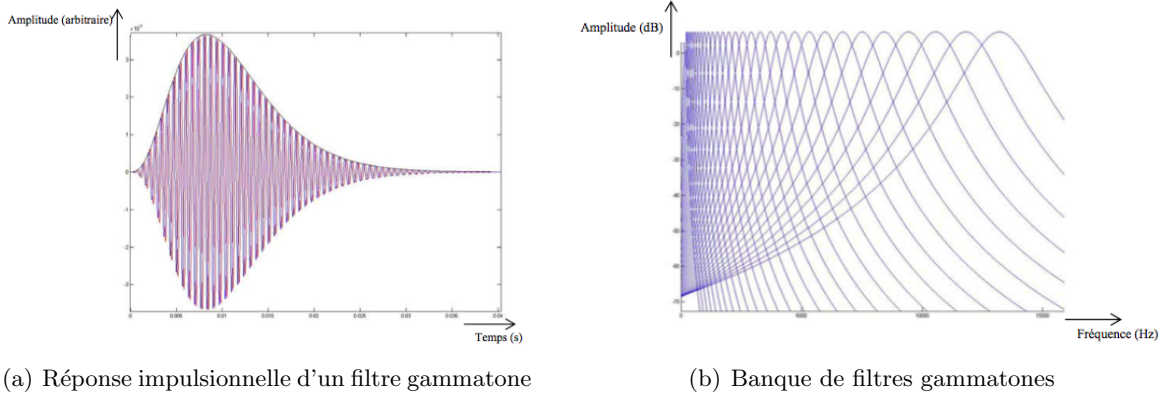


FIG. 2.12 – [Gnansia 2005] - Illustration des caractéristiques temporelles et fréquentielles des filtres gammatone.

filtres gammatones augmente avec la fréquence de telle sorte qu'elle représente environ 15% de leur fréquence centrale [Gnansia 2005]. Cette caractéristique importante représente une des particularités de notre système auditif, dont la discrimination fréquentielle s'avère meilleure pour les basses fréquences. Cependant, si le modèle (2.10) s'avère plausible pour de faibles niveaux sonores, l'utilisation de filtres auditifs plus développés est nécessaire pour modéliser les phénomènes non linéaires constatés à l'intérieur de la cochlée. Par exemple, la bande passante ainsi que la fréquence centrale des filtres s'avère être fonction de l'intensité sonore. [Irino 1997] propose d'introduire un terme analogue à une modulation de fréquence dans l'expression de la réponse impulsionnelle des filtres gammatone afin de produire un spectre en amplitude asymétrique par rapport à la fréquence centrale. Appelé "filtre gammachirp", ce type de filtre auditif est aujourd'hui le plus souvent utilisé pour modéliser la décomposition fréquentielle opérée par notre oreille interne, mais n'a pas encore été exploité en robotique.

2.1.5 Conclusion

Les méthodes biomimétiques présentent des résultats contrastés. À partir d'un nombre très réduit de microphones, la communauté robotique a su proposer des solutions originales pour localiser et extraire des sons de l'environnement d'un robot. Seulement, nous avons illustré toute la difficulté à reproduire les capacités auditives de l'homme à partir de mécanismes bioinspirés. Ainsi, l'exploitation des indices interauraux nécessite une modélisation très précise des modifications induites par la présence de la tête. Les indices interspectraux sont quant à eux en l'état actuel très difficilement exploitables. Enfin, l'effet de précédence est à l'origine de résultats de localisation intéressants, mais fournis à une cadence peu élevée. Dans tous les cas, la définition de modèles précis de la propagation s'avère indispensable pour décrire avec justesse les évolutions des différents indices auditifs. De plus, nous avons montré le manque de robustesse de toutes ces techniques aux variations de l'environnement acoustique. Beaucoup de modèles, validés expérimentalement en chambre anéchoïque, ne permettent pas d'aboutir à une localisation précise sans connaissance à priori de l'environnement dans lequel le robot sera amené à se déplacer. L'ensemble de ces éléments nous poussent à mettre de côté dans un premier temps l'ensemble de ces approches biomimétiques pour nous focaliser sur les techniques issues de la communauté du Traitement du Signal.

2.2 Les méthodes d'antennerie

Nous venons de détailler dans la partie précédente l'ensemble des méthodes *bio-inspirées* utilisées jusqu'ici en robotique. Ces méthodes présentent l'avantage indéniable de ne nécessiter qu'un nombre très réduit de microphones (de deux à quatre en règle générale), facilitant grandement leur implémentation sur une plateforme robotique. Cependant, elles paient parfois le prix de leur simplicité géométrique par une certaine complexité algorithmique, qui nécessite alors du matériel spécifique pour pouvoir les implémenter en temps réel [Kuratori 2005]. Dès lors, pourquoi se limiter à l'utilisation de systèmes biomimétiques ? Travailler avec plus de deux microphones peut permettre en effet d'augmenter la résolution de la localisation, de rendre le système plus robuste aux différents bruits, etc. Nous convergeons donc tout naturellement vers l'utilisation d'antennes de microphones, domaine déjà largement étudié par les communautés du Traitement du Signal et de l'Acoustique.

Les antennes sont utilisées depuis longtemps dans des applications aussi variées que la radio-astronomie, les radars, les systèmes à sonar, les télécommunications, etc. Leur exploitation en robotique est au contraire restée longtemps assez confidentielle. Cependant, ces cinq dernières années ont vu l'adoption progressive par la communauté robotique de méthodes éprouvées pour la localisation de source sonore. Après avoir introduit les notations de ce chapitre, nous présentons la méthode haute-résolution la plus commune en traitement du signal appelée MUSIC. Cet exemple illustre efficacement les limites imposées par un traitement temps réel des informations sonores. Puis, dans une seconde partie, nous abordons les méthodes visant à déterminer les délais temporels causés par la propagation entre de multiples microphones. Ici, l'utilisation de la redondance de l'information permet d'aboutir à une localisation plus précise et plus robuste aux conditions expérimentales. Enfin, nous terminons par l'ensemble des techniques déjà utilisées en robotique basées sur la formation de voie : la simplicité des algorithmes et de leurs implémentations font de celle-ci le premier candidat idéal à une utilisation sur un robot.

2.2.1 Notations

Considérons que S sources sonores ponctuelles émettent depuis les points repérés par $\mathbf{r}_1^s, \mathbf{r}_2^s, \dots, \mathbf{r}_S^s$, dans un système de coordonnées sphériques tel que $\mathbf{r}_s^s = (r_s, \theta_s, \psi_s)$. Les quantités r_s , θ_s et ψ_s représentent donc respectivement la distance à l'origine, l'azimut et l'élévation de la $s^{\text{ième}}$ source (voir figure 2.13). Rappelons que l'hypothèse de dépendance temporelle harmonique (1.27) est supposée vérifiée, de sorte que n'importe quel signal spatio-temporel $y(\mathbf{r}, t)$ s'écrit sous la forme

$$y(\mathbf{r}, t) = Y(\mathbf{r}, k)e^{jkct}, \quad (2.11)$$

où $Y(\mathbf{r}, k)$ décrit l'amplitude complexe de $y(\mathbf{r}, t)$ ne dépendant que de l'espace, avec k le nombre d'onde ou fréquence. Si les signaux mis en jeu sont large bande, alors par application du principe de superposition, il vient

$$y(\mathbf{r}, t) = \frac{c}{2\pi} \int_{-\infty}^{\infty} Y(\mathbf{r}, k)e^{jkct} dt. \quad (2.12)$$

Notons que la paire $(y(\mathbf{r}, t); Y(\mathbf{r}, k))$ définit une Transformée de Fourier.

Dans toute la suite, nous considérons une antenne constituée de N microphones placés aux positions $\mathbf{r}_n^m$, $n = 1, \dots, N$. Nous faisons l'hypothèse que l'ensemble de ces transducteurs est omnidirectionnel, et possède la même réponse en fréquences $a(k)$. La sélection de transducteurs de bonne qualité et éventuellement appariés garantit la validité de cette simplification. Nous

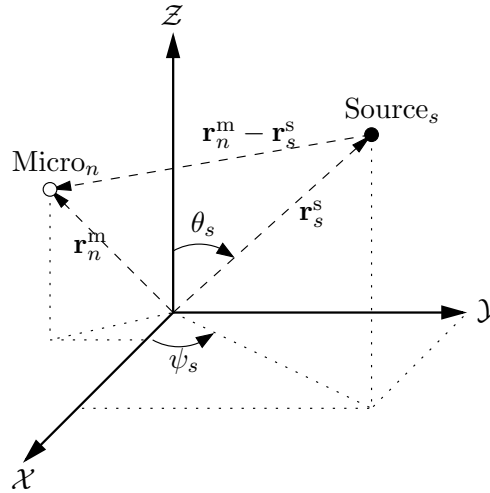


FIG. 2.13 – Organisation d'une antenne de microphone.

avons démontré que le potentiel des vitesses $\Phi(\mathbf{r}, t)$ peut être décrit, dans le cas d'une source de fréquence k située en $\mathbf{r}_s^s$ par l'équation (1.53) page 15. De cette relation, il découle immédiatement que le potentiel des vitesses $\Phi_n(t)$ mesuré en $\mathbf{r} = \mathbf{r}_n^m$, position du $n^{\text{ième}}$ microphone, s'exprime par $\Phi_n(t) \triangleq \phi_n(\mathbf{r}_n^m, k)e^{jkct} = -\frac{1}{4\pi\|\mathbf{r}_n^m - \mathbf{r}_s^s\|} e^{jkc(t - \frac{\|\mathbf{r}_n^m - \mathbf{r}_s^s\|}{c})}$. Si $\Phi_0(t)$ désigne le potentiel en $\mathbf{r} = \mathbf{0}$, i.e. en l'origine du repère, alors il vient

$$\Phi_n(t) = \frac{\|\mathbf{r}_s^s\|}{\|\mathbf{r}_n^m - \mathbf{r}_s^s\|} \Phi_0\left(t - \frac{\|\mathbf{r}_n^m - \mathbf{r}_s^s\|}{c} + \frac{\|\mathbf{r}_s^s\|}{c}\right). \quad (2.13)$$

Sous ces notations, en considérant que la réponse en fréquence $a(k) = 1$ et que les microphones considérés sont sensibles aux variations de pression au cours du temps, le signal sonore $m_n(t)$ causé par les S sources sonores et reçu par le $n^{\text{ième}}$ microphone est donc donné par la relation

$$m_n(t) = \mathcal{K} \sum_{s=1}^S \Phi_n(t),$$

où $\mathcal{K}$ désigne une constante exprimant la relation de proportionnalité existant entre le potentiel des vitesses et le signal $s_s^0(t)$, i.e. $\Phi_0(t) = \mathcal{K}s_s^0(t)$. Afin de simplifier les notations, nous supposons dans la suite que $\mathcal{K} = 1$. L'introduction de (2.13) dans l'équation précédente donne alors

$$m_n(t) = \sum_{s=1}^S \alpha_n(\mathbf{r}_s^s) s_s^0\left(t - \tau_n(\mathbf{r}_s^s)\right) + b_n(t), \quad (2.14)$$

où $s_s^0(t)$ désigne le signal fictif reçu à l'origine O et dû à la seule $s^{\text{ième}}$ source, et le bruit additif $b_n(t)$ modélise la présence de sources parasites dans l'environnement ainsi que le bruit électronique en sortie des microphones. En d'autres termes, $m_n(t)$ est une combinaison linéaire des signaux $s_s^0(t)$ amplifiés d'un facteur $\alpha_n(\mathbf{r})$ et décalés temporellement de délais $\tau_n(\mathbf{r})$, avec

$$\tau_n(\mathbf{r}) = \frac{\|\mathbf{r}_n^m - \mathbf{r}\|}{c} - \frac{\|\mathbf{r}\|}{c}, \quad (2.15)$$

$$\alpha_n(\mathbf{r}) = \frac{\|\mathbf{r}\|}{\|\mathbf{r}_n^m - \mathbf{r}\|}. \quad (2.16)$$

La transformée de Fourier de l'équation (2.14) conduit à l'expression

$$M_n(k) = \sum_{s=1}^S V_n(\mathbf{r}_s^s, k) S_s^0(k) + B_n(k), \quad (2.17)$$

où $M_n(k)$, $S_s^0(k)$ et $B_n(k)$ désignent respectivement les transformées de Fourier des signaux $m_n(t)$, $s_s^0(t)$ et $b_n(t)$. De plus,

$$V_n(\mathbf{r}, k) = \|\mathbf{r}\| e^{jk\|\mathbf{r}\|} \frac{e^{-jk\|\mathbf{r}_n^m - \mathbf{r}\|}}{\|\mathbf{r}_n^m - \mathbf{r}\|}, \quad (2.18)$$

et constitue le $n^{\text{ième}}$ élément du *vecteur d'antenne* $\mathbf{V}(\mathbf{r}, k) \triangleq (V_1(\mathbf{r}, k), \dots, V_N(\mathbf{r}, k))^T$. De la même façon, nous noterons $\mathbf{M}(k) \triangleq (M_1(k), \dots, M_N(k))^T$ le vecteur appelé *vecteur des observations* regroupant les transformées de Fourier des signaux reçus. De cette façon, l'équation (2.17) peut être ré-écrite matriciellement sous la forme

$$\mathbf{M}(k) = \mathcal{V}(k) \mathbf{S}^0(k) + \mathbf{B}(k), \quad (2.19)$$

avec $\mathbf{S}^0(k) \triangleq (S_1^0(k), \dots, S_S^0(k))^T$ et $\mathbf{B}(k) \triangleq (B_1(k), \dots, B_N(k))^T$. La matrice $\mathcal{V}(k)$, appelée *matrice d'antenne*, satisfait donc

$$\mathcal{V}(k) \triangleq (\mathbf{V}(\mathbf{r}_1^s, k), \dots, \mathbf{V}(\mathbf{r}_S^s, k)). \quad (2.20)$$

L'équation (2.17) nous indique que l'élément $V_n(\mathbf{r}_s^s, k)$ du vecteur d'antenne $\mathbf{V}(\mathbf{r}_s^s, k)$ peut être vu comme la fonction de transfert, lorsque seule la source s est active, entre le signal présent en sortie du micro n et le signal fictif mesuré en O . Ainsi, $V_n(\mathbf{r}_s^s, k)$ peut être aussi interprété comme étant la fonction de transfert *normalisée* entre la source s et le capteur n . De ce fait, $V_n(\mathbf{r}_s^s, k)$ est solution de l'équation d'Helmholtz (1.29), et il est effectivement possible d'exprimer (2.18) sous une forme plus complexe analogue à (1.47) décrivant des fronts d'onde sphériques. L'expression 2.18 peut être simplifiée en considérant cette fois des fronts d'onde plans, i.e. en admettant que la distance à la source tend vers l'infini, auquel cas la courbure du front d'onde tend vers zéro. Notons cependant que si ce cas limite permet de passer rigoureusement des ondes sphériques aux ondes planes, des fronts d'onde plans peuvent être envisagés dès lors que les distances des microphones aux sources excèdent une borne inférieure d_r . Cette distance est fonction de la fréquence et de la taille de l'antenne, et sera discutée en détail dans le chapitre suivant. Retenons simplement qu'elle permet de discriminer deux domaines de proximité au capteur : dans celui dit du *champ proche*, pour des distances $r \leq d_r$, les fronts d'ondes *devront* être considérés comme sphériques, alors que dans celui dit du *champ lointain*, pour des distances $d > d_r$, les fronts d'ondes *pourront* être considérés comme plans. Dans la suite, toutes les grandeurs relatives au champ lointain porteront en exposant le signe ∞ et ne seront plus fonction de la distance r , de telle sorte que le vecteur d'antenne $\mathbf{V}^\infty(\mathbf{r}, k) = \mathbf{V}^\infty(\theta, \psi, k)$ en champ lointain s'écrit $\mathbf{V}^\infty(\theta, \psi, k) \triangleq (V_1^\infty(\theta, \psi, k), \dots, V_N^\infty(\theta, \psi, k))^T$, chaque élément $V_n^\infty(\theta, \psi, k)$ satisfaisant $V_n^\infty(\theta, \psi, k) = V_n^\infty(\mathbf{r}, k) \triangleq \lim_{r \rightarrow \infty} V_n(\mathbf{r}, k)$.

A partir de maintenant, pour simplifier l'étude, nous supposons que l'antenne considérée est linéaire et que les N microphones qui la constituent sont placés aux abscisses $z_1, \dots, z_N$ de l'axe $\mathcal{Z}$, cf. figure 2.14. Le problème présentant une symétrie de révolution autour de l'axe $\mathcal{Z}$,

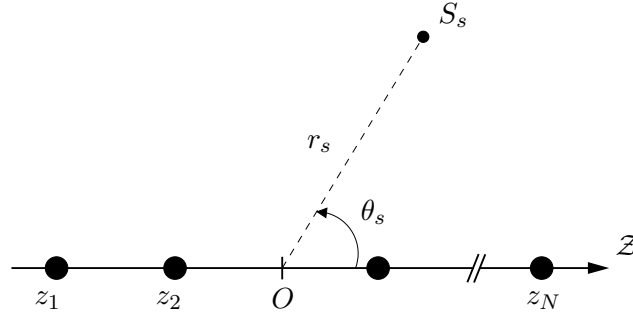


FIG. 2.14 – Organisation de l'antenne linéaire de microphones. Chacun des N microphones est placé à l'abscisse z_n , tandis que la $s^{\text{ième}}$ source émet depuis le point S situé en (r_s, θ_s) .

toutes les caractéristiques sont invariantes par rapport aux angles ψ , de sorte que le vecteur $\mathbf{r}$ peut être réduit à ses composantes $\mathbf{r} = (r, \theta)$. En outre, la distance $\|\mathbf{r}_n^m - \mathbf{r}_s^s\|$ entre la $s^{\text{ième}}$ source et le $n^{\text{ième}}$ microphone se simplifie en

$$\|\mathbf{r}_n^m - \mathbf{r}_s^s\| = \sqrt{r_s^2 + z_n^2 - 2r_s z_n \cos \theta_s}. \quad (2.21)$$

L'expression (2.18) du $n^{\text{ième}}$ élément du vecteur d'antenne devient alors

$$V_n(\mathbf{r}, k) = V_n(r, \theta, k) = \frac{r e^{jkr}}{\sqrt{r^2 + z_n^2 - 2r z_n \cos \theta}} e^{-jk\sqrt{r^2 + z_n^2 - 2r z_n \cos \theta}}. \quad (2.22)$$

Enfin, en champ lointain, (2.22) se particularise en

$$V_n^\infty(\theta, k) = e^{-jkz_n \cos \theta}. \quad (2.23)$$

Du fait que sous ces hypothèses toutes les sources se situent à une distance infinie, l'expression (2.14) du signal $m_n(t)$, noté maintenant $m_n^\infty(t)$ reçu par le $n^{\text{ième}}$ microphone se transforme en

$$m_n^\infty(t) = \sum_{s=1}^S s_s^0 \infty \left(t - \tau_n^\infty(\theta_s) \right) + b_n(t), \quad (2.24)$$

où $s_s^0 \infty(t)$ désigne le signal fictif en O causé par la $s^{\text{ième}}$ source située en champ lointain. Le délai lié à la propagation $\tau_n(\mathbf{r})$, d'expression (2.16) en champ proche, s'écrit maintenant $\tau_n^\infty(\mathbf{r}) = \tau_n^\infty(\theta)$ en champ lointain et s'exprime selon

$$\tau_n^\infty(\theta) = \frac{z_n}{c} \cos \theta, \quad (2.25)$$

comme le laissait supposer (2.23).

2.2.2 Une méthode haute résolution : MUSIC

La méthode d'analyse spectrale paramétrique MUSIC, pour Multiple Signal Classification, proposée dans [Schmidt 1979], permet une amélioration significative de la localisation d'une source. A ce titre, elle fait partie des méthodes dites "à haute résolution", dont elle est certainement la plus célèbre. Nous proposons tout d'abord de la décrire, puis de détailler son utilisation dans le contexte de la robotique.

2.2.2.1 Description de l'approche

On admet que la scène sonore comprend S sources sonores faible bande, de même fréquence k ainsi qu'une antenne de N microphones telle que celle représentée figure 2.14, avec $S < N$. Les sources sont supposées assimilables à des processus centrés, stationnaires et ergodiques. Dans la suite, nous omettons la dépendance à la fréquence k afin de simplifier les notations. Rappelons ici l'équation (2.19), permettant d'exprimer les signaux reçus sur les microphones, en fonction des signaux sources et du bruit

$$\mathbf{M} = \mathcal{V}\mathbf{S}^0 + \mathbf{B},$$

où $\mathcal{V}$ désigne la matrice d'antenne de dimension $(N \times S)$ définie par (2.20). Le bruit additif $\mathbf{B} = \mathbf{B}(k)$ est supposé spatialement et temporellement blanc, stationnaire, et de même puissance σ_b^2 sur chacun des microphones, de telle sorte que sa matrice de covariance Γ_B , de dimension $N \times N$ vérifie

$$\Gamma_B \triangleq E[\mathbf{B}\mathbf{B}^H] = \sigma_b^2 \mathcal{I}_N, \quad (2.26)$$

où $\mathcal{I}_N$ désigne la matrice identité de dimension N et $E[\cdot]$ l'espérance mathématique. Ce bruit est en outre supposé indépendant des sources, auquel cas

$$E[\mathbf{S}^0 \mathbf{B}^H] = \mathbf{0}. \quad (2.27)$$

Le vecteur $\mathbf{G} = \mathcal{V}\mathbf{S}^0 = (\mathbf{V}(\mathbf{r}_1^s) | \dots | \mathbf{V}(\mathbf{r}_S^s))\mathbf{S}^0$, de dimension $(N \times 1)$, appartient au sous-espace $\mathbb{G}_S$ de $\mathbb{C}^N$ engendré par les vecteurs d'antenne $\mathbf{V}(\mathbf{r}_s^s)$ ($s = 1 \dots S$), également appelé *sous-espace signal*. Nous noterons $\mathbb{G}_B$ le sous-espace supplémentaire, appelé aussi *sous-espace*

bruit. Dans la suite, nous supposons que les signaux sources ne sont pas corrélés, ce qui implique que leur matrice de covariance $\Gamma_{S^0} \triangleq E[\mathbf{S}^0 \mathbf{S}^{0H}]$ est non singulière. Sous cette hypothèse, et en considérant que les vecteurs d'antenne $\mathbf{V}(\mathbf{r}_s^s)$, $s = 1, \dots, S$, forment une famille libre, le sous-espace signal $\mathbb{G}_S$ est de dimension égale à S .

L'algorithme MUSIC est basé sur la décomposition en éléments propres de la matrice de covariance $\Gamma_M \triangleq E[\mathbf{M}\mathbf{M}^H]$ des signaux issus des microphones, aussi appelée matrice interspectrale. Celle-ci s'écrit, en raison de l'hypothèse d'indépendance (2.27)

$$\Gamma_M = \Gamma_S + \Gamma_B, \text{ avec } \Gamma_S = \mathcal{V} \Gamma_{S^0} \mathcal{V}^H \text{ et } \Gamma_B = \sigma_b^2 \mathcal{I}_N. \quad (2.28)$$

Du fait que la matrice $\mathcal{V}$ est supposée de rang S , la matrice hermitienne $\Gamma_S \in \mathbb{C}^{N \times N}$, semi-définie positive par construction, est de rang $S < N$. Il en résulte que son noyau est de dimension $N - S$. Par conséquent, la décomposition en éléments propres de Γ_S conduit à S valeurs propres réelles strictement positives $\lambda_n, n = 1, \dots, S$, correspondant aux S sources, ainsi qu'à $(N - S)$ valeurs propres nulles $\lambda_n, n = S + 1, \dots, N$. Notons $\mathbf{U}_n$ le vecteur propre à droite associé à la valeur propre $\lambda_n, n = 1, \dots, N$. La matrice interspectrale Γ_M est elle aussi une matrice semi-définie positive par construction, et admet donc N valeurs propres réelles positives ou nulles associées à des vecteurs propres pouvant être choisis orthogonaux entre eux. D'après (2.28), les vecteurs propres de Γ_S sont aussi vecteurs propres de Γ_M , mais avec comme valeur propre l_n associée

$$l_n = \lambda_n + \sigma_b^2, \quad n \in [1, \dots, N]. \quad (2.29)$$

L'ordonnancement des valeurs propres de Γ_M dans l'ordre croissant permet donc de distinguer les deux sous-espaces vectoriels $\mathbb{G}_S$ et $\mathbb{G}_B$: le sous-espace signal $\mathbb{G}_S$ est engendré par les vecteurs $\mathbf{U}_n$ ($n \in [1, \dots, S]$) associés aux S plus grandes valeurs propres de Γ_M , et le sous-espace bruit $\mathbb{G}_B$ est engendré par les vecteurs $\mathbf{U}_n$ ($n \in [S+1, \dots, N]$) associés aux $N - S$ valeurs propres de Γ_M égales à σ_b^2 .

C'est sur cette propriété qu'est fondée l'estimation de l'angle d'incidence – voire de la distance – des sources par la méthode MUSIC. En effet, d'après la propriété d'orthogonalité des vecteurs propres de Γ_M , tout vecteur $\mathbf{X}$ du sous-espace bruit $\mathbb{G}_B$ est orthogonal à tout vecteur de l'espace signal $\mathbb{G}_S$. Un tel vecteur $\mathbf{X}$ est donc en particulier orthogonal à tous les vecteurs d'antenne $\mathbf{V}(\mathbf{r}_1^s), \dots, \mathbf{V}(\mathbf{r}_S^s)$, i.e.

$$\forall \mathbf{X} \in \mathbb{G}_B, \quad \forall s \in [1, \dots, S], \quad \mathbf{V}(\mathbf{r}_s^s)^H \mathbf{X} = 0. \quad (2.30)$$

Il s'en suit également que les vecteurs d'antenne $\mathbf{V}(\mathbf{r}_s^s)$ vérifient

$$\forall s \in [1, \dots, S], \quad \sum_{i=S+1}^N |\mathbf{V}(\mathbf{r}_s^s)^H \mathbf{U}_i|^2 = 0. \quad (2.31)$$

L'idée de la méthode MUSIC est donc d'exploiter le fait que les vecteurs d'antenne, aux directions des différentes sources prises séparément, sont orthogonaux à l'espace bruit déterminé par la décomposition spectrale de la matrice de covariance des signaux. En pratique, Γ_M ne peut être qu'approximée à partir des enveloppes complexes déterminées aux instants d'échantillonnage t_l , avec $l \in [1, \dots, L]$, e.g. en définissant son estimée,

$$\hat{\Gamma}_M = \frac{1}{L} \sum_{l=1}^L \mathbf{m}(t_l) \mathbf{m}(t_l)^H, \text{ avec } \mathbf{m}(t) = (m_1(t), \dots, m_N(t))^T, L \geq N. \quad (2.32)$$

Les vecteurs propres $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_N$ de $\hat{\Gamma}_M$ et l'espace bruit estimé $\hat{\mathbb{G}}_B$ ne vérifient pas parfaitement les propriétés théoriques (2.30) et (2.31). Cependant, la fonction

$$P(\mathbf{r}) = \frac{1}{\sum_{i=S+1}^N |\mathbf{V}(\mathbf{r})^H \hat{\mathbf{U}}_i|^2} = \frac{1}{\mathbf{V}(\mathbf{r})^H \Pi_B \mathbf{V}(\mathbf{r})}, \quad (2.33)$$

appelée *pseudo-spectre* peut être utilisée pour estimer les positions des sources $\hat{\mathbf{r}}_s^s$, $s = 1, \dots, S$ à partir de ses maxima. La matrice $\Pi_B = \sum_{i=S+1}^N \hat{\mathbf{U}}_i \hat{\mathbf{U}}_i^H$ est quant à elle nommée *projecteur dans l'espace bruit*. Finalement la méthode MUSIC peut être synthétisée en les étapes suivantes :

1. Acquisition de L N -uplets $\mathbf{m}(t_1), \dots, \mathbf{m}(t_L)$ par les microphones ;
2. Calcul de l'estimée $\hat{\Gamma}_M$ de la matrice interspectrale,
3. Décomposition de $\hat{\Gamma}_M$ en les valeurs propres $l_1 \geq l_2 \dots \geq l_N$ et les vecteurs propres $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_N$;
4. Séparation des espaces signal $\hat{\mathbb{G}}_S$ et bruit $\hat{\mathbb{G}}_B$, et calcul du projecteur Π_B ;
5. Définition du pseudo-spectre $P(\mathbf{r})$, e.g. sur un ensemble fini de valeurs de $\mathbf{r}$;
6. Et enfin, détection des positions en tant que les maxima de $P(\mathbf{r})$.

2.2.2.2 Exploitation en robotique

A notre connaissance, l'unique implémentation de MUSIC en robotique a été proposée dans [Asano 1999], sur le robot Jijo-2 présenté figure 2.15(a). Celui-ci est équipé de 8 microphones répartis sur sa circonférence. La localisation de sources telles que la voix est effectuée en adaptant l'algorithme présenté précédemment au cas large bande. Dans un premier temps, des pseudo-spectres $P(\mathbf{r}, k)$ sont déterminés pour toutes les fréquences k comprises dans un domaine $[k_L; k_H]$ sélectionné par une transformée de Fourier. La localisation est ensuite basée sur le pseudo-spectre moyen

$$\bar{P}(\mathbf{r}) = \sum_{k=k_L}^{k_H} P(\mathbf{r}, k) = \sum_{k=k_L}^{k_H} \frac{1}{\mathbf{V}(\mathbf{r}, k)^H \Pi_B \mathbf{V}(\mathbf{r}, k)}. \quad (2.34)$$

Néanmoins, [Asano 1999] propose de remplacer dans les calculs des pseudo-spectres le vecteur $\mathbf{V}(\mathbf{r}, k)$ défini en (2.18) par

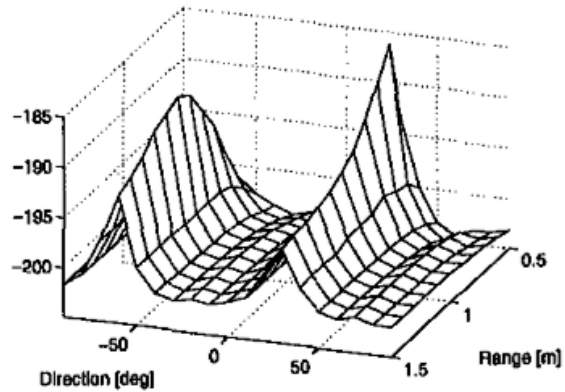
$$\bar{\mathbf{V}}(\mathbf{r}, k) = (e^{-jk\tau_1(\mathbf{r})}, \dots, e^{-jk\tau_N(\mathbf{r})}), \quad (2.35)$$

de façon à ne prendre en compte que la différence en phase entre chacun des microphones. Au final, les signaux reçus sur les microphones, échantillonnés à 16 kHz, sont d'abord décomposés dans le domaine fréquentiel par une FFT calculée sur 1024 points, à partir de fenêtres d'observation de 512 points décalées successivement de 160 points. L'estimation de la covariance (2.32) s'effectue sur $L = 16$ valeurs, et conduit au résultat de localisation présenté figure 2.15(b), où $f_L = 500$ Hz et $f_H = 3$ kHz. Les deux sources semblent correctement localisées, à la fois en azimut et en distance, mais aucun détail n'est donné sur la précision obtenue.

Si ces résultats permettent d'estimer à la fois l'azimut et la distance des sources, c'est au prix de calculs complexes qui ne peuvent probablement pas être effectués en temps réel. [Asano 1999] ne détaille absolument pas l'implémentation de l'algorithme, et il est fort probable que ses résultats soient obtenus hors ligne à partir de relevés expérimentaux. En effet, MUSIC étendu de cette façon au cas large bande nécessite, pour chacune des fréquences constituant les signaux, de calculer les matrices de covariance et d'en effectuer la décomposition en éléments propres.



(a) Robot Jijo-2, équipé de 8 microphones



(b) Résultats de localisation obtenus avec MUSIC

FIG. 2.15 – [Asano 1999] - Utilisation de l'algorithme haute résolution MUSIC en robotique.

Compte tenu des domaines de fréquence mis en jeu ici et de la fréquence d'échantillonnage retenue, cela reviendrait ici à calculer environ 160 décompositions en éléments propres de matrices de dimension 8×8 en seulement $160ms$. A moins d'utiliser une CPU puissante difficilement embarquable sur un robot, une telle charge de calcul n'est pas envisageable. C'est pour cette raison que MUSIC n'a quasiment jamais été utilisé en robotique, malgré la précision des résultats obtenus. Cependant, des résultats récents issus de la communauté du traitement du signal permettent aujourd'hui d'envisager l'utilisation de MUSIC pour des signaux large bande, tout en réduisant drastiquement la puissance de calcul nécessaire. Nous discuterons de ces extensions dans le chapitre 4 de cette thèse.

2.2.3 Localisation par corrélation

Tout comme un son arrive à nos deux oreilles décalé temporellement à cause de la propagation, l'échantillonnage spatial opéré par une antenne de microphones est à l'origine de décalages temporels, aussi appelés *Time Delay Of Arrival* ou TDOA, entre chacun des signaux disponibles en sortie des microphones. Les méthodes de localisation de source présentées ici visent à exploiter l'information spatiale incluse dans ces délais pour remonter à l'origine de la source sonore. Pour cela, deux étapes sont indispensables :

- La première étape consiste à déterminer les TDOAs entre chaque paire de microphones. Nous présentons dans une première partie la méthode la plus immédiate, basée sur les intercorrélations entre les signaux mesurés.
- La seconde étape repose sur l'expression théorique, à partir d'un modèle de propagation, des TDOAs en fonction de la position de la source. A partir de leurs mesures expérimentales, il est alors possible de réaliser la localisation par inversion de modèle. Nous présentons dans une seconde partie les solutions géométriques proposées par les roboticiens à ce problème.

Notons la ressemblance immédiate entre la notion de TDOA, habituellement utilisée en antennerie, et celle de l'ITD (ou IPD dans le cas monofréquentiel), réservée plutôt aux approches bio-inspirées. Ces deux grandeurs traduisent pourtant la même réalité physique, si bien que l'utilisation de TDOAs pourrait être perçue comme une approche biomimétique. Certains modèles laissent penser que le cerveau effectue une opération de corrélation, au moyen d'un mécanisme de lignes à retard neuronales, pour déterminer cet indice interaural [Purves 2004]. Cependant, la méthode de calcul des TDOAs présentée ici, de même que son exploitation pour la localisation, est très éloignée des opérations effectuées effectivement par le cerveau. C'est pourquoi nous avons choisi de classer ces approches dans la partie relative à l'antennerie.

2.2.3.1 Détermination des TDOAs

Généralités La méthode la plus classique pour la détermination du décalage temporel entre deux signaux $m_i(t)$ et $m_j(t)$ perçus par deux microphones i et j , repose sur la définition de la fonction d'intercorrélation $R_{ij}(\tau)$ dont l'expression est

$$R_{ij}(\tau) = E[m_i(t)m_j(t - \tau)]. \quad (2.36)$$

Supposons, pour simplifier les notations, qu'une seule source sonore soit présente dans l'environnement et émette un signal $s(t)$. D'après la formulation analytique des signaux $m_i(t)$ et $m_j(t)$ issue de (2.14), la fonction d'intercorrélation $R_{ij}(\tau)$ est reliée à l'autocorrélation $R_{ss}(\tau)$ de la source, l'intercorrélation $R_{b_ib_j}$ du bruit et le TDOA $\Delta T_{ij} = \tau_i - \tau_j$ entre les micros i et j par

$$R_{ij}(\tau) = \alpha_i \alpha_j R_{ss}(\tau - \Delta T_{ij}) + R_{b_ib_j}(\tau). \quad (2.37)$$

Par conséquent, ΔT_{ij} est la valeur de τ qui maximise la fonction $R_{ij}(\tau)$. Seulement, la définition (2.36) requiert une durée d'observation des signaux infinie. En pratique, seule une estimée $\hat{R}_{ij}(\tau)$ peut être calculée, e.g. sur une fenêtre d'observation centrée sur l'instant t_0 et de durée T par [Omologo 1994]

$$\hat{R}_{ij}(\tau) = \frac{1}{T} \int_{t_0 - \frac{T}{2}}^{t_0 + \frac{T}{2}} m_i(t)m_j(t - \tau)dt. \quad (2.38)$$

Comme les signaux issus des microphones sont généralement numérisés à une période d'échantillonnage T_e , l'observation sur une durée T revient donc à travailler sur un nombre fini P d'échantillons tel que $T = PT_e$. L'équation (2.38) devient alors

$$\hat{R}_{ij}[\tau'] = \frac{1}{P} \sum_{p=0}^{P-1} m_i[p]m_j[p - \tau'], \quad (2.39)$$

avec $m_i[p] = m_i(pT_e)$ et $\hat{R}_{ij}[\tau'] = \hat{R}_{ij}(\tau'T_e)$. Le problème du calcul de la fonction d'intercorrélation par (2.39) réside dans sa complexité en $\mathcal{O}(P^2)$. Cependant, il demeure possible d'exhiber une approximation dans le domaine fréquentiel au moyen de la transformée de Fourier inverse de l'interspectre, réduisant ainsi sa complexité à $\mathcal{O}(P \log_2 P)$ [Valin 2003]. Cette approximation dans le domaine spectral s'exprime par

$$\hat{R}_{ij}[\tau'] = \sum_{k=0}^{P-1} M_i[k]M_j^*[k]e^{j2\pi k\tau'/P}, \quad (2.40)$$

où $M_i[k]$ désigne la transformée de Fourier discrète du signal numérique $m_i[p]$ définie par $M_i[k] = \sum_{p=0}^{P-1} m_i[p] e^{-j(2\pi/P)kp}$, $k \in \mathbb{Z}$. La détermination des TDOAs ΔT_{ij} peut donc reposer sur (2.40). Or, la fonction d'intercorrélation dépend fortement des caractéristiques temporelles et fréquentielles des signaux. En particulier, pour des signaux composés majoritairement de basses fréquences, les échantillons adjacents sont fortement corrélés, entraînant des "pics" pouvant être larges dans la fonction $\hat{R}_{ij}[\tau']$. L'augmentation de la durée d'observation T peut permettre de résoudre ce problème. Plus généralement, il s'avère que le choix de la valeur T est critique quant aux performances des méthodes de détermination des TDOAs ΔT_{ij} à partir des pics de la fonction d'intercorrélation [Omologo 1994].

Une première amélioration [Silverman 1992] suggère l'introduction de l'*intercorrélation normalisée* (Normalized Cross-Correlation, ou NCC)

$$\hat{R}_{ij}^{(NCC)}[\tau'] \propto \frac{\hat{R}_{ij}[\tau']}{\sqrt{\sum_{p=0}^P m_i^2[p]} \sqrt{\sum_{p=0}^P m_j^2[p - \tau']}}. \quad (2.41)$$

Cette normalisation par les énergies des signaux sur la fenêtre d'observation permet de faire ressortir sensiblement le pic de corrélation. Or, si la source sonore est périodique de période T_k (e.g. la voix, pour les sons voisés), des micropériodicités apparaissent également dans la fonction d'intercorrélation normalisée. Ces multiples pics de corrélations, situés en $\tau = \Delta T_{ij} + nT_k$, $n \in \mathbb{Z}$, rendent alors la détection du véritable TDOA très difficile.

Généralisation Pour résoudre ce problème, [Omologo 1994] rappelle les méthodes d'*intercorrélation généralisée* (Generalized Cross-Correlation, ou GCC), introduites pour la première fois dans [Knapp 1976]. Celles-ci adjoignent à la définition (2.40) un terme de pondération fréquentielle discret $\Psi[k]$, conduisant à une nouvelle fonction d'intercorrélation $R_{ij}^{(GCC)}[\tau']$ dont l'expression est

$$R_{ij}^{(GCC)}[\tau'] = \sum_{k=0}^{P-1} \Psi[k] M_i[k] M_j^*[k] e^{j2\pi k\tau'/P}. \quad (2.42)$$

Toute la difficulté réside donc dans le choix de la fonction de pondération $\Psi(k)$. [Knapp 1976] propose différentes alternatives qui aboutissent à de multiples améliorations de la résolution de la fonction d'intercorrélation. La plus célèbre, appelée Phase Transform (PHAT), consiste à blanchir les spectres des signaux reçus au moyen de la fonction $\Psi^{(PHAT)}[k]$ définie par [Omologo 1997]

$$\Psi^{(PHAT)}[k] = \frac{1}{|M_i[k]| |M_j[k]|}. \quad (2.43)$$

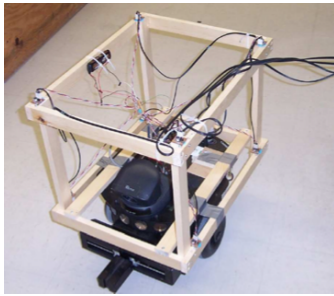
Le blanchiment ainsi réalisé permet donc de ne prendre en compte que la phase des signaux reçus dans le calcul de la fonction d'intercorrélation, en donnant la même importance à chacune des fréquences les constituant. Toutefois, cette pondération ne différencie pas les fréquences émises par la source et le bruit. Ceci augmente la sensibilité de la méthode au bruit et rend la détection de la voix plus difficile.

En guise de solution, [Valin 2006] définit une nouvelle fonction de pondération pénalisant les fréquences où le rapport signal sur bruit est le plus faible. Appelée *Reliability-Weighted Phase Transform* (RWPHAT), cette méthode aboutit à la fonction de pondération $\Psi^{(RWPHAT)}[k]$

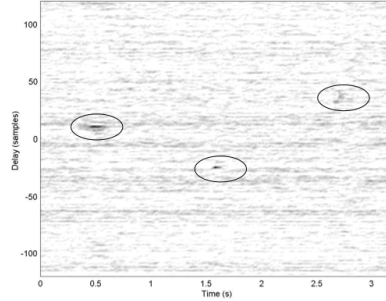
$$\Psi^{(RWPHAT)}[k] = \zeta_i[k]\zeta_j[k], \text{ avec } \zeta_n[k] = \frac{\varepsilon_n[k]}{\varepsilon_n[k] + 1}, \quad (2.44)$$

où $\varepsilon_n[k]$ désigne une estimée a priori du rapport signal sur bruit au niveau du $n^{\text{ième}}$ microphone pour la fréquence k/PT_e , évaluée sur une fenêtre d'observation. Cette pondération permet de masquer les fréquences dominées par le bruit, et produit une fonction d'intercorrélation $R^{(RWPHAT)}[\tau']$ moins bruitée que celle obtenue avec (2.43), comme l'atteste la figures 2.16.

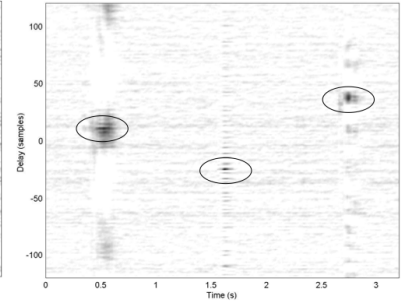
Notons enfin que d'autres méthodes cherchent à améliorer les propriétés de la fonction d'intercorrélation en présence de réverbération dans l'environnement. Citons par exemple [Ma 2004] qui propose une autre pondération adaptative conduisant à une meilleure robustesse vis à vis de ces perturbations que PHAT pour des temps de réverbération importants en environnement très bruité (RSB < 0 dB).



(a) Robot Pioneer 2, équipé de son antenne de microphones



(b) Fonction d'intercorrélation obtenue par PHAT



(c) Fonction d'intercorrélation obtenue par RWPHAT

FIG. 2.16 – [Valin 2003] - Robot Pioneer2. Comparaison des résultats de corrélation obtenus avec PHAT et RWPHAT.

Exploitation en robotique L'utilisation des méthodes avancées de corrélation présentées précédemment est assez nouvelle en robotique. Jusqu'à très récemment, le décalage temporel entre deux signaux a souvent été déterminé soit par la mesure des points de passage à zéro des signaux issus des microphones, soit par l'utilisation de la fonction d'intercorrélation classique. Citons par exemple [Cavaco 1999], qui propose de décomposer fréquemment les signaux au moyen d'une banque de filtres, puis de déterminer l'ITD entre deux microphones par la mesure des points de passage à zéro. [Okuyama 2002] exploite l'intercorrélation (2.39) de façon à mesurer les TDOAs entre quatre microphones, positionnés sur les sommets d'un tétraèdre. L'originalité réside ici dans la sélection de la fenêtre d'observation : plutôt que de calculer l'intercorrélation sur la totalité des signaux, un simple seuil permet de détecter les zones temporelles libres d'écho pendant lesquelles les TDOAs sont déterminés. Dans la même veine, [Murray 2005] calcule le TDOA entre deux microphones à partir d'une simple corrélation afin de déterminer l'azimut de la source.

En parallèle, [Wang 2004] propose d'utiliser PHAT sur une antenne de microphones disposés sur les murs d'une pièce devant être visitée par un robot-guide. Le but est de localiser ce robot grâce au discours enregistré qu'il émet depuis un haut parleur. La localisation acoustique est alors fusionnée avec les informations perçues par ses capteurs proprioceptifs, de façon à le recaler sur la carte de l'environnement. L'antenne utilisée est de grande taille (sa plus grande dimension

est de 3.2 m) et comporte 24 microphones régulièrement espacés. La pièce dans laquelle évolue le robot, ainsi que la plateforme elle-même sont visibles sur la figure 2.17. [Valin 2003] suggère une méthode de corrélation analogue à RWPHAT pour déterminer les TDOAs entre chacun des microphones situés sur les arêtes d'un cube transporté par la plateforme mobile Pioneer 2 présentée figure 2.16(a).

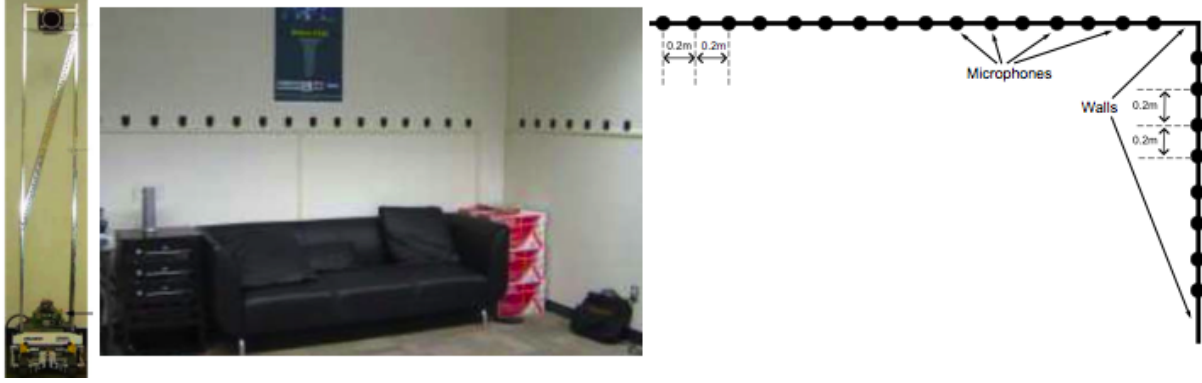


FIG. 2.17 – [Wang 2004] - Présentation du robot guide, et détails sur l'antenne de microphone.

2.2.3.2 Utilisation des TDOAs pour la localisation

Supposons maintenant que les TDOAs entre les différents microphones constituant l'antenne aient été obtenus par une des méthodes présentées précédemment. La seconde étape consiste à exploiter ces délais de façon à localiser la source. La méthode la plus immédiate consiste à considérer un dipôle constitué de deux microphones espacés d'une distance d . Sous ces conditions, et en considérant des fronts d'ondes plans, l'azimut de la source θ_s s'exprime en fonction du TDOA ΔT_{12} , d'après (2.25), par

$$\theta_s = \sin^{-1} \left(\frac{c\Delta T_{12}}{d} \right). \quad (2.45)$$

Cette simple considération géométrique est utilisée par [Murray 2005] pour établir l'azimut de la source sonore, qui alimente ensuite un réseau de neurones réalisant le suivi de celle-ci.

Le principe précédent peut être étendu simplement de façon à déterminer les coordonnées cartésiennes de la source à localiser. Notons $\mathbf{r}_s = (u, v, w)$ la position inconnue de cette source et $\mathbf{r}_n = (x_n, y_n, z_n)$ les positions connues de chacun des microphones constituant l'antenne. Supposons que la propagation s'effectue en espace libre. Dans une telle situation, les fronts d'ondes parvenant à chacun des microphones de l'antenne sont sphériques, de centre la source sonore. Sous l'hypothèse où chacune de ces sphères ne supporte qu'un seul des N microphones, les coordonnées (u, v, w) de la source et sa distance d au premier récepteur sont reliées aux TDOAs ΔT_{1i} entre le $i^{\text{ième}}$ et le premier microphone par

$$\forall i \in [1, \dots, N], (x_i - u)^2 + (y_i - v)^2 + (z_i - w)^2 = (d + c\Delta T_{1i})^2, \quad (2.46)$$

Quelques manipulations sur (2.46) permettent alors d'aboutir au système

$$\begin{pmatrix} 2(x_1 - x_2) & 2(y_1 - y_2) & 2(z_1 - z_2) & -2c\Delta T_{12} \\ 2(x_1 - x_3) & 2(y_1 - y_3) & 2(z_1 - z_3) & -2c\Delta T_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 2(x_1 - x_N) & 2(y_1 - y_N) & 2(z_1 - z_N) & -2c\Delta T_{1N} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ d \end{pmatrix} = \begin{pmatrix} c^2\Delta T_{12}^2 + x_1^2 + y_1^2 + z_1^2 - x_2^2 - y_2^2 - z_2^2 \\ c^2\Delta T_{13}^2 + x_1^2 + y_1^2 + z_1^2 - x_3^2 - y_3^2 - z_3^2 \\ \vdots \\ c^2\Delta T_{1N}^2 + x_1^2 + y_1^2 + z_1^2 - x_N^2 - y_N^2 - z_N^2 \end{pmatrix}, \quad (2.47)$$

qui, une fois "inversé", conduit aux inconnues (u, v, w, d) . Cette méthode est proposée dans [Mahajan 2001] pour la mesure de temps de vol d'ondes ultrasoniques, mais est tout à fait envisageable dans le domaine des fréquences audibles. Notons qu'il est nécessaire que l'antenne comprenne au moins 5 microphones pour que 2.47 ne soit pas sous-déterminé. Un plus grand nombre de microphones peut être introduit de façon à estimer également la vitesse du son c . Toutefois, rien ne garantit le bon conditionnement des matrices mises en jeux, si bien que l'inversion du système conduit parfois, malgré des valeurs de TDOA proches, à des positions estimées très différentes.

C'est pour cette raison que [Valin 2003] propose un modèle plus simple, analogue à celui utilisé dans [Okuyama 2002], et faisant l'hypothèse d'ondes planes. Constatant que le vecteur unitaire $\nu = (u', v', w')$ pointant en direction de la source (supposée donc à l'infini), et le vecteur $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ reliant le microphone i au microphone j , satisfont

$$\forall i \in [1, \dots, N], \quad \nu \cdot \mathbf{r}_{i1} = c\Delta T_{i1}, \quad (2.48)$$

les variables u', v', w' s'obtiennent par "inversion" du système de $N - 1$ équations

$$\begin{pmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \\ \vdots & \vdots & \vdots \\ x_N - x_1 & y_N - y_1 & z_N - z_1 \end{pmatrix} \begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \begin{pmatrix} c\Delta T_{12} \\ c\Delta T_{13} \\ \vdots \\ c\Delta T_{1N} \end{pmatrix}. \quad (2.49)$$

Le principal avantage de (2.49) sur (2.47) réside dans la forme de la matrice à inverser. Celle-ci ne dépend plus que de la position des microphones, et peut par conséquent être ajustée de façon à garantir son bon conditionnement. De plus, une fois la géométrie du capteur décidée, le résultat de son inversion est constant et peut donc être mémorisé pour réduire la charge de calcul nécessaire à la localisation. Cependant, n'oublions pas que le modèle de propagation utilisé pour (2.49) suppose que les fronts d'onde parvenant à l'antenne sont plans.

Bien que les localisations obtenues soient de bonne qualité, elles dépendent fortement du type de signal émis par la source. Evidemment, l'étape critique se situe lors de la détermination des TDOAs, et aucune localisation précise ne peut être obtenue sans en effectuer une mesure de qualité. Ainsi, RWPHAT ne peut être exploité pour déterminer les TDOAs d'une source faible bande : les informations de position sont en effet contenue sur un nombre réduit de fréquence, et le blanchiment des transformées de Fourier rend la fonction de corrélation fortement bruitée. Cependant, pour des sons large bande, [Valin 2003] couple RWPHAT et (2.49) et aboutit à une précision angulaire de l'ordre de 3° pour une distance comprise entre 3 et 5 mètres. [Okuyama 2002] rapporte des erreurs angulaires plus fortes, d'environ 5° . L'emploi de la corrélation classique pour la détermination des TDOAs peut expliquer ce résultat moyen, malgré l'isolation des zones libres d'écho dans les signaux. La méthode pour remonter à la localisation de la source depuis les TDOAs n'est toutefois pas précisée par l'auteur. Enfin, [Mahajan 2001] montre que l'estimation

de la distance d par inversion de (2.47), conduit à une précision d'environ quelques millimètres, mais pour des signaux ultrasonores.

2.2.4 Formation de voie

De toutes les méthodes issues du traitement du signal, les approches par formation de voie sont probablement les plus utilisées en robotique. Leur simplicité ainsi que leur faible coût calculatoire les rend en effet a priori particulièrement bien adaptées à ce contexte. Pour autant, comme nous le verrons par la suite, leurs performances en localisation et en résolution dépendent très fortement du capteur sonore utilisé, en particulier de sa taille et du nombre de ses microphones. Nous nous proposons de débiter cette partie par quelques définitions et généralités sur la formation de voie, puis ensuite de parcourir une grande partie des résultats l'exploitant pour la localisation de source sonore en robotique mobile.

2.2.4.1 Principe de la formation de voie

La formation de voie est une méthode permettant la focalisation d'une antenne de microphones dans une direction spécifique de l'espace. Les signaux sonores provenant d'une direction spatiale d'intérêt sont ainsi amplifiés, tout en atténuant les autres directions d'incidence, dans l'objectif de localiser ou d'isoler (extraire) une ou plusieurs sources sonores. Typiquement, une formation de voie exploite les signaux $m_n(t)$, $n = 1, \dots, N$ des N microphones d'une antenne en les combinant linéairement par l'intermédiaire de filtres de réponse impulsionnelle $w_n(t)$, de façon à obtenir un signal de sortie $y(t)$ résultat d'une sommation constructive des signaux. Ce principe est illustré figure 2.18.

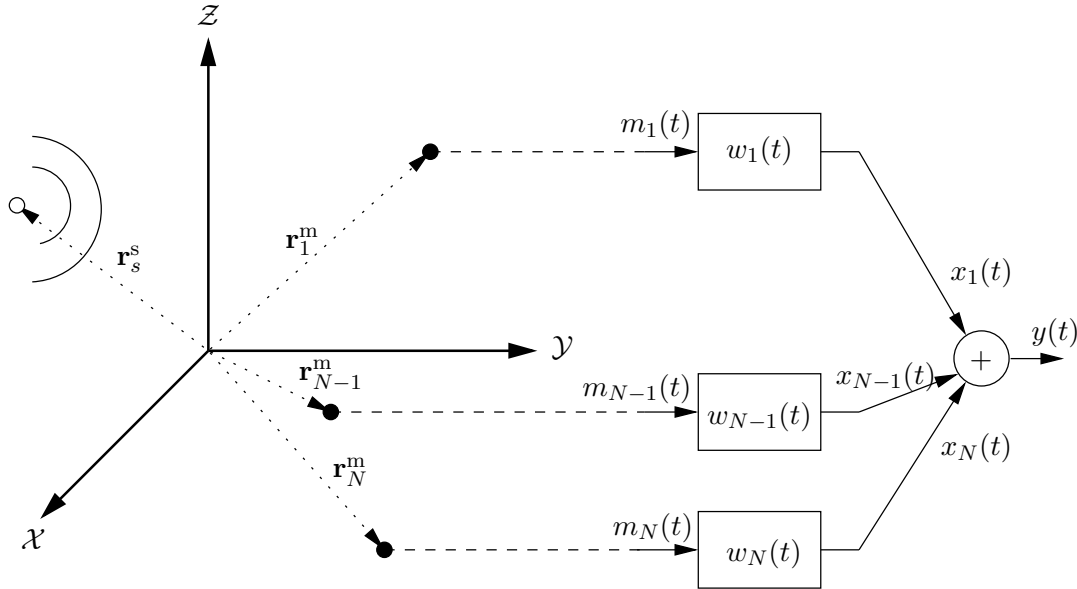


FIG. 2.18 – Principe de la formation de voie

Nous pouvons donc écrire

$$y(t) = \sum_{n=1}^N w_n(t) * m_n(t) \quad (2.50)$$

où $*$ désigne le produit de convolution. Si nous notons $S_s^0(k)$, $B_n(k)$ et $Y(k)$ les transformées de Fourier de $s^0(t)$, $b_n(t)$ et $y(t)$, alors (2.50) se transforme à l'aide de (2.17) en

$$Y(k) = \sum_{s=1}^S D(\mathbf{r}_s, k) S_s^0(k) + \sum_{n=1}^N W_n(k) B_n(k), \quad (2.51)$$

où la *réponse spatio-fréquentielle* $D(\mathbf{r}, k)$ de la formation de voie, aussi appelée *diagramme de directivité*, s'exprime par

$$D(\mathbf{r}, k) = \sum_{n=1}^N W_n(k) V_n(\mathbf{r}, k), \quad (2.52)$$

avec $W_n(k) = \int_{-\infty}^{\infty} w_n(t) e^{-jkct} dt$ la fonction de transfert du filtre attaché au $n^{\text{ième}}$ microphone.

Ainsi, tout comme un système linéaire est représenté par sa réponse en fréquence, une formation de voie est caractérisée par sa réponse spatio-fréquentielle $D(\mathbf{r}, k)$. Cette fonction des variables d'espace et de la fréquence peut être assimilée à une fonction de transfert normalisée entre une source située en $\mathbf{r}$ et la sortie de la formation de voie $y(t)$, et indique quelles zones de l'espace sont amplifiées ou atténuées. Remarquons que l'équation (2.52) fait appel aux éléments $V_n(\mathbf{r}, k)$; la réponse d'antenne ainsi définie est donc valable à la fois en champ proche et en champ lointain. Cependant, une forme plus simple $D^\infty(\theta, \psi, k)$ peut être exhibée en considérant des fronts d'ondes plans grâce au passage à la limite $D^\infty(\theta, \psi, k) = \lim_{r \rightarrow \infty} D(\mathbf{r}, k)$ déjà utilisé dans les parties précédentes.

Dans le cas de l'antenne linéaire présentée figure 2.14, pour laquelle chacun des N microphones est placé selon l'axe $\mathcal{Z}$ aux abscisses $(z_1, \dots, z_N)$, la réponse s'écrit [Van Veen 1988] [Krim 1996]

$$D^\infty(\theta, k) = \sum_{n=1}^N W_n(k) e^{-jkz_n \cos \theta}. \quad (2.53)$$

L'immense majorité des approches de la robotique basées sur la formation de voie formule cette hypothèse simplificatrice d'ondes planes. L'équation (2.53) est donc importante et il est opportun de caractériser dès maintenant l'efficacité du filtrage spatial ainsi obtenu. Dans ce but, afin de simplifier la présentation, nous supposons dans la suite que les N microphones alignés selon l'axe $\mathcal{Z}$ sont régulièrement espacés d'une distance d , de sorte que

$$z_n = \left(n - \frac{N+1}{2}\right)d, n = 1, \dots, N. \quad (2.54)$$

Dans ce cas, la longueur de l'antenne L est donc $L = (N-1)d$.

Considérons dans un premier temps que les fonctions de transfert $W_n(k)$ des filtres placés en aval des microphones de l'antenne vérifient

$$W_n(k) = 1, \forall k \in \mathbb{R}. \quad (2.55)$$

Nous envisagerons très bientôt d'autres choix pour $W_n(k)$. Sous l'hypothèse simplificatrice (2.55), l'expression (2.53) devient

$$D^\infty(\theta, k) = \sum_{n=1}^N e^{-jk(n - \frac{N+1}{2})d \cos \theta} = \frac{\sin(k \frac{Nd}{2} \cos \theta)}{\sin(k \frac{d}{2} \cos \theta)} = \frac{\sin(\frac{\pi f}{c} Nd \cos \theta)}{\sin(\frac{\pi f}{c} d \cos \theta)}. \quad (2.56)$$

Etude de l'influence des paramètres de l'antenne Plusieurs observations utiles pour une exploitation en robotique peuvent être établies en traçant le diagramme d'antenne $D^\infty(\theta, k)$ dans les configurations suivantes :

1. variation du nombre de microphones N , la longueur L de l'antenne et la fréquence k (ou f) étant fixées ;
2. variation de la longueur L , pour N et k constants ;
3. variation de la fréquence k pour N et L constants.

Une telle étude, très classique, est menée entre autres dans [McCowan 2001]. La figure 2.19(a) présente le module de la réponse de l'antenne pour le premier de ces scénarii, avec $f = 1$ kHz et $L = 0.7$ m. Nous observons que l'antenne est globalement sensible à des ondes provenant de la direction $\theta = 90^\circ$. Le choix de filtres passe-tout (2.55) conduit en effet naturellement à un déphasage nul entre les signaux issus des microphones, et donc à leur sommation constructive, si et seulement si la source se situe en face de l'antenne.

Le principal effet de l'augmentation du nombre de microphones pour une taille d'antenne fixe est la diminution du niveau des lobes secondaires. Dans cette situation, l'espacement entre chacun des microphone diminue, de sorte que la fréquence d'échantillonnage spatial augmente. Ainsi, de façon à rejeter au maximum les signaux sonores provenant de directions autres que celle d'intérêt, il est nécessaire de travailler avec une antenne constituée d'un grand nombre de microphones. Naturellement en pratique, les dimensions géométriques minimales des microphones viennent limiter en pratique leur nombre.

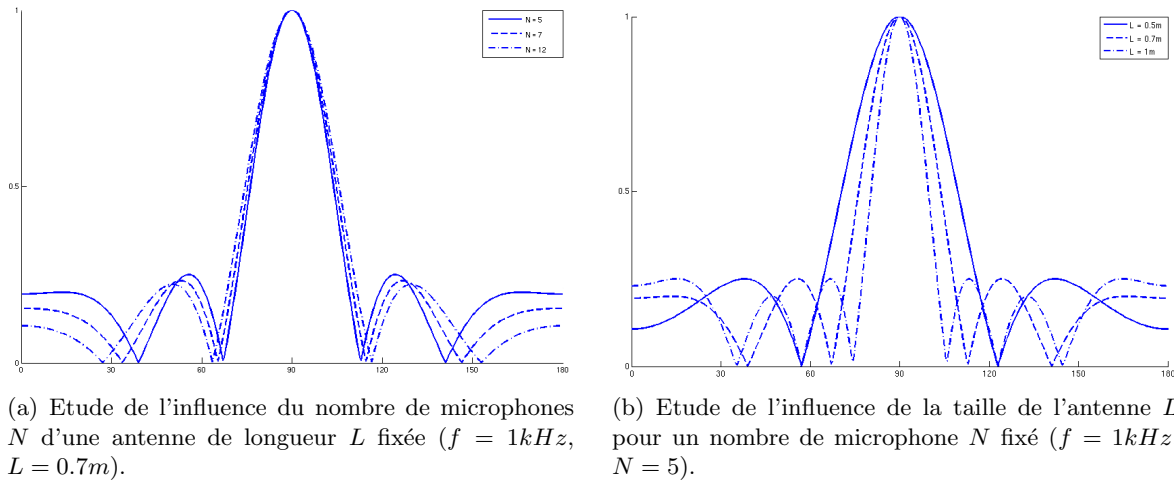


FIG. 2.19 – Réponse normalisée de l'antenne pour différentes valeurs de N et L .

Le diagramme d'antenne pour le scénario 2 est représenté figure 2.19(b). Nous remarquons que la largeur du lobe principal diminue lorsque la taille de l'antenne augmente. Cette constatation est capitale : elle indique qu'il faudrait embarquer une antenne de très grande dimension afin de privilégier finement une direction de l'espace. Evidemment, les dimensions de l'antenne se trouvent souvent limitées par la géométrie de la plateforme mobile, et la résolution obtenue sera donc d'autant plus mauvaise que le capteur embarqué sera de petite taille. Plus généralement, un simple examen de l'équation (2.56) montre que la largeur du lobe principal varie de manière décroissante par rapport au produit fL . En résumé, pour une fréquence donnée, les deux caractéristiques fondamentales d'un diagramme d'antenne, à savoir la largeur du lobe

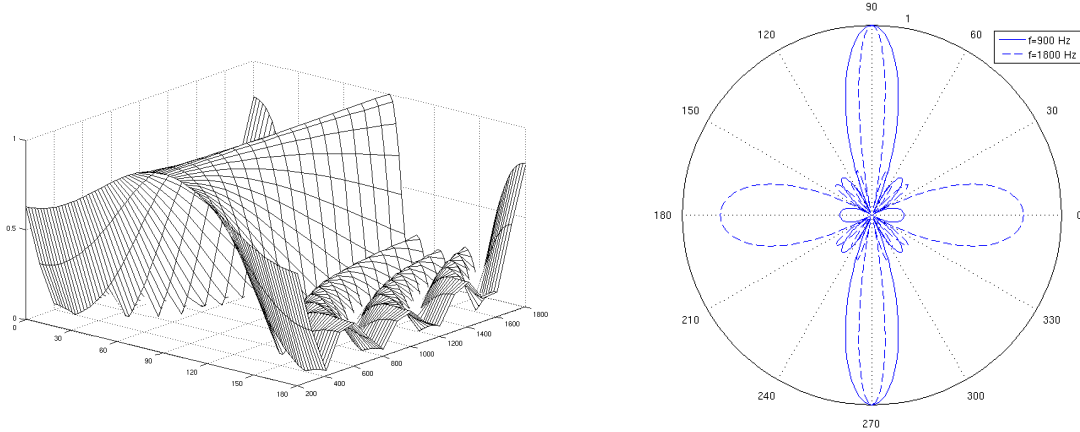
principal et le niveau des lobes secondaires, sont respectivement conditionnés par l'écartement des microphones et leur nombre. Ces propriétés sont à mettre en parallèle de celles constatées en traitement du signal, lors de l'acquisition de signaux : la taille de l'antenne joue le rôle de la fenêtre d'observation, tandis que l'espacement entre les microphones est analogue à la période d'échantillonnage. Cette analogie permet d'aboutir aux mêmes conclusions dans les domaines temporels (signaux échantillonnés) et spatial (formation de voie).

Enfin, le résultat du troisième et dernier scénario est représenté sur la figure 2.20(a). Ici, la longueur de l'antenne ainsi que le nombre de microphones sont maintenus constants, tandis que la fréquence f varie. Nous constatons que la largeur du lobe principal varie avec la fréquence. Ce phénomène pressenti précédemment est à l'origine d'une très faible résolution spatiale des basses fréquences : les ondes de grande longueur sont en quelque sorte sur échantillonnées, et ne peuvent par conséquent pas être localisées précisément. Nous retrouvons là encore une analogie avec le traitement de signaux numériques : l'observation sur une courte durée d'un signal basse fréquence ne permet pas la détermination précise de sa fréquence. Un second phénomène important touche cette fois les hautes fréquences : celles-ci sont sujettes à du repliement, à l'origine de l'apparition de multiples lobes de même amplitude que le lobe principal. L'échantillonnage de l'onde par l'antenne de microphones doit en effet obéir à un théorème de Shannon spatial, selon lequel l'espacement maximal d entre deux microphones successifs doit satisfaire

$$d < d_{max} = \frac{\lambda_{min}}{2} = \frac{c}{2f_{max}}, \quad (2.57)$$

où f_{max} désigne la fréquence maximale émise par la source. Le repliement est illustré figure 2.20(b) pour une antenne admettant $N = 5$ microphones espacés de $d = 17.5$ cm, soit une longueur totale $L = 0.7$ m. Dans ce cas, la fréquence de Shannon (2.57) est $f_{max} \approx 970$ Hz. Ainsi, pour $f = 900$ Hz (tracé en trait continu), le diagramme de directivité ne présente qu'un seul lobe principal en $\theta = 90^\circ$ (et en $\theta = 270^\circ$ pour des raisons de symétrie de révolution autour de l'axe de l'antenne linéaire). Par contre, pour $f = 1800$ Hz, deux lobes supplémentaires apparaissent : le filtrage spatial obtenu ne permet pas de distinguer des ondes en provenance de la direction d'écoute de celles perçues en raison du repliement. Ce cas est évidemment à éviter, et il faut donc veiller à espacer les microphones en fonction de la plus haute fréquence devant être traitée.

Conséquences pour la robotique L'étude précédente montre que pour obtenir une antenne capable d'"écouter" finement dans une direction, celle-ci doit être de grande taille et comporter un nombre élevé de microphones. Mais malgré ces précautions élémentaires, son diagramme évolue selon la fréquence et perd énormément en résolution pour les plus basses d'entre elles. Pour éviter ce phénomène, une première solution pourrait consister à ne traiter que les hautes fréquences, en veillant tout de même à respecter le théorème de Shannon. Ces constatations sont en contradiction avec les contraintes de la robotique. Pour être embarquable, l'antenne doit en effet être de petite dimension ; le filtrage spatial des basses fréquences sera donc de très mauvaise qualité. Pour limiter la consommation, il faut un capteur équipé d'un nombre réduit de microphones ; l'efficacité du filtrage, en terme de réjection des directions d'arrivées parasites, sera donc limitée. Enfin, dans le cas où la source sonore est un locuteur, les signaux reçus par l'antenne sont majoritairement basses fréquences ; leur exploitation pour la localisation s'avère donc difficile a priori. Toutefois, il ne faut pas perdre de vue que le filtrage spatial effectué par l'antenne est très peu coûteux en terme de puissance de calcul, et c'est principalement pour cette raison que la formation de voie est aujourd'hui la technique la plus utilisée en robotique.



(a) Réponse normalisée de l'antenne en fonction de θ et f
($N = 5$, $L = 0.7m$)

(b) Illustration de l'effet du repliement spatial

FIG. 2.20 – Etude de l'influence de la fréquence, pour N et L fixés.

2.2.4.2 Exploitation des techniques de formation de voie pour la localisation de source en robotique

Les propriétés du filtrage spatial énoncées précédemment peuvent être utilisées afin de localiser une source sonore dans l'environnement. Le principe le plus commun en robotique est extrêmement simple : il consiste à polariser l'antenne de microphones successivement dans chacune des directions à explorer, afin d'obtenir une carte de puissance acoustique de l'environnement. Celle-ci présente alors un pic dans la direction d'émission de la source sonore à localiser. Ce mode de fonctionnement est totalement analogue à celui d'un radar passif, à la différence qu'ici l'antenne de microphone est polarisée électroniquement dans une direction, et peut rester donc fixe par rapport à la plateforme mobile.

La construction de la carte d'énergie nécessite de discrétiser les directions d'écoute possibles. Pour cela, [Valin 2004] propose d'utiliser une grille icosaoèdre uniforme posée sur une sphère, telle que celle présentée figure 2.24(a). D'autres discrétisations sont envisageables, selon la forme du capteur et le nombre de points de tests, conduisant à un compromis entre la puissance de calcul nécessaire et la résolution souhaitée. Ne perdons pas de vue cependant que la résolution de la localisation, si elle dépend effectivement du pas de la grille de discrétisation, est fortement liée à la forme de la réponse d'antenne pour les raisons énoncées précédemment (limitation de la taille et du nombre de microphones).

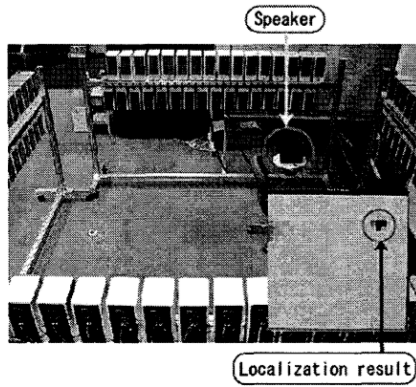
Performances des formations de voie La construction de la carte d'énergie acoustique de l'environnement nécessite donc de former le diagramme d'antenne de façon à la rendre sensible à une seule direction d'intérêt θ_0 . Les filtres $w_n(t)$ présentés figure 2.18 page 52 permettent cette opération. Dans la partie précédente, nous avons montrés que des filtres "passe-tout" conduisent à privilégier la direction $\theta_0 = 90^\circ$. La même idée peut être utilisée pour une autre direction θ_0 : il suffit que les filtres $w_n(t)$ compensent les retards τ_n liés à la propagation pour que les signaux perçus en provenance de la direction θ_0 soient remis en phase avant leur sommation. Ainsi, dans le cas de l'antenne linéaire représentée figure 2.14, et sous l'hypothèse d'ondes planes, les fonctions de transfert $W_n(k)$ s'écrivent

$$W_n(k) = e^{jkz_n \cos \theta_0}. \quad (2.58)$$

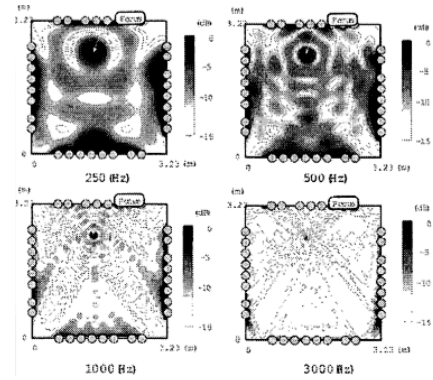
En combinant (2.58) avec (2.53), la réponse d'antenne en champ lointain s'exprime donc comme

$$D^\infty(\theta) = \frac{\sin\left(\frac{\pi f}{c} N d (\cos \theta_0 - \cos \theta)\right)}{\sin\left(\frac{\pi f}{c} d (\cos \theta_0 - \cos \theta)\right)}. \quad (2.59)$$

La formation de voie obtenue, appelée *filter and sum beamforming*, ou *formation de voie conventionnelle*, est de très loin la plus utilisée en robotique. Or, elle souffre de tous les défauts explicités auparavant, et en particulier de la mauvaise résolution des basses fréquences. Ainsi, bien que [Tamai 2004] utilise une antenne, inenvisageable en robotique, équipée de 128 microphones répartis dans une pièce (figure 2.21(a)), la mauvaise focalisation pour les grandes longueur d'onde oblige l'auteur à filtrer l'ensemble des fréquences inférieures à 500 Hz. Ce phénomène est illustré figure 2.21(b), représentant la réponse de l'antenne pour différentes fréquences. Nous retrouvons clairement le même type de courbe que sur la figure 2.20(a). [Mattos 2004] arrive



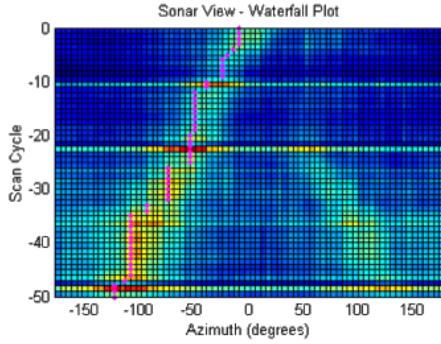
(a) Photographie de l'antenne. Chaque côté fait 3.2m de long.



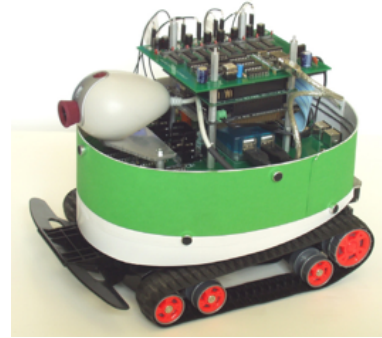
(b) Diagramme d'antenne simulé pour différentes fréquences.

FIG. 2.21 – [Tamai 2004] - Localisation de source à partir d'une antenne de 128 microphones.

à la même conclusion par simulations concernant l'antenne implémentée sur la petite plateforme mobile présentée figure 2.22(b). Ici, les 8 microphones sont insérés sur une ceinture entourant le robot, mais leur emplacement est décidé de manière empirique à partir d'un logiciel de simulation développé pour tracer la réponse d'antenne. Celle-ci laisse paraître un lobe principal suffisamment fin pour les fréquences supérieures à 1 kHz, tandis que les fréquences inférieures à 800 Hz demeurent totalement inexploitable pour la localisation. [Tamai 2005] propose l'antenne composée de trois anneaux de microphones reproduite figure 2.23(a). 8 microphones sont placés sur la circonférence du capteur faisant 44 cm de diamètre, tandis que 24 autres récepteurs sont répartis sur 3 anneaux de 11 cm de rayon. L'antenne est donc constituée de 32 microphones, intervenant dans une formation de voie analogue aux précédentes. La figure 2.23(b) montre la réponse obtenue. Comme prévu, la directivité du capteur est très mauvaise en basses fréquences, si bien que malgré la taille importante du capteur et le nombre élevé de ses microphones, seules les fréquences situées entre 1 kHz et 2 kHz sont exploitables pour la localisation. Au delà, de multiples lobes apparaissent dans le diagramme d'antenne en raison du repliement.



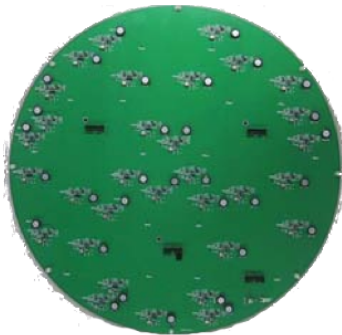
(a) Résultats de localisation en azimuth.



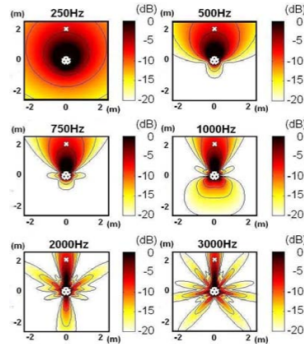
(b) Robot EvBot-II équipé de son antenne de 8 microphones

FIG. 2.22 – [Mattos 2004] - Représentation du système de localisation proposé

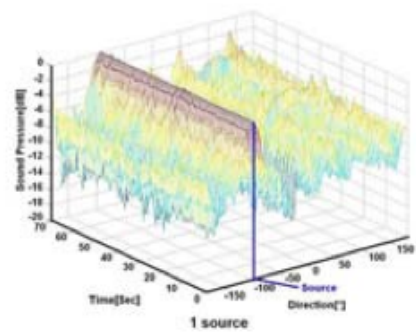
Résultats de localisation Une fois l'antenne polarisée dans une direction souhaitée grâce à un choix judicieux de filtres $w_n(t)$, il est nécessaire de calculer l'énergie acoustique issue de chacune de ces directions. Nous aboutissons alors à une "carte sonore" de l'environnement utilisée pour la localisation. Ce calcul peut être effectué simplement en intégrant, sur une fenêtre temporelle finie, le carré du signal de sortie de la formation de voie. [Valin 2004] propose une autre méthode de détermination de cette énergie, à partir d'un calcul de corrélation plus facilement implémentable en temps réel. Les cartes d'énergie obtenues sont semblables à celle présentée figure 2.22(a) [Mattos 2004]. Ici, un son d'hélicoptère est émis depuis un haut-parleur et déplacé autour de la plateforme. La carte obtenue permet d'évaluer correctement le déplacement de la source, mais les mauvaises réponses d'antenne obtenues avec la configuration spatiale des microphones retenue produit très clairement de fausses détections. Une source virtuelle semble présente dans l'environnement, et se déplace symétriquement par rapport à la véritable source à détecter.



(a) Capteur circulaire de 32 microphones



(b) Réponses de l'antenne pour différents fréquences



(c) Résultats de localisation pour une source sonore

FIG. 2.23 – [Tamai 2005] - Capteur circulaire proposé et résultats de localisation

Les résultats de localisation en azimuth obtenus dans [Tamai 2005] sont visibles figure 2.23. La source sonore, de la musique jouée depuis un haut parleur, est correctement localisée, avec une erreur moyenne d'environ 3° . [Tamai 2005] montre ensuite que pour la même source, la carte de l'environnement déterminée à partir de seulement 8 des 32 microphones de son antenne

possède présente de forts lobes secondaires à l'origine de potentielles fausses détections. Cette constatation s'explique par l'utilisation d'une formation de voie conventionnelle, et est justifiée par la figure 2.19. Les performances en extraction de deux sources sonores sont ensuite étudiées. La méthode reste très simple et est basée sur l'hypothèse très restrictive selon laquelle les deux sources à extraire ne possèdent pas le même contenu fréquentiel. La focalisation de l'antenne dans les deux directions d'émission estimées lors de la localisation, permet de récupérer deux signaux. Une simple soustraction spectrale permet alors d'en extraire les signaux des sources, malgré un filtrage spatial imparfait réalisé par la formation de voie.

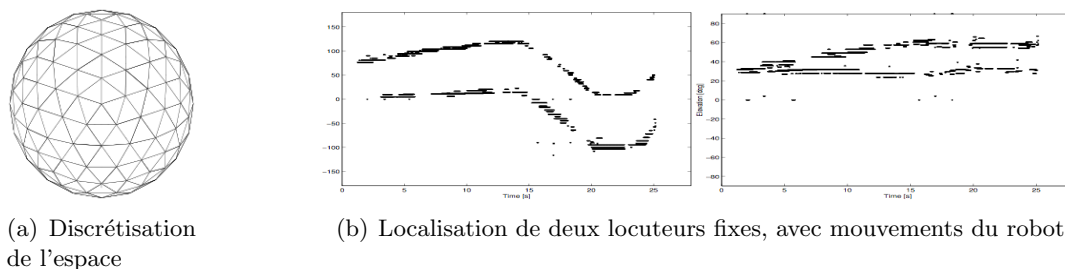
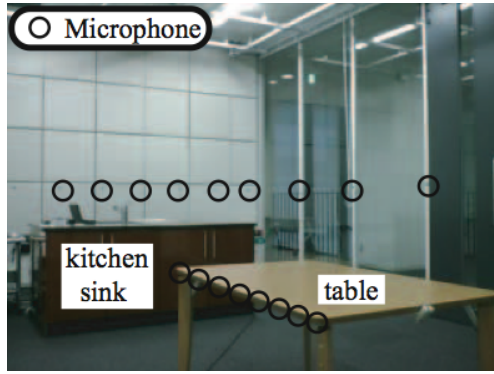


FIG. 2.24 – [Valin 2004] - Résultats de localisation

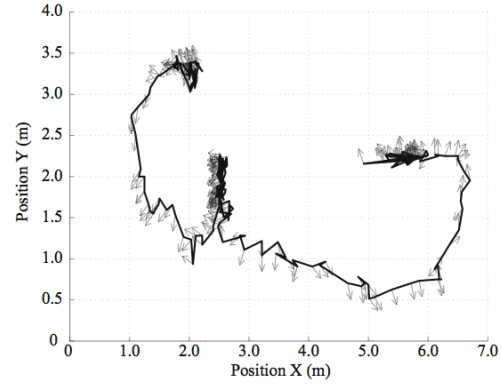
[Valin 2004] aboutit aux résultats de localisation présentés figure 2.24(b), pour deux locuteurs parlant simultanément pendant un mouvement du robot. Afin d'éviter les fausses détections rencontrées dans [Mattos 2004], [Valin 2004] propose d'effectuer un post-filtrage probabiliste de la carte d'énergie, basé sur deux estimateurs simples à court et moyen terme. Ce filtre permet alors un lissage de la localisation au cours du temps. Les résultats obtenus montrent une bonne robustesse du système vis à vis des bruits émis par les moteurs au cours du mouvement, tout en restant raisonnables en terme de puissance de calcul. Malgré tout, ceux-ci sont effectués sur un ordinateur déporté équipé d'une carte d'acquisition, dont ils nécessitent environ 30% de CPU (AthlonXP 2000+). Compte tenu de la méthode de calcul de la puissance (analogue à un calcul de corrélation généralisée RWPHAT), le système n'est pas capable de localiser des signaux faible bande, mais fonctionne parfaitement avec des signaux vocaux.

Autres utilisations de la formation de voie en robotique D'autres travaux proposent d'exploiter les propriétés de filtrage spatial d'une formation de voie pour estimer certaines propriétés de la source ou pour aboutir à des systèmes de reconnaissance vocale performants. Parmi ceux-ci, citons [Nakadai 2005], qui propose une formation de voie conventionnelle modifiée de façon à prendre en compte la directivité de la source, afin d'en estimer la position et la direction d'émission. De plus, l'estimation de la directivité permet de discriminer une voix émise depuis un haut-parleur d'une voix réelle émise par une personne. La méthode proposée s'appuie sur une antenne composée de 64 microphones placés à l'intérieur d'une pièce, présentée sur la figure 2.25(a). Les résultats de localisation sont visibles figure 2.25(b). La source sonore est localisée avec une précision moyenne de $0.25m$, tandis que l'angle d'émission de la source est déterminé avec une erreur d'environ 10° .

[Mizoguchi 2004] propose d'utiliser la formation de voie en émission, pour générer un son à un endroit précis de l'espace. Ici, la formation de voie a pour rôle de déphaser le signal source au niveau de chacun des haut-parleurs de l'antenne, de façon à provoquer des interférences destructrices dans les zones de l'espace où l'antenne n'est pas focalisée. Le procédé est représenté sur la figure 2.26(a), sur laquelle figure l'antenne linéaire utilisée, munie de 16 haut-parleurs



(a) Dispositif expérimental proposé



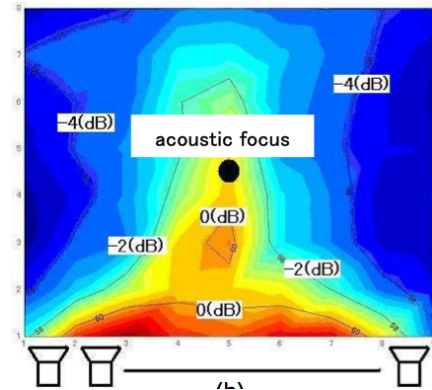
(b) Estimation de la position du robot

FIG. 2.25 – [Nakadai 2005] - Résultats de localisation, obtenus dans une pièce acoustiquement préparée.

régulièrement espacés. La figure 2.26(b) représente le champ de pression acoustique mesuré dans la pièce équipée de l'antenne. Celle-ci permet bien de privilégier une zone de l'espace, mais ne fournit que 4 dB d'atténuation dans les zones non désirées. Même si cette atténuation est sensible à l'oreille, elle ne suffit pas pour garantir qu'un message envoyé par cette antenne ne soit perçu que par son destinataire.



(a) Dispositif expérimental



(b) Carte de puissance de l'environnement sonore

FIG. 2.26 – [Mizoguchi 2004] - Utilisation d'une formation de voie pour l'émission directive de sons dans l'environnement.

Enfin, citons [Choi 2003] qui propose une méthode de formation de voie adaptative visant à optimiser le filtrage spatial vis à vis de perturbations acoustiques (écho, sources parasites). L'amélioration du filtrage permet une augmentation sensible des taux de reconnaissance vocale. Sur le même type d'application, [Ohashi 2005] propose l'utilisation de deux formations de voie : la première permet d'isoler le signal vocal émis par un locuteur, tandis que la seconde cherche au contraire à estimer l'ensemble du bruit ambiant de façon ensuite à le soustraire au signal extrait. Là encore, la robustesse de la reconnaissance vis à vis du bruit est améliorée, et ce même avec

un nombre de microphones faible.

2.3 Conclusion

Nous venons de parcourir tout au long de ce chapitre une grande majorité des méthodes de localisation proposées ces dix dernières années par la communauté Robotique. Cette étude a été pour nous l'occasion de distinguer deux types d'approches. Tout d'abord, nous avons présenté les techniques bioinspirées, ayant pour objectif de reproduire artificiellement le fonctionnement de notre système auditif. Nous avons pu constater toute la difficulté à exploiter des indices acoustiques élémentaires, et la nécessaire prise en compte du rôle primordial de la tête dans le processus de localisation. Pour cela, différents modèles de propagations ont été proposés, et nous avons constaté que les plus simples d'entre eux ne suffisaient à expliquer les courbes expérimentales relevées en chambre anéchoïque. Puis nous nous sommes attachés à décrire les techniques issues des communautés du Traitement du Signal et de l'Acoustique, en veillant systématiquement à évaluer leur applicabilité dans un contexte robotique. La méthode haute résolution MUSIC étendue naïvement au cas large bande ne permet pas d'envisager une utilisation en temps réel ; nous allons étudier dans le chapitre 4 une extension récente de cette méthode réduisant significativement le coût calculatoire. Nous avons ensuite introduit les techniques de localisation par corrélation, conduisant à des estimations précises de la position de la source sonore. Celles-ci nécessitent cependant très souvent l'hypothèse d'ondes planes afin de limiter la complexité algorithmique. Enfin, nous avons présenté les approches exploitant les propriétés de filtrage spatial d'une antenne de microphones obtenues à partir de formations de voie. Cette solution s'est avérée être la moins complexe d'un point de vue calculatoire et demeure aujourd'hui la méthode la plus utilisée dans la littérature. Cependant, nous avons montré que de telles méthodes étaient incapables de localiser précisément des signaux basse fréquence à partir d'un capteur de petite taille constitué d'un faible nombre de microphones. C'est pourquoi nous proposons dans le chapitre suivant d'améliorer le comportement de l'antenne pour ces mêmes fréquences.

Chapitre 3

Synthèse d'une formation de voie large bande

Nous avons présenté dans le chapitre précédent un panorama complet des méthodes de localisation de source sonore envisagées en robotique. Nous avons constaté que parmi celles "empruntées" au Traitement du Signal, les approches par formation de voie sont les plus communes. Cependant, une grande majorité des algorithmes proposés repose sur une formation de voie dite "conventionnelle", consistant à compenser les retards entre les microphones liés à la propagation. Or, un tel choix conduit à une très mauvaise résolution de la localisation des basses fréquences. Pour traiter ce problème, de nombreux auteurs se sont contentés de mettre en œuvre un filtrage des signaux perçus de façon à n'en conserver que les harmoniques les plus élevées. Si une telle solution conduit effectivement à une localisation plus précise, elle ne prend pas en compte les composantes les plus énergétiques des signaux, et donc les plus à même d'être exploitées. C'est dans ce but que nous proposons ci-après une méthode de formation de voie originale visant à conférer à l'antenne de bonnes propriétés de filtrage spatial des basses fréquences.

Ce chapitre est organisé comme suit. Nous précisons dans une première partie les notations utilisées ainsi que la nature du problème de synthèse. Dans une deuxième partie, nous proposons une solution simple et immédiate, résultant d'une extension large bande naïve d'une méthode existante de synthèse en champ lointain. Cette première étape nous permet de mettre en évidence la nécessaire gestion d'un compromis entre la qualité de la réponse d'antenne obtenue et certains paramètres de synthèse. Puis, dans une troisième section, nous étudions l'influence de la distance sur la forme du diagramme de directivité. Il apparaît en particulier qu'une formation de voie dimensionnée sous l'hypothèse de champ lointain ne peut pas être utilisée pour la localisation d'une source sonore basse fréquence située près de l'antenne. Enfin, nous présentons dans une quatrième et dernière partie une méthode de synthèse originale, basée sur la représentation modale des réponses d'antenne et valide à la fois pour le champ proche et le champ lointain. Cette stratégie s'avère plus efficace que la précédente en terme de conditionnement mathématique du problème et de qualité des résultats obtenus.

3.1 Généralités et notations

Le vocable *formation de voie*, ou *beamforming* en anglais, désigne une stratégie de filtrage spatial de l'information acoustique par une antenne de microphones. Le réseau de transducteurs réalise un échantillonnage spatial de l'onde sonore, multidimensionnel et non uniforme dans le

cas général, et les signaux qu'ils délivrent sont ensuite traités afin de polariser le capteur dans une direction spécifique de l'espace. La formation de voie est souvent le premier élément d'une chaîne complète de traitements visant à estimer les paramètres de la source sonore d'intérêt, tels que la distance, l'azimut, l'élévation ou le contenu fréquentiel. Nous allons l'utiliser dans la suite pour déterminer les positions angulaires de sources sonores situées dans l'environnement.

Ce chapitre reprend les notations introduites dans le chapitre précédent. Ainsi, nous considérons une antenne comprenant N microphones repérés par les vecteurs $\mathbf{r}_1^m, \dots, \mathbf{r}_N^m$. Les sources à localiser, au nombre de S , sont quant à elles situées aux positions $\mathbf{r}_1^s, \dots, \mathbf{r}_S^s$. Nous supposons également que la formation de voie est réalisée numériquement, i.e. les signaux issus des microphones sont échantillonnés temporellement à une fréquence f_e préalablement à leur traitement sur un ordinateur numérique. Ce type d'implémentation est aujourd'hui le plus courant, à l'exception du cas des signaux très haute fréquence qui exige une réalisation analogique, souvent par l'intermédiaire de lignes à retard.

3.1.1 Formation de voie faible bande

Un signal peut être considéré comme *faible bande* lorsque la gamme de ses fréquences utiles est faible devant sa fréquence centrale. D'après [Van Veen 1988], la nature faible bande d'un signal est aussi fonction de sa durée d'observation, au sens où le produit du temps d'observation et de la bande passante (Time Bandwidth Product, ou TBWP) intervient dans sa caractérisation. En bande de base, c'est à dire ici dans le domaine des fréquences audio, ce critère peut être simplifié, et nous commettrons l'abus de langage consistant à assimiler un signal faible bande à un signal monochromatique de fréquence f_0 ou k_0 . L'étude d'une formation de voie faible bande est donc conceptuellement simplifiée puisque la dépendance à la fréquence temporelle de la réponse d'antenne ne doit plus être explicitée. Sous sa forme générale, celle-ci peut être représentée par la figure 3.1, où les sorties monochromatiques $m_n(t)$ des N microphones, $n = 1, \dots, N$, après numérisation à une période d'échantillonnage T_e , sont combinées linéairement par l'intermédiaire de coefficients complexes $w_n^{[k_0]}$. Sur cette figure, $m_n[l]$, $l \in \mathbb{Z}$, représente le signal numérique défini par $m_n[l] = m_n(lT_e)$.

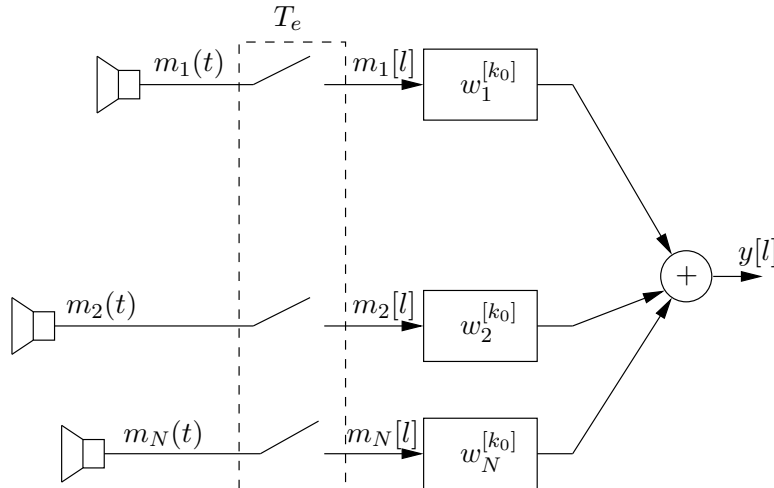


FIG. 3.1 – Principe de la formation de voie faible bande

Ainsi, d'après cette même figure, la sortie $y(lT_e)$ de la formation de voie s'écrit

$$y[l] = y(lT_e) = \sum_{n=1}^N w_n^{[k_0]} m_n[l]. \quad (3.1)$$

Supposons qu'une seule source faible bande, située en $\mathbf{r}^s$ et émettant à la fréquence k_0 , soit présente dans la scène. Alors, le signal numérique $s^0[l]$ reçu en l'origine du repère de l'espace (voir chapitre 2, page 41) est de la forme

$$s^0[l] = e^{jk_0 clT_e}. \quad (3.2)$$

Par conséquent, d'après l'équation (2.14), le signal numérique $m_n[l]$ reçu par le $n^{\text{ième}}$ microphone admet pour expression, en négligeant pour l'instant le terme de bruit,

$$m_n[l] = \alpha_n(\mathbf{r}^s) s^0(lT_e - \tau_n(\mathbf{r}^s)). \quad (3.3)$$

L'introduction de (3.3) dans (3.1) conduit alors à l'expression

$$y[l] = D_{[k_0]}(\mathbf{r}^s) s^0[l], \quad (3.4)$$

dans laquelle nous reconnaissons la réponse d'antenne faible bande $D_{[k_0]}(\mathbf{r})$ définie par

$$D_{[k_0]}(\mathbf{r}) = \sum_{n=1}^N w_n^{[k_0]} V_n(\mathbf{r}, k_0). \quad (3.5)$$

Notons que l'expression (3.5) est évidemment un cas particulier de l'expression générale (2.52) page 53, obtenue dans le chapitre précédent : seule la fréquence k_0 est considérée, et les signaux des microphones sont combinés linéairement par de simples coefficients complexes. L'équation (3.5) peut être ré-écrite matriciellement sous la forme

$$D_{[k_0]}(\mathbf{r}) = \mathbf{W}_{[k_0]}^T \mathbf{V}(\mathbf{r}, k_0), \quad (3.6)$$

où $\mathbf{V}(\mathbf{r}, k_0) = [\mathbf{V}_1(\mathbf{r}, k_0) | \dots | \mathbf{V}_N(\mathbf{r}, k_0)]$. Le vecteur $\mathbf{W}_{[k_0]}$, appelé *vecteur de poids*, est défini par

$$\mathbf{W}_{[k_0]} \triangleq (w_1^{[k_0]}, w_2^{[k_0]}, \dots, w_N^{[k_0]})^T. \quad (3.7)$$

Dans le cas particulier d'une antenne linéaire analogue à celle représentée figure 2.13 page 40, pour laquelle les N microphones sont placés aux abscisses $z_1, \dots, z_N$ le long de l'axe $\mathcal{Z}$, la réponse faible bande se simplifie en

$$D_{[k_0]}(\mathbf{r}) = D_{[k_0]}(r, \theta) = \sum_{n=1}^N w_n^{[k_0]} \frac{e^{jk_0(r - \sqrt{r^2 + z_n^2 - 2rz \cos \theta})}}{\sqrt{r^2 + z_n^2 - 2rz \cos \theta}}. \quad (3.8)$$

Si la source se situe suffisamment loin pour considérer que les fronts sont plans au voisinage de l'antenne, alors l'expression en champ lointain du diagramme de directivité devient

$$D_{[k_0]}^\infty(\mathbf{r}) \triangleq \lim_{r \rightarrow \infty} D_{[k_0]}(r, \theta) = D_{[k_0]}^\infty(\theta) = \sum_{n=1}^N w_n^{[k_0]} V_n^\infty(\theta, k_0) = \sum_{n=1}^N w_n^{[k_0]} e^{-jk_0 z_n \cos \theta}. \quad (3.9)$$

Les deux expressions précédentes sont les plus utilisées dans la littérature pour aborder le problème de la synthèse de formation de voie faible bande, dont l'objectif est alors de déterminer les N coefficients complexes $w_1^{[k_0]}, \dots, w_N^{[k_0]}$.

3.1.2 Formation de voie large bande

Un signal est dit large bande si son contenu fréquentiel ne peut pas être assimilé à une fréquence unique. Il en est ainsi de la voix, qui couvre la bande 20 Hz – 20 kHz. Notre objectif étant de localiser un signal vocal, il nous faudra donc envisager une formation de voie large bande, capable de traiter l'ensemble des fréquences constituant le signal acoustique. Deux implémentations classiques sont envisageables.

3.1.2.1 Implémentation dans le domaine fréquentiel

L'idée est ici de décomposer dans le domaine fréquentiel les signaux reçus sur les microphones de l'antenne, et de traiter chaque "bin" fréquentiel obtenu comme une composante faible bande, cf. figure 3.2. En pratique, les Q composantes fréquentielles $M_n(k_q)$, $q = 1, \dots, Q$ sont déterminées pour chaque $n^{\text{ième}}$ microphone, $n = 1, \dots, N$, au moyen d'une transformée de Fourier discrète implémentée sur ordinateur via l'algorithme de calcul rapide FFT (Fast Fourier Transform). Une formation de voie faible bande distincte, identique à celle présentée au §3.1.1, est ensuite appliquée à chaque N -uplet $(M_1(k_q), \dots, M_N(k_q))$ de façon à délivrer en sortie le résultat du filtrage spatial relatif à la fréquence k_q , $q = 1, \dots, Q$. Finalement, une transformée de Fourier inverse, sous forme IFFT, permet la reconstruction de la série temporelle correspondant au filtrage spatial de l'onde large bande.

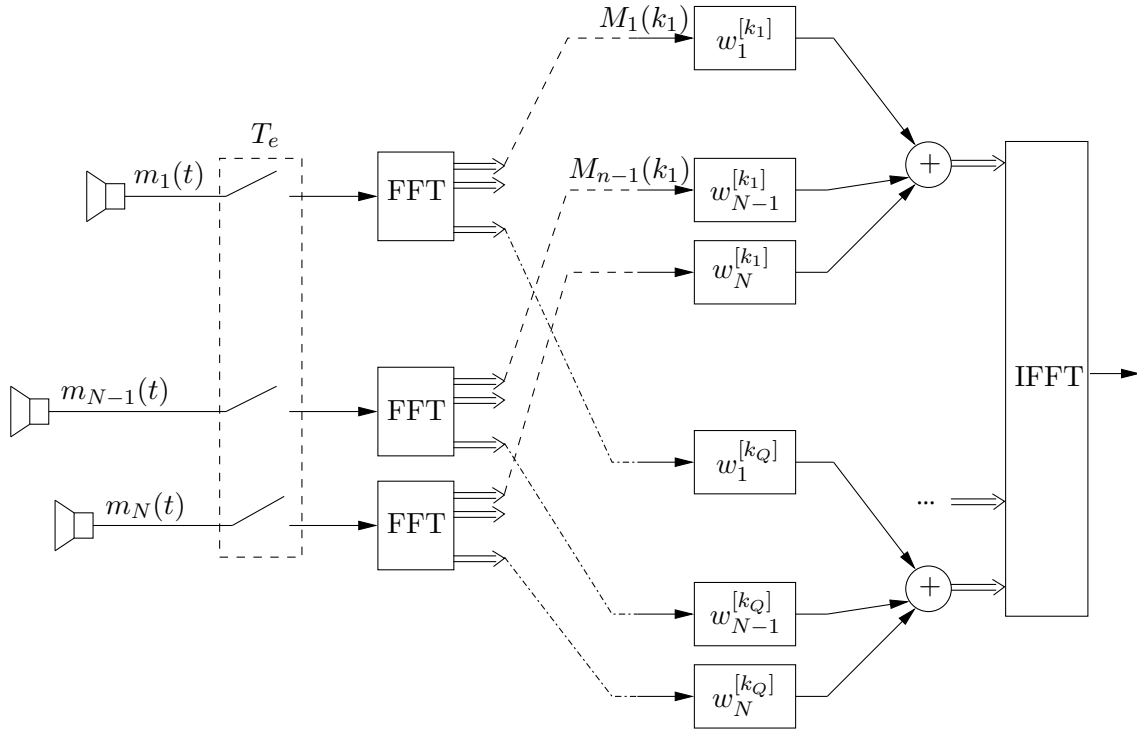


FIG. 3.2 – Principe de la formation de voie large bande, implémentée dans le domaine fréquentiel

Pour cette implémentation dans le domaine fréquentiel, la réponse d'antenne $D(\mathbf{r}, k)$ est l'agrégation des diagrammes d'antenne élémentaires $D_{[k_q]}(\mathbf{r})$ relatifs aux fréquences k_q , $q = 1, \dots, Q$. Ceux-ci s'écrivent, d'après l'équation (3.5)

$$D_{[k_q]}(\mathbf{r}) = \sum_{n=1}^N w_n^{[k_q]} V_n(\mathbf{r}, k_q) = \mathbf{W}_{[k_q]}^T \mathbf{V}(\mathbf{r}, k_q), \quad (3.10)$$

ou bien, dans le cas particulier du champ lointain (3.9),

$$D_{[k_q]}^\infty(\theta) = \mathbf{W}_{[k_q]}^T \mathbf{V}^\infty(\theta, k_q). \quad (3.11)$$

L'étape de synthèse large bande consiste donc en la détermination de NQ coefficients $w_1^{[k_1]}$, $\dots$, $w_N^{[k_1]}$, $\dots$, $w_1^{[k_Q]}$, $\dots$, $w_N^{[k_Q]}$. A ce stade, un certain nombre de remarques peuvent être effectuées. Imaginons le cas où seule la bande de fréquences 400 Hz – 3 kHz du signal acoustique est considérée. Si ce signal est échantillonné à une fréquence $f_e = 10$ kHz et décomposé fréquemment par une FFT effectuée sur 512 points, alors il peut être représenté par environ 133 fréquences. Toutes ces composantes sont autant de formations de voie à déterminer. Bien que l'implémentation dans le domaine fréquentiel simplifie le problème en le transformant en de multiples cas faible bande, elle requiert par conséquent une forte parallélisation des opérations. Une sélection intelligente des composantes utiles à la localisation permettrait néanmoins une nette réduction des besoins en puissance de calcul, mais le problème de la définition d'un critère de sélection satisfaisant demeure.

3.1.2.2 Implémentation dans le domaine temporel

Plutôt que de procéder à la décomposition fréquentielle des signaux perçus, l'approche par formation de voie dans le domaine temporel consiste en l'insertion d'un filtre en aval de chacun des microphones, cf. figure 3.3. De cette façon, le traitement s'effectue globalement, sans nécessiter de transformée de Fourier. Le plus souvent, des filtres numériques à réponse impulsionnelle finie (RIF) sont utilisés, de fonction de transfert

$$W_n(k) = \sum_{q=0}^Q w_{nq} e^{-jkqT_e}, \quad (3.12)$$

où w_{nq} désigne le $q^{\text{ième}}$ coefficient du $n^{\text{ième}}$ filtre d'ordre Q , lequel est attaché au $n^{\text{ième}}$ microphone de l'antenne. On retrouve ainsi la configuration déjà présentée dans le chapitre 2 selon laquelle, d'après l'équation (2.52) page 53, la réponse d'antenne $D_{\text{LB}}(\mathbf{r}, k)$ est de la forme

$$D_{\text{LB}}(\mathbf{r}, k) = \sum_{n=1}^N W_n(k) V_n(\mathbf{r}, k). \quad (3.13)$$

En insérant l'équation (3.12) dans (3.13), nous obtenons l'expression générale du diagramme de directivité

$$D_{\text{LB}}(\mathbf{r}, k) = \sum_{n=1}^N \sum_{q=0}^Q w_{nq} e^{-jkqT_e} V_n(\mathbf{r}, k), \quad (3.14)$$

ou bien, sous forme matricielle,

$$D_{\text{LB}}(\mathbf{r}, k) = \mathbf{W}_{\text{LB}}^T \mathbf{V}_{\text{LB}}(\mathbf{r}, k), \quad (3.15)$$

avec

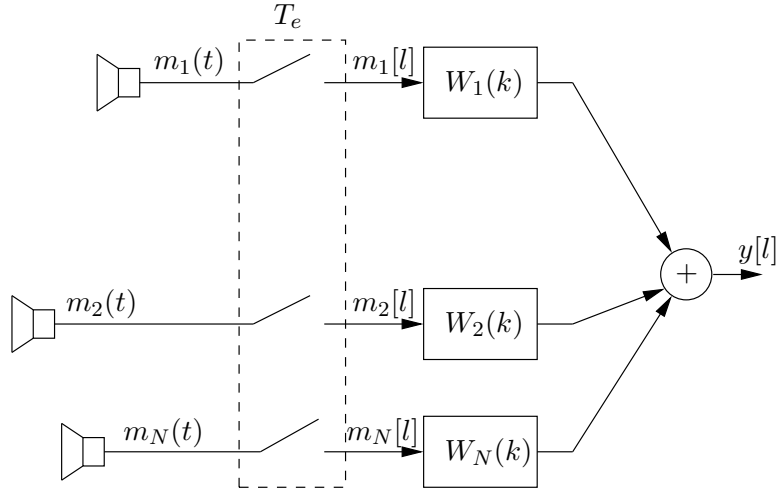


FIG. 3.3 – Principe de la formation de voie large bande, implémentée dans le domaine temporel

$$\mathbf{V}_{\text{LB}}(\mathbf{r}, k) = \mathbf{V}(\mathbf{r}, k) \otimes \mathbf{V}_{\text{FIR}}(k), \quad (3.16)$$

et

$$\mathbf{W}_{\text{LB}} = \begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{ik} \\ \vdots \\ w_{NQ} \end{pmatrix}, \quad \mathbf{V}_{\text{FIR}}(k) = \begin{pmatrix} 1 \\ e^{-jkcT_e} \\ e^{-jkc2T_e} \\ \vdots \\ e^{-jkcQT_e} \end{pmatrix}.$$

L'opération $\otimes$ désigne le produit de Kronecker : étant données deux matrices $\mathcal{X}$ et $\mathcal{Y}$ de dimensions respectives $p \times q$ et $t \times l$, $\mathcal{X} \otimes \mathcal{Y}$ est la matrice de dimension $pt \times ql$ formée des pq sous matrices $\mathcal{Z}_{i,j} = \mathcal{X}_{i,j} \mathcal{Y}$, pour $(i, j) \in 1, \dots, p \times 1, \dots, q$. Dans le cas particulier de l'antenne linéaire présentée figure 2.13 page 40, dont les N microphones sont placés aux abscisses $z_1, \dots, z_N$ le long de l'axe $\mathcal{Z}$, la réponse large bande en champ lointain devient

$$D_{\text{LB}}^{\infty}(\mathbf{r}, k) = D_{\text{LB}}^{\infty}(\theta, k) = \sum_{n=1}^N \sum_{q=0}^Q w_{nq} e^{-jkcqT_e} e^{-jkz_n \cos \theta} = \mathbf{W}_{\text{LB}}^T \mathbf{V}_{\text{LB}}^{\infty}(\theta, k), \quad (3.17)$$

avec

$$\mathbf{V}_{\text{LB}}^{\infty}(\theta, k) = \mathbf{V}^{\infty}(\theta, k) \otimes \mathbf{V}_{\text{FIR}}(k). \quad (3.18)$$

Cette expression est certainement la plus utilisée dans la littérature pour la synthèse large bande dans le domaine temporel. Ici, l'objectif est de déterminer les $(Q+1)$ coefficients $w_{n0}, \dots, w_{nQ}$ de chacun des N filtres RIF réalisant la formation de voie. Nous proposons dans ce chapitre une nouvelle méthode de synthèse du vecteur $\mathbf{W}_{\text{LB}}$ de façon à conférer à l'antenne de bonnes propriétés de filtrage spatial, y compris si sa dimension est limitée par le contexte robotique.

3.1.3 Définition du problème de synthèse

Comme mentionné au §2.2.4.2, utiliser la formation de voie en robotique revient à exploiter les capacités de filtrage spatial d'une antenne pour dresser une carte d'énergie acoustique de l'environnement. Nous avons vu que cette carte était ensuite exploitée par les roboticiens pour estimer l'azimut de la source à partir de sa valeur maximale. Son calcul est effectué en polarisant électroniquement l'antenne dans toutes les positions potentielles à tester, et en déterminant l'énergie du signal de sortie de chacune des formation de voie, le plus souvent sur une fenêtre d'observation. Cette polarisation est obtenue en synthétisant une formation de voie aux caractéristiques adéquates pour chacune des positions hypothésées.

De nombreuses méthodes de synthèse ont déjà été proposées dans la littérature, chacune visant à conférer à l'antenne de bonnes caractéristiques de filtrage spatial, à rendre la réalisation moins sensible vis à vis d'éventuelles interférences, à aboutir à des synthèses optimales en terme de robustesse au bruit, etc. Après avoir précisé un certain nombre de grandeurs, nous définissons ci-après un diagramme de directivité idéal permettant d'améliorer sensiblement la qualité de la localisation par rapport aux utilisations précédentes de la formation de voie en robotique.

3.1.3.1 Remarques générales

Il convient de préciser un certain nombre de grandeurs nécessaires à la suite du chapitre. Nous choisissons de limiter le traitement à la bande de fréquences [400 Hz; 3 kHz]. Ce domaine est très voisin de celui utilisé pour toutes les communications téléphoniques. Il avait été choisi historiquement afin de préserver l'intelligibilité du signal vocal tout en occupant la bande spectrale la plus réduite possible. Un tel choix impose immédiatement une condition fondamentale sur la dimension du capteur. Comme indiqué par l'équation (2.57) page 55, l'espacement entre deux microphones successifs doit être suffisamment petit pour vérifier le théorème de Shannon spatial. Dans le cas d'une antenne linéaire, à la fréquence maximale $f_{max} = 3$ kHz, cette équation définit la distance maximale entre deux microphones adjacents comme $d_{max} = 5.66$ cm, et donc la taille maximale admissible du capteur pour un nombre de microphones donné. La tentation est alors grande de réduire la dimension de l'antenne. Le capteur étant alors de petite taille, son embarquabilité est facilitée, et le théorème de Shannon nous garantit l'absence de repliement. Pour autant, il ne faut pas perdre de vue que la résolution des basses fréquences est liée à la dimension de l'antenne. C'est pour cette raison que nous chercherons toujours à nous rapprocher de la limite de Shannon en plaçant nos microphones exactement à la valeur préconisée par l'équation (2.57).

Nous considérons dans la suite de ce chapitre une antenne acoustique constituée de N microphones omnidirectionnels de fonction de transfert unité. Sa configuration reste celle représentée figure 2.13; l'antenne est donc linéaire, et les microphones régulièrement espacés. Cependant, le formalisme que nous présentons reste valable quel que soit l'espacement entre chacun des microphones. Le choix d'une antenne linéaire peut paraître discutable, et reste ici, il faut l'avouer, un choix *par défaut*. A notre connaissance, il existe très peu d'études sur la forme à conférer à une antenne pour obtenir un diagramme de directivité donné. Mais l'objectif de notre étude est davantage de déterminer une méthode de synthèse de formation de voie valable quelle que soit la disposition des microphones plutôt que d'établir une forme d'antenne optimale. Enfin, compte tenu des contraintes d'embarquabilité sur le robot destiné à recevoir notre antenne, nous avons choisi d'imposer à celle-ci une taille maximale de l'ordre de 40cm. Comme chacun des microphones se doit d'être espacé de 5.66 cm, leur nombre sera fixé à $N = 8$.

Finalement, du fait de l'utilisation d'une antenne linéaire et pour des raisons de symétrie,

seul l'azimut θ et la distance r constituent le vecteur des variables d'espace. Nous ferons souvent l'hypothèse de champ lointain, selon laquelle les fronts d'ondes parvenant à l'antenne sont plans. Sous cette hypothèse, la distance r tend vers l'infini, et seul l'azimut θ entre alors en jeux. Nous étudierons en détail l'influence de la distance ainsi que les conséquences d'une mauvaise hypothèse sur la forme des fronts d'onde dans la section suivante.

3.1.3.2 Définition du comportement angulaire de référence

Comme nous l'avons illustré dans le chapitre précédent, beaucoup de méthodes basées sur la formation de voie en robotique ne sont pas adaptées au traitement de signaux basse fréquence. Cela conduit beaucoup d'auteurs à filtrer l'ensemble des fréquences inférieures à environ 1 kHz de façon à pouvoir localiser efficacement les sources sonores de l'environnement. Seulement, dans le cas d'un signal vocal, se priver de ces fréquences revient à en ignorer les composantes fréquentielles les plus puissantes. Nous allons chercher dans ce chapitre à déterminer un ensemble de coefficients ou de filtres visant à conférer à l'antenne de bonnes propriétés de filtrage spatial sur l'intégralité du domaine fréquentiel d'intérêt.

Une réponse d'antenne *parfaite* et idéale pour la localisation pourrait se présenter sous la forme d'un dirac $\delta(\theta - \theta_c)$. A un tel diagramme correspond une polarisation parfaite dans l'unique direction θ_c . Seulement, une telle fonction de directivité nécessiterait un capteur de dimension infinie. Un diagramme d'antenne *idéal*, mais plus réaliste car probablement plus atteignable, pourrait être représenté par la figure 3.4. Centrée en $\theta = \theta_c$, la formation de voie ainsi faite possède un comportement invariant en fréquence : c'est là une des caractéristiques essentielles que nous allons tenter d'atteindre. Finalement, nous venons de définir une réponse d'antenne de référence que nous notons $D_d(\theta, k)$, entièrement paramétrées par les trois angles θ_c, θ_p et θ_s , définissant chacun respectivement la direction de polarisation et l'équivalent de bandes passante et coupées angulaires.

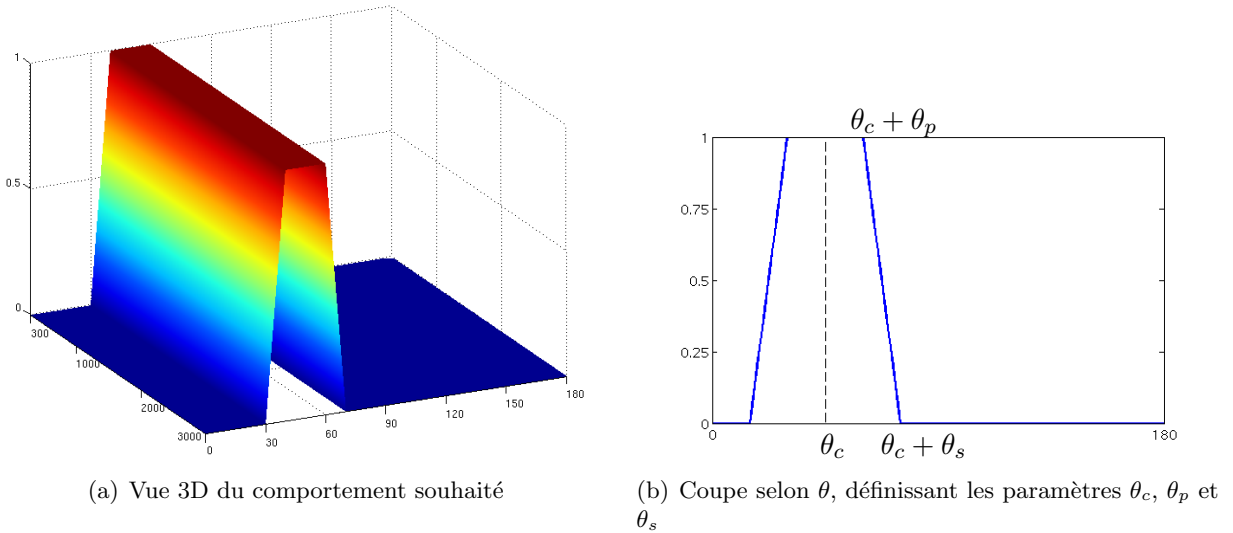


FIG. 3.4 – Définition du comportement de référence

3.2 Synthèse naïve large bande par optimisation convexe

L'ensemble de cette section repose sur l'hypothèse de champ lointain. Nous supposons donc que le diagramme de référence présenté précédemment, noté $D_d^\infty(\theta, k)$, est défini pour une distance théorique infinie. Nous avons vu également que deux implémentations d'une formation de voie large bande sont envisageables. Cependant, pour ces deux cas, le problème de synthèse peut se résumer en :

Déterminer le vecteur $\mathbf{W}$ tel que $D^\infty(\theta, k) \approx D_d^\infty(\theta, k)$ pour tous les $\theta \in [0; 180^\circ]$ et $k \in [k_{\min}; k_{\max}]$.

L'objectif est donc de minimiser l'écart entre deux fonctions de θ et k , à priori sur un continuum de valeurs. Cependant, nous considérons la version "échantillonnée" suivante de ce problème de synthèse :

Déterminer le vecteur $\mathbf{W}$ tel que $D^\infty(\theta, k) \approx D_d^\infty(\theta, k)$ pour tous les $\theta \in \Theta = [\theta_1, \dots, \theta_I]$ et $k \in K = [k_1, \dots, k_J]$,

où I et J désignent respectivement le nombre de contraintes angulaires et fréquentielles. La résolution d'un tel problème, étant donné un diagramme d'antenne désiré, a été proposée dans le cadre de l'optimisation convexe par [Wang 2003], pour le cas faible bande. Après avoir rappelé la méthodologie proposée, cette même méthode sera étendue au cas large bande.

3.2.1 Optimisation et programmation semi-définie

Nous proposons dans cette partie de balayer très rapidement quelques définitions et propriétés en rapport avec l'optimisation convexe. Soit une fonction ϕ de $\mathbb{R}^m$ dans $\mathbb{R}$. Un problème d'optimisation quelconque, de dimension finie, peut s'écrire sous la forme

$$\min_{x \in \mathcal{C}} \phi(x), \quad (3.19)$$

où le vecteur $x \in \mathbb{R}^m$ regroupe les variables de décision – ou d'optimisation –, $\mathcal{C} \subset \mathbb{R}^m$ désigne l'ensemble des contraintes, i.e. des valeurs admissibles de x , et ϕ est le critère, ou coût. Notons que lorsque $\mathcal{C} = \mathbb{R}^m$ (réciproquement $\mathcal{C} \neq \mathbb{R}^m$), alors (3.19) définit un problème d'optimisation non contraint (réciproquement sous contraintes).

Dans le cas général, résoudre (3.19) est un problème complexe pouvant admettre plusieurs minima locaux. Néanmoins, il existe des cas où l'optimum est global. Il en est ainsi du cas convexe, i.e. lorsque la fonction objectif ϕ ainsi que l'ensemble $\mathcal{C}$ sont tous deux convexes. Rappelons qu'un ensemble $\mathcal{C}$ est dit convexe si et seulement si le segment reliant deux points quelconques de $\mathcal{C}$ est lui même contenu dans $\mathcal{C}$. Cette notion a été généralisée en analyse, et une fonction ϕ est dite à son tour convexe si et seulement si l'ensemble $\{x : \phi(x) \leq \alpha\}$ est convexe quel que soit α , i.e. si et seulement si quels que soient x et y , quel que soit $\lambda \in [0; 1]$, $\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y)$. Citons quelques fonctions convexes de la variable vectorielle x couramment utilisées : les fonctions affines $a^T x + b$, où a est un vecteur et b un scalaire ; les fonctions quadratiques $x^T A x$, avec A une matrice semi-définie positive ; les normes de $\|x\|$, incluant la norme Euclidienne.

Lorsqu'un problème d'optimisation est convexe, tout minimum local de la fonction coût est en réalité globalement optimal. Examinons maintenant une certaine classe de problèmes d'optimisation convexe appelés *programmes semi-définis*. Tout d'abord, définissons une *inégalité matricielle affine* (Linear Matrix Inequality) comme une inégalité de la forme

$$\mathbf{F}(\mathbf{x}) \triangleq \mathbf{F}_0 + \sum_{i=1}^m x_i \mathbf{F}_i \geq 0, \quad (3.20)$$

où le vecteur $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$ regroupe les variables de décision, $\mathbf{F}_i \in \mathbb{R}^{n \times n}$, $i = 0, \dots, m$, désignent m matrices symétriques données, et " $\mathbf{F}(\mathbf{x}) \geq 0$ " signifie que la combinaison matricielle affine $\mathbf{F}(\mathbf{x})$ est symétrique et semi-définie positive. On note immédiatement que l'ensemble $\mathcal{C} = \{\mathbf{x} : \mathbf{F}(\mathbf{x}) \geq 0\}$ est convexe. Sous sa forme générale, un programme semi-défini concerne la minimisation d'un critère affine sous des contraintes LMI, i.e., le vecteur $\mathbf{c}$ étant donné,

$$\begin{array}{ll} \text{minimiser} & \mathbf{c}^T \mathbf{x} \\ \text{sous} & \mathbf{F}(\mathbf{x}) \geq 0. \end{array} \quad (3.21)$$

Les programmes LMI sont solubles numériquement en temps polynomial avec une précision arbitraire, au moyen d'algorithmes dédiés. C'est pourquoi un problème plus général – e.g. d'Automatique, de Traitement du Signal, d'optimisation – est supposé résolu dès lors qu'il peut être retranscrit sous cette forme. De nombreuses approches pour l'analyse et la commande des systèmes repose sur ce paradigme depuis l'ouvrage [Boyd 1994]. Le lecteur intéressé trouvera des applications au Traitement du Signal dans [Balakrishnan 1998] et [Luo 2003].

Le problème de la synthèse d'antenne par optimisation convexe peut être en général formulé de la façon suivante [Wang 2003]

$$\begin{array}{ll} \text{minimiser} & \varepsilon \\ \text{sous} & \mathbf{W}^T \mathbf{A}_i (\mathbf{W}^T \mathbf{A}_i)^H - 2\text{Re}\{\mathbf{W}^T \mathbf{B}_i\} + \mathbf{D}_i \leq \varepsilon, \quad i = 1, \dots, L, \end{array} \quad (3.22)$$

avec $\mathbf{W} \in \mathbb{C}^N$ la variable de décision et $\mathbf{A}_i \in \mathbb{C}^N$, $\mathbf{B}_i \in \mathbb{C}^N$, $\mathbf{D}_i \in \mathbb{R}^N$, $\varepsilon \in \mathbb{R}$ des données. On montre que (3.22), possédant un objectif linéaire sous des contraintes quadratiques et appelé programme sur le cône du second ordre (Second Order Programm, ou SOCP), est en fait une forme particulière de (3.21).

Tous les résultats de synthèse que nous allons présenter dans ce manuscrit ont été obtenus avec le solveur SDPT3, dans sa version 3.2 [Tutuncu 2006]. Celui-ci est interfacé avec Matlab, via le programme Yalmip [Löfberg 2004] : le problème d'optimisation, saisi en syntaxe Matlab, est interprété par Yalmip de façon à le traduire sous la forme requise par SDPT3.

3.2.2 Méthode faible bande

De nombreuses méthodes de synthèse sont proposées dans la littérature. L'une des méthodes les plus célèbres a été formulée par Dolph [Dolph 1946], à partir de la théorie des polynômes de Chebyshev. Le diagramme d'antenne obtenu, dit "de Dolph-Chebyshev", possède le niveau minimum de lobes secondaires pour une largeur donnée du lobe principal. Hélas, cette méthode requiert des microphones régulièrement espacés. D'autres solutions consistent en une formulation quadratique du problème de synthèse [Tseng 1992], pour laquelle l'objectif est très souvent de minimiser l'erreur quadratique entre la réponse d'antenne souhaitée et celle recherchée. Parmi les approches basées sur l'optimisation convexe citons par exemple [Lebret 1997] ou [Scholnik 2001] qui proposent tous des solutions SOCP pour la formation de voie faible bande. Récemment, Wang et Al. ont étendu la synthèse de formation de voie en gain seulement dans le cas d'une antenne linéaire aux micros non régulièrement espacés [Wang 2003], tout en prenant en compte des contraintes convexes permettant de robustifier la réponse d'antenne aux erreurs en gain et phase des microphones. Nous proposons ici de tester l'extension immédiate large bande de cette

approche pour une antenne de petite dimension. Mais détaillons auparavant les résultats obtenus dans le cas faible bande.

3.2.2.1 Présentation du problème d'optimisation

Considérons une réponse d'antenne désirée $D_d^\infty(\theta, k_0)$ en champ lointain à la fréquence de référence f_0 ou k_0 , cf. par exemple la figure 3.4. [Wang 2003] propose le problème d'optimisation suivant :

$$\begin{array}{ll} \text{minimiser} & \varepsilon \\ \text{sous} & |D_{[k_0]}^\infty(\theta) - D_d^\infty(\theta, k_0)|^2 \leq \varepsilon, \quad \forall \theta \in \Theta \end{array} \quad (3.23)$$

où $\Theta = [\theta_1, \dots, \theta_I]$ désigne l'ensemble discret sur lequel l'optimisation "min-max" est effectuée. Le problème (3.23) peut être reformulé sous une forme SOCP analogue à (3.22) à partir de (3.6), i.e.,

$$\begin{array}{ll} \min & \varepsilon \\ \text{sous} & \mathbf{W}_{[k_0]}^T \mathbf{V}^\infty(\theta, k_0) (\mathbf{W}_{[k_0]}^T \mathbf{V}^\infty(\theta, k_0))^H - 2\text{Re}\{\mathbf{W}_{[k_0]}^T \mathbf{V}^\infty(\theta, k_0) D_d^\infty(\theta)^H\} + |D_d^\infty(\theta)|^2 \leq \varepsilon, \quad \forall \theta \in \Theta, \end{array} \quad (3.24)$$

Outre leurs propriétés intéressantes relativement à leur résolution, les problèmes (3.23) et (3.24) peuvent intégrer facilement des contraintes supplémentaires, possiblement antagonistes, dès lors que celles-ci peuvent être transcrites sous une forme LMI. Il est par exemple possible de limiter ce que nous appellerons, par abus de langage, la *puissance du vecteur de poids* $\mathbf{W}_{[k_0]}$ notée G_B , au deçà d'un seuil $\delta > 0$, en ajoutant au problème d'optimisation précédent la contrainte

$$G_B = \sum_{i=1}^N |w_i^{[k_0]}|^2 = \mathbf{W}_{[k_0]}^H \mathbf{W}_{[k_0]} \leq \delta, \quad (3.25)$$

(3.25) peut ensuite être mis sous forme LMI et ajoutée à (3.24) pour obtenir un nouveau problème d'optimisation convexe. Notons que cette dernière contrainte est fondamentale pour garantir une bonne robustesse de l'antenne au bruit. En effet, dans le cas où une seule source sonore est présente dans la scène, nous avons vu, d'après l'équation (2.51), que le spectre $S(k)$ de la sortie de la formation de voie s'écrit

$$S(k) = \mathbf{W}(k)^T \mathbf{V}(\mathbf{r}, k) S^0(k) + \mathbf{W}(k)^T \mathbf{B}(k), \quad (3.26)$$

avec $S^0(k)$ le signal reçu à l'origine du repère et $\mathbf{B}(k) \triangleq (B_1(k), \dots, B_N(k))^T$ le vecteur des bruits présents sur les microphones. La covariance de $S(k)$ est donc de la forme

$$E[S(k)S(k)^H] = \mathbf{W}^T \mathbf{V} (\mathbf{W}^T \mathbf{V})^H E[S^0(k)S^0(k)^H] + \mathbf{W}^T E[\mathbf{B}\mathbf{B}^H] (\mathbf{W}^T)^H, \quad (3.27)$$

sous l'hypothèse d'indépendance entre le bruit et le signal. Si le bruit est blanc spatialement et de même puissance sur chacun de smicrophones, alors $E[\mathbf{B}\mathbf{B}^H] = \sigma^2 \mathcal{I}$, de sorte que le terme $G_B = \mathbf{W}^H \mathbf{W}$ vient multiplier sa covariance du bruit dans (3.27). C'est pourquoi G_B est traditionnellement appelé *gain de bruit blanc* (White Noise Gain) de l'antenne, et constitue un critère important qu'il convient de limiter.

3.2.2.2 Résultats d'optimisation

Nous proposons dans cette partie quelques résultats d'optimisation obtenus dans le cas faible bande, directement issus de la méthode d'optimisation (3.23) proposée dans [Wang 2003]. Ceux-ci justifient l'importance de la contrainte (3.25) relative à la limitation de la puissance des coefficients. Pour les exemples de synthèse qui suivent, nous avons fixé les valeurs $\theta_p = 10^\circ$ et $\theta_s = 25^\circ$ du comportement de référence $D_d^\infty(\theta)$ représenté figure 3.4(b). L'ensemble Θ intervenant dans la définition de (3.23) correspond à une discrétisation des angles d'incidence de l'onde entre 0 et 180° par pas de 1° . Par conséquent, (3.23) comporte 181 contraintes.

Optimisation sans limite sur G_B Nous présentons ici les résultats d'optimisation obtenus sans la contrainte additionnelle (3.25). Considérons dans un premier temps que $f_0 = 3\text{kHz}$, et $\theta_c = 60^\circ$. La figure 3.5 représente deux solutions du problème d'optimisation, pour deux valeurs différentes de N . Nous pouvons constater que les deux réponses possèdent bien un lobe principal

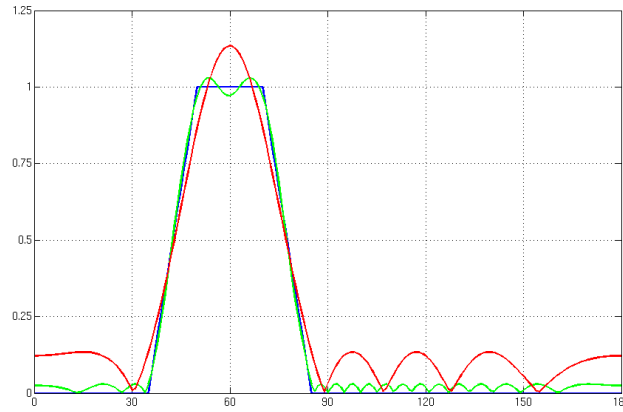


FIG. 3.5 – Résultat d'optimisation faible bande, pour $f_0 = 3\text{kHz}$ ($N = 8$: rouge, $N = 25$: vert). En bleu est représenté $D_d^\infty(\theta)$

dans la direction attendue, tout en atténuant les directions périphériques. Evidemment, la forme trapézoïdale du comportement de référence n'est pas respectée car l'utilisation de seulement $N = 8$ microphones ne fournit pas suffisamment de degrés de liberté pour l'approximer. Comme l'atteste la courbe verte, augmenter le nombre de microphone permet de résoudre ce problème, tout en atténuant davantage le niveau des lobes secondaires. Cette remarque est essentielle : comme nous devons limiter la taille de l'antenne, la réponse pouvant être obtenue possédera toujours une forme "douce" analogue à celle représentée en rouge sur la figure 3.5. Remarquons également que la réponse d'antenne dépasse la valeur maximale égale à 1 définie pour $D_d^\infty(\theta)$; cependant, cette valeur maximale peut être ajustée par une simple normalisation des coefficients réalisant la formation de voie. Sans limitation de la valeur de G_B , le problème d'optimisation retourne un vecteur $\mathbf{W}_{[k_0]}$ tel que $\mathbf{W}_{[k_0]}^H \mathbf{W}_{[k_0]} = 0.18$. Pour cette fréquence, le problème d'optimisation fournit donc un jeu de coefficients de puissance admissible et permettant de réaliser une formation de voie polarisant l'antenne dans la direction $\theta_c = 60^\circ$.

Optimisation sous la contrainte $G_B \leq \delta$ Le problème d'optimisation (3.24) est maintenant résolu en y adjoignant la contrainte (3.25). Dans un premier temps, afin de mettre en avant uni-

quement l'effet de cette contrainte supplémentaire, nous avons fixé le nombre N de microphone à $N = 25$. La figure 3.6 regroupe les résultats de synthèse obtenus pour une fréquence $f_0 = 1\text{kHz}$ et pour différentes valeurs de δ . Nous pouvons constater que plus la contrainte sur la puissance des

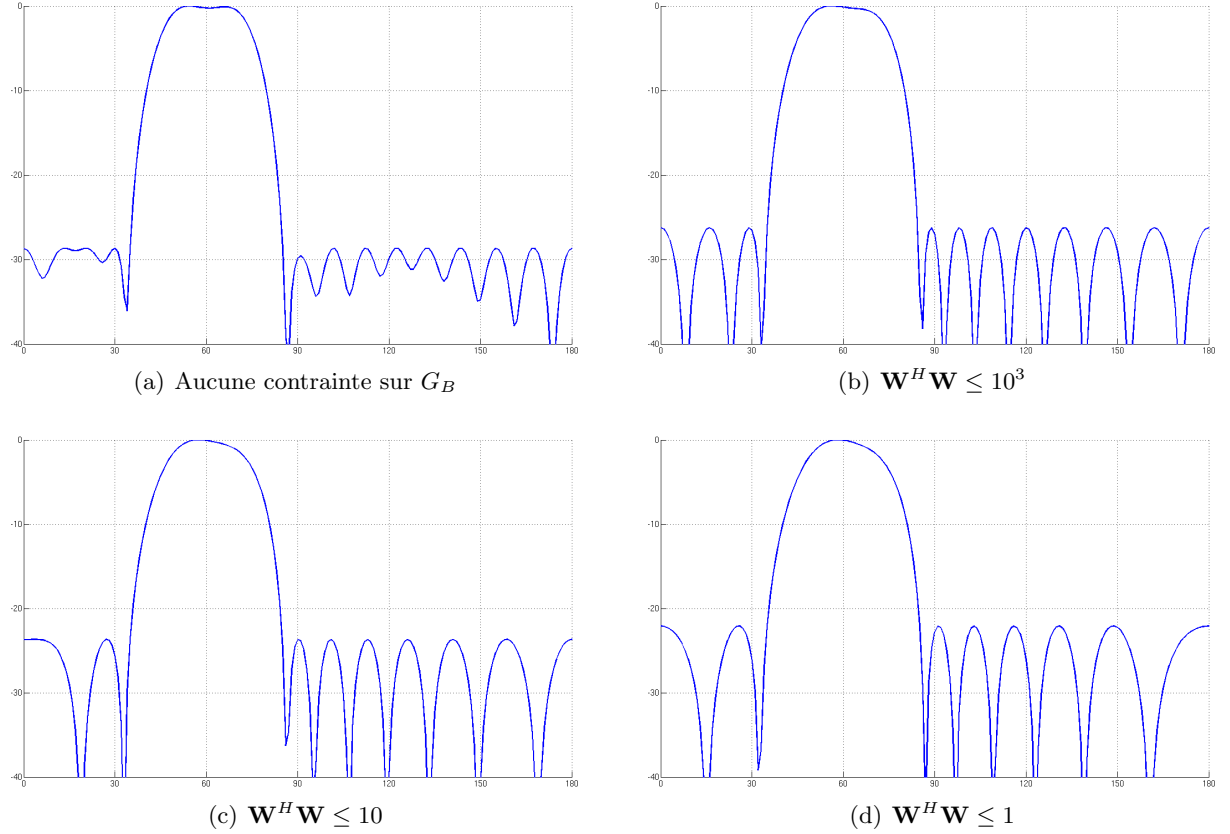


FIG. 3.6 – Résultats d'optimisation, pour $N = 25$ et différentes contraintes sur G_B

coefficients est forte (faibles valeurs de δ), plus le niveau des lobes secondaires augmente. Ainsi, sans contrainte supplémentaire, l'atténuation dans les directions non souhaitées atteint environ 28.7dB pour une valeur de $G_B = 3.9 \cdot 10^{11}$, tandis que pour $G_B \leq 1$, cette même atténuation n'est plus que de 22.1dB.

Intéressons nous maintenant à l'utilisation de cette méthode de synthèse dans le cadre d'une formation de voie large bande. Pour cela, nous proposons dans la suite d'étendre le problème d'optimisation (3.23) au cas multifréquences, pour les deux types d'implémentations envisagées au §3.1.2.

3.2.3 Méthode large bande, dans le domaine fréquentiel

3.2.3.1 Présentation du problème d'optimisation

L'implémentation dans le domaine fréquentiel d'une formation de voie large bande nécessite le traitement individuel de chacune des Q composantes fréquentielles k_q , $q \in [1, \dots, Q]$, des signaux reçus sur les microphones. Le problème de synthèse dans un tel cas peut donc se résumer en de multiples synthèses faible bande indépendantes, pour lesquelles la réponse d'antenne souhaitée $D_d^\infty(\theta, k_q)$ est identique. Nous proposons donc de résoudre le problème d'optimisation

$$\begin{array}{ll} \text{minimiser} & \varepsilon_q \\ \text{sous} & |\mathbf{W}_{[k_q]}^T \mathbf{V}^\infty(\theta, k_q) - D_d^\infty(\theta, k_q)|^2 \leq \varepsilon_q, \quad \forall \theta \in \Theta, \quad \forall q \in [1, \dots, Q], \end{array} \quad (3.28)$$

dans lequel le vecteur $\mathbf{W}_{[k_q]}$ contient les N coefficients complexes $w_n^{[k_q]}$ à déterminer. Ce nouveau problème d'optimisation est donc identique à (3.23), seule la dimension fréquentielle ayant été ajoutée. Au besoin, nous pouvons à nouveau limiter la puissance G_B des éléments du vecteur $\mathbf{W}_{[k_q]}$ en ajoutant la contrainte convexe supplémentaire

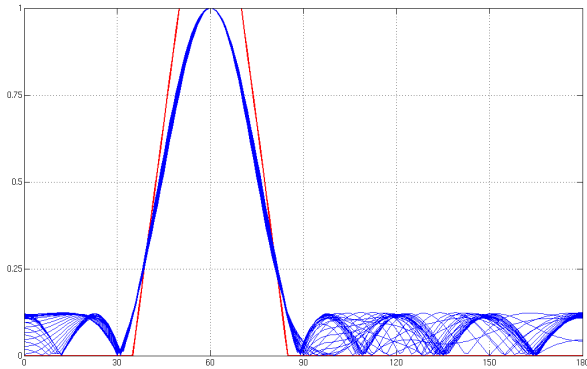
$$G_B = \mathbf{W}_{[k_q]}^H \mathbf{W}_{[k_q]} \leq \delta_q; \quad (3.29)$$

où δ_q désigne la valeur maximale admise de G_B pour la fréquence k_q . En pratique, une même contrainte δ est appliquée à l'ensemble des fréquences, de sorte que $\delta_q = \delta$, $q = 1, \dots, Q$.

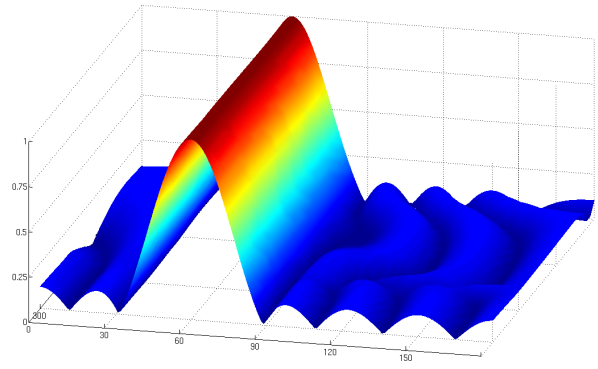
3.2.3.2 Résultats d'optimisation

Tout comme précédemment, nous proposons de résoudre (3.28) en fixant les valeurs des paramètres angulaires du comportement désiré $D_d^\infty(\theta, k_q)$ à $\theta_c = 60^\circ$, $\theta_p = 10^\circ$ et $\theta_s = 25^\circ$ pour l'ensemble des fréquences k_q .

Optimisation sans contrainte sur G_B Examinons les solutions obtenues lors de la résolution de (3.28) pour différentes fréquences comprises entre 400Hz et 3kHz par pas de 100Hz. L'ensemble des réponses d'antenne obtenues pour ces $Q = 26$ fréquences est visible sur la figure 3.7(a). Nous constatons que les réponses d'antenne sont toutes très semblables et que leurs lobes



(a) Superposition des diagrammes d'antenne obtenus pour l'ensemble des fréquences en fonction de θ



(b) Concaténation de toutes les réponses d'antenne dans le domaine fréquentiel

FIG. 3.7 – Formations de voie faibles bandes pour $f \in [300 : 100 : 3000]$

principaux se superposent presque parfaitement. Ce comportement est donc invariant en fréquence, les plus basses d'entre elles possédant maintenant la même résolution que les plus hautes. La figure 3.7(b) représente la réponse d'antenne large bande obtenue en concaténant chacun des diagrammes monofréquentiels précédents. Le comportement fréquentiel est très proche du comportement de référence représenté figure 3.4 et notre objectif semble donc être atteint. Or, si nous traçons la valeur du gain G_B correspondant à chacune des formations de voie synthétisées, nous nous apercevons que celle-ci prend des valeurs gigantesques rendant la synthèse totalement

inexploitable en environnement réel. Comme l'atteste la figure 3.8, la puissance des coefficients optimaux explose quasi-exponentiellement dans les basses fréquences pour atteindre des valeurs

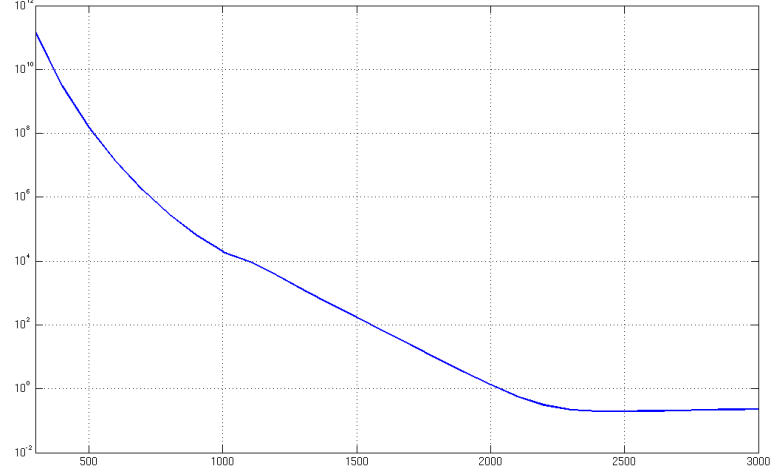
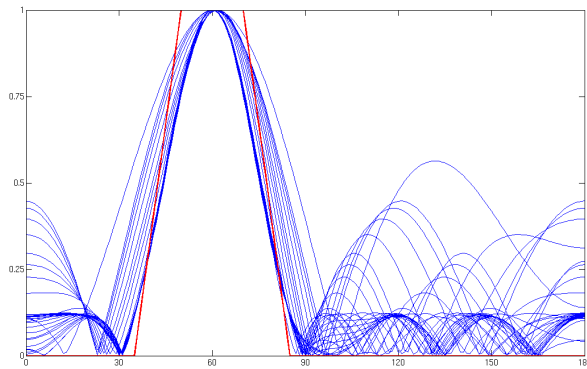


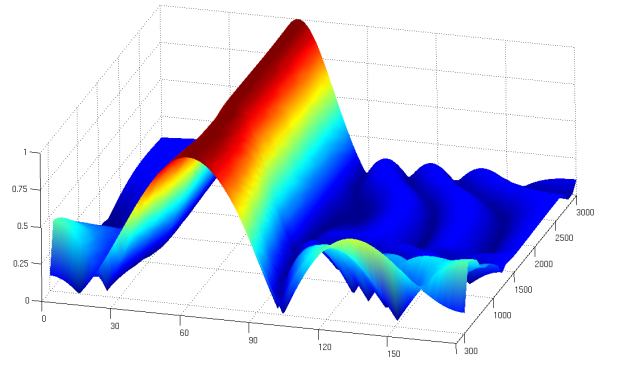
FIG. 3.8 – Evolution de $G_B = \mathbf{W}^H \mathbf{W}$ en fonction de la fréquence utilisée pour la synthèse faible bande.

de l'ordre de 10^{12} pour une fréquence de 400Hz. Finalement, le gain en largeur de lobe a été obtenu au détriment de la robustesse vis à vis du bruit. Cette constatation fondamentale reste valable dans toute la suite, et un compromis devra être effectué entre la résolution améliorée des basses fréquences et un gain G_B trop important. Nous avons vu que nous pouvions ajouter simplement la contrainte (3.29) visant à limiter la puissance des coefficients. Examinons maintenant son influence.

Optimisation sous la contrainte $G_B \leq \delta$ Comme précédemment, intéressons nous maintenant à une synthèse large bande sur la bande de fréquences $[400; 3000]$ Hz par pas de 100Hz. Nous avons choisi ici d'imposer $\mathbf{W}^H \mathbf{W} \leq 100$ pour chacune des 26 fréquences de synthèse. La figure 3.7(a) présente l'ensemble des réponses d'antenne normalisées obtenues. Nous voyons apparaître



(a) Superposition des diagramme d'antenne obtenus pour l'ensemble des fréquences en fonction de θ .



(b) Concaténation de toutes les réponses d'antenne dans le domaine fréquentiel.

FIG. 3.9 – Formations de voie faibles bandes pour $f \in [300 : 100 : 3000]$, avec $\mathbf{W}^H \mathbf{W} \leq 100$

des lobes secondaires d'un niveau très important pour les plus basses fréquences, résultat d'une très forte contrainte sur la puissance des coefficients, ainsi qu'une augmentation sensible de la largeur du lobe principal. Le comportement invariant en fréquence que nous avons obtenu sans contrainte sur G_B n'est donc plus vérifié. Par contre, la réponse d'antenne demeure quasiment inchangée pour les plus hautes fréquences du domaine d'optimisation et présente, dans la bande de fréquence 1 – 3kHz, un comportement correct ainsi qu'une valeur de G_B raisonnable. La figure 3.7(b) représente la concaténation de toutes les réponses d'antenne faible bande obtenues. Clairement, une telle formation de voie est difficilement utilisable tant le filtrage spatial opéré dans les basses fréquences est de mauvaise qualité. Examinons maintenant si une implémentation dans le domaine temporel permet d'améliorer le comportement de l'antenne.

3.2.4 Méthode large bande, dans le domaine temporel

Nous cherchons depuis le début de ce chapitre à synthétiser une formation de voie admettant un diagramme de directivité invariant en fréquence. Ce problème a déjà été traité dans la communauté Acoustique, dans le but par exemple d'exploiter l'information sonore pour la visio-conférence [Khalil 1994]. La dimension du capteur n'a alors que peu d'importance, puisque les microphones sont fixés aux murs de la pièce, de sorte que les antennes peuvent atteindre plusieurs mètres. Dans ce cadre, [Goodwin 1993] et [Chou 1995] proposent un capteur constitué de l'imbrication de plusieurs antennes, chacune étant adaptée à la réception d'une certaine gamme de longueur d'onde. Le nombre de microphones nécessaires étant très élevé, l'application de telles stratégies à la robotique pose cependant problème. D'une manière générale, la plupart des méthodes de synthèse des filtres permettant de réaliser une formation de voie large bande implémentée dans le domaine temporel imposent une position particulière de microphones. Ainsi, un positionnement logarithmique respectant une certaine loi fonction des fréquences mises en jeu est proposé dans [Ward 1996] et [Ward 1998]. Si les résultats montrent effectivement un comportement invariant sur de grandes plages de fréquences, le capteur se doit d'être composé d'une trentaine de microphones. De telle méthodes de synthèse ne sont donc pas utilisables dans notre cas, c'est pourquoi nous proposons d'étendre la méthode précédente issue de [Wang 2004] pour la détermination de filtres RIF.

3.2.4.1 Formulation du problème d'optimisation

Considérons une réponse d'antenne de référence $D_d^\infty(\theta, k)$, analogue à celle représentée sur la figure 3.4(a). Nous avons tenté d'approximer une telle réponse dans la partie précédente en la décomposant dans le domaine fréquentiel et en effectuant de multiples synthèses faible bande. Dans cette partie, nous proposons de déterminer les N filtres numériques RIF de fonction de transfert $W_n(k)$, $n = 1, \dots, N$ permettant d'approximer au mieux la fonction $D_d^\infty(\theta, k)$. Nous avons vu au §3.1.2.2 que la réponse $D_{LB}^\infty(\theta, k)$ pouvait s'écrire sous la forme

$$D_{LB}^\infty(\theta, k) = \sum_{n=1}^N \sum_{q=0}^Q w_{nq} e^{-jkcqT_e} e^{-jkz_n \cos \theta} = \mathbf{W}_{LB}^T \mathbf{V}_{LB}^\infty(\theta, k), \quad (3.30)$$

où $\mathbf{V}_{LB}^\infty(\theta, k)$ est défini selon (3.16). Nous proposons dans cette partie d'étendre le problème d'optimisation présenté dans [Wang 2003] et représenté par (3.24) au cas large bande pour la synthèse des filtres FIR réalisant la formation de voie. L'objectif ici est donc de déterminer le vecteur $\mathbf{W}_{LB}$ contenant les NQ coefficients des filtres FIR d'ordre Q placés en amont de chacun des N microphones de l'antenne. Nous proposons de résoudre le problème de minimisation

$$\begin{array}{ll} \text{minimiser} & \varepsilon \\ \text{sous} & |D_{\text{LB}}^\infty(\theta, k) - D_d^\infty(\theta, k)|^2 \leq \varepsilon, \forall \theta \in \Theta, \forall k \in K. \end{array} \quad (3.31)$$

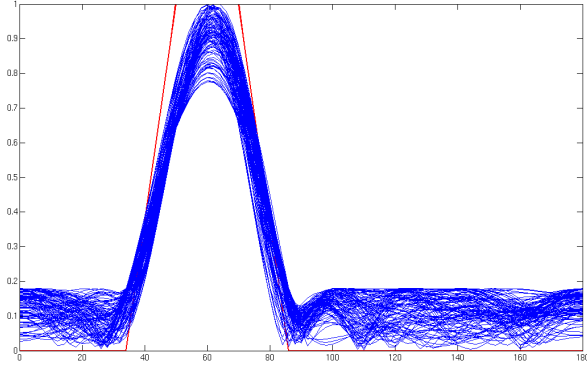
Comparé au problème d'optimisation faible bande précédent, seule la dimension fréquentielle a été ajoutée. Le problème reste bien évidemment convexe et permet donc systématiquement la convergence de sa solution numérique vers l'optimum global. Les deux ensembles Θ et K définissent une grille sur laquelle est minimisé l'écart entre le comportement souhaité D_d^∞ et celui recherché D_{LB}^∞ , et le nombre de contraintes du problème est lié au nombre de points de cette grille. Dans la suite, nous considérerons que Θ est l'ensemble des angles de θ compris entre 0 et 180° par pas de 2°. Nous préciserons plus tard les fréquence constituant l'ensemble K . Comme précédemment, nous seront amenés à limiter la puissance des coefficients des filtres RIF, e.g. en ajoutant la contrainte supplémentaire suivante au problème (3.31)

$$\mathbf{W}_{\text{LB}}^H \mathbf{W}_{\text{LB}} \leq \delta. \quad (3.32)$$

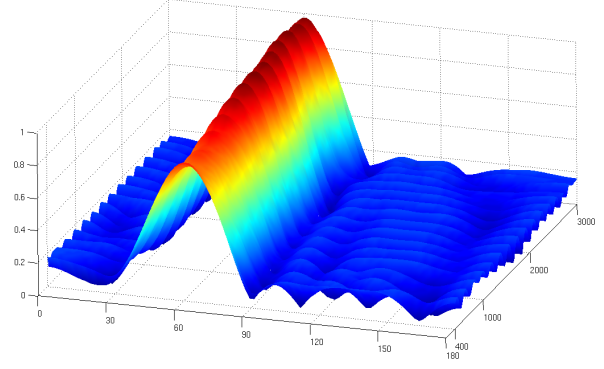
3.2.4.2 Résultats d'optimisation

Intéressons nous dans un premier temps au cas où la puissance des coefficients n'est pas contrainte. Fixons arbitrairement l'ordre des filtres à déterminer à $Q = 35$. L'influence de Q sera examinée plus en détails dans la section suivante. Concernant la définition de la grille $\Theta \times K$, si le choix du pas angulaire, ici de 2°, n'a que peu d'influence sur les résultats d'optimisation en raison du faible nombre de microphones mis en jeux, la définition du domaine K est beaucoup plus critique. Compte tenu de la nature discrète des contraintes fréquentielles, aucune garantie ne peut être fournie quant au comportement de la réponse d'antenne entre chacun des points de la grille. Une solution serait d'imposer un maillage fin selon les fréquence, mais ceci augmenterait très sensiblement le nombre de contraintes du problème (3.31), et rendrait donc sa résolution lente et possiblement mal conditionnée. A l'inverse, augmenter le pas en fréquences n'est pas envisageable car le relâchement des contraintes sur de larges bandes de fréquence conduit à une réponse d'antenne s'éloignant trop du gabarit. Il faut donc proposer un ensemble K constitué d'un nombre d'éléments réduit mais garantissant un comportement fréquentiel correct sur toute la plage des fréquences d'intérêt, fixée ici à [400Hz, 3kHz]. La figure 3.10 représente la réponse d'antenne obtenue après résolution de (3.31). L'optimisation a été effectuée par pas de 100Hz pour les fréquences comprises entre 400Hz et 2900Hz, et par pas de 25Hz entre 2900Hz et 3kHz : le pas en fréquence a dû être affiné pour les plus hautes fréquences afin de garantir la forme de la réponse d'antenne sur tout le domaine d'intérêt. La fréquence d'échantillonnage a été fixée à $f_e = 8\text{kHz}$. Le problème (3.31) admet donc 280 inconnues w_{nq} et environ 2650 contraintes angulaires et fréquentielles. Ce nombre conséquent de contraintes conduit à de longs temps de résolution, de l'ordre de 45 mins sur un ordinateur Pentium IV 2 GHz. Le diagramme d'antenne obtenu possède une largeur de lobe quasiment invariante en fréquence, au prix de légères oscillations du niveau du lobe principal. Comparativement à la formation de voie conventionnelle, les basses fréquences sont tout de même bien mieux résolues, et là encore notre objectif semble être atteint. Hélas, la puissance des coefficients des filtres atteint ici $5.7 \cdot 10^8$, et rend inexploitable une telle formation de voie en environnement réel. Il est donc nécessaire de limiter cette puissance en adjoignant la contrainte (3.32) au problème d'optimisation.

Pour les mêmes ensembles Θ et K que précédemment, nous avons limité la puissance des coefficients de sorte que $\mathbf{W}_{\text{LB}}^H \mathbf{W}_{\text{LB}} = 6 \cdot 10^3$. La figure 3.11 représente la réponse d'antenne obtenue après résolution du problème d'optimisation. Nous constatons que la limitation de la valeur



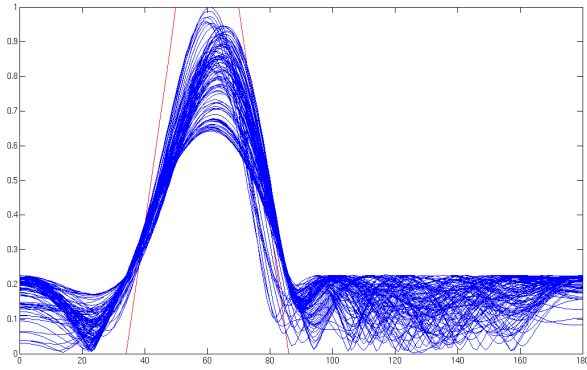
(a) Superposition des diagrammes d'antenne obtenus pour l'ensemble des fréquences en fonction de θ .



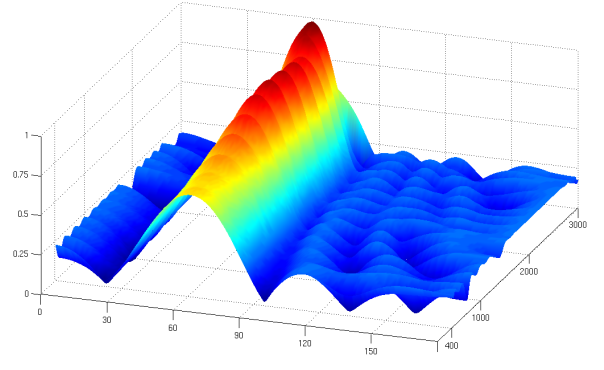
(b) Représentation de la réponse d'antenne dans le domaine fréquentiel.

FIG. 3.10 – Formations de voie faible bande pour $f \in [400; 3000]$ Hz, sans limite sur le vecteur $\mathbf{W}_{LB}$.

de $\mathbf{W}_{LB}^H \mathbf{W}_{LB}$ vient principalement modifier le comportement basse fréquence de la formation de voie, en agissant directement sur la largeur du lobe principal du diagramme d'antenne. Le phénomène d'oscillation du niveau du lobe principal est toujours présent, voire même accentué, et vient détériorer sensiblement l'efficacité du filtrage spatial. S'ajoute une déviation angulaire du maximum d'écoute, rendant sensible l'antenne à une direction θ_c respectée à plus ou moins 5° . Cette déviation, bien que non négligeable, est toutefois à relativiser car touchant principa-



(a) Superposition des diagrammes d'antenne obtenus pour l'ensemble des fréquences en fonction de θ .



(b) Représentation de la réponse d'antenne dans le domaine fréquentiel.

FIG. 3.11 – Formations de voies faibles bandes pour $f \in [400; 3000]$ Hz, avec $\mathbf{W}_{LB}^H \mathbf{W}_{LB} = 6.10^3$.

lement les dernières hautes fréquences, comme l'atteste la figure 3.11(b). Intéressons nous par ailleurs aux diagrammes de Bode des filtres numériques synthétisés, reportés sur la figure 3.12. Le domaine fréquentiel sur lequel est effectué l'optimisation y est représenté en blanc. Les fréquences grisées ne sont donc pas contraintes et agissent comme des degrés de libertés lors de la synthèse. Précisons que ces fréquences devront être évidemment éliminées par un dispositif annexe de filtrage (cf. chapitre 4). Le gain des filtres décroît d'environ 35dB à -20dB selon les fréquences croissantes. Evidemment, les plus forts gains sont obtenus pour les basses fréquences, laissant présager que ce sont principalement ces mêmes fréquences qui seront les plus sensibles

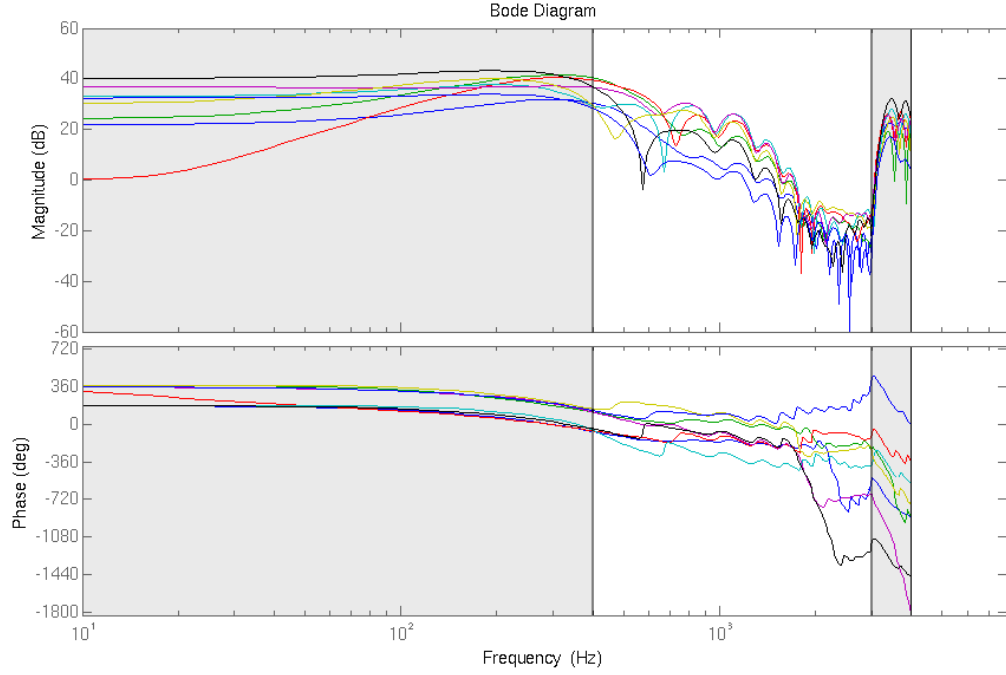


FIG. 3.12 – Diagramme de Bode des 8 filtres réalisant la formation de voie dont le diagramme d’antenne est visible figure 3.11.

au bruit. La phase quant à elle n’évolue quasiment pas sur une grande gamme de fréquences : n’oublions pas en effet que la synthèse d’antenne effectuée ici porte sur le gain et la phase du comportement désiré, phase fixée arbitrairement à 0° pour l’ensemble des fréquences et des angles sur lequel est effectuée l’optimisation. Il est probable qu’une telle phase soit fortement contraignante car difficilement atteignable, et de fait vienne perturber le comportement en gain de la réponse d’antenne. Dès lors, nous pouvons rechercher quelle serait la meilleure phase à donner au comportement désiré de l’antenne de façon que le diagramme en gain se rapproche autant possible de l’objectif.

3.2.4.3 Synthèse large bande en gain seulement

Le diagramme de référence est en réalité une fonction à valeurs complexes, définissant un gain et une phase pour chaque couple $(\theta, k) \in \mathbb{R}^2$. Pourtant, nous n’avons jusqu’à maintenant représenté que le gain du diagramme d’antenne. En effet, l’information de phase n’est pas utile au processus de localisation postérieur. Comme indiqué précédemment, le fait de fixer à zéro la phase du nombre complexe $D_d(\theta)$ pourrait trop fortement contraindre le problème d’optimisation et dégrader notablement les performances en gain du diagramme obtenu. Pour résoudre ce problème, [Wang 2003] propose dans le cas faible bande de relâcher la contrainte de phase en posant

$$\begin{array}{ll} \text{minimiser} & \varepsilon \\ \text{sous} & ||D_{k_0}^\infty(\theta) - |D_d^\infty(\theta)||^2 \leq \varepsilon, \forall \theta \in \Theta. \end{array} \quad (3.33)$$

Dans le cas particulier où les microphones sont régulièrement espacés, des techniques de synthèse de filtres numériques en gain uniquement peuvent être immédiatement adaptées à ce nouveau problème [Rossignol 2001]. Bien que le problème (3.33) sorte du cadre de l'optimisation convexe, des techniques de factorisation spectrale permettent d'en convexifier les contraintes mais restent là encore applicables uniquement pour des récepteurs régulièrement espacés [Wu 1996]. Une solution est donc proposée dans [Wang 2003] et basée sur l'algorithme suivant dit de décomposition croisée

Etape 1. W_1 étant fixé, résoudre le problème d'optimisation

$$\begin{aligned} & \text{minimiser} && \varepsilon \\ & \text{sous} && ((\mathbf{W}_1 + \mathbf{W}_2)^T \mathbf{V}(\theta, k_0))((\mathbf{W}_1 + \mathbf{W}_2)^T \mathbf{V}(\theta, k_0))^H \leq (|D_d(\theta) + \varepsilon|)^2, \\ & && 4\text{Re}\{\mathbf{V}^T \mathbf{W}_1 (\mathbf{V}^T \mathbf{W}_2)^H\} \geq \max(0, |D_d(\theta) - \varepsilon|), \quad \forall \theta \in \Theta, \end{aligned} \quad (3.34)$$

en la variable de décision W_2 .

Etape 2. Pour la valeur de W_2 ainsi déterminée, résoudre (3.34), W_1 étant la nouvelle variable d'optimisation.

Etape 3. Répéter les étapes 1 et 2 jusqu'à convergence.

Etape 4. Le vecteur $\mathbf{W}$ optimal est donné par $\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2$.

Le problème (3.34) est dit multiconvexe en $\mathbf{W}_1$ et $\mathbf{W}_2$, au sens où si l'une de ces variables est fixée, il est convexe et peut donc être résolu en utilisant le même type de solveur que précédemment. Toutefois, bien que l'étape 3 garantisse la convergence vers un optimum local, celui-ci peut ne pas être global en raison de l'initialisation arbitraire du vecteur $\mathbf{W}_1$ au tout début de l'algorithme. Evidemment, cette initialisation est délicate et se résume en la détermination d'une phase permettant au diagramme de directivité en gain de mieux s'approcher de la référence. Une amélioration de cet algorithme, que nous détaillerons pas ici, est proposée par l'auteur, et consiste à alterner des résolutions de (3.33) à l'aide de l'algorithme proposé, avec des résolutions de (3.23) garantissant la nature convexe du problème global d'optimisation.

Nous avons appliqué cette méthode au cas large bande, pour la synthèse de filtres numériques RIF. Comme précédemment, nous avons simplement ajouté la dimension fréquentielle à chacun des problèmes d'optimisation intervenant dans la résolution. Nous présentons sur la figure 3.13 une succession de résultats obtenus au fur et à mesure des différentes itérations. Nous constatons l'émergence du lobe principal d'écoute ainsi que l'atténuation de plus en plus forte dans les autres directions. Or, compte tenu du nombre de contraintes mises en jeu et du caractère récursif de l'algorithme, la résolution sous Matlab de ce nouveau problème est extrêmement longue, si bien que le résultat présenté ici a nécessité deux jours complets de calculs et a donc été interrompu avant obtention de l'optimum global. Clairement, cette méthode d'optimisation apparaît intéressante dans le cas faible bande uniquement, mais s'avère totalement inexploitable dans le cas large bande, quelle que soit sa méthode d'implémentation (temporelle ou fréquentielle).

3.2.4.4 Etude de l'influence des paramètres de synthèse

Revenons maintenant à l'optimisation du diagramme de directivité en gain et phase, et étudions en détail l'influence de chacun des paramètres de synthèse du problème (3.31). Supposons pour l'instant qu'aucune contrainte sur la puissance $\mathbf{W}_{LB}^H \mathbf{W}_{LB}$ ne soit prise en compte lors de

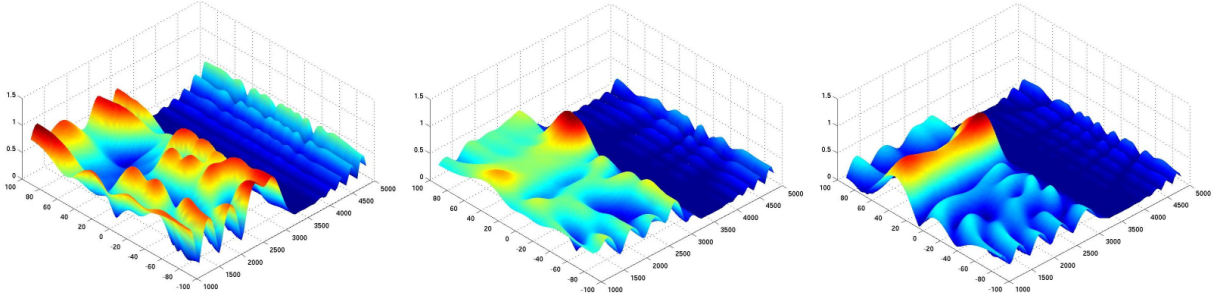
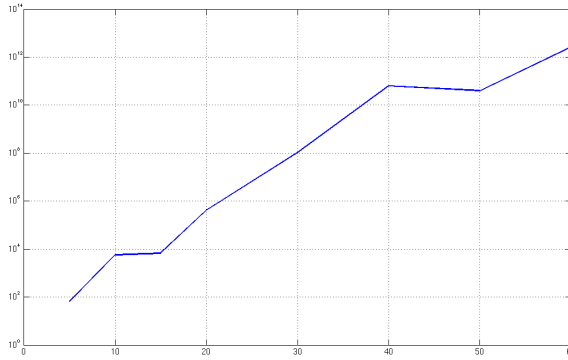
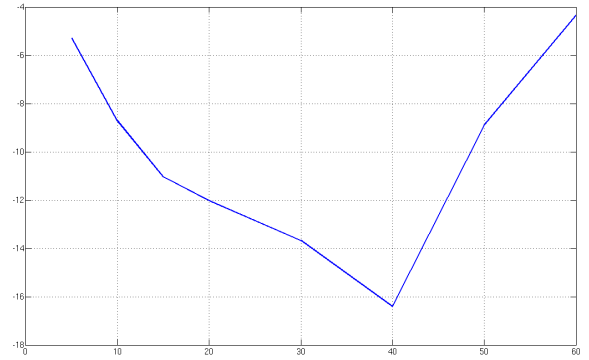


FIG. 3.13 – Réponse d’antenne pour différentes itérations lors de la synthèse convexe en gain seulement.

la résolution. La figure 3.14(a) représente l’évolution de cette puissance en fonction de l’ordre Q des filtres, pour Q compris entre 5 et 60. Nous observons une explosion pouvant atteindre 10^{12} pour l’ordre le plus élevé. Cependant, cette tendance s’accompagne d’une amélioration de la réponse d’antenne pour des ordres Q inférieurs à 40. En effet, comme l’indique la figure 3.14(b),



(a) Evolution de la puissance des coefficients $W^H W$ en fonction de l’ordre Q des filtres numériques.



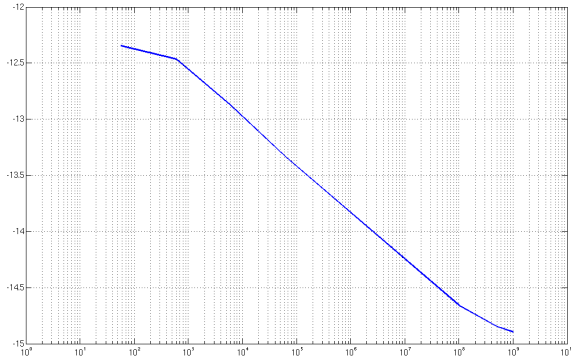
(b) Evolution du niveau des lobes secondaires, exprimé en dB, en fonction de l’ordre Q des filtres numériques.

FIG. 3.14 – Etude de l’influence de l’ordre des filtres RIF sur les résultats d’optimisation, sans limite δ , pour $N = 8$, $\theta_c = 60^\circ$, $\theta_p = 10^\circ$ et $\theta_s = 26^\circ$.

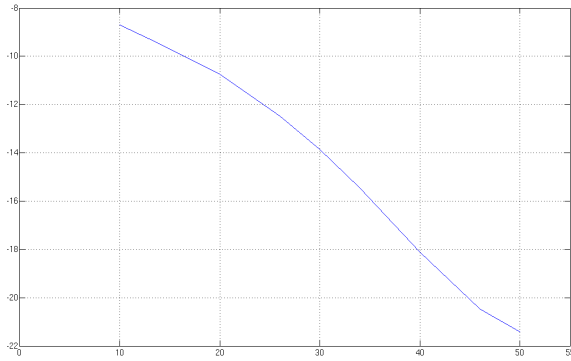
le niveau des lobes secondaires diminue rapidement lorsque l’ordre augmente. Cependant, nous pouvons remarquer que cette tendance s’inverse pour des ordres supérieurs à 40 : en effet, le solveur connaît des problèmes numériques pour ces valeurs et n’est pas capable de déterminer un vecteur $\mathbf{W}_{LB}$ satisfaisant l’ensemble des contraintes, le nombre d’inconnues devenant vraisemblablement trop grand. Dans ce cas, le vecteur $\mathbf{W}_{LB}$ obtenu contient alors un grand nombre de valeurs nulles. Au final, il est donc nécessaire de sélectionner un ordre Q suffisamment élevé pour obtenir des lobes secondaires de faibles amplitudes.

Il convient également de limiter $\mathbf{W}_{LB}^H \mathbf{W}_{LB}$ au moyen de (3.32). Considérons donc maintenant que nous avons fixé la valeur $Q = 35$, et examinons l’influence de la limite supérieure δ de $\mathbf{W}_{LB}^H \mathbf{W}_{LB}$. La figure 3.15(a) suggère naturellement que plus la valeur de δ est faible, plus le niveau des lobes secondaires augmente. Nous avons déjà décrit ce comportement auparavant, visible figure 3.11, souvent accompagné d’oscillations du niveau du lobe principal. Cependant,

cette perte d'efficacité du filtrage spatial peut être compensée en élargissant la taille du lobe



(a) Evolution du niveau des lobes secondaires, exprimé en dB, en fonction de la limite δ sur la puissance des coefficients.



(b) Evolution du niveau des lobes secondaires, exprimé en dB, en fonction de la largeur du lobe principal θ_s du comportement désiré $D_d(\theta)$.

FIG. 3.15 – Influence de la limite δ

principal. En effet, pour une contrainte sur la puissance des coefficients fixée à $\delta = 1000$, et toujours avec $Q = 35$, la figure 3.15(b) nous indique que plus le lobe principal est large, plus le niveau des lobes secondaires est faible. Ce comportement est principalement lié au faible nombre de microphones utilisé lors de la synthèse, et nous indique qu'il faut effectuer un compromis entre un lobe principal plus fin et un fort niveau d'atténuation dans les directions non souhaitées.

3.2.5 Conclusion

Nous avons présenté dans cette partie une première méthode naïve de synthèse de formation de voie large bande. Cette extension immédiate d'un résultat issu du traitement du signal permet de mettre en évidence toute la difficulté à traiter le contenu basse fréquence des signaux sonores à partir d'une antenne de petite dimension. Ainsi, nous avons montré qu'une diminution sensible de la largeur du lobe principal de la réponse d'antenne pour les basses fréquences est accompagnée systématiquement d'une augmentation de l'amplification du bruit blanc. Si effectivement cette perte de robustesse vis à vis du bruit peut paraître rédhibitoire, il ne faut pas perdre de vue que la formation de voie conventionnelle, pour ces mêmes fréquences, reste très peu utilisable puisque le filtrage spatial opéré par l'antenne s'avère de très mauvaise qualité, comme illustré figure 3.16(b). Hélas, la résolution de (3.23) s'avère parfois très longue, compte tenu du nombre très important de contraintes qu'elle comporte, et tenter d'étendre l'optimisation pour y inclure éventuellement des contraintes supplémentaires de robustesse vis à vis des incertitudes sur les caractéristiques de microphones paraît impossible. Pour résoudre ce problème, nous proposons dans la suite de ce chapitre un nouveau formalisme permettant de réduire drastiquement le nombre de contraintes.

Nous pouvons également remarquer, à travers ces premiers exemples de synthèse, l'émergence d'une forme privilégiée du lobe principal. Il est certain que le comportement désiré $D_d^\infty(\cdot)$ tel que défini dans cette partie serait difficile à atteindre compte tenu du faible nombre de microphones. Les formes quasi-rectangulaires paramétrées par les angles θ_p et θ_s ne sont en effet pas adaptées à une antenne de faible dimension. Elargir le lobe principal permet cependant d'améliorer les résultats, comme en témoigne la synthèse visible figure 3.16(a) : au prix d'une légère déviation angulaire du niveau maximal, le comportement de la formation de voie est quasiment invariant

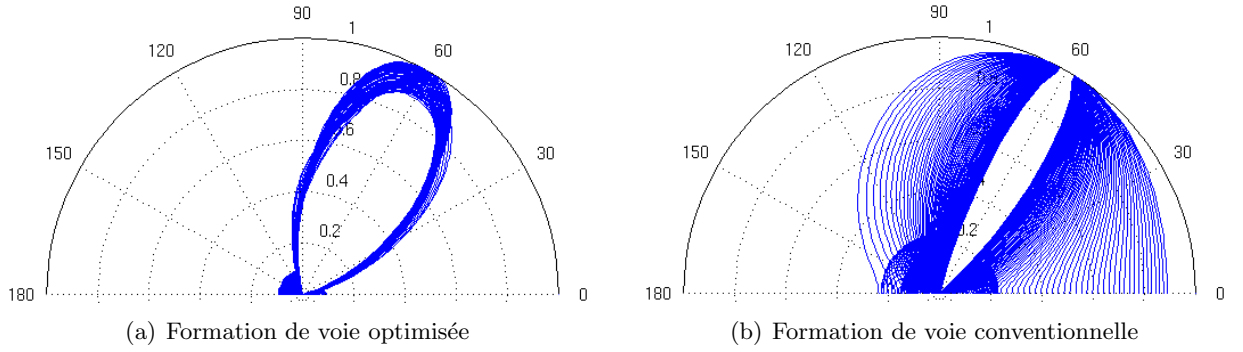


FIG. 3.16 – Comparaison formation de voie conventionnelle / optimisée, pour $f \in [400; 3000]\text{Hz}$, avec $G_B \leq 1000$, $Q = 35$ et $\theta_s = 46^\circ$.

en fréquence et présente un faible niveau des lobes secondaires. Enfin, l'ensemble des résultats présentés jusqu'ici est basé sur l'hypothèse de champ lointain. Nous avons en effet systématiquement travaillé sur les expressions en l'infini des réponses d'antenne et donc supposé que les sources à localiser étaient suffisamment lointaines pour assimiler les fronts d'ondes parvenant au capteur à des plans. Nous proposons donc d'étudier maintenant les conséquences d'une fausse hypothèse sur la distance à la source afin de vérifier si une formation de voie large bande quasi-invariante en fréquence et synthétisée en champ lointain, conserve de bonnes propriétés de filtrage spatial lorsque la source à localiser se situe dans le champ proche.

3.3 Influence de la distance : du champ lointain au champ proche

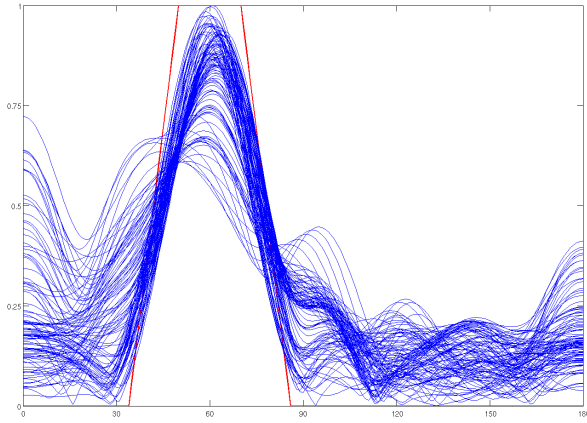
Tous les résultats de synthèse présentés jusqu'ici reposent sur l'hypothèse de champ lointain. Cette hypothèse, systématiquement invoquée en robotique, permet de considérer une expression simplifiée de la réponse d'antenne. Pourtant, de nombreux résultats de localisation présentés dans le chapitre 2 de ce manuscrit sont obtenus en imaginant la source sonore à localiser dans un rayon de 1 à 3 mètres. A de telles distances, l'hypothèse de champ lointain est-elle encore vérifiée ? Nous proposons de répondre à cette question dans cette partie en étudiant l'influence de la distance sur la forme de la réponse d'antenne. Cette étude, comme nous le verrons, est assez immédiate mais n'a jamais été proposée dans la communauté robotique à notre connaissance. Elle apparaît cependant fondamentale dans le cadre de l'utilisation d'un capteur sonore embarqué pour l'interaction homme - robot par exemple. Pour illustrer cette remarque, considérons un scénario d'interaction très simple, mettant en jeu un robot équipé d'une caméra pour la reconnaissance et le suivi de visage, ainsi que d'un capteur sonore. Compte tenu du champ visuel souvent restreint des caméras (généralement pas plus de 25° en azimut autour de l'axe optique), il est peu probable que deux personnes en interaction à mi-distance avec le robot se projettent dans la même image dès lors qu'elles sont suffisamment distantes l'une de l'autre. Dès lors, comment détecter que l'un de ces deux tuteurs parle au robot pour ensuite braquer la caméra dans sa direction ? Le capteur sonore peut permettre de détecter la présence d'un signal sonore en provenance d'une certaine direction, à condition que l'algorithme soit adapté à cette distance proche d'interaction homme-robot. A notre connaissance, cette considération n'a jamais été approfondie en robotique.

Soit $\mathbf{W}_{LB}^\infty$ le vecteur contenant, dans le cas d'une implémentation dans le domaine temporel,

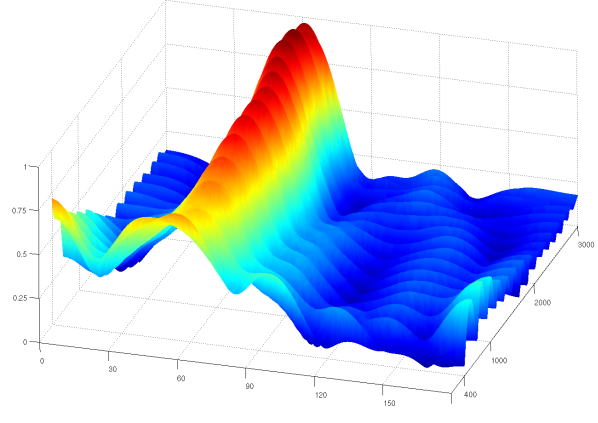
l'ensemble des coefficients des filtres réalisant la formation de voie large bande en champ lointain caractérisé par le diagramme $D_{LB}^\infty(\theta, k)$. Examiner l'influence de la distance r revient à tracer la réponse d'antenne $D^{\infty \rightarrow r}(\mathbf{r}, k) = D^{\infty \rightarrow r}(r, \theta, k)$ définie, pour ce vecteur $\mathbf{W}_{LB}^\infty$, par

$$D^{\infty \rightarrow r}(\mathbf{r}, k) = \mathbf{W}_{LB}^{\infty T} \mathbf{V}_{LB}(\mathbf{r}, k), \quad (3.35)$$

lorsque r varie. Par exemple, considérons la réponse d'antenne représentée de la figure 3.10, obtenue sans limiter la puissance des coefficients de filtres. Pour une distance théorique infinie à la source, la réponse d'antenne possède de bonnes caractéristiques spatiales, et conduit à une amélioration notable de la résolution des basses fréquences. Exploitions maintenant la formule 3.35 pour déduire le comportement de l'antenne à une distance de 1 m, représentative d'une distance classique d'interaction entre deux personnes [Hall 1966]. A une telle distance, la réponse d'antenne devient celle reportée figure 3.17. Très clairement, l'amélioration recherchée –et obtenue– en champ lointain est totalement perdue en utilisant cette même antenne à une distance de 1 m. Comme nous pouvons le constater, les caractéristiques spatiales de la réponse d'antenne sont no-



(a) Superposition des diagrammes d'antenne obtenus pour l'ensemble des fréquences en fonction de θ .



(b) [Représentation de la réponse d'antenne dans le domaine fréquentiel.

FIG. 3.17 – Réponse d'antenne optimisée, synthétisée en champ lointain, et utilisée à une distance de 1m.

tablement dégradées, en particulier dans les plus basses fréquences. Les lobes secondaires voient leur niveau augmenter à tel point qu'il n'est parfois plus possible de distinguer le lobe principal d'écoute. Examinons maintenant l'évolution de cette déformation en fonction de la distance, en traçant la réponse d'antenne $P^{\infty \rightarrow r}(\mathbf{r}, k)$ pour la fréquence $f = 400\text{Hz}$ et différentes valeurs de r . Les tracés obtenus sont visibles sur la figure 3.18 pour 4 valeurs différentes de la distance à la source. A une distance de 5m, le niveau des lobes secondaires commence à augmenter légèrement ; cependant, la forme du lobe principal reste quasiment inchangée. Ce n'est par contre pas le cas à 2m, distance à laquelle le lobe principal voit sa largeur augmenter significativement. De la même façon, le niveau des lobes secondaires augmente nettement dans les deux directions périphériques $\theta = 0^\circ$ et $\theta = 180^\circ$. Enfin, à une distance de 1.25m, la formation de voie devient inexploitable, tant le lobe principal est noyé dans les lobes secondaires devenus extrêmement importants.

Cette étude rapide nous montre donc les limites de l'hypothèse de champ lointain. Si elle permet de simplifier énormément le problème de la propagation en assimilant l'ensemble des

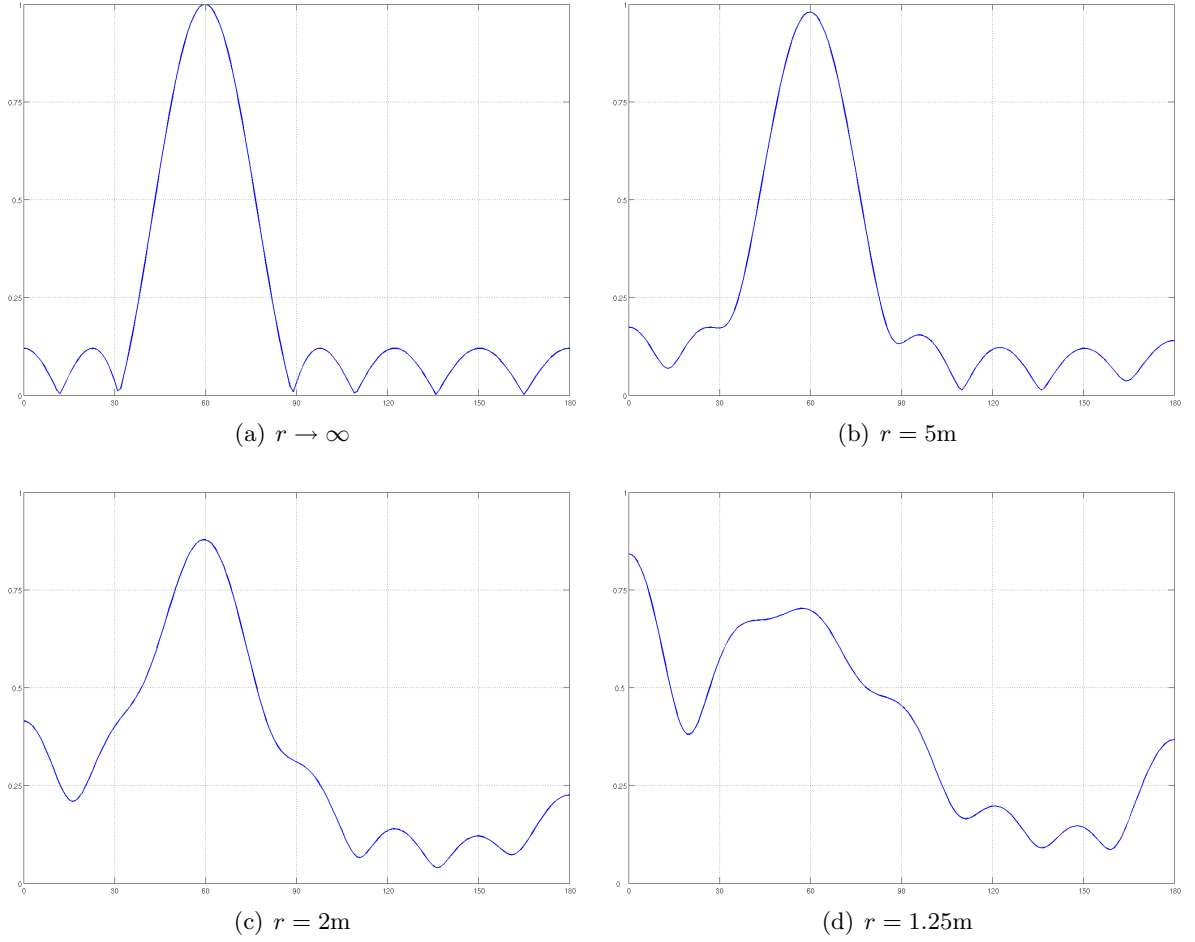


FIG. 3.18 – Evolution de la réponse d’antenne faible bande pour la fréquence $f = 400\text{Hz}$ pour différentes distances r à l’antenne.

fronts d’onde à des plans, elle ne permet pas de travailler sur des distances proches de l’antenne avec des sources comportant des basses fréquences. Dès lors, à partir de quelle distance est-il possible de considérer le champ lointain ? La communauté Acoustique utilise classiquement comme critère la distance de Rayleigh, exprimant la distance d_R au delà de laquelle cette hypothèse est valide, et donnée par

$$d_R = \frac{2L^2}{\lambda}, \quad (3.36)$$

où L désigne la plus grande dimension de l’antenne, et λ la longueur d’onde considérée. Cette expression permet de mettre en évidence que cette distance limite est une fonction de la fréquence et de la qualité de l’échantillonnage spatial opéré par l’antenne à travers sa plus grande dimension L . Cependant, elle ne suffit pas totalement à décrire les phénomènes liés au champ proche : en effet, si nous traçons la réponse d’antenne d’une formation de voie conventionnelle à une distance de 1 m, celle-ci reste quasiment inchangée quelle que soit la fréquence, si bien qu’il est possible de supprimer totalement cette classification proche/lointain. Cette constatation semble indiquer que c’est bien le caractère invariant en fréquence de la réponse d’antenne optimisée qui est à

l'origine des déformations constatées en champ proche, et cette invariance n'est pas conservée lors de l'utilisation de l'antenne pour localiser une source située à une distance ne vérifiant pas l'hypothèse locale d'ondes planes. Idéalement, il serait donc nécessaire de travailler avec une formation de voie dont les caractéristiques spatiales sont invariantes vis à vis de la distance à la source. Une antenne linéaire de petite dimension ne permet pas d'obtenir facilement cette invariance, car seules des antennes à deux dimensions peuvent éventuellement atteindre cette caractéristique par construction [Abhayapala 1999], au prix d'un encombrement et d'un nombre de microphones plus grand, incompatible avec une utilisation en robotique. [Ward 1997] tente tout de même de conférer à une antenne linéaire une telle propriété, en cherchant minimiser l'écart entre les réponses d'antenne exprimées en champ proche et lointain, mais au prix de lobes secondaires parfois élevés. Il est donc nécessaire de résoudre le problème d'optimisation précédent pour différentes distances en travaillant cette fois sur la réponse d'antenne $D(\mathbf{r}, \theta)$, dépendant explicitement de r . Or, comme nous l'avons vu dans la partie précédente, la méthode de synthèse utilisée jusqu'ici est constituée d'un très grand nombre de contraintes, à l'origine de problèmes numériques et de délais de résolution parfois très longs. Nous proposons dans la partie suivante une méthode de synthèse originale, adaptée spécifiquement à la formation de voie, permettant de réduire nettement le nombre de contraintes et valable à la fois en champ proche ou en champ lointain.

3.4 Synthèse modale large bande par optimisation convexe

La méthode de synthèse proposée ci-après est basée sur la décomposition modale des réponses d'antenne sur la base des harmoniques sphériques. L'emploi de ce formalisme adapté spécifiquement à la formation de voie permet le traitement simultané du champ proche ou du champ lointain, et son exploitation pour la synthèse à travers un problème d'optimisation convexe constitue une contribution originale. Nous présentons dans une première partie les éléments théoriques relatifs à l'analyse modale des diagrammes de directivité, en résumant les articles [Kennedy 1998] et [Kennedy 1999]. Ensuite, à partir de cette écriture, nous posons un nouveau problème d'optimisation convexe permettant de déterminer les coefficients ou filtres réalisant la formation de voie. Enfin, nous comparons les résultats de synthèse avec ceux obtenus lors de la résolution du problème précédent.

3.4.1 Éléments d'analyse modale

3.4.1.1 Généralités

Nous avons redémontré dans le chapitre 1 la solution de l'équation d'onde sphérique. Nous sommes parvenus à l'expression suivante du potentiel des vitesses $\phi(\mathbf{r}, k)$

$$\phi(\mathbf{r}, k) = \sum_{m=0}^{\infty} \sum_{n=-m}^m h_m^{(2)}(kr) P_m^n(\cos \theta) (\mathcal{A}_{mn} \cos(n\psi) + \mathcal{B}_{mn} \sin(n\psi)), \quad (3.37)$$

où $\mathcal{A}_{mn}$ et $\mathcal{B}_{mn}$ désignent deux constantes réelles d'intégration. Dans la suite, nous manipulerons la représentation complexe de cette solution, qui s'écrit sous la forme

$$\phi(\mathbf{r}, k) = \sum_{m=0}^{\infty} \sum_{n=-m}^m \alpha_{mn}(k) M_{mn}(\mathbf{r}, k), \quad (3.38)$$

où les coefficients complexes $\alpha_{nm}(k)$, appelés *coefficients modaux*, sont fonction de la fréquence k , et

$$M_{mn}(\mathbf{r}, k) \triangleq h_m^{(2)}(kr) Y_{nm}(\theta, \psi), \quad (3.39)$$

désignent les modes de la solution de l'équation d'Helmholtz (1.29). Les fonctions

$$Y_{nm}(\theta, \psi) \triangleq \sqrt{\frac{2m+1}{4\pi} \frac{(m-|n|)!}{(m+|n|)!}} P_m^n(\cos \theta) e^{jn\psi}, \quad (3.40)$$

pour $n = -m, \dots, m$ et $n \in \mathbb{Z}^+$ sont connues dans la littérature sous le nom d'*harmoniques sphériques*, et constituent une base orthonormale. En effet, à partir des propriétés d'orthogonalité des polynômes de Legendre $P_m^n(\cdot)$ et des exponentielles, apparaissant tous deux dans l'expression (3.40), et selon le produit scalaire usuellement défini sur l'ensemble des fonctions de puissance finie, nous avons

$$\int_0^{2\pi} \int_0^\pi Y_{mm}(\theta, \psi) Y_{m'n'}^*(\theta, \psi) \sin \theta d\theta d\psi = \delta_{nn'} \delta_{mm'}, \quad (3.41)$$

où δ_{mn} désigne le symbole de Kronecker prenant la valeur 0 (respectivement 1) si $n \neq m$ (respectivement $n = m$). Une formation de voie consiste en une combinaison linéaire des signaux reçus sur chacun des microphones. Or, ces mêmes signaux vérifient l'équation d'Helmholtz (1.29) et peuvent donc être formulés de manière analogue à (3.38). Par conséquent, compte tenu de la nature linéaire des traitements effectués par une formation de voie, n'importe quelle réponse d'antenne peut aussi s'exprimer sous la forme (3.38). Cette remarque fondamentale nous permet donc d'introduire un nouveau cadre pour l'expression des réponses d'antenne, à l'origine des développements théoriques suivant.

3.4.1.2 Expression modale d'une formation de voie

Dans la littérature, une formation de voie est généralement formalisée différemment selon que l'on considère le champ proche ou le champ lointain. A chacun de ces cas correspondent une écriture des méthodes de synthèse spécifiques. Les harmoniques sphériques permettent cependant d'unifier l'expression de la réponse d'antenne pour toutes les distances et toutes les fréquences.

Du fait que les fonctions de Hankel de seconde espèce admettent comme forme asymptotique

$$h_m^{(2)}(kr) = j^{m+1} \frac{1}{kr} e^{-jkr} \text{ lorsque } r \rightarrow \infty, \quad (3.42)$$

les modes (3.39) décroissent vers zéro lorsque r tend vers l'infini. Or, ce passage à la limite permet précisément de passer du champ proche au champ lointain. Pour conserver cette propriété dans l'expression unifiée du diagramme de directivité, normalisons les modes $M_{mn}(\mathbf{r}, k)$ en

$$\bar{M}(\mathbf{r}, k) = r e^{ikr} M(\mathbf{r}, k) = R_m(r, k) Y_{mn}(\theta, \psi), \text{ où } R_m(r, k) \triangleq r e^{ikr} h_m^{(2)}(kr). \quad (3.43)$$

Ainsi, l'amplitude du mode normalisé $\bar{M}(\mathbf{r}, k)$ prend une valeur finie lorsque r tend vers l'infini, puisque $\lim_{r \rightarrow \infty} R_m(r, k) = \frac{j^{m+1}}{k}$. Notons que la dépendance angulaire des modes $\bar{M}(\mathbf{r}, k)$ est entièrement déterminée par les fonctions $Y_{mn}(\theta, \phi)$, puisque $R_m(r, k)$ n'est pas fonction des angles θ et ϕ . Les harmoniques sphériques définissent donc un ensemble de formes élémentaires permettant de générer par combinaison linéaire n'importe quel diagramme d'antenne. La figure

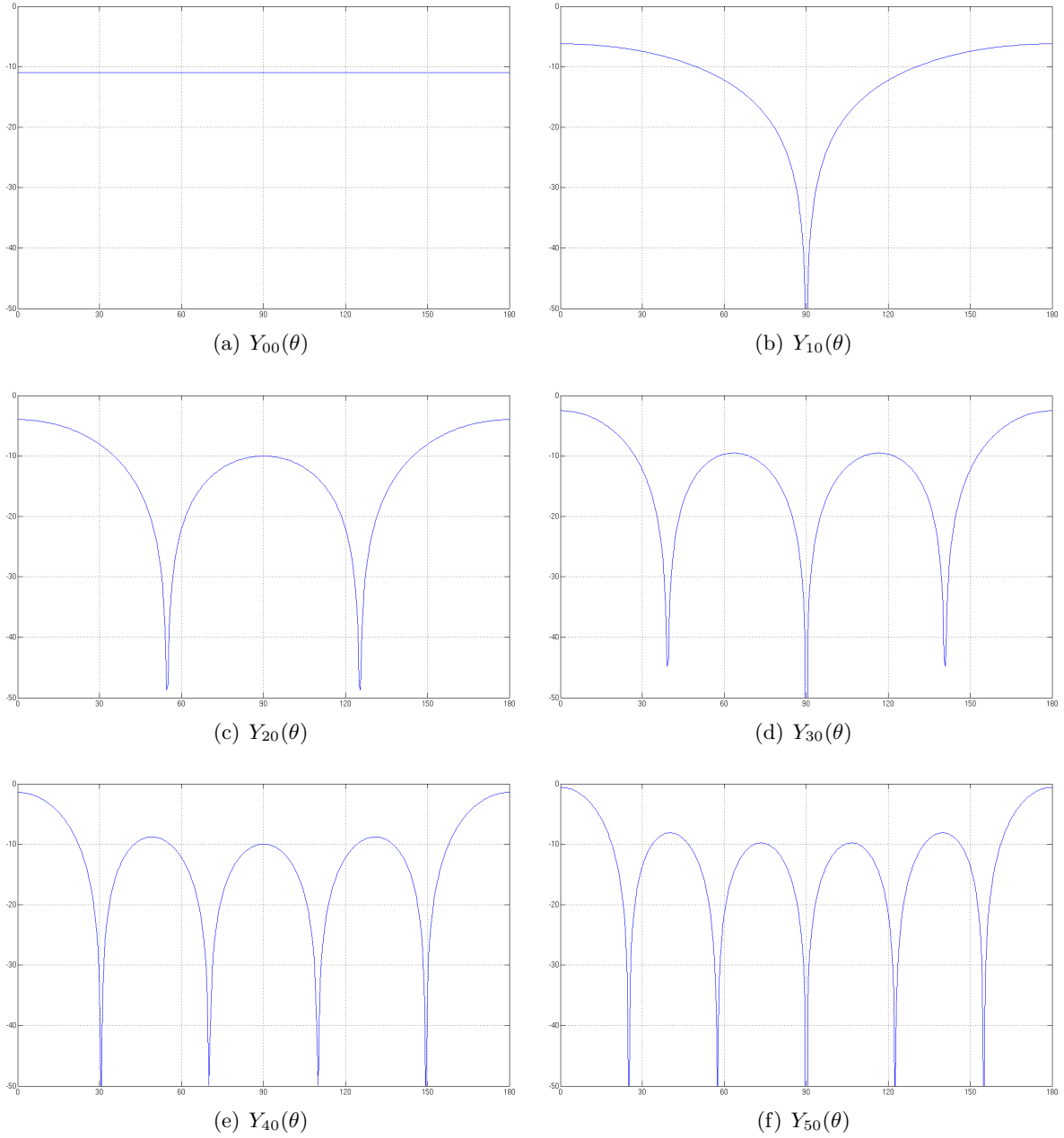


FIG. 3.19 – Tracés de l'amplitude des harmoniques sphériques $Y_{mn}(\theta)$, $n = 0$, en fonction de θ

3.19 représente les tracés de quelques harmoniques d'ordre faible, pour $n = 0$. Finalement, n'importe quel diagramme d'antenne $D(\mathbf{r}, k)$ peut donc s'écrire sous la forme

$$D(\mathbf{r}, k) = \sum_{m=0}^{\infty} \sum_{n=-m}^m \alpha_{mn}(k) R_m(r, k) Y_{mn}(\theta, \psi) , \quad (3.44)$$

et par conséquent être caractérisé par les constantes complexes $\alpha_{mn}(k)$. On montre en effet, à partir de la propriété d'orthogonalité (3.41), que ces coefficients modaux peuvent s'exprimer

sous la forme

$$\alpha_{mn}(k) = \frac{1}{R_m(r, k)} \int_0^{2\pi} \int_0^\pi D(\mathbf{r}, k) Y_{m'n'}^*(\theta, \psi) \sin \theta \, d\theta d\psi. \quad (3.45)$$

Les équations (3.44) et (3.45) définissent une transformation orthogonale analogue à la série de Fourier. Cette transformation en possède les mêmes propriétés importantes, telles que la linéarité et une relation de Parseval. Cette dernière est fondamentale et nous permettra d'approximer la somme infinie (3.44) par une somme finie restreinte à M modes.

3.4.1.3 Relation de Parseval

Approximons une réponse d'antenne $D(\mathbf{r}, k)$, dont l'expression rigoureuse est donnée par (3.44), par une somme finie de termes modaux. Cette approximation, notée $\hat{D}(\mathbf{r}, k)$, peut être écrite, pour une valeur M donnée

$$\hat{D}(\mathbf{r}, k) = \sum_{m=0}^M \sum_{n=-m}^m \hat{\alpha}_{mn}(k) R_m(r, k) Y_{mn}(\theta, \psi). \quad (3.46)$$

Cherchons maintenant quels sont les $\hat{\alpha}_{mn}(k)$ qui minimisent l'erreur quadratique ε_M entre la représentation exacte $D(\mathbf{r}, k)$ et $\hat{D}(\mathbf{r}, k)$. De manière analogue à l'approximation du minimum d'erreur quadratique obtenue par troncature d'une série de Fourier, il est montré dans [Kennedy 1998] que cette erreur est minimale lorsque $\hat{\alpha}_{mn}(k) = \alpha_{mn}(k)$, pour $m = 0, \dots, M$ et $n = -m, \dots, m$, et vaut

$$\varepsilon_M = \int_0^{2\pi} \int_0^\pi |D(\mathbf{r}, k) - \hat{D}(\mathbf{r}, k)|^2 \sin \theta \, d\theta d\phi = \sum_{m=M+1}^\infty \sum_{n=-m}^m |\alpha_{mn}(k) R_m(r, k)|^2. \quad (3.47)$$

Ainsi, l'équation (3.47) précise la valeur de l'erreur résultant de la troncature d'ordre M de la somme infinie (3.44), d'autant plus faible que le nombre de coefficients modaux M est important. Cependant nous pouvons déjà intuitivement, tant la ressemblance avec la transformation de Fourier utilisée en analyse fréquentielle est grande, que seuls les coefficients d'ordres faibles sont importants dans l'approximation de forme générale de la réponse d'antenne, les ordres plus élevés étant seulement nécessaires à la descriptions des détails fins. Pour illustrer cette remarque, nous avons tracé sur la figure 3.20, en trait épais, la réponse d'antenne $D^\infty(\theta)$ donnée par

$$D^\infty(\theta) = \sum_{n=1}^N e^{-jk_0 z_n \cos \theta}, \quad (3.48)$$

pour l'antenne linéaire considérée jusqu'ici, à la fréquence $f_0 = 3\text{kHz}$. Sans détailler pour l'instant la méthode de calcul des coefficients modaux $\alpha_m(k_0)$, nous avons superposé sur cette même figure la représentation de $\hat{D}^\infty(\theta, k_0)$ pour différentes valeurs de M . Nous pouvons constater que le tracé exact et celui obtenu avec seulement 15 termes sont quasiment confondus, et que visiblement l'erreur commise en tronquant la somme infinie est extrêmement faible. Evidemment, même si nous ne l'avons pas représentée ici, la phase est elle aussi parfaitement décrite à partir de ces mêmes 15 coefficients. Ainsi, n'importe quelle formation de voie peut être convenablement décrite par un nombre fini de coefficients modaux. C'est cette propriété qui est la clé de la nouvelle méthode de synthèse à venir : plutôt que d'optimiser sur une grille venant discrétiser les valeurs désirées de la réponse d'antenne, nous poserons le problème dans l'espace modal en

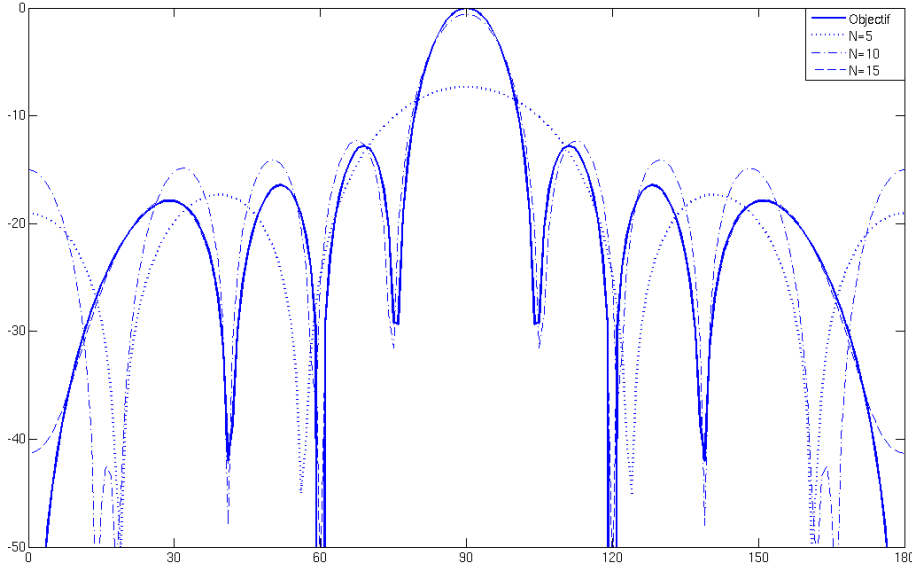


FIG. 3.20 – Approximation par un nombre fini M de coefficients modaux de la réponse d'antenne donnée par (2.56), pour une antenne de $N = 8$ microphones à la fréquence de 3kHz.

représentant par leurs coefficients modaux le comportement de référence $D_d(\cdot)$ et la réponse d'antenne recherchée.

3.4.2 Calculs des coefficients modaux

Le calcul des coefficients modaux $\alpha_{mn}(k)$, étant donné une réponse d'antenne $D(\mathbf{r}, k)$, nécessite le calcul de l'intégrale (3.45). Pour une réponse d'antenne quelconque, et compte tenu de la forme de cette intégrale, ce calcul ne peut être effectué que numériquement. Cependant, nous avons jusqu'à présent découplé le diagramme d'antenne $D(\cdot)$, fonction mathématique des variables d'espace et du temps, de la façon pratique dont il est obtenu, à savoir à travers un réseau de capteurs discrets répartis dans l'espace. Il est montré dans [Abhayapala 1999] que la réponse d'une telle antenne "physique" peut également s'exprimer sous une forme modale. Nous rappelons ici ce résultat essentiel, puis en déduisons l'expression générale des coefficients modaux. Nous nous intéresserons ensuite au cas particulier des antennes linéaires.

3.4.2.1 Dans le cas général

Comme indiqué dans le §2.2.4, le diagramme d'antenne $D(\mathbf{r}, k)$ d'une formation de voie réalisée à partir de N microphones situés en $\mathbf{r}_1^m, \dots, \mathbf{r}_N^m$ s'écrit sous la forme

$$D(\mathbf{r}, k) = \sum_{n=1}^N W_n(k) V_n(\mathbf{r}, k), \quad (3.49)$$

où $W_n(k)$ désigne la fonction de transfert du filtre placé en aval du $n^{\text{ième}}$ microphone. Il est rappelé dans [Abhayapala 1999] que l'élément $V_n(\mathbf{r}, k)$ peut s'exprimer également sous la forme d'une somme pondérée de modes $\bar{M}(\mathbf{r}, k)$; en effet, nous avons [Ziomek 1995]

$$\begin{aligned}
 V_n(\mathbf{r}, k) &= V_n(r, \theta, \Phi, k) \\
 &= \frac{\|\mathbf{r}\|}{\|\mathbf{r}_n^m - \mathbf{r}\|} e^{jk_0(\|\mathbf{r}\| - \|\mathbf{r}_n^m - \mathbf{r}\|)} \\
 &= \sum_{m=0}^{\infty} \sum_{n=-m}^m \left[-j4\pi j_m(k\|\mathbf{r}_n^m\|) Y_{mn}^*\left(\frac{\mathbf{r}_n^m}{\|\mathbf{r}_n^m\|}\right) \right] R_m(\|\mathbf{r}\|, k) Y_{mn}\left(\frac{\mathbf{r}}{\|\mathbf{r}\|}\right), \quad (3.50)
 \end{aligned}$$

pour $r = \|\mathbf{r}\| \geq \|\mathbf{r}_n^m\|$, $n \in [1, \dots, N]$. La fonction $j_m(x)$, appelée *fonction de Bessel sphérique*, est définie par

$$j_m(x) \triangleq \sqrt{\frac{\pi}{2x}} J_{m+\frac{1}{2}}(x), \quad (3.51)$$

où $J_m(x)$ désigne la fonction de Bessel de première espèce. En combinant l'équation (3.50) avec (3.49), il vient alors immédiatement

$$D(\mathbf{r}, k) = \sum_{m=0}^M \sum_{n=-m}^m \left[-j4\pi \sum_{n=1}^N W_n(k) j_m(k\|\mathbf{r}_n^m\|) Y_{mn}^*\left(\frac{\mathbf{r}_n^m}{\|\mathbf{r}_n^m\|}\right) \right] R_m(\|\mathbf{r}\|, k) Y_{mn}\left(\frac{\mathbf{r}}{\|\mathbf{r}\|}\right). \quad (3.52)$$

Ainsi, par identification avec (3.44), nous pouvons exhiber l'expression générale des coefficients modaux $\alpha_{mn}(k)$

$$\alpha_{mn}(k) = -j4\pi \sum_{n=1}^N W_n(k) j_m(k\|\mathbf{r}_n^m\|) Y_{mn}^*\left(\frac{\mathbf{r}_n^m}{\|\mathbf{r}_n^m\|}\right). \quad (3.53)$$

Comme nous pouvions nous y attendre, les coefficients $\alpha_m(k)$ dépendent uniquement des paramètres de l'antenne (vecteur position des microphones $\mathbf{r}_n^m$ et filtres W_n) et de la fréquence. En résumé, nous retiendrons que

$$\begin{aligned}
 D(\mathbf{r}, k) &= \sum_{n=1}^N W_n(k) V_n(\mathbf{r}, k) = \sum_{m=0}^{\infty} \sum_{n=-m}^m \alpha_{mn}(k) R_m(r, k) Y_{mn}(\theta, \psi), \\
 \text{avec } \alpha_{mn}(k) &= -j4\pi \sum_{n=1}^N W_n(k) j_m(k\|\mathbf{r}_n^m\|) Y_{mn}^*\left(\frac{\mathbf{r}_n^m}{\|\mathbf{r}_n^m\|}\right)
 \end{aligned}$$

3.4.2.2 Pour une antenne linéaire

Soit une antenne linéaire analogue à celle utilisée précédemment, constituée de N microphones situés aux abscisses $z_1, \dots, z_N$ de l'axe $\mathcal{Z}$. Pour des raisons de symétrie, la réponse d'un tel réseau de microphones est nécessairement invariante selon ϕ . Ainsi, l'expression modale générale (3.44) se simplifie en faisant disparaître la dépendance par rapport à l'angle ϕ pour devenir

$$D(\mathbf{r}, k) = \sum_{m=0}^{\infty} \alpha_m(k) R_m(\mathbf{r}, k) Y_m(\theta), \quad (3.54)$$

avec

$$Y_m(\theta) \triangleq Y_{m0}(\theta, \phi) = \sqrt{\frac{2m+1}{4\pi}} P_m(\cos \theta), \text{ avec } P_m(\cos \theta) \triangleq P_m^0(\cos \theta). \quad (3.55)$$

Evidemment, la propriété d'orthogonalité (3.41) des harmoniques sphériques reste valide dans le cas $n = 0$, et permet d'aboutir à la nouvelle expression intégrale des coefficients modaux $\alpha_m(k)$

$$\alpha_m(k) = \frac{2\pi}{R_m(r, k)} \int_0^\pi D(\mathbf{r}, k) Y_m(\theta) \sin \theta \, d\theta. \quad (3.56)$$

Utilisons maintenant l'expression générale de la réponse d'une formation de voie (3.49), et insérons-la dans l'intégrale précédente. Rappelons la propriété suivante des fonctions de Legendre

$$\sum_{n=-m}^n Y_{mn}^*\left(\frac{\mathbf{r}}{\|\mathbf{r}\|}\right) Y_{mn}\left(\frac{\mathbf{r}_n}{\|\mathbf{r}_n\|}\right) = \frac{2m+1}{4\pi} P_m(\cos \gamma), \quad (3.57)$$

pour laquelle, dans le cas d'une antenne linéaire, $\gamma = |\theta|$. La combinaison de (3.50) et (3.41) permet d'aboutir à l'expression

$$\alpha_m(k) = -2jk \sqrt{\pi(2m+1)} \sum_{n=1}^N W_n(k) j_m(kz_n). \quad (3.58)$$

Evidemment, cette nouvelle équation, valable uniquement pour une antenne linéaire, est un cas particulier de (3.53) démontrée pour une configuration d'antenne quelconque. L'équation (3.58) permet donc, connaissant les positions z_n des microphones de l'antenne linéaire et les filtres W_n réalisant la formation de voie, de transformer une représentation classique de sa réponse en sa forme modale. Evidemment, comme précédemment, les coefficients $\alpha_m(k)$ sont indépendants de la distance r , et l'expression (3.58) est valable en champ proche comme en champ lointain. Il est d'ailleurs fondamental de remarquer qu'une fois les coefficients modaux $\alpha_m(k)$ déterminés à partir de la configuration du réseau de microphones, il est possible de calculer la réponse d'antenne pour n'importe quelle distance r grâce à l'équation (3.54). Celle-ci nécessite rigoureusement la connaissance d'une infinité de coefficients modaux ; cependant, la propriété de Parseval évoquée précédemment permet de ne retenir qu'un nombre restreint de ces coefficients complexes. C'est ainsi que la réponse d'antenne *exacte* représentée figure (3.20) a été approximée par une somme pondérée de modes $\bar{M}(\mathbf{r}, k)$ constituée de seulement 15 termes calculés à partir de l'équation (3.58). Pour résumer, pour une antenne linéaire, nous avons

$$D(\mathbf{r}, k) = \sum_{m=0}^{\infty} \alpha_m(k) R_m(r, k) Y_m(\theta),$$

avec $\alpha_m(k) = -2jk \sqrt{\pi(2m+1)} \sum_{n=1}^N W_n(k) j_m(kz_n).$

3.4.3 Synthèse modale large bande dans le domaine fréquentiel

Nous venons de montrer que n'importe quelle réponse d'antenne peut être décomposée dans la base modale et fidèlement représentée par un nombre fini de coefficients complexes $\alpha_m(k)$,

$m = 0, \dots, M$. Nous proposons dans la suite d'exploiter cette propriété pour définir une nouvelle méthode de synthèse. Nous supposons toujours que l'antenne est linéaire ; cependant, la méthode présentée reste valable pour d'autres configurations.

3.4.3.1 Présentation du problème de synthèse

Comme précédemment, l'implémentation dans le domaine fréquentiel d'une formation de voie large bande nécessite le traitement individuel des Q composantes fréquentielles k_q , $q \in [1, \dots, Q]$, des signaux issus des microphones. Chacune de ces formations de voie, de réponse $D_{[k_q]}(\mathbf{r}) = D_{[k_q]}(r, \theta)$, peut être représentée par ses coefficients modaux $\alpha_m(k_q)$. Leur expression est donnée par (3.58), qui s'écrit sous la forme matricielle

$$\alpha_m(k_q) = -2jk\sqrt{\pi(2m+1)} \mathbf{W}_{[k_q]}^T \mathbf{J}_m(k_q), \quad (3.59)$$

avec

$$\mathbf{J}_m(k) \triangleq (j_m(kz_1), \dots, j_m(kz_N))^T. \quad (3.60)$$

L'objectif de la méthode de synthèse que nous proposons est de déterminer le vecteur $\mathbf{W}_{[k_q]}$ contenant les N coefficients complexes $w_n^{[k_q]}$, $n \in 1, \dots, N$, permettant d'approximer au mieux la réponse d'antenne de référence $D_d(r, \theta, k_q)$ représentée par ses $M+1$ coefficients modaux $\alpha_0^d(k_q), \dots, \alpha_M^d(k_q)$. Pour cela, nous proposons de résoudre le problème d'optimisation convexe

$$\begin{array}{ll} \text{minimiser} & \varepsilon_q \\ \text{sous} & |\alpha_m(k_q) - \alpha_m^d(k_q)|^2 \leq \varepsilon_q, \quad \forall m \in [0, \dots, M], \end{array} \quad (3.61)$$

de manière indépendant pour chacune des fréquences k_q . Notre objectif est donc de minimiser l'écart, en gain et phase, entre les représentations modales complexes des réponses d'antenne désirées et effectives. Comparativement au problème d'optimisation (3.28), il n'existe plus de grille Θ définissant autant de contraintes que de valeurs angulaires discrètes. Ainsi, les 181 contraintes angulaires utilisées pour obtenir la réponse d'antenne visible figure 3.7 peuvent être remplacées par $M+1$ contraintes modales, avec $M+1 \ll 181$ grâce à la relation de Parseval. Cependant, il faut veiller à ce que les coefficients modaux non contraints $\alpha_m(k_q)$, pour $m > M$, possèdent une influence négligeable sur l'expression *exacte* (3.54), laquelle requiert un nombre infini de coefficients complexes. Ceci revient à s'assurer que l'erreur définie par (3.47) est suffisamment faible pour la valeur de M retenue. Comme précédemment, il est possible et nécessaire de limiter la puissance des coefficients réalisant la formation de voie en ajoutant la contrainte convexe additionnelle

$$G_B = \mathbf{W}_{[k_q]}^H \mathbf{W}_{[k_q]} \leq \delta_q, \quad (3.62)$$

où δ_q indique la valeur maximale autorisée du gain de bruit blanc G_B de l'antenne pour la fréquence k_q . La résolution du problème (3.61) nécessite la connaissance des coefficients modaux α_m^d du comportement de référence D_d que nous cherchons à obtenir. Nous explicitons leur calcul ci-après.

3.4.3.2 Calcul des coefficients modaux du comportement de référence

D'une manière générale, tout diagramme d'antenne $D(\mathbf{r}, k)$ est périodique en θ et ϕ , de périodes respectives $T_\theta = 2\pi$ et $T_\phi = 2\pi$. Nous pouvons donc le décomposer en une série de

Fourier selon ces deux variables angulaires. Ainsi, sans perdre en généralité, nous proposons d'écrire, pour une distance r donnée

$$D(\mathbf{r}, k) = D(r, \theta, \phi, k) = \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} a_{l_1, l_2}(r, k) e^{j l_1 \theta} e^{j l_2 \psi}, \quad (3.63)$$

où les coefficients de Fourier $a_{l_1, l_2}(r, k)$, avec $(l_1, l_2) \in \mathbb{Z}^{+2}$, satisfont

$$a_{l_1, l_2}(r, k) = \frac{1}{(2\pi)^2} \int_{\theta=0}^{2\pi} \int_{\phi=0}^{2\pi} D(r, \theta, \psi, k) e^{-j l_1 \theta} e^{-j l_2 \phi} d\theta d\psi. \quad (3.64)$$

La dépendance à la distance et à la fréquence des coefficients $a_{l_1, l_2}(\cdot)$ traduit le fait qu'à priori n'importe quelle formation de voie possède un diagramme d'antenne dont la forme change selon la valeur de r et k .

L'expression (3.63) se simplifie dans le cas d'une antenne linéaire, car la dépendance à l'angle ψ disparaît. De plus, le diagramme devient paire en θ par construction, si bien que (3.63) s'écrit

$$D(\mathbf{r}, k) = D(r, \theta, k) = \sum_{l=0}^{\infty} a_l(r, k) \cos(l\theta), \quad (3.65)$$

avec

$$a_l(r, k) = \frac{1}{\pi} \int_0^{2\pi} D(r, \theta, k) \cos(l\theta) d\theta. \quad (3.66)$$

Le calcul des coefficients modaux $\alpha_m(k)$ de la réponse $D(r, \theta, k)$ d'une antenne linéaire, nécessite de résoudre l'intégrale (3.56). Pour cela, nous proposons de combiner (3.65) avec (3.56), ce qui conduit à

$$\alpha_m(r, k) = \frac{\sqrt{\pi(2m+1)}}{R_m(r, k)} \sum_{l=0}^{\infty} a_l(r, k) g_m(l), \quad (3.67)$$

avec

$$g_m(l) = \int_0^{\pi} \cos(l\theta) P_m(\cos \theta) \sin \theta d\theta. \quad (3.68)$$

Ainsi, à partir des coefficients de la série de Fourier $a_l(r, k)$ de n'importe quelle forme de diagramme $D(r, \theta, k)$ d'une antenne linéaire, nous sommes en mesure de déterminer les coefficients modaux $\alpha_m(r, k)$ sous réserve de pouvoir calculer la nouvelle intégrale (3.68). Nous avons montré, par récurrence sur m en prenant en compte les propriétés des polynômes de Legendre, que $g_m(l)$ s'écrit sous la forme

$$g_m(l) = \frac{l^2 - (m-2)^2}{l^2 - (m+1)^2} g_{m-2}(l), \quad (3.69)$$

avec

$$g_0(l) = \begin{cases} -\frac{1 + (-1)^l}{l^2 - 1} & , \text{ pour } l \neq 1 \\ 0 & , \text{ sinon} \end{cases} \quad (3.70)$$

et

$$g_1(l) = \begin{cases} \frac{-1 + (-1)^l}{l^2 - 4} & , \text{ pour } l \neq 2 \\ 0 & , \text{ sinon.} \end{cases} \quad (3.71)$$

De plus, nous pouvons montrer que $g_m(l)$ vérifie

$$\forall (m, l) \in \mathbb{Z}^2 \quad , \quad \begin{cases} g_{2m}(2l + 1) = 0 \\ g_{2m+1}(2l) = 0 \end{cases} \quad (3.72)$$

$$\forall m > l \quad , \quad g_m(l) = 0. \quad (3.73)$$

Nous venons ainsi de déterminer la décomposition modale du diagramme de directivité de n'importe quelle antenne linéaire. Celle-ci est en particulier valable pour le comportement de référence $D_d(\theta)$ défini précédemment à travers la figure 3.4. Cependant, nous avons remarqué que ses formes trapézoïdales en faisaient un objectif inatteignable lors de la synthèse, compte tenu du petit nombre de microphone constituant le capteur. Ainsi, nous suggérons de bénéficier de la représentation (3.63) pour limiter $D_d(\theta)$ à une série de Fourier admettant L termes, et proposer une fonction objectif aux formes plus douces, plus compatibles avec les dimensions de l'antenne. Un tel choix présente de plus une conséquence très intéressante : en effet, l'équation (3.73) implique que l'ensemble des coefficients modaux $\alpha_m(r, k)$ tels que $m > L$ sont nuls. En d'autres termes, si le comportement désiré est décrit par une série de Fourier constituée de $L + 1$ termes, alors sa représentation modale *exacte* est elle aussi constituée de $M + 1 = L + 1$ coefficients modaux. Nous avons illustré l'ensemble de ces remarques sur les figures 3.21(a) et 3.21(b). Sur la première est rappelée la forme utilisée précédemment, ainsi que sa série de Fourier limitée à $L + 1 = 8$ coefficients. La forme de la réponse d'antenne ainsi obtenue, représentée en rouge, est beaucoup plus proche des diagrammes de directivité obtenus par optimisation, et laisse présager par conséquent d'un meilleur comportement de la synthèse modale. La figure 3.21(b) représente

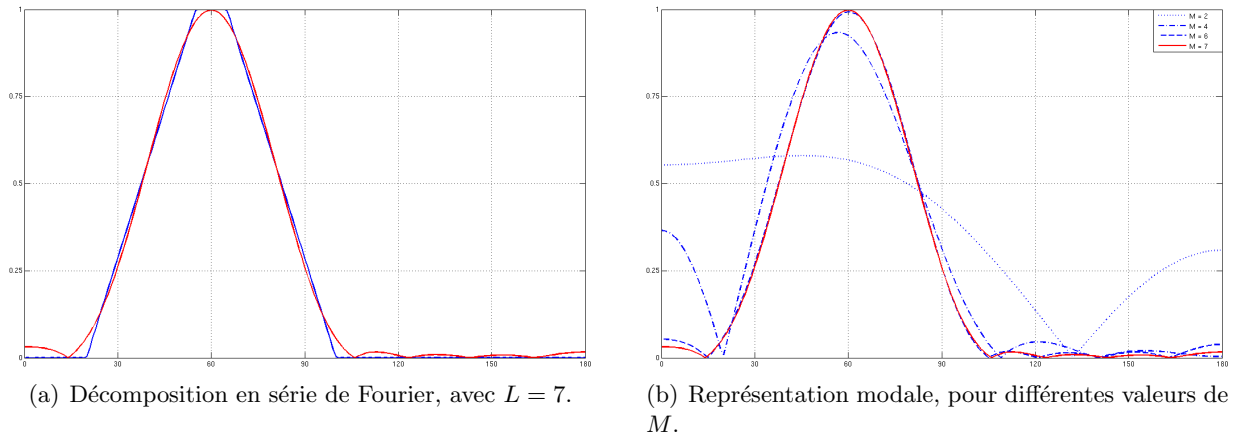


FIG. 3.21 – Décomposition modale de la réponse d'antenne de référence.

cette même série de Fourier tronquée, mais reproduite à partir de sa représentation modale. Nous constatons que le tracé obtenu à partir des $M + 1 = 8$ coefficients modaux se superpose exactement. Nous venons donc de déterminer une représentation d'une réponse d'antenne de

référence $D_d(r, \theta, k)$ possédant un nombre fini de coefficients modaux $\alpha_m^d(r, k)$ définis par

$$D_d(r, \theta, k) = \sum_{l=0}^L a_l(r, k) \cos(l\theta), \quad (3.74)$$

$$\text{avec } \alpha_m^d(r, k) = \frac{\sqrt{\pi(2m+1)}}{R_m(r, k)} \sum_{l=0}^L a_l(r, k) g_m(l). \quad (3.75)$$

L'équation précédente nous indique que les coefficients modaux sont des fonction de la fréquence k et de la distance r . Pourtant, nous avons pris soin d'indiquer précédemment que les coefficients modaux donnés par (3.58) étaient indépendant de la distance. En effet, il faut bien comprendre que nous parlons de deux types de décomposition modales différentes : la première est liée à une configuration d'antenne donnée, tandis que la seconde est relative à une fonction $D_d(\cdot)$ spécifiant la réponse d'antenne souhaitée *pour une fréquence et une distance donnée*.

3.4.3.3 Résultats d'optimisation

L'implémentation large bande dans la domaine fréquentiel nécessite la synthèse indépendante de Q formations de voie, pour des fréquences f_q que nous disposons entre 400Hz et 3kHz avec un pas de 100Hz. Nous proposons dans cet exemple de synthétiser une réponse d'antenne invariante en fréquence, pour une distance $r = r_0 = 1\text{m}$. Pour cela, nous définissons le diagramme d'antenne $D_d(r, \theta, k_q) = D_d(r_0, \theta)$ souhaité selon la figure 3.4, les coefficients de la série de Fourier a_l étant par conséquent égaux à

$$\begin{aligned} a_0(r_0) &= \frac{\theta_p + \theta_s}{\pi}, \text{ et} \\ a_l(r_0) &= \frac{4 \cos(l\theta_0)(\cos(l\theta_p) - \cos(l\theta_s))}{\pi l^2(\theta_s - \theta_p)}, \quad \forall l > 0, \end{aligned} \quad (3.76)$$

pour $\theta_s \neq \theta_p$. Dans la suite, nous fixons les paramètres angulaires $\theta_c = 60^\circ$, $\theta_p = 5^\circ$ et $\theta_s = 40^\circ$. De façon à définir un comportement de référence plus adapté aux dimensions du problème (faible nombre de microphones et petite taille de l'antenne), nous ne conservons que les $L+1 = 8$ premiers termes de la série de Fourier de la réponse d'antenne de référence, si bien que la fonction $D_d(r_0, \theta)$ est maintenant définie par (3.74). Nous pouvons donc en calculer la représentation modale à partir de (3.75), qui renvoie les valeurs des 8 coefficients modaux $\alpha_0^d(r_0), \dots, \alpha_7^d(r_0)$, et $\alpha_8^d(r_0), \dots, = \alpha_\infty^d(r_0) = 0$. Lors de la résolution du problème d'optimisation (3.61) pour $M = L = 7$, rien ne garanti que les coefficients $\alpha_m(k_q)$ admettent eux une valeur nulle pour $m > M = L$. Ainsi, une fois la minimisation effectuée et le vecteur $\mathbf{W}_{[k_q]}$ déterminé, la réponse d'antenne optimisée est tracée à l'aide de l'équation (3.49) de façon à s'affranchir d'éventuelles erreurs liées à la troncation de la somme infinie 3.4.2.2.

Optimisation sans contraintes sur G_B Intéressons nous dans un premier temps au cas où la contrainte sur la puissance G_B n'est pas prise en compte. Les résultats d'optimisation pour chacune des fréquences de synthèse sont reportés figure 3.22(a). Nous pouvons constater que le comportement angulaire obtenu est quasiment invariant en fréquence, avec cependant une très légère déviation du maximum du lobe principal de l'ordre de 1° . Le niveau des lobes secondaires, comparativement aux exemples de synthèse en champ lointain précédents, est beaucoup plus faible. Ce phénomène peut s'expliquer par l'emploi d'une fonction D_d possédant un lobe principal

plus large et d'une forme bien plus douce. La figure 3.22(b) présente la concaténation de chacune

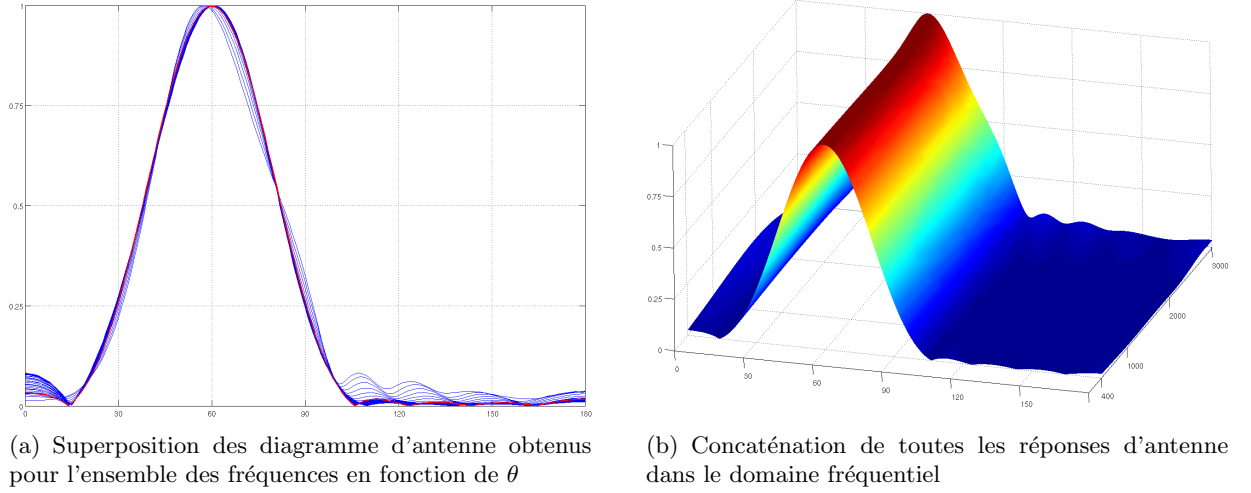


FIG. 3.22 – Formations de voie faible bande pour $f \in [400 : 100 : 3000]$, pour une distance $r_0 = 1\text{m}$

des réponses d'antenne dans le domaine fréquentiel. Visiblement, la formation de voie large bande obtenue possède un meilleur comportement en basse fréquences que dans le cas conventionnel. Cependant, si nous traçons la puissance des coefficients réalisant chacune des formations de voie faible bande en fonction de la fréquence, nous constatons que celle-ci admet de très fortes valeurs aux basses fréquences, de l'ordre de 10^8 pour une fréquence de 400 Hz. Même si cette valeur demeure très importante, elle reste tout de même 10000 fois moins grande que celle obtenue lors de l'exemple de synthèse en champ lointain précédent. Finalement, nous aboutissons à nouveau à la même constatation : il est nécessaire d'effectuer un compromis entre une résolution améliorée des basses fréquences et un gain G_B trop important.

Optimisation sous contraintes sur G_B Ajoutons maintenant la contrainte (3.62) au problème d'optimisation (3.61), avec $\delta_q = \delta = 100$ pour chacune des fréquences de synthèse. La figure 3.24 présente l'ensemble des réponses d'antennes obtenues. Nous perdons en partie l'invariance en fréquence recherchée. Bien que les hautes fréquences ne soient pas sensiblement modifiées, les diagrammes d'antenne associés aux basses fréquences voient leur lobe principal se déformer nettement. La direction d'écoute θ_c n'est plus respectée, et le niveau des lobes secondaire augmente. Cependant, comparativement à la figure 3.9 obtenue en champ lointain, ces déformations sont significativement moins importantes, de sorte que le filtrage spatial opéré par cette formation de de voie large bande reste exploitable. Cette remarque est d'autant plus vraie si nous cherchons à utiliser en champ proche une formation de voie large bande synthétisée en champ lointain, comme en témoigne la figure 3.17. Examinons maintenant le cas où la formation de voie est implémentée dans le domaine temporel.

3.4.4 Synthèse modale large bande dans le domaine temporel

Plutôt que de traiter indépendamment chacune des composantes fréquentielles des signaux reçus sur les N microphones de l'antenne, nous proposons ici de déterminer les fonctions de transfert $W_n(k)$ des N filtres numériques réalisant la formation de voie large bande. Comme

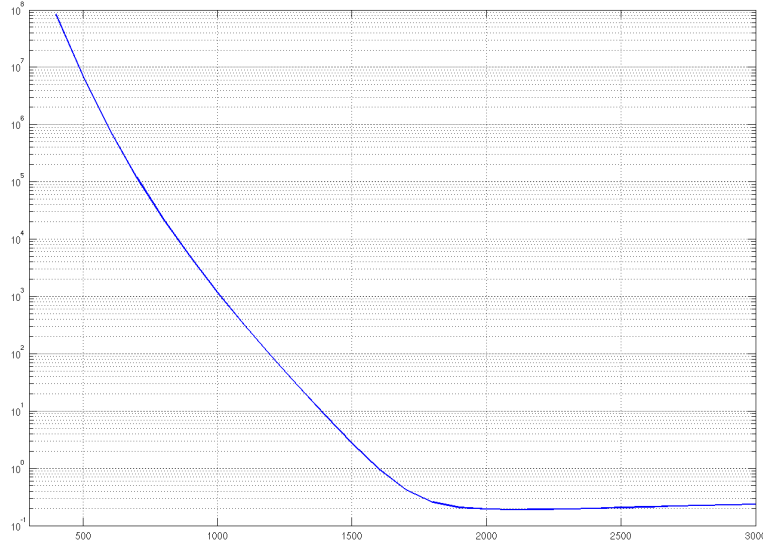


FIG. 3.23 – Evolution de $G_B = \mathbf{W}^H \mathbf{W}$ en fonction de la fréquence utilisée pour la synthèse modale faible bande.

précédemment, nous choisissons d'utiliser des filtres numériques à réponse impulsionnelle finie, d'ordre Q , de fonction de transfert (3.12). L'objectif du problème de synthèse que nous proposons ici est donc de déterminer les NQ coefficients réels w_{nq} .

3.4.4.1 Présentation du problème de synthèse

Considérons dans un premier temps que la réponse de référence $D_d(r, \theta, k)$ soit définie par la figure 3.4. Ce diagramme d'antenne, étant donné sa série de Fourier de coefficient $a_l(r, k)$ donnés par (3.76), est décomposable dans l'espace modal à partir de l'équation (3.75). De la même façon, nous pouvons calculer les coefficients modaux $\alpha_m(k)$ grâce à l'équation (3.4.2.2). Compte tenu de l'expression (3.12) des fonctions de transfert des filtres utilisés, les coefficients $\alpha_m(k)$ ont donc ici pour expression

$$\alpha_m(k) = -2jk\sqrt{\pi(2m+1)} \sum_{n=1}^N \sum_{q=0}^Q w_{nq} e^{-jkcqT_e} j_m(kz_n), \quad (3.77)$$

ou bien, sous forme matricielle,

$$\alpha_m(k) = -2jk\sqrt{\pi(2m+1)} \mathbf{W}_{\text{LB}}^T \mathbf{V}_{\text{Modal}}(k, m), \quad (3.78)$$

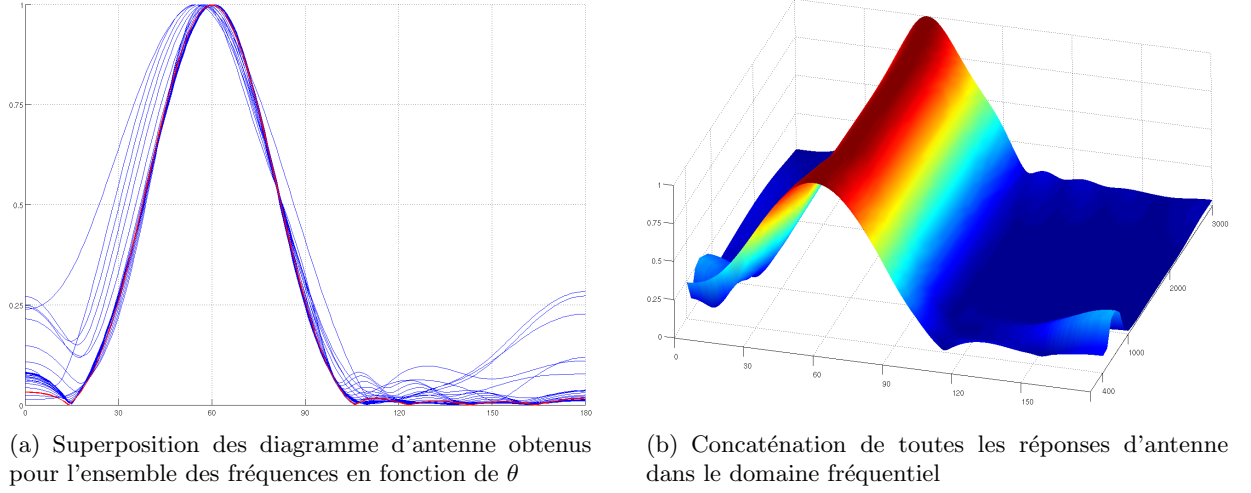
avec

$$\mathbf{V}_{\text{Modal}}(k, m) = \mathbf{J}_m(k) \otimes \mathbf{V}_{\text{FIR}}, \quad (3.79)$$

et, bien sûr,

$$\mathbf{V}_{\text{FIR}} = (1, e^{-jkcT_e}, \dots, e^{-jkcQT_e})^T. \quad (3.80)$$

La méthode de synthèse proposée consiste à déterminer le vecteur $\mathbf{W}_{\text{LB}}$ permettant d'approximer la réponse d'antenne désirée $D_d(r_0, \theta, k)$ définie pour une distance r_0 , et représentée


 FIG. 3.24 – Formations de voie faibles bandes pour $f \in [400 : 100 : 3000]$, avec $G_B = 100$.

par ses $M + 1$ coefficients modaux $\alpha_0^d(r_0, k), \dots, \alpha_M^d(r_0, k)$, via la résolution du problème d'optimisation convexe

$$\begin{array}{ll} \text{minimiser} & \varepsilon \\ \text{sous} & |\alpha_m(k) - \alpha_m^d(r_0, k)|^2 \leq \varepsilon, \forall m \in [0, \dots, M], \forall k \in K, \end{array} \quad (3.81)$$

où K désigne le domaine de fréquences sur lequel est effectuée la minimisation. Rigoureusement, (3.81) consiste en la minimisation de l'écart pire cas entre deux fonctions de k sur le continuum K . En pratique, nous échantillonnons dans le domaine fréquentiel (3.81), de sorte que K désigne un ensemble discret de fréquences. Comparons maintenant le problème (3.81) à l'extension large bande (3.31) étudiée dans le cas du champ lointain. Celle-ci a nécessité une discrétisation à la fois du domaine angulaire et des fréquences, et donc à un nombre de contraintes conséquent, à l'origine d'un temps de résolution très long (45 min environ) et de problèmes numériques liés au solveur. Avec cette nouvelle méthode de synthèse dans l'espace modal, les 91 contraintes angulaires sont remplacées par seulement $M + 1$ contraintes modales, avec $M + 1 \ll 91$. Par exemple, dans la suite, nous aurons $M = 7$, soit environ 12 fois moins de contraintes. Dans ces conditions, le temps de résolution devient raisonnable (pas plus de 5 minutes), et les problèmes numériques disparaissent. Néanmoins, la puissance des éléments du vecteur $\mathbf{W}_{LB}$ devra être limitée en adjoignant à (3.81) la contrainte

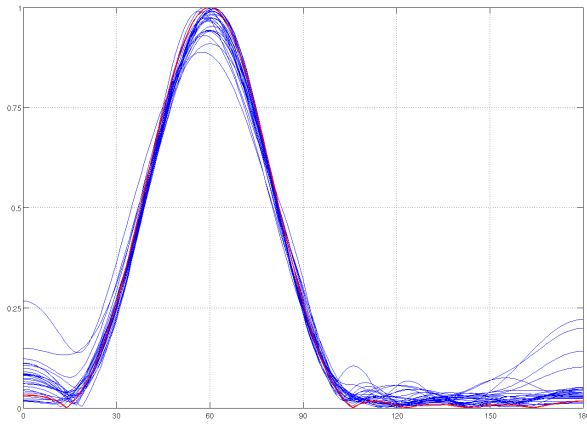
$$\mathbf{W}_{LB}^H \mathbf{W}_{LB} \leq \delta. \quad (3.82)$$

3.4.4.2 Résultats d'optimisation

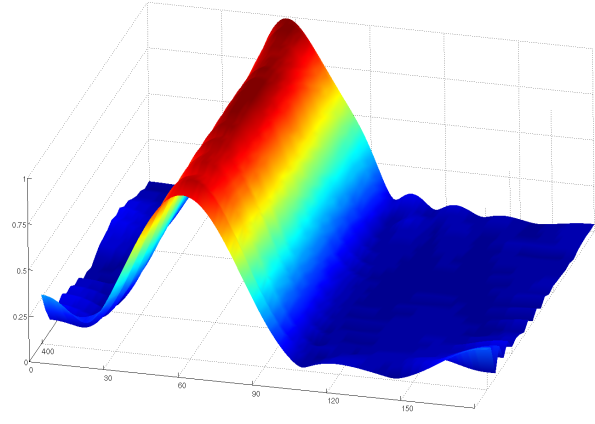
Dans toute la suite, les angles paramétrant $D_d(r_0, \theta, k)$ sont fixés à $\theta_c = 60^\circ$, $\theta_p = 5^\circ$, $\theta_s = 40^\circ$. Sa décomposition en série de Fourier demeure donnée par (3.76), dont nous ne conservons que les $L + 1 = 8$ premiers termes $a_l(r_0)$ pour chacune des fréquence appartenant à l'ensemble K . Celui-ci est composé de 26 fréquences comprises entre 400Hz et 3kHz par pas de 100Hz. Le problème d'optimisation (3.81) est résolu en minimisant l'écart entre $M + 1 = 8$ coefficients modaux, pour une distance de synthèse $r_0 = 1\text{m}$. Enfin, l'ordre des filtres $W_n(k)$ est fixé à

$Q = 35$, et la fréquence d'échantillonnage à $f_e = 8$ kHz. Examinons maintenant les réponses d'antenne obtenue avec ou sans la contrainte (3.82).

La figure 3.25 présente le tracé de la réponse d'antenne obtenue après optimisation dans l'espace modal, pour des fréquences comprises entre 400Hz et 3kHz par pas de 50Hz. Comme précédemment, nous devons nous assurer que la réponse d'antenne, contrainte sur un nombre fini de fréquences constituant la grille K , est acceptable entre chacune de ces fréquences de synthèse. Il s'avère que le diagramme d'antenne résultant présente des propriétés de filtrage spatial satisfaisantes, avec en particulier un niveau de lobes secondaires faible. Pour les plus basses fréquences, nous pouvons cependant constater que le niveau de ces lobes secondaires est légèrement plus élevé dans les deux directions périphériques 0° ou 180° . Le lobe principal



(a) Coupe selon multiples fréquences de la réponse d'antenne optimisée obtenue à une distance $r_0 = 1$ m.



(b) Vue 3D de la réponse d'antenne optimisée

FIG. 3.25 – Formations de voie large bande pour $f \in [400 : 50 : 3000]$.

possède une forme quasiment invariante en fréquence, significativement plus homogène que celle obtenue en champ lointain précédemment. En particulier, les oscillations du niveau maximal du lobe principal sont beaucoup plus faibles. Enfin, n'oublions pas que nous avons effectué ici une synthèse en champ proche ; rigoureusement, il nous faudrait donc comparer ce nouveau résultat avec le diagramme d'antenne $D^{\infty \rightarrow r_0}(\mathbf{r}, k)$ reporté figure 3.17. Nous sommes donc en mesure d'améliorer très nettement le comportement de l'antenne pour des distances proches du capteur, tout en continuant à exploiter le contenu basse fréquence des signaux à localiser. Evidemment, comme nous n'avons pas contraint la puissance des éléments du vecteur $\mathbf{W}_{\text{LB}}$, celle-ci atteint une valeur très importante de l'ordre de $2 \cdot 10^7$. Il est donc nécessaire d'ajouter la contrainte (3.82) au problème d'optimisation.

Pour les mêmes paramètres que dans l'exemple précédent, nous avons limité $\mathbf{W}_{\text{LB}}^H \mathbf{W}_{\text{LB}}$ à une valeur maximale $\delta = 10^3$. La figure 3.26 présente le diagramme d'antenne obtenu après résolution de (3.81). Nous pouvons constater que la limitation sur $\mathbf{W}_{\text{LB}}^H \mathbf{W}_{\text{LB}}$ a pour principal effet de dégrader l'invariance en fréquence. De plus, la position du maximum du lobe principal varie légèrement, en particulier pour les basses fréquences. Cependant, le niveau des lobes secondaires demeure suffisamment faible pour pouvoir affirmer que le filtrage spatial reste de meilleure qualité que celui opéré par une formation de voie conventionnelle. Nous avons tracé sur la figure 3.27 les diagrammes de Bode des 8 filtres numériques réalisant la formation de voie. Ceux-ci présentent une amplification allant d'environ -25 dB à 30 dB dans la bande des fréquence utilisées pour la

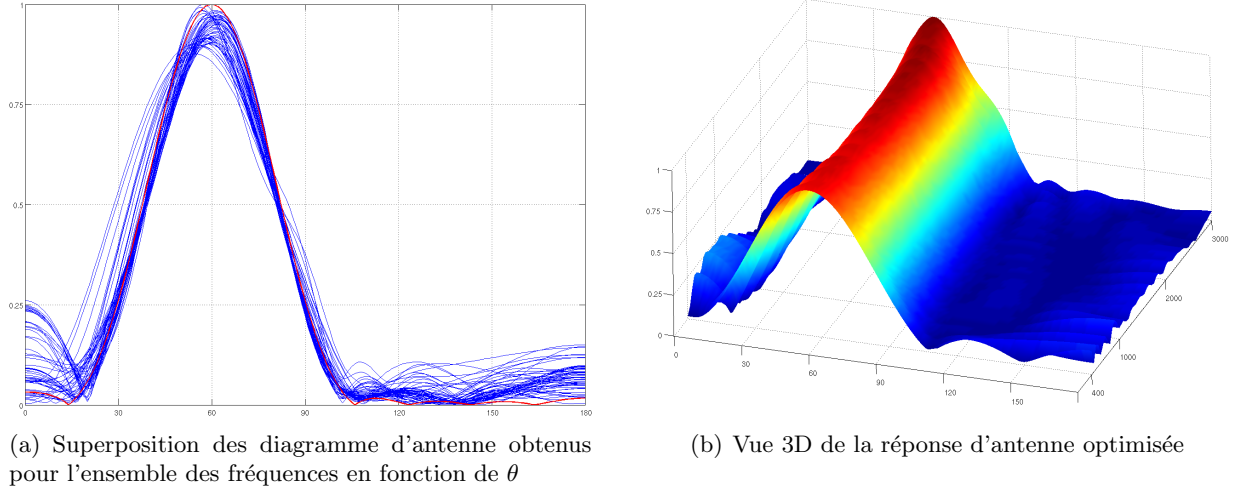


FIG. 3.26 – Formations de voie large bande pour $f \in [400 : 50 : 3000]$, avec $G_B = 1000$.

synthèse et représentée en blanc. Naturellement, les plus forts gains sont localisés aux plus basses fréquences, laissant présager que ces mêmes fréquences seront les plus affectées par la présence de bruit blanc spatial et temporel dans la scène sonore.

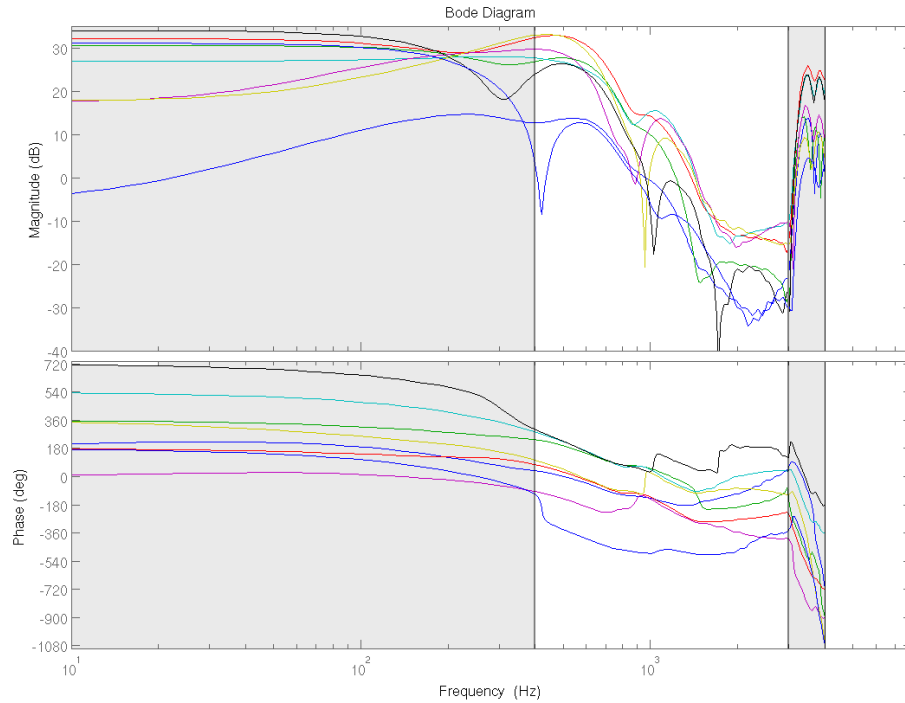


FIG. 3.27 – Diagramme de Bode des 8 filtres réalisant la formation de voie dont le diagramme d'antenne est visible figure 3.26.

3.5 Conclusion

Nous avons cherché dans ce chapitre à synthétiser des diagrammes d'antenne invariant en fréquences, avec pour objectif d'affiner la largeur de leur lobe principal pour les plus basses fréquences. Pour cela, nous avons proposé deux méthodes d'optimisation convexe. La première est issue d'une extension immédiate de travaux de la communauté du Traitement du Signal, et conduit à minimiser l'écart pire cas sur une grille entre deux fonction de θ et k . Le nombre important de contraintes est à l'origine de durées de résolution importantes, et conduit à certains problèmes numériques. Nous avons alors proposé une nouvelle méthode de synthèse décrite dans l'espace modal des formations de voie. Ce formalisme conduit à une baisse sensible de nombre de contrainte, et permet de poser directement le problème de synthèse à une distance r fixée. Enfin, nous nous sommes intéressé à l'effet du champ proche sur les diagrammes de directivité synthétisés, et montré qu'il était nécessaire d'adapter la formation de voie à la distance effective d'émission de la source. Nous proposons maintenant d'exploiter l'ensemble de ces résultats pour la localisation dans le chapitre suivant.

Chapitre 4

Exploitation de formations de voie large bande pour la localisation

Nous avons proposé dans le chapitre précédent une nouvelle méthode de synthèse de formations de voie. Basée sur la représentation modale des diagrammes de directivité, elle permet de déterminer les coefficients de filtres numériques afin d'obtenir par exemple une réponse d'antenne invariante en fréquence, au prix d'une amplification plus importante du bruit présent dans la scène sonore. Nous proposons dans ce chapitre d'exploiter ce résultat pour la localisation de source, à travers deux algorithmes différents ayant pour but l'estimation de la distance r entre le capteur et la source, et/ou l'azimut θ . Cependant, avant même d'évoquer d'algorithme de localisation, nous devons disposer des signaux sonores reçus par chacun des microphones de l'antenne. Plus précisément, un système d'acquisition est indispensable pour tester et valider chacun des algorithmes en environnement réel. L'un des principaux objectifs de cette thèse est donc aussi la conception en elle même du capteur, du choix des microphones jusqu'au dispositif de traitement des données.

Ce chapitre est organisé de la façon suivante. Nous détaillons dans une première partie la conception électronique du capteur sonore, en nous attachant tout particulièrement à décrire les fonctions et modèles de chacun de ses éléments. Puis, nous proposons sur cette base une première méthode simple de localisation, basée sur les propriétés de filtrage spatial atteintes dans le chapitre précédent, pour la détermination de l'azimut θ de la source. Enfin, nous décrivons dans une troisième partie un algorithme haute résolution basé sur une extension large bande de MUSIC pour laquelle notre méthode de synthèse modale s'avère particulièrement bien adaptée. Les performances de chacune de ces méthodes de localisation sont illustrées par de nombreuses simulations effectuées sous Matlab.

4.1 Conception de la chaîne d'acquisition et traitement

L'acquisition de signaux audio ne pose traditionnellement pas de problème. Leur nature basse fréquence n'impose en effet aucune condition de bande passante sur les composants, et les fondeurs proposent aujourd'hui une immense gamme de produits spécialisés pour le traitement audio. Nous pouvons distinguer trois éléments fondamentaux nécessaires :

- le dispositif de conversion des ondes sonores en impulsions électriques ou, beaucoup plus simplement, les microphones,
- le système de mise en forme des signaux, suivi d'un étage de conversion analogique numérique pour un traitement discret des données,

- le dispositif de traitement à proprement parler, embarquant les algorithmes permettant la localisation.

L’acquisition de microphones de bonne qualité est probablement le premier critère important pour espérer obtenir des signaux acoustiques représentatifs de la scène sonore. Nous proposons de décrire dans une première partie le type de microphone retenus. Puis vient le dispositif d’acquisition, dont le rôle est de mettre en forme les signaux issus de chacun des microphones à travers différents étages de gain et de filtrage. De nombreuses cartes de ce type existent dans le commerce, équipées parfois de fonctionnalités avancées (jeux de Codecs pour la compression audio, programmation dynamique des paramètres, etc.) et intégrant un dispositif de traitement à base de composants spécialisés. Nous avons dans un premier temps retenu une solution industrielle, proposée par la société Bittware, offrant l’avantage de se présenter comme une interface matérielle contrôlable depuis Matlab. Nous avons finalement opté pour la conception complète d’une telle carte qui sera présentée dans une seconde partie. Une fois les signaux issus des microphones filtrés et numérisés, il faut les traiter pour être capable d’estimer la position du signal acoustique. Une première solution est évidemment de transférer l’ensemble des données numériques directement sur un ordinateur disposant d’une grande puissance de calcul. Notre objectif est de concevoir un capteur sonore autonome, capable de communiquer avec la CPU du robot pour délivrer les résultats d’estimation de position. Cela implique donc l’utilisation de composants intégrables spécialisés, parmi lesquels nous pouvons citer le DSP ou le FPGA. Nous avons étudié chacune de ces solutions pour finalement retenir un FPGA de dernière génération de la société Xilinx. Nous proposons d’en étudier les caractéristiques dans une troisième partie. Enfin, nous décrivons dans une dernière partie le logiciel embarqué sur FPGA prenant en charge la programmation des paramètres de la carte d’acquisition.

4.1.1 Choix des microphones

Un microphone est un transducteur acousto-électrique. A ce titre, il a pour rôle de transformer l’information acoustique, présente sous la forme de variations de pression, en une information électrique (courant ou tension) exploitable par n’importe quel système moderne d’acquisition. Historiquement, nous devons un des premiers microphones utilisables à Alexander Graham Bell, inventeur du téléphone en 1876. Son principe de fonctionnement est aujourd’hui toujours le même : une partie mobile, le plus souvent une membrane, est mise en mouvement par les ondes de pression, et ces oscillations mécaniques sont ensuite converties en un signal électrique acheminé vers un amplificateur auquel le microphone est raccordé. Cependant, tous les microphones ne captent pas le son de la même façon. La technologie utilisée pour capter les variations de pression, ainsi que l’environnement dans lequel est placée la membrane (capsule), viennent modifier leurs caractéristiques, en terme de directivité ou de bande passante. Aujourd’hui, les grands constructeurs de microphones (GRAS, SCHOEPS, SCHUR, etc ...) proposent une gamme très étendue de produits, spécialisés pour l’acquisition de sons d’instruments, de la voix, ou autre. Il existe donc plusieurs modes de classification des microphones, selon leur principe de transduction, leur sensibilité au mode d’incidence de la pression, ou de leur directivité [Asch 2006]. Pour notre application, nous proposons de travailler avec des microphones omnidirectionnels, puisque nous ne pouvons pas privilégier sans connaissance à priori une direction spécifique de l’espace. Le choix de la méthode de transduction est par contre plus spécifique. Il existe plusieurs technologies différentes de conversion, basées majoritairement sur des effets mécaniques ou électrostatiques [Jouhaneau 1999]. Nous avons fait ici le choix de travailler avec des microphones électrostatiques pour leur qualité de restitution ainsi que pour l’existence de modèles de petite

taille embarquables sur un robot. Cependant, il nous faudra faire particulièrement attention à la sensibilité des microphones retenus, celle-ci étant d'autant plus faible que le microphone est de petite taille. La sensibilité d'un microphone s'exprime en mV/Pa et permet de déterminer le niveau de tension disponible pour un niveau sonore L donné, exprimé en dB (ou dB_{SPL}) selon la relation

$$L = 20 \log \frac{p}{p_0}, \quad (4.1)$$

où p désigne la pression mesurée en Pascal, et $p_0 = 2.10^5$ Pa est la pression de référence. Pour information, voici un tableau récapitulant les niveaux sonores exprimés en dB pour différents types de sources sonores

Niveau sonore	dB
Seuil d'audition	0 dB
Chuchotement	30 dB
Conversation	60 dB
Circulation urbaine	90 dB
Concert	120 dB
Moteur à réaction (à 10m)	150 dB

Si la notion de sensibilité est directement reliée au niveau de tension disponible en sortie des microphones, elle permet aussi de définir l'intensité acoustique minimale pouvant être mesurée. En effet, l'agitation thermique des molécules d'air est suffisante pour générer un signal de sortie extrêmement faible, même dans des conditions de silence parfait. Assimilable à un bruit thermique, la tension causée par ce phénomène est traditionnellement de l'ordre de $5\mu\text{V}$ [GRAS 2004], et aucun signal acoustique à l'origine d'une tension plus faible ne peut être détecté. Ainsi, un microphone possédant une sensibilité de 50 mV/Pa sera capable de mesurer des sons d'une intensité minimale de 14dB, là où un autre microphone d'une sensibilité de seulement 4mV/Pa ne pourra mesurer de sons inférieurs à 36dB. En pratique, chaque microphone est nécessairement connecté à un pré-amplificateur chargé aussi très souvent de l'alimentation du capteur, introduisant lui-même un certain niveau de bruit qu'il faudra espérer minimum pour la qualité de la chaîne d'amplification.

Au final, nous avons choisi d'équiper notre antenne de 8 microphones quart de pouce, de type 40PQ, de la marque GRAS. Ces microphones sont particulièrement bien adaptés à une utilisation en antennerie puisque leur caractéristiques sont aisément reproductibles sur un grand nombre de capteurs. Un autre avantage important pour notre application est l'intégration de l'étage de pré-amplification à l'intérieur du corps du microphone, si bien que celui-ci ne nécessite plus qu'une alimentation externe en courant. Enfin, nous avons fait le choix d'un appariement en phase de façon à garantir une réponse des microphones identique à 1° prêt. Le microphone retenu est visible sur la figure 4.1, où sont aussi reproduits un certain nombre de données techniques issues des caractéristiques fournies par le constructeur. Nous disposons donc maintenant de microphones capables de délivrer une tension proportionnelle à la variation de pression acoustique et d'une sensibilité de 8mv/Pa. Or nous souhaitons localiser des locuteurs s'adressant à un robot. D'après le tableau précédent, l'intensité sonore dans une telle situation est d'environ 60dB pour une conversation d'un niveau sonore normal. Le niveau de pression, selon l'équation (4.1), est donc d'environ 20mPa, à l'origine d'une tension de sortie de l'ordre de 0.16mV. Cette tension est donc extrêmement faible, et il s'avère nécessaire d'amplifier et de traiter chacun des



Values quoted for 23 °C and 4 mA supply

Nominal Sensitivity:

at 250Hz 8 mV/Pa

Frequency Response:

±1 dB 100 Hz - 3 kHz

±2 dB 100 Hz - 10 kHz

Lower Limit of Dynamic Range:

Thermal noise <30 dBA re. 20 µPa

Phase Match:

100 Hz - 3 kHz ±3°

3 kHz - 5 kHz ±5°

FIG. 4.1 – Photo du type de microphone retenu et principales caractéristiques techniques.

signaux de sortie des microphones constituant l'antenne. C'est le rôle de la carte d'acquisition que nous allons détailler maintenant.

4.1.2 Dispositif d'acquisition

La conception de la carte d'acquisition que nous allons présenter dans cette partie constitue une des contributions importantes de ce travail. Notre objectif est de proposer une solution matérielle capable de traiter des signaux audio suffisamment génériques pour pouvoir être adaptée éventuellement à d'autres type de microphones ou de signaux acoustiques. Ainsi, nous avons choisi de rendre programmable logiciellement le plus de fonctionnalités possible de cette carte, à partir d'un pilote open-source pouvant s'insérer aisément dans l'architecture logicielle de contrôle des plateformes mobiles. Si un tel cahier des charges s'avère ambitieux, il permet de garantir une certaine pérennité de la réalisation, tout en garantissant une exploitation facilitée par les futurs utilisateurs de cette nouvelle modalité auditive. Nous proposons de présenter dans une première partie quelques généralités sur les dispositifs d'acquisition, avant de détailler chacun des éléments constitutifs de notre carte électronique. Notre objectif est d'aboutir à un modèle simple, sous forme de fonction de transfert, permettant ensuite une simulation fiable du dispositif expérimental utilisé pour la localisation.

4.1.2.1 Description d'une chaîne d'acquisition classique

Le but d'une mesure est de faire correspondre à une grandeur physique que l'on désire connaître, une grandeur compréhensible par le système d'acquisition. De nos jours, il s'agit presque uniquement de signaux électriques qui ont remplacé les systèmes mécaniques et optiques d'autrefois. Ainsi le capteur délivre un signal qui est amplifié et converti généralement en tension par un conditionneur. Le système d'acquisition n'a plus alors à traiter qu'un courant électrique ou une tension dont la valeur est directement reliée à l'amplitude du phénomène vu par le capteur et cela, idéalement, quel que soit son type. Le conditionneur joue dans cette succession de traitements un rôle primordial puisqu'il est le premier élément à amplifier et/ou filtrer la donnée à acquérir. Dans notre cas, le choix de ce conditionneur est très restreint, puisqu'il a pour rôle supplémentaire d'alimenter chacun des microphones de l'antenne avec le courant continu nécessaire au bon fonctionnement du pré-amplificateur intégré à chacun des récepteurs. Sur les conseils d'industriels spécialisés en acoustique, nous avons choisi de travailler avec le conditionneur 8 voies visible sur la figure 4.2 pouvant être rendu autonome par un simple branchement sur batterie. Au final, la carte d'acquisition aura donc pour rôle de traiter le signal

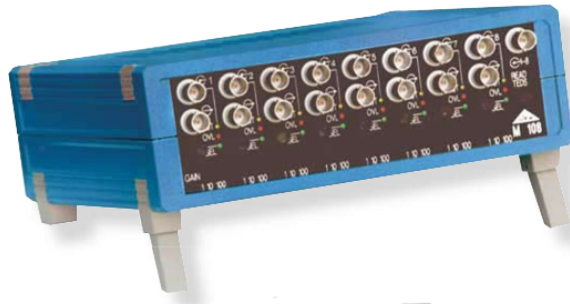


FIG. 4.2 – Photo du conditionneur. Assez imposant, il devra nécessairement être embarqué sur le robot.

sortant de ce conditionneur, sachant que celui-ci est déjà capable de fournir une amplification de 1, 10, ou 100.

Nous avons déjà précisé que nous souhaitons effectuer un traitement numérique des données acoustiques. Cela nécessite donc une conversion analogique-numérique des tensions disponibles en sortie du conditionneur, et impose donc la présence de certains éléments sur la carte d'acquisition. Les traitements réalisés par celle-ci peuvent être classés en trois catégories :

- Amplification : Nous l'avons vu, les tensions sur lesquelles nous allons travailler possèdent une très faible amplitude. Il est donc impératif d'obtenir des niveaux de tension suffisants pour effectuer une conversion numérique de qualité sur la totalité de l'échelle de tension disponible.
- Filtrage : Nous avons précisé dans le chapitre précédent que nous souhaitons manipuler une gamme de fréquences comprise entre 400Hz et 3kHz. Nous allons donc filtrer les signaux issus des microphones de façon à échantillonner uniquement la gamme de fréquence d'intérêt pour notre application. De plus, la présence d'une conversion analogique - numérique nécessite la présence d'un filtre passe-bas indispensable pour éviter tout phénomène de repliement.
- Conversion analogique-numérique : C'est là l'étape essentielle effectuée sur la carte d'acquisition. Le résultat de cette conversion est ensuite transmis au composant effectuant les calculs relatifs à la localisation de la source sonore.

Nous proposons maintenant de détailler rapidement chacune de ces trois catégories, à travers la description des composants retenus pour la réalisation de la carte d'acquisition.

4.1.2.2 Choix des composants

Filtre passe-haut L'ensemble microphone + conditionneur possède une bande passante telle que le signal disponible est composé de très basses fréquences (de l'ordre du Hz parfois) venant modifier sensiblement la composante continue du signal à échantillonner. En pratique, nous constatons que la partie informative du signal possède une amplitude extrêmement faible qui se situe autour de cette valeur moyenne mouvante. Il est donc nécessaire de supprimer les très basses fréquences du signal à traiter avant de l'amplifier de façon à éviter tout risque de saturation. Nous proposons donc l'emploi d'un filtre analogique passe-haut, composé d'une double cellule à capacité et résistance, de fonction de transfert

$$H_{PH}(p) = \frac{(RCp)^2}{1 + 3RCp + (RC)^2p^2}, \quad (4.2)$$

où p désigne la variable de Laplace, $R = 10 \text{ k}\Omega$ et $C = 0.68\mu\text{F}$. Le filtre proposé est donc un filtre d'ordre 2 possédant une fréquence de coupure à -3 dB de l'ordre de 60 Hz . Remarquons que ce filtre ne prend pas en charge l'éventuel filtrage des fréquences inférieures à 400 Hz de façon à éventuellement en disposer en cas de besoin pour la localisation. Cet étage de filtrage est ensuite suivi d'un étage de gain fixe réalisé à partir d'un amplificateur d'instrumentation INA128 de Texas Instruments. Ce type de composant est particulièrement bien adapté à l'amplification de signaux de très faible amplitude, et présente une impédance d'entrée très importante indispensable au bon conditionnement des signaux. Nous avons choisi de fixer un gain égal à $G_{PH} = 10$, réglable très précisément par l'intermédiaire de potentiomètres à forte démultiplication. Compte tenu de la faible valeur du gain G_{PH} , nous allons négliger la bande passante de cet amplificateur (produit gain-bande de 10 MHz) et supposer qu'il est assimilable à un gain pur.

Gain Réglable Une fois filtré et amplifié, le signal disponible pour un niveau sonore de 60 dB atteint maintenant environ 160mV . Une telle tension n'est évidemment pas encore suffisante pour un échantillonnage efficace, et il est donc nécessaire de l'amplifier une seconde fois. Fixons comme objectif une tension maximale de l'ordre de 2.5V permettant d'atteindre, comme nous le verrons, la pleine échelle de la conversion analogique-numérique. Le gain à apporter est alors de l'ordre 24 dB . Cependant, cette valeur de gain approximative vient d'être estimée pour un cas particulier, et changera nécessairement selon l'application audio envisagée. Pour éviter de câbler définitivement une telle valeur, nous proposons l'emploi d'un gain programmable (PGA, pour Programmable Gain Amplifier) permettant de choisir à tout moment la valeur d'amplification à appliquer sur chacune des voix. Nous avons choisi pour cela le PGA2311 de Burr-Brown, fournissant un gain G_{PGA} pouvant aller de -95.5 dB à 30 dB par pas de 0.5 dB , avec une précision de 0.05 dB . Le choix de la valeur d'amplification s'effectue à travers une interface numérique série à trois fils, que nous exploitons plus tard pour pouvoir programmer logiciellement le composant. Cet amplificateur dispose de deux entrées, aussi un seul composant est capable de traiter deux voies d'acquisition simultanément.

Filtre anti-repliement Nous disposons maintenant d'une tension possédant un niveau suffisant pour pouvoir être échantillonnée. Cependant, une telle opération nécessite de limiter la bande passante du signal à acquérir de façon à respecter le théorème de Shannon et éviter tout phénomène de repliement. Pour cela, nous proposons l'utilisation d'un filtre passe bas à capacités commutées MAX293 de Maxim, traditionnellement utilisé pour le filtrage basse fréquence. Basé sur le principe de la commutation d'une capacité par plusieurs interrupteurs (transistors), ce type de filtre permet une modification à tout moment de sa fréquence de coupure en agissant uniquement sur la fréquence d'une horloge présente en entrée du composant. La relation entre la fréquence de coupure f_c et la fréquence d'horloge f_H est fixée par le constructeur, et vérifie ici $f_H = 100f_c$. Ainsi, si nous sommes capable de générer une horloge de fréquence adaptable par voie logicielle, nous pouvons modifier à notre guise la fréquence de coupure passe-bas de chacune des voies d'acquisition. Finalement, le filtre retenu est de type elliptique, d'ordre 8, et possède un niveau d'atténuation de 80 dB pour une bande de transition de 1.5 . Sa fonction de transfert, que nous noterons $H_{CC}^{f_H}(p)$, est représentée sous la forme d'un diagramme de Bode dans la documentation constructeur et sera approximée par le résultat d'une synthèse analogique sous Matlab d'un filtre possédant les mêmes caractéristiques. Si le principe de la commutation

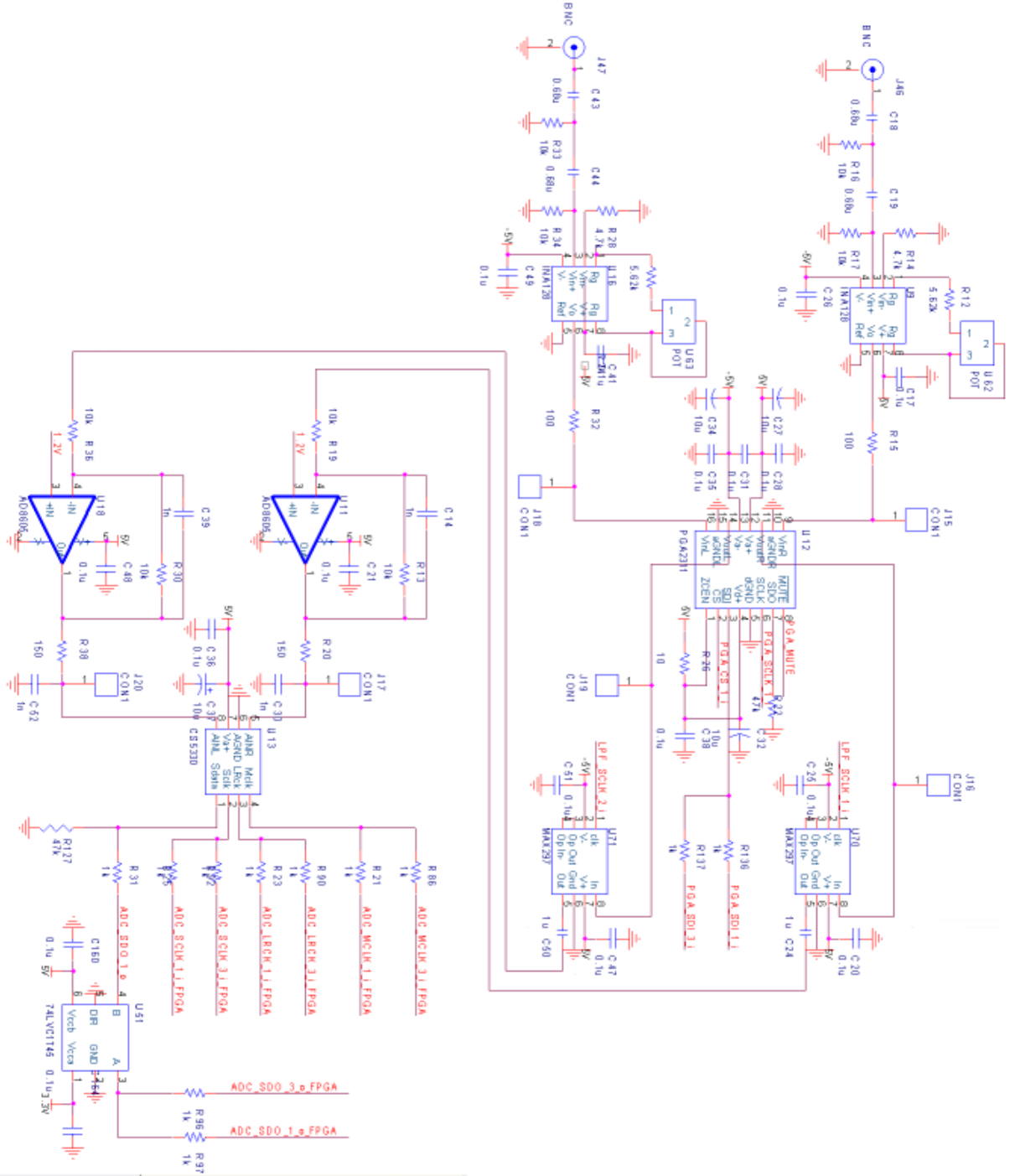


FIG. 4.3 – Détails de la chaîne d'acquisition. 2 voies sont représentées, la carte complète comprenant 4 modules identiques.

permet de choisir simplement, selon l'application, la fréquence de coupure, il est souvent à l'origine d'un bruit d'horloge venant parasiter la tension de sortie du filtre. Il est donc nécessaire de couper ces parasites de fréquence f_H de façon à ne pas polluer le signal échantillonné. Pour cela,

nous avons placé deux filtres analogiques en cascade dont le rôle est d'assurer une atténuation minimale de l'ordre de 20 dB (recommandation constructeur) pour cette fréquence. Le premier filtre est actif et comprend en particulier un amplificateur opérationnel faible bruit, le second est passif et est réalisé à partir d'une simple cellule résistance/capacité. Au final, l'étage de filtrage anti repliement possède une fonction de transfert $H_{AR}(p)$ donnée par

$$H_{AR}(p) = \frac{R_A C_1 p}{(1 + R_A C_1 p)(1 + R_A C_2 p)} \frac{1}{1 + R_B C_3 p} H_{CC}^{f_H}(p), \quad (4.3)$$

avec $R_A = 10 \text{ k}\Omega$, $R_B = 150 \Omega$, $C_1 = 1 \mu\text{F}$, et $C_2 = C_3 = 1 \text{ nF}$. Le premier filtre actif est un filtre passe-bande de fréquences de coupure $f_{c1} = 16 \text{ Hz}$ et $f_{c2} = 16 \text{ kHz}$. Le second filtre présente une coupure à $f_{c3} = 1 \text{ MHz}$, conseillée par la documentation constructeur du convertisseur.

Conversion analogique-numérique L'étage de conversion est le dernier élément fondamental de la chaîne d'acquisition. Il a pour rôle la transformation d'une grandeur analogique continue en une donnée numérique image de l'amplitude du signal d'entrée à un instant donné, et à une cadence f_e appelée fréquence d'échantillonnage. Il existe un grand nombre de méthodes de conversion, privilégiant la rapidité, la précision, ou le coût du composant ; citons par exemple les convertisseurs flash, à approximations successives, à rampes ou à sur-échantillonnage [Asch 2006]. Nous nous sommes plus particulièrement intéressés à cette dernière catégorie, idéale pour la conversion de signaux dans la bande audio. Sans trop détailler leur mode de fonctionnement, de tels convertisseurs sont basés sur le principe de la modulation sigma-delta visant à coder les variations d'un signal sur un unique bit, dont la valeur change à une fréquence très élevée vis à vis du signal traité. Leur principal avantage réside dans leur bruit de quantification extrêmement réduit permettant d'envisager des fortes résolutions de conversion [Aziz 1996]. En pratique, le signal est échantillonné sur un bit à une fréquence très élevée vis à vis de la fréquence de Shannon. Puis différents filtres numériques décimateurs sont appliqués jusqu'à retrouver la fréquence d'échantillonnage souhaitée initialement, sur un nombre de bits plus importants (jusqu'à 24 bits pour les convertisseurs à haute résolution) [Park 1992]. Une conséquence immédiate est que ce type de convertisseur analogique numérique devient assimilable à un filtre analogique de fonction de transfert notée $H_{CAN}(p)$. Celui-ci peut être perçu comme un filtre anti-repliement intégré et permet éventuellement d'alléger les contraintes fortes sur la largeur de la bande de transition du véritable filtre limitant la bande passante placé en amont du convertisseur. Nous proposons d'utiliser ici le convertisseur analogique-numérique CS5330 de Cirrus Logic, possédant une résolution de 18 bits, et introduisant un filtrage de 80 dB pour toute fréquence supérieure à $f_e/2$. Bien que l'expression exacte de la fonction de transfert $H_{CAN}(p)$ ne soit pas fournie dans la documentation constructeur, nous l'approximerons par une synthèse analogique effectuée sous Matlab à partir des paramètres fournis. Le composant retenu est entièrement paramétrable par l'intermédiaire d'une liaison série permettant entre autre de choisir la fréquence d'échantillonnage entre 2 kHz et 50 kHz. Ainsi, il sera possible de programmer logiquement la valeur souhaitée de la cadence d'échantillonnage selon le type de signal audio utilisé.

4.1.2.3 Modélisation de la chaîne d'acquisition

Nous disposons maintenant des caractéristiques de chacun des composants retenus. Nous sommes donc capable de tracer les diagrammes de Bode des étages présentés précédemment, et surtout d'obtenir le comportement fréquentiel de la chaîne d'acquisition complète. Nous avons représenté sur la figure 4.4 les gains de chacun de ces éléments en fonction de la fréquence, pour une fréquence de coupure du filtre à capacités commutées $f_c = 3 \text{ kHz}$ et une fréquence

d'échantillonnage $f_e = 8 \text{ kHz}$. Notons que d'après les documentations constructeur, le conver-

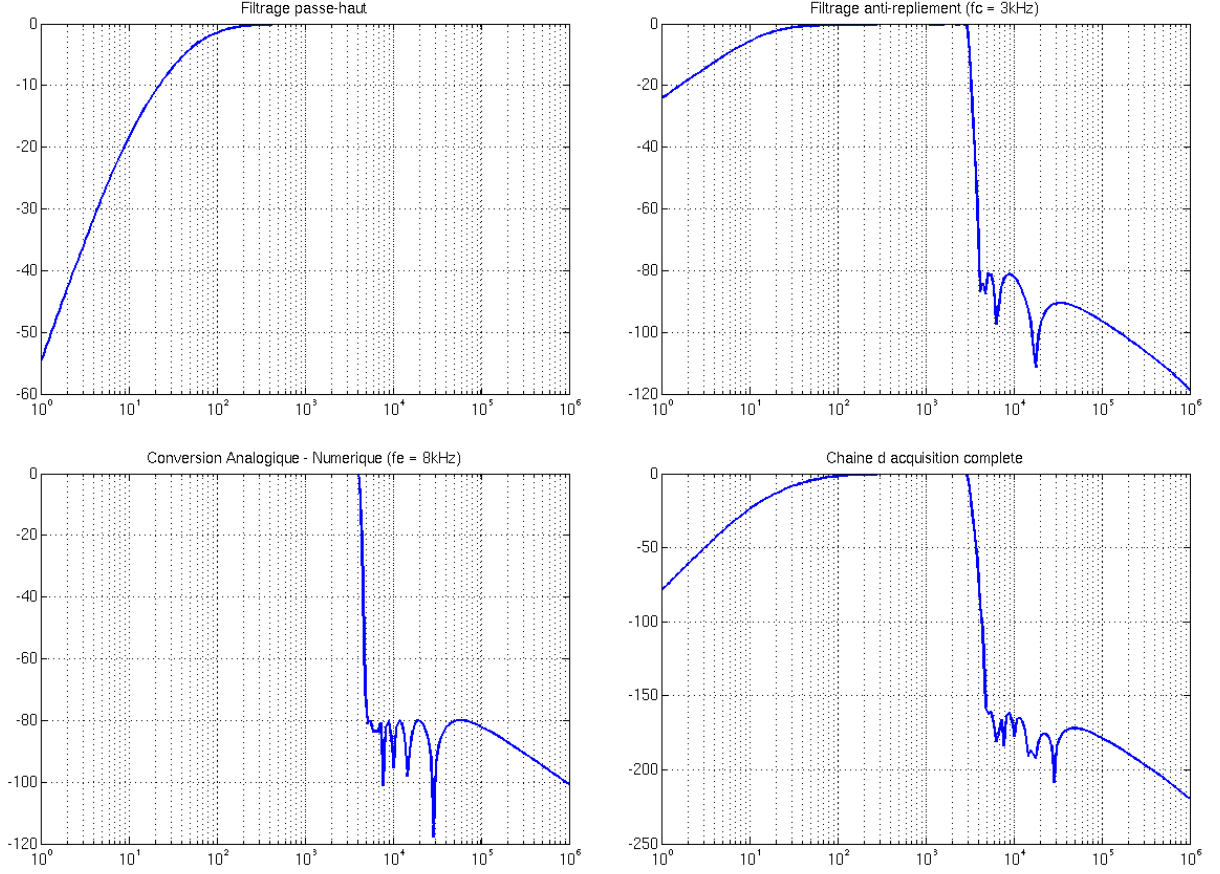


FIG. 4.4 – Diagrammes de Bode (gain seulement) des différents éléments de la chaîne d'acquisition.

tisseur numérique-analogique ainsi que le filtre à capacités commutées possèdent tous deux des caractéristiques fréquentielles qui sont respectivement normalisées aux fréquences f_e et f_c . La forme de leur diagramme ne change donc pas. Au final, nous pouvons constater que le signal échantillonné possède ici une largeur de bande comprise entre environ 100 Hz et 3 kHz et se trouve fortement filtré pour les fréquences supérieures, avec une atténuation minimale fournie par la chaîne d'acquisition de l'ordre de 160 dB. Revenons enfin sur les calculs d'amplitudes que nous avons débutés à la fin de la partie 4.1.1. Comme indiqué, nous disposons d'une tension de l'ordre de 0.16 mV en sortie des microphones choisis pour un niveau sonore de 60 dB. Après amplification par 100 du conditionneur, suivie de celle de l'amplificateur d'instrumentation réglée à 10, puis du gain programmable fournissant 24 dB, la tension disponible à l'entrée de l'étage de conversion atteint environ au maximum 5V pic à pic : le convertisseur retenu admet justement une pleine échelle de cette valeur, validant donc le choix de l'ensemble des gains de la chaîne d'acquisition.

4.1.3 Traitement numérique embarqué

Nous disposons maintenant d'un système capable de filtrer et convertir numériquement les signaux émis par chacun des microphones de l'antenne. Il nous faut maintenant traiter ces informations pour pouvoir décider d'où proviennent les sons captés dans l'espace. Pour cela, il est tout à fait envisageable d'utiliser un ordinateur personnel qui, connecté à la carte d'acquisition par l'intermédiaire d'une liaison à déterminer, prendrait en charge l'ensemble des calculs nécessaires. Si cette solution permet d'envisager l'implémentation d'algorithmes lourds sur une CPU récente, elle ne convient certainement pas à une utilisation embarquée sur une plateforme mobile. Nous avons donc opté pour l'utilisation de calculateurs embarqués, faibles consommateurs d'énergie, et facilement intégrables au dispositif présenté précédemment. Aujourd'hui, deux types de composants, plus ou moins spécialisés, sont traditionnellement utilisés pour le traitement de signaux numériques. Nous les présentons rapidement dans les deux sections suivantes, avant de détailler la solution retenue.

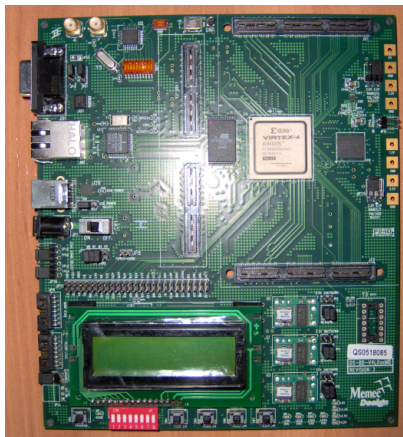
4.1.3.1 Le DSP

Le DSP (Digital Signal Processor) est un composant électronique optimisé pour le traitement du signal numérique. C'est en réalité un processeur dont l'architecture a été adaptée en profondeur de façon à le rendre capable d'effectuer des opérations complexes en un minimum de cycles. Le calcul élémentaire le plus utilisé en traitement numérique est la multiplication et addition, appelée aussi MAC (Multiplication and Accumulation). Le DSP est capable d'effectuer un tel traitement en un seul et unique cycle d'horloge, le rendant particulièrement adapté aux opérations de filtrage [Eyre 2000]. Ajoutons qu'il existe également des DSPs permettant des calculs à virgule fixe ou flottante. Un tel composant apparaît donc parfaitement approprié à l'implémentation de formations de voie large bande. Nous avons donc opté dans un premier temps pour une solution industrielle à base de DSP. Celle-ci était censée autoriser une programmation depuis un outil de génération automatique de code dédié à partir de schémas Simulink. Ainsi, une fois la procédure de localisation simulée et validée, il était théoriquement possible d'un simple clic de tester en pratique les algorithmes. Suite à de nombreuses déconvenues avec notre interlocuteur industriel, nous avons finalement opté pour une solution "maison" basée sur un FPGA de dernière génération.

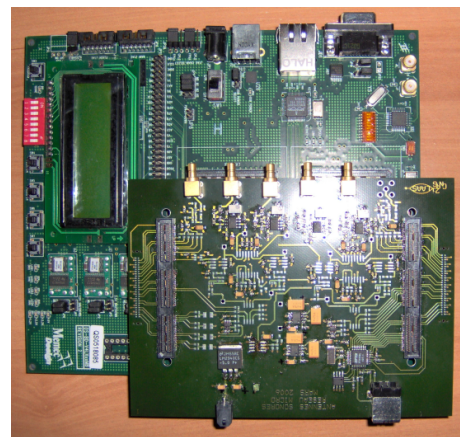
4.1.3.2 Le FPGA

Le FPGA (Field-Programmable Gate Array) est un circuit logique programmable, composé de nombreuses cellules logiques élémentaires librement assemblables. Dans un FPGA, ces cellules sont constituées d'une table de correspondance (LUT ou Look-Up Table), permettant l'implémentation d'équations logiques entre ses entrées, et d'une bascule jouant le rôle de mémoire. Ces blocs élémentaires sont connectés entre eux par l'intermédiaire d'une matrice de routage entièrement configurable. L'architecture interne d'un FPGA n'est donc pas optimisée pour une application particulière et se veut, au contraire, extrêmement générique. Pendant longtemps, l'enchevêtrement de l'ensemble des blocs et les technologies de fabrication ont limité la puissance de traitement des composants programmables. Aujourd'hui, le FPGA tend à se spécialiser en incluant, en plus du réseau logique programmable, des unités de calcul spécialisées pour tout type de traitement. En particulier, il existe maintenant des FPGAs dédiés au traitement numérique du signal intégrant des multiplieurs/accumulateurs pré-câblés et capables de rivaliser avec les meilleurs DSPs. D'une manière générale, les composants programmables sont particulièrement adaptés à tous types de traitements fortement parallèles [Campbell 1999]. La formation

de voie peut être vue comme telle, puisque requérant un traitement parallèle de chacune des voies d'acquisition. Combiné à l'exploitation de multiplieurs logiques, le FPGA semble donc aujourd'hui le composant le mieux adapté pour notre application. Cependant, si l'aspect matériel est fondamental pour espérer obtenir de bonnes performances, il ne faut pas perdre de vue que le FPGA requiert une programmation en VHDL (Very High speed integrated circuit Hardware Description Language). Il s'agit d'un langage de description du matériel destiné à décrire le comportement et/ou l'architecture du système numérique. Plutôt bas niveau, il est à opposer au langage C traditionnellement utilisé pour la programmation des DSPs mais permettant une description algorithmique haut-niveau des traitements à effectuer. De tous les composants programmables récents du commerce, nous avons retenus le FPGA Virtex 4 SX35 de la marque Xilinx pour ses très bonnes capacités en calcul numérique, mais aussi car il embarque 192 multiplieurs logiques capables d'effectuer chacun une opération MAC en un cycle d'horloge. Plutôt que de concevoir une carte embarquant ce FPGA, nous avons choisi de travailler à partir d'une carte d'évaluation du composant, fournissant suffisamment de connecteurs pour y raccorder la carte d'acquisition présentée précédemment. Finalement, la carte d'évaluation ainsi que la carte d'acquisition réalisée par nos soins sont visibles sur la figure 4.5. Compte tenu de la méthode de



(a) Carte d'évaluation FPGA Virtex 4 Xilinx.



(b) Carte d'acquisition conçue au LAAS (module de 4 voies) connecté au FPGA.

FIG. 4.5 – Photos des cartes d'acquisition et traitement

connexion des deux cartes retenues, nous avons dû scinder la partie acquisition en deux modules de 4 voies chacun, venant s'empiler sur les connecteurs disponibles.

4.1.4 Aspects logiciels

Terminons la présentation de la conception du matériel en présentant rapidement le fonctionnement du logiciel que nous avons intégré au FPGA. Comme nous l'avons précisé, notre objectif est de pouvoir commander sous forme logicielle les fonctionnalités de la carte d'acquisition (choix des fréquences de coupure et échantillonnage, choix des gains). Nous souhaitons ajouter la possibilité de router les données audio échantillonnées vers un ordinateur de façon à pouvoir d'abord tester hors-ligne nos algorithmes. Pour assurer un débit de données suffisant, nous avons choisi d'équiper la carte d'acquisition d'une liaison USB, dont le protocole est géré nativement par un composant de la société FTDI venant s'interfacer avec le FPGA. Une telle

solution présente deux avantages : il n'est pas nécessaire de coder intégralement le protocole USB, celui-ci étant pris en charge par ce composant additionnel, et FTDI fournit une librairie open-source permettant l'écriture d'un pilote en C sur l'ordinateur hôte. Au final, nous proposons donc l'architecture logicielle visible sur la figure 4.6, sur laquelle sont reportées les deux cartes présentées figure 4.5(b). Evidemment, le composant central de la carte d'évaluation reste le FPGA Virtex 4, qui prend en charge de manière modulaire chacune des fonctionnalités à contrôler. Ainsi, les blocs USB, LPF et PGA ont respectivement pour rôle le contrôle de l'USB, la programmation de la fréquence de coupure des filtres à capacités commutées et le choix de la valeur des amplifications des gains programmables. L'ensemble de ces blocs est relié à un contrôleur général ("Master_Controller"), véritable chef d'orchestre codé sous forme de machine à états finis par l'intermédiaire de deux bus d'adresses et de données. Deux blocs distincts se chargent du fonctionnement de la conversion analogique-numérique. Le premier, nommé MCLK, génère les horloges utiles aux convertisseurs selon la fréquence d'échantillonnage souhaitée ; le second, nommé ADC_Data, récupère les données numériques issues de la conversion. Enfin, nous avons besoin de stocker en mémoire un certain nombre de paramètres. Pour cela, nous avons codé un contrôleur nommé Flash permettant d'exploiter la mémoire de type flash placée sur la même carte que le FPGA. L'architecture modulaire proposée ici permet donc d'ajouter simplement de nouveaux blocs de façon à prendre en charge d'autres fonctionnalités de la carte d'évaluation (comme l'afficheur LCD par exemple), ou plus généralement encore des blocs algorithmiques venant traiter les données numériques pour la localisation.

4.1.5 Conclusion

Nous venons de présenter rapidement le travail de conception matérielle et logicielle nécessaire à l'obtention de signaux acoustiques destinés à être traités pour la localisation. Après avoir choisi l'ensemble des composants et conçu intégralement la carte de mise en forme et de conversion des signaux issus des microphones, nous avons proposé l'utilisation d'un FPGA de dernière génération pour le traitement numérique temps-réel de ces données. Ce travail nous permet maintenant d'envisager d'embarquer sur une des plateformes mobiles du LAAS-CNRS un capteur audio constitué de 8 microphones. Cependant, l'intégration et la validation du capteur ne sont malheureusement pas encore terminées à ce jour et nous ne disposons donc pas de véritables données expérimentales relevées depuis l'antenne. Ainsi, dans toute la suite, nous proposerons de simuler les signaux des microphones ainsi que le comportement de la chaîne d'acquisition par l'intermédiaire de son modèle présenté figure 4.4, de façon à déterminer les performances des méthodes de localisation que présentons maintenant.

4.2 Formation de voie large bande optimisée et cartes de puissance

Nous proposons dans cette section d'utiliser les formations de voies, obtenues par optimisation convexe dans le chapitre précédent, pour le tracé d'une carte de puissance de l'environnement acoustique de l'antenne. Notre objectif est de simuler le processus de localisation au complet, en tenant compte des caractéristiques fréquentielles de la chaîne d'acquisition. Pour cela, nous expliquons dans une première partie comment les filtres déterminés par résolution du problème de synthèse peuvent être effectivement utilisés pour réaliser en pratique la formation de voie désirée. Puis nous détaillons dans une seconde partie quelques résultats de localisation.

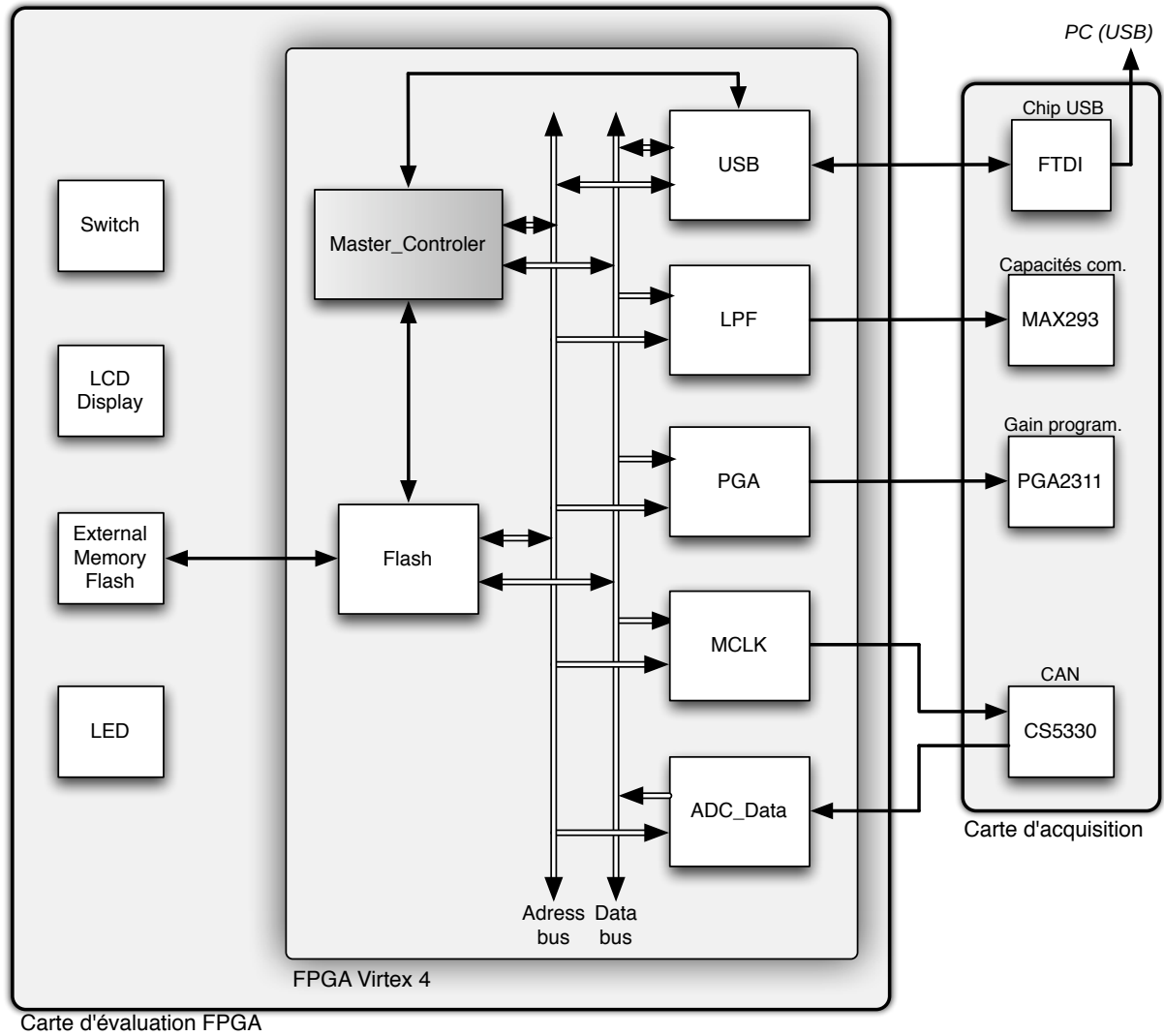


FIG. 4.6 – Représentation de l'architecture logicielle implémentée dans le FPGA. Chacun de blocs prend en charge une fonctionnalité de la carte d'acquisition ou de la carte d'évaluation.

4.2.1 Chaîne d'acquisition et formation de voie

Nous avons présenté dans le chapitre 3 différentes méthodes de synthèse permettant de conférer à l'antenne une meilleur directivité pour les grandes longueur d'onde. Ces méthodes nécessitent la discrétisation de l'espace des fréquences compris entre deux valeurs $f_{\min}$ et $f_{\max}$ utilisées pour la localisation. Dans notre cas, nous avons fixé $f_{\min} = 400$ Hz et $f_{\max} = 3$ kHz. En dehors de ces limites, la réponse d'antenne obtenue est non contrainte et conduit le plus souvent à des filtres possédant de forts gains dans ces zones. Ce phénomène a pu être constaté figure 3.27 page 103 où le gain des 8 filtres obtenus après optimisation atteint environ 35 dB dans les domaines non contraints. Nous avons représenté sur la figure 4.7 les conséquences sur le diagramme d'antenne d'un tel phénomène, pour une synthèse comprise cette fois entre 500 Hz et 1 kHz. Comme attendu, le diagramme d'antenne présente à son tour un très fort gain pour les fréquences n'appartenant pas au domaine d'optimisation. Une telle constatation, si elle n'a rien

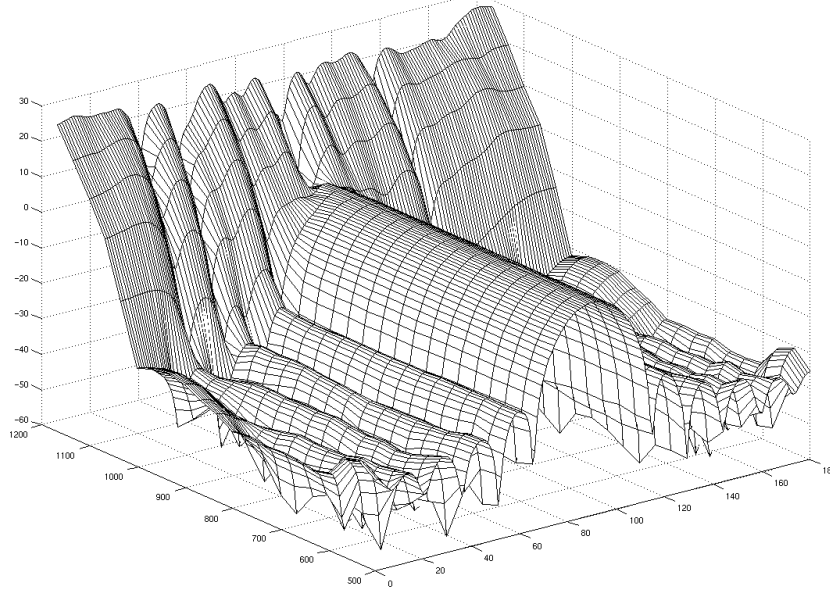


FIG. 4.7 – Phénomène d’explosion numérique en dehors du domaine d’optimisation.

d’étonnant, est cependant fondamentale pour pouvoir espérer utiliser en pratique une formation de voie déterminée par optimisation. Nous proposons donc ici d’étudier comment intervient une réponse d’antenne $D(\mathbf{r}, k)$ synthétisée sur une grille en fréquences dans le comportement de la totalité de la chaîne d’acquisition représentée figure 4.8. Celle-ci est constituée principalement de quatre éléments. Nous retrouvons en effet les microphones de fonction de transfert $M(p)$, la carte d’acquisition de fonction de transfert $ACQ(p)$, ainsi que les N filtres numériques de fonction de transfert $W_n(z)$ réalisant la formation de voie. Nous avons ajouté à cette liste un filtre numérique passe-haut $NUM(z)$ ayant pour but de couper toute fréquence inférieure à $f_{\min}$, un tel filtrage n’étant pas effectué par la carte d’acquisition. Enfin, comme indiqué figure 4.8, nous supposons que le comportement fréquentiel de chacune des voies d’acquisition est identique. Le tri des microphones effectué par leur fabricant, ainsi que l’utilisation de composants possédant des caractéristiques hautement reproductibles, garantissent une telle hypothèse.

Dans la suite, nous notons $x^*(t)$ la contrepartie à temps continu du signal numérique $x[l]$ dont les valeurs sont déterminées par échantillonnage d’un signal continu $x(t)$. En d’autres termes, cela revient à considérer que $x^*(t)$ est la modulation $x^*(t) = \text{ll}_{T_e}(t)x(t)$ de $x(t)$ par le peigne de Dirac $\text{ll}_{T_e}(t) = \sum_m \delta(t - mT_e)$. Les transformées de Fourier respectives des signaux $x(t)$ et $x^*(t)$, appelées aussi spectres par abus de langage, sont notées $X(f)$ et $X^*(f)$, et vérifient $X^*(f) = \frac{1}{T_e} \sum_m X(f - \frac{m}{T_e})$. La formation de voie étant réalisée numériquement, le signal de sortie $s[l]$ est un signal numérique tandis que le signal d’entrée $e(t)$, signal acoustique émis par la source, est un signal analogique. Il n’existe donc pas de fonction de transfert entre les signaux $s^*(t)$ et $e(t)$. Cependant, nous pouvons écrire pour chacune des voies d’acquisition,

$$\begin{cases} X_n^*(f) &= \frac{1}{T_e} \sum_m ANA_n(f - \frac{m}{T_e})E(f - \frac{m}{T_e}), \\ S_n^*(f) &= DIG_n(f)X_n^*(f), \end{cases} \quad (4.4)$$

avec

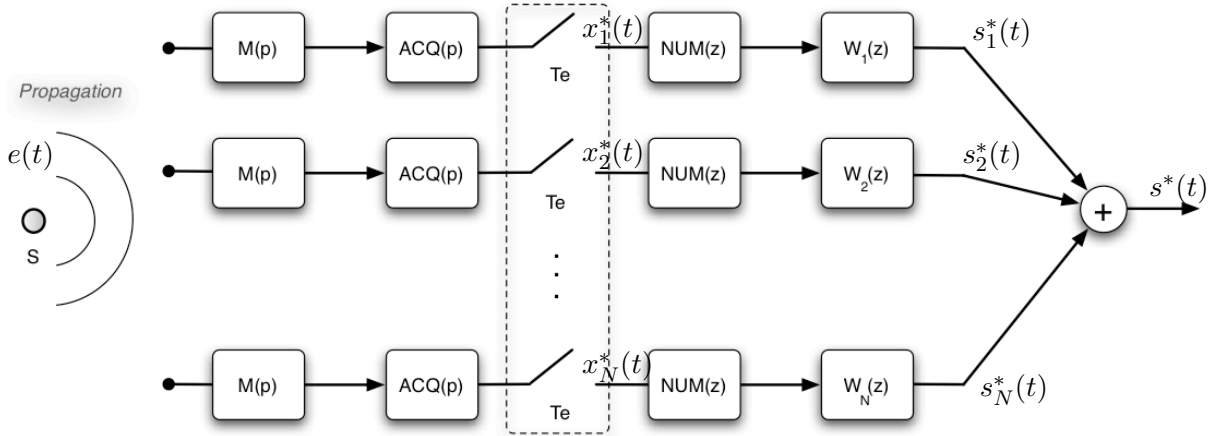


FIG. 4.8 – Réalisation pratique d'une formation de voie dans une direction de l'espace.

$$\begin{cases} ANA_n(f) &= ACQ(f)M(f)V_n(\mathbf{r}, f), \\ DIG_n(f) &= W_n(e^{j2\pi f T_e})NUM(e^{j2\pi f T_e}). \end{cases} \quad (4.5)$$

Les équations (4.4) - (4.5) indiquent que la transformée de Fourier $X_n^*(f)$ est obtenue par repliement et périodisation du spectre du signal $e(t)$ émis la source sonore, préalablement propagé, converti par le $n^{\text{ième}}$ microphone et filtré par la chaîne d'acquisition. Comme nous avons $DIG_n(f) = DIG_n(f - \frac{m}{T_e})$, $\forall m \in \mathbb{Z}$, nous pouvons combiner les équations précédentes afin d'obtenir le spectre $S^*(f)$ du signal de sortie de la formation de voie, qui vérifie

$$S^*(f) = \sum_{n=1}^N S_n^*(f) = \frac{1}{T_e} \sum_m \sum_{n=1}^N DIG_n(f - \frac{m}{T_e}) ANA_n(f - \frac{m}{T_e}) E(f - \frac{m}{T_e}) \quad (4.6)$$

$$= \frac{1}{T_e} \sum_m S^\sharp(f - \frac{m}{T_e}), \quad (4.7)$$

avec

$$S^\sharp(f) = \left[\sum_{n=1}^N W_n(f) V_n(\mathbf{r}, f) \right] NUM(e^{j2\pi f T_e}) ACQ(f) M(f) E(f). \quad (4.8)$$

Analysons maintenant l'équation (4.8) et ses conséquences sur le bon fonctionnement du système complet, et imaginons dans un premier temps que nous ayons $M(f) = ACQ(f) = NUM(e^{j2\pi f T_e}) = 1$. Dans une telle situation, l'équation (4.7) nous indique que le spectre de la source $E(f)$, après filtrage par la réponse d'antenne optimisée $D(\mathbf{r}, f) = \sum_{n=1}^N W_n(f) V_n(\mathbf{r}, f)$, est replié et périodisé de façon à obtenir le spectre du signal de sortie $S^*(f)$. Il faut donc veiller à ce que ce repliement ne vienne pas modifier les caractéristiques attendues du diagramme de directivité dans la bande de fréquence d'intérêt. Or, le phénomène d'explosion numérique que connaît la réponse d'antenne pour les fréquences non contraintes ne permet pas d'atteindre cet objectif. Il est donc absolument nécessaire de filtrer l'ensemble des fréquences n'appartenant pas au domaine contraint de sorte que le repliement et la périodisation de $S^\sharp(f)$ ne posent aucun

problème. Ceci est rendu possible par la présence des filtres de la chaîne d'acquisition qui, en plus de limiter la bande passante des signaux en vue de leur échantillonnage, permettent de contrer les forts gains des filtres numériques réalisant la formation de voie. Or nous avons vu dans le chapitre précédent que ce phénomène était intimement lié à la puissance des coefficients de ces filtres, puissance que nous avons limité à une valeur δ pour borner le gain de bruit blanc de l'antenne. Nous voyons donc apparaître ici un second critère justifiant de cette limitation lors de la synthèse. Ainsi, sans cette contrainte supplémentaire, la réponse d'antenne présente des gains tels que l'ensemble des filtres utilisés jusqu'ici ne suffisent pas à empêcher tout problème de repliement. Au final, pour qu'une formation de voie optimisée soit utilisable en pratique, il faut veiller à ce que le diagramme d'antenne possède un gain inférieur à l'inverse de l'atténuation apportée par la chaîne d'acquisition pour les fréquences non contraintes du problème d'optimisation. Cette conclusion est assez inédite et prouve que s'il faut évidemment s'intéresser en priorité au résultat d'optimisation dans la zone sous contraintes, il est nécessaire pour une implémentation pratique de contrôler le comportement de l'antenne pour l'ensemble des fréquences non contraintes.

4.2.2 Résultats de localisation

Nous sommes donc capables maintenant d'utiliser les résultats de synthèse du chapitre précédent pour la localisation, sous réserve qu'ils respectent les conditions énumérées précédemment. En robotique, la formation de voie est souvent utilisée pour tracer une carte de puissance acoustique de l'environnement. Celle-ci est obtenue par polarisations successives de l'antenne dans chacune des directions à tester. La courbe de puissance résultante est ensuite exploitée pour estimer les azimuts des sources par détection de ses maxima. Nous avons illustré dans le chapitre 2 toute la difficulté à travailler avec des signaux basse fréquence à partir d'une telle méthode de localisation et exploitant une formation de voie conventionnelle. La largeur du lobe, importante pour les grandes longueur d'onde, rend en effet délicate l'estimation de l'azimut, et interdit comme nous allons le voir la détection simultanée de multiples sources sonores. Nous avons justement cherché à améliorer le comportement basse fréquence de notre antenne, de sorte que nous puissions maintenant les exploiter pour la localisation. Nous proposons donc dans cette partie d'examiner les nouvelles cartes de puissance, obtenues à partir de ces formations de voie optimisées, et calculées à partir de la simulation sous Matlab/Simulink de la propagation et de la chaîne d'acquisition.

La structure simulée est la suivante. Le signal reçu par chacun des 8 microphones constituant l'antenne est déterminé selon l'équation (2.14) page 41, puis filtré par les éléments de la chaîne d'acquisition représentés figure 4.8 avant d'être échantillonné. Ces signaux sont ensuite exploités en parallèle par 91 formations de voie, chacune correspondant à une polarisation dans des directions θ_c comprises entre 0° et 180° , par pas de 2° . Nous préciserons plus loin les paramètres retenus lors de leurs synthèses. Nous disposons donc de 91 signaux numériques en sortie de ces formations de voie, correspondant aux signaux perçus par l'antenne en provenance de chacune de ces directions. Nous calculons alors leur puissance sur une fenêtre d'observation de durée DT_e , et reportons ce résultat dans un diagramme polaire. Ce calcul est répété au cours du temps, et conduit à autant de courbes de puissance que de fenêtres d'observation. Examinons maintenant les résultats de localisation obtenus selon cette procédure, pour différents types de source, en champ proche ou lointain.

4.2.2.1 Localisation de signaux sinusoïdaux

Nous proposons dans cette première partie de comparer les résultats de localisation obtenus à partir de formations de voie conventionnelles ou optimisées, et pour des environnements supposés non bruyants. Nous étudierons un peu plus loin l'influence du bruit blanc spatial sur la localisation. Pour cet exemple, la synthèse en champ lointain est effectuée selon la première méthode large bande décrite dans ce manuscrit, page 76, et visant à déterminer les coefficients des filtres numériques d'ordre $Q = 50$ réalisant la formation de voie pour une fréquence d'échantillonnage $f_e = 8$ kHz. La puissance de ces coefficients est limitée à 5.10^3 lors de la résolution du problème d'optimisation. La figure 4.9 présente les diagrammes d'antenne des formations de voie conventionnelle et optimisée, pour la direction de polarisation $\theta_c = 90^\circ$ et 26 fréquences comprises dans la bande 400 Hz – 3 kHz. Le résultat obtenu montre une réponse d'antenne possédant un lobe principal d'une largeur quasi-constante selon la fréquence, au prix d'une légère oscillation de son amplitude. Nous nous attendons donc à bénéficier directement de cette amélioration de la résolution des basses fréquences sur la carte de puissance. Pour le vérifier, nous simulons

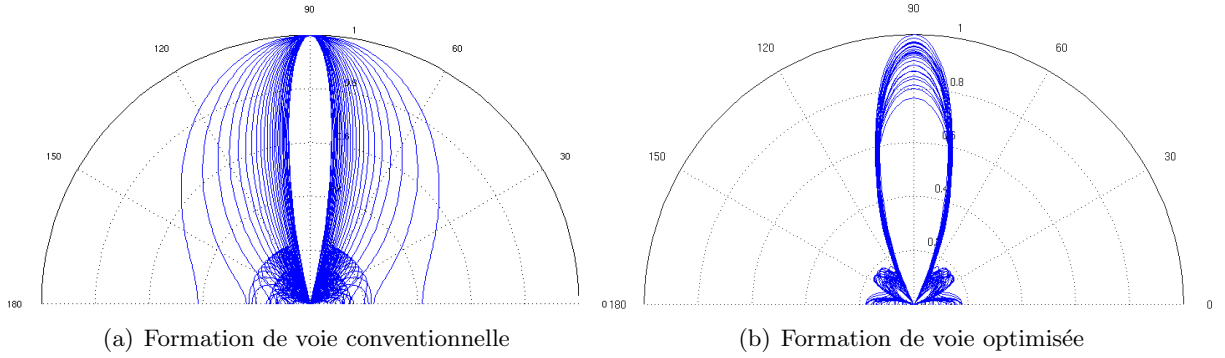


FIG. 4.9 – Comparaison des réponses d'antenne d'une formation de voie conventionnelle et optimisée, en champ lointain, pour la direction $\theta = 90^\circ$.

la présence de 2 sources sonores, situées en champ lointain, et émettant depuis les directions $\theta_1 = 70^\circ$ et $\theta_2 = 120^\circ$. La première source émet deux signaux sinusoïdaux de fréquence 400 Hz et 2.8 kHz, tandis que la seconde émet un signal composé des fréquences 400 Hz, 550 Hz, 1 kHz et 2.5 kHz, dont les amplitudes ont été ajustées de telle sorte qu'il possède une puissance environ 3.5 fois plus forte que celui émis par la première source. La puissance des sorties des 91 formations de voie a été calculée sur une fenêtre de $D = 500$ points, et conduit à la carte de puissance représentée figure sur la 4.10 en rouge, sur laquelle est également tracé en bleu le résultat pour une formation de voie conventionnelle. Notons que ces courbes ont été normalisées par rapport à leur plus forte valeur ; cette normalisation est systématiquement effectuée dans la suite afin d'améliorer la lisibilité des résultats. Compte tenu de la présence de basses fréquences dans les deux sources sonores, une telle formation de voie conduit à une carte acoustique ne comprenant qu'un seul maximum situé en $\theta = 72^\circ$. De plus, le lobe possède une forme telle qu'il est impossible de détecter une seconde source dans la scène, la première plus puissante cachant littéralement sa présence. L'utilisation de la formation de voie optimisée permet de résoudre ce problème, puisque nous pouvons distinguer sur la nouvelle carte de puissance un deuxième lobe lié à la seconde source. Les directions estimées à partir des maxima de cette courbe correspondent de plus exactement aux véritables directions d'émission simulées. Enfin, l'estimation des ordres de grandeur des puissances est respectée, le lobe pointant dans la direction θ_1 étant,

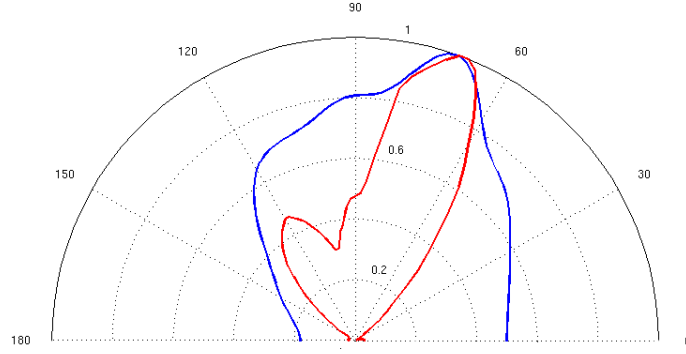


FIG. 4.10 – Carte de puissance acoustique, pour une formation de voie conventionnelle (bleu) et optimisée (rouge).

comme attendu, environ 3 fois plus grand que le second.

La simulation que nous proposons, si elle permet d'illustrer efficacement les limites de la formation de voie conventionnelle, suppose la présence de sources émettant des signaux composés de sommes de sinus. Ce type d'onde sonore est rare dans un environnement réel, c'est pourquoi nous proposons dans la suite de travailler systématiquement avec au moins une source sonore émettant un signal vocal enregistré depuis un des microphones de l'antenne par un oscilloscope numérique.

4.2.2.2 Influence de la distance

Nous proposons dans cette deuxième partie d'étudier l'influence de la distance sur la carte de puissance utilisée pour la localisation. Pour cela, nous simulons un environnement constitué de 2 sources sonores large bande dont les tracés temporels sont présentés figure 4.11. La première est le signal vocal d'un homme prononçant la phrase "ceci est un test de localisation de source sonore", émis dans un premier temps en champ lointain depuis la direction $\theta_V = 55^\circ$. La seconde, extraite d'une piste musicale, est simulée comme émettant en champ lointain depuis la direction $\theta_B = 90^\circ$ et joue le rôle d'une source de bruit. Nous supposons cette fois encore l'absence de bruit blanc spatial. Nous avons simulé le processus de localisation, à partir des formations de

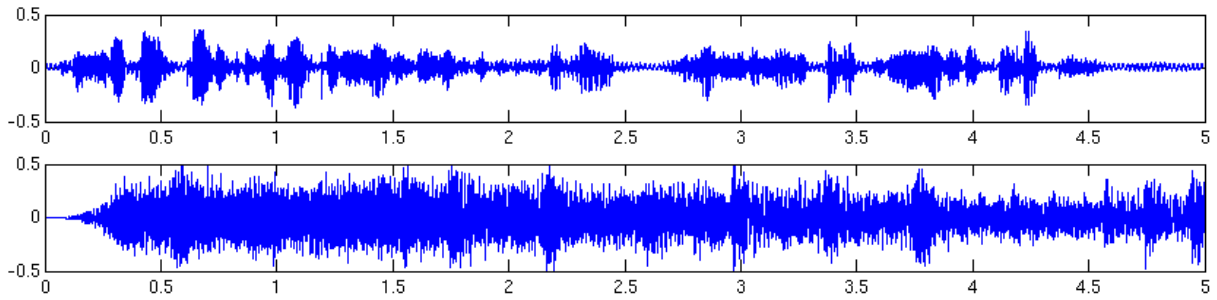


FIG. 4.11 – Signaux émis par les deux sources sonores à localiser : au dessus : signal vocal, enregistré depuis un microphone, dessous : extrait de musique.

voie optimisées, sur une durée de 5 secondes pour des fenêtres d'observation des 91 directions d'une durée $T_{\text{obs}} = 500T_e$, correspondant à une fréquence potentielle de localisation de 16 Hz.

A ce terme, nous disposons de 80 cartes de puissance, toutes regroupées sur la figure 4.12. Nous pouvons y constater que deux sources sonores sont bien détectées comme provenant des

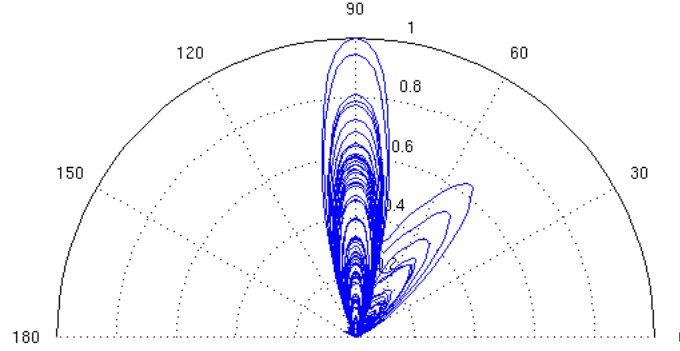


FIG. 4.12 – Carte de puissance obtenue à partir d’une formation de voie optimisée, synthétisée en champ lointain, pour des sources effectivement situées en champ lointain.

direction simulées. Le signal musical émettant de manière continue au cours du temps, nous pouvons remarquer que le lobe pointant dans la direction $\theta = \theta_B$ est plus marqué et plus souvent localisé que le signal vocal. Pour autant, et bien que celui-ci soit nettement moins puissant, la direction θ_V est correctement estimée, et l’utilisation de la formation de voie optimisée permet d’en distinguer la provenance.

Intéressons nous maintenant au cas où les source à localiser se situent à une distance de 1m de l’antenne. Nous avons montré dans le chapitre 3 que les basses fréquences pour une telle distance ne pouvaient plus être modélisées dans le champ lointain, et étaient à l’origine d’une déformation marquée de la réponse d’antenne. Ainsi, la réponse d’antenne figure 4.9(b), synthétisée en champ lointain, se transforme en celle représentée figure 4.13 pour une distance de 1m. Nous pouvons constater que les déformations induites par le champ proche modifient fortement l’efficacité du filtrage spatial des basses fréquences, tandis que les plus hautes, à la même distance, peuvent être encore considérés comme étant en champ lointain. De telles dégradations du diagramme

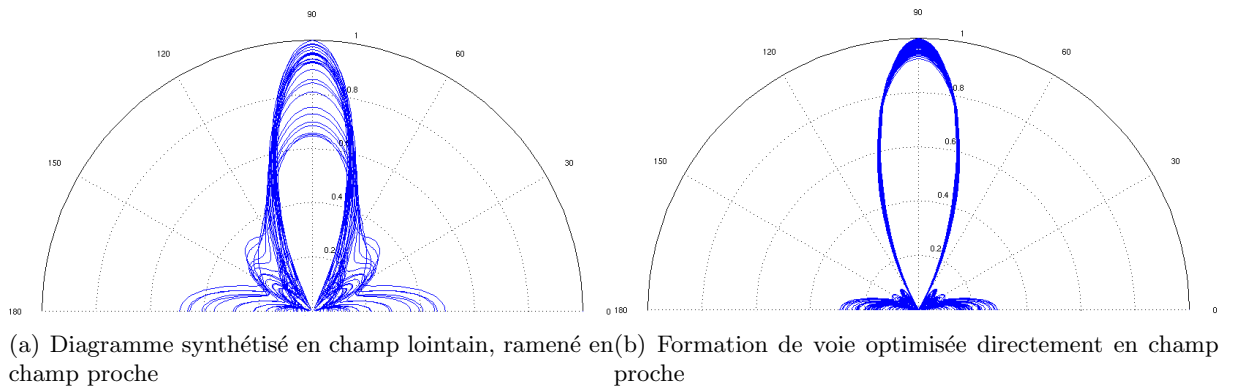


FIG. 4.13 – Comparaison des réponses d’antenne d’une formation de voie optimisée : à gauche, synthétisée en champ lointain mais utilisée à une distance de 1 m, à droite synthétisée par notre méthode modale, directement à 1 m.

de directivité peuvent être à l’origine de déformations de la carte de puissance, de sorte qu’il

devienne impossible de distinguer deux sources sonores se situant à des distances proches de l'antenne. La présence de ces dégradations est confirmée lorsque nous utilisons l'ensemble des 91 formations de voie synthétisées en champ lointain pour calculer la carte de puissance acoustique de l'environnement. Si nous considérons les deux mêmes sources sonores que précédemment, mais émettant toutes deux depuis une distance de 1m avec les mêmes azimuts θ_B et θ_V , la carte de puissance normalisée figure 4.12 se transforme en celle représentée figure 4.14. Nous pouvons y constater que les déformations du diagramme de directivité en champ proche conduisent, comme attendu, à des lobes de puissance plus larges et d'amplitude plus faible que précédemment. La

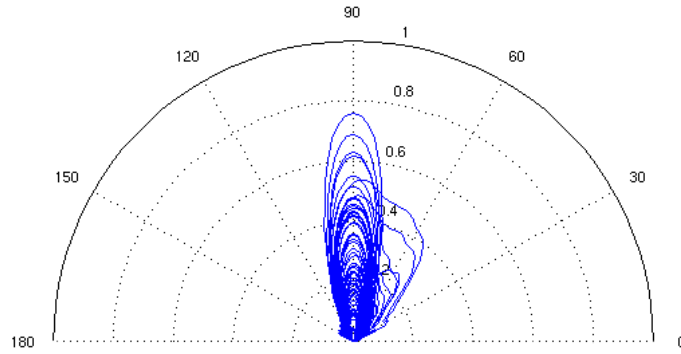


FIG. 4.14 – Carte de puissance obtenue à partir d'une formation de voie synthétisée en champ lointain, pour des sources situées en champ proche (1m)

qualité de la localisation en est fortement dégradée, puisqu'il devient très difficile d'estimer la direction d'émission du signal vocal. Les deux sources sont même parfois à l'origine d'un seul et même lobe de puissance pointant dans une fausse direction. Ces simulations montrent que si l'on utilise en champ proche une formation de voie synthétisée sous l'hypothèse de champ lointain, alors nous perdons tout le bénéfice de la méthode d'optimisation sur la résolution des basses fréquences.

Pour résoudre ce problème, nous proposons d'utiliser la seconde méthode d'optimisation proposée dans le chapitre 3, page 95, basée sur la représentation modale des diagrammes d'antenne, pour obtenir à une distance de 1 m les mêmes caractéristiques que la formation de voie précédente obtenue en champ lointain. Nous avons à nouveau effectué la synthèse de 91 formations de voie, en imposant que la puissance des coefficients des filtres RIF d'ordre $Q = 50$ soit inférieure à 10^3 . Nous obtenons après résolution, pour la direction $\theta_c = 90^\circ$ le diagramme d'antenne représenté sur la figure 4.13(b). Le gain en terme d'efficacité du filtrage spatial est évident et le comportement fréquentiel redevient quasi-invariant. Nous pouvons constater cependant la présence de lobes secondaires dans les deux directions périphériques $\theta = 0^\circ$ ou 180° pour les basses fréquences. L'amélioration de la réponse d'antenne permet alors d'envisager un meilleur comportement de la méthode de localisation. Pour le vérifier, nous avons simulé les mêmes conditions sonores que précédemment, mais en utilisant les formations de voie nouvellement synthétisées pour le calcul de la carte de puissance. La carte de puissance obtenue est représentée à la figure 4.15, sur laquelle nous pouvons à nouveau discerner nettement la présence de deux sources sonores. Comparativement à la figure 4.14, les deux lobes de puissance pointent bien dans les deux directions simulées, rendant possible la localisation des deux sources situées en champ proche.

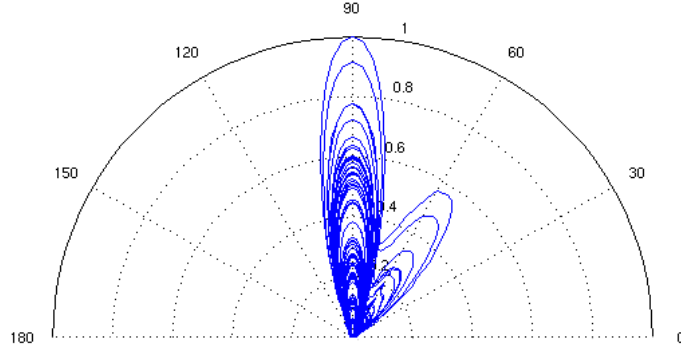


FIG. 4.15 – Carte de puissance obtenue à partir d’une formation de voie synthétisée à une distance de 1m, pour deux sources situées à la même distance.

4.2.2.3 Influence du bruit

Nous avons successivement montré, en supposant l’absence de bruit dans l’environnement, l’avantage d’utiliser une formation de voie invariante en fréquence ainsi que l’importance d’être capable de prendre en compte la présence éventuelle de sources sonores en champ proche. Nous proposons maintenant d’étudier les conséquences d’un bruit blanc spatial sur la carte de puissance. Pour cela, nous avons simulé une source sonore située à une distance de 1m de l’antenne, et émettant un bruit blanc depuis la direction $\theta = 60^\circ$. Celui-ci a été simulé en ajoutant à chacun des signaux reçus par les microphones un bruit blanc indépendant de même puissance qui a été ajusté pour obtenir différentes valeurs de Rapport Signal sur Bruit (RSB). Nous exploitons les formations de voie synthétisées précédemment pour une distance de 1m (voir figure 4.13(b) pour la direction $\theta_c = 90^\circ$), et calculons les cartes de puissance en intégrant les sorties de ces 91 formations de voie sur 300 points. Nous obtenons les cartes de puissance représentées sur la figure 4.16, calculées pour différentes valeurs du RSB. En l’absence totale de bruit blanc, i.e.

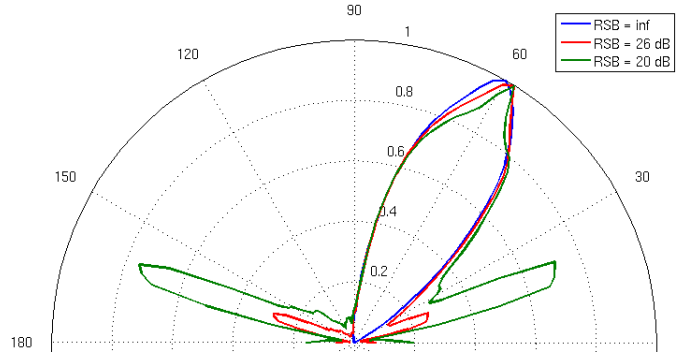


FIG. 4.16 – Tracé de la carte de puissance pour différentes valeur du rapport signal sur bruit, pour une source émettant depuis la direction $\theta = 60^\circ$, en champ lointain.

pour un RSB infini, la source sonore est correctement localisée avec un lobe pointant exactement dans la direction attendue. Pour des valeurs de RSB supérieures à environ 30 dB, la carte de puissance n’est pas modifiée de manière visible par la présence de bruit blanc spatial dans l’environnement. Par contre, les valeurs inférieures à 30 dB sont à l’origine de l’apparition de lobes

secondaires dans la carte de puissance, dont l'amplitude est d'autant plus élevées que le RSB est faible. Ce phénomène s'accompagne d'une légère déviation du lobe principal de l'ordre de 4° au maximum légèrement plus marquée dans les directions périphériques. La présence de ces lobes supplémentaires pour de faibles RSB peut s'expliquer par la présence de lobes secondaires dans les réponses d'antenne dont le niveau est cumulé par chacune des formations de voie, et amplifié par le bruit blanc spatial provenant de toutes les directions. Cependant, selon l'exploitation de la carte de puissance, il est envisageable de filtrer ces lobes supplémentaires en remarquant qu'ils apparaissent systématiquement aux mêmes angles θ , quel que soit l'angle d'incidence de la source.

Finalement, nous avons simulé la localisation des deux sources sonores représentées figure 4.11, lorsque celles-ci émettent depuis les directions $\theta_V = 120^\circ$ et $\theta_B = 60^\circ$ à une distance de 1m et en présence d'un bruit blanc spatial à l'origine d'un RSB de 28 dB mesuré relativement à la puissance du signal reçu sur le premier microphone. Les sorties de chacune des formations de voie sont intégrées sur 300 points, correspondant à une fréquence de localisation de l'ordre de 26 Hz. Les deux cartes de puissance normalisée, respectivement obtenues pour une formation de voie conventionnelle et optimisée, sont représentées à la figure 4.17. La présence de bruit blanc

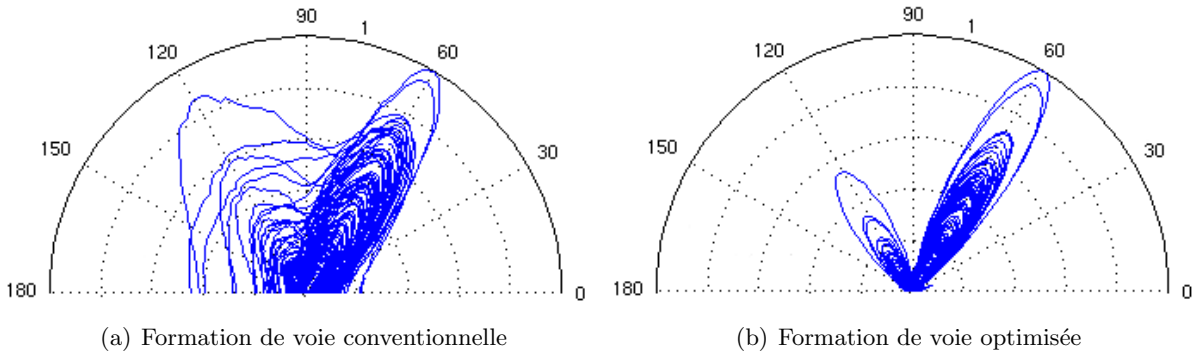


FIG. 4.17 – Carte de puissance, pour une formation de voie en champ lointain conventionnelle ou optimisée, en présence d'un bruit blanc (RSB = 28 dB).

spatial est à l'origine de légères déformations de la carte de puissance lors de l'exploitation de la formation de voie conventionnelle. Nous retrouvons aussi de larges lobes ne permettant pas de discriminer clairement la direction d'arrivée du signal vocal. La formation de voie optimisée, pour la valeur de RSB retenue, ne présente pas de lobes importants liés à la présence de bruit blanc, et permet de distinguer très nettement deux directions d'arrivée pointant, avec une erreur maximale de l'ordre de 2° , dans les deux directions d'incidence des deux sources.

4.2.3 Conclusion

Nous avons proposé dans cette section d'exploiter les formations de voie optimisées, synthétisées dans le chapitre précédent, afin de tracer une carte de puissance acoustique de l'environnement. Cette utilisation est la plus classique en robotique, puisqu'elle ne requiert que peu de puissance de calcul, et permet d'atteindre une cadence de localisation de l'ordre de la cadence vidéo. Un telle méthode de localisation s'avère donc particulièrement bien adaptée aux contraintes liées à la robotique mobile. Cependant, nous avons montré que les formations de voie synthétisées voyaient leur sensibilité au bruit augmentée, de sorte que la localisation soit plus ardue

pour des rapports signal sur bruit inférieurs à 25 dB. L'exploitation de signaux basse fréquence pose aussi le problème du modèle de propagation, à choisir entre celui du champ proche ou lointain. Bien que fondamental, ce problème n'avait jamais été soulevé en robotique. Nous avons montré les conséquences sur la localisation de l'utilisation de formations de voie synthétisées en champ lointain, mais exploitées pour déterminer la positions de sources situées en champ proche. Adapter la formation de voie à la distance effective d'émission permet d'aboutir à une localisation en azimuth plus précise et de discerner éventuellement la présence de deux sources proches. Or, ceci nécessite la connaissance à priori de cette distance qui fait pourtant partie à part entière du problème de localisation. C'est évidemment là une limite très importante de cette méthode. Nous proposons dans la partie suivante d'étudier une extension large bande récente de l'algorithme haute résolution MUSIC, permettant l'estimation à la fois de l'azimut et de la distance pour chacune des sources sonores.

4.3 Formation de voie large bande optimisée et méthodes haute résolution

4.3.1 Rappels sur MUSIC et notations

Nous proposons ici de rappeler très brièvement quelques notations introduites dans le chapitre 2 ainsi que le principe de l'algorithme MUSIC faible bande. Nous supposons que la scène sonore est constituée de S sources sonores faible bande situées en $\mathbf{r}_1^s, \dots, \mathbf{r}_S^s$, de même fréquence k , et d'une antenne de N microphones telle que $S < N$. Dans la suite, la notation k sera sous entendue. Nous avons vu que le vecteur des observations $\mathbf{M} = (M_1, \dots, M_N)^T$ vérifie l'équation (2.19)

$$\mathbf{M} = \mathcal{V}\mathbf{S}^0 + \mathbf{B}, \quad (4.9)$$

avec $\mathcal{V} = (\mathbf{V}(\mathbf{r}_1), \dots, \mathbf{V}(\mathbf{r}_S))^T$, et où $\mathbf{B}$ désigne un bruit additif spatialement et temporellement blanc, stationnaire et de même puissance σ_b^2 sur chacun des microphones. Le détail des autres notations a été introduit page 41. La matrice de covariance Γ_M des signaux reçus sur chacun des microphones, appelée aussi matrice interspectrale, s'écrit alors, sous l'hypothèse d'indépendance entre le signal et le bruit,

$$\Gamma_M = \mathcal{V}\Gamma_{S^0}\mathcal{V}^H + \Gamma_B = \Gamma_S + \Gamma_B. \quad (4.10)$$

Nous avons vu que la décomposition en valeurs et vecteurs propres de la matrice hermitienne Γ_M permet de distinguer deux sous-espaces vectoriels supplémentaires : le premier, appelé sous-espace signal $\mathbb{G}_S$, est engendré par les vecteurs propres $\mathbf{U}_s$, $s \in [1, \dots, S]$ associés aux S plus grandes valeurs propres l_s , tandis que le second, appelé sous-espace bruit $\mathbb{G}_B$, est engendré par les vecteurs propres $\mathbf{U}_b$, $b \in [S+1, \dots, N]$, associés aux $(N-S)$ plus petites valeurs propres $l_b = \sigma_b^2$. Cette propriété permet d'affirmer que tout vecteur de l'espace bruit est orthogonal à tout vecteur de l'espace signal, et en particulier aux vecteurs $\mathbf{V}(\mathbf{r}_1^s), \dots, \mathbf{V}(\mathbf{r}_S^s)$. De ce fait, la fonction

$$P(\mathbf{r}) = \frac{1}{\mathbf{V}(\mathbf{r})^H \Pi_B \mathbf{V}(\mathbf{r})}, \text{ avec } \Pi_B = \sum_{i=S+1}^N \mathbf{U}_i \mathbf{U}_i^H, \quad (4.11)$$

appelée pseudo-spectre MUSIC, admet S valeurs infinies en $\mathbf{r} = \mathbf{r}_s$, $s \in [1, \dots, S]$. En pratique, la matrice interspectrale Γ_M ne peut être qu'estimée, si bien que les vecteurs propres $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_N$

de $\hat{\Gamma}_M$ ne vérifient pas parfaitement les relations précédemment mentionnées. Néanmoins, la fonction $P(\mathbf{r})$ peut être utilisée pour estimer les positions des sources $\hat{\mathbf{r}}_s, s \in 1, \dots, S$ à partir de ses maxima.

4.3.2 Extensions de MUSIC au cas de sources large bande

L'algorithme MUSIC rappelé précédemment est valable uniquement pour des sources sonores faible bande. Cette partie généralise la méthode au problème de la localisation de sources large bande. L'approche utilisée consiste systématiquement à partitionner l'espace des fréquences en Q bandes étroites centrées sur les fréquences f_q ou $k_q, q \in [1, \dots, Q]$. Cette décomposition dans l'espace des fréquences s'effectue en pratique par l'intermédiaire d'une transformée de Fourier discrète, implémentée sous forme FFT.

Nous avons présenté dans le chapitre 2 une première extension élémentaire, déjà utilisée en robotique, de l'algorithme MUSIC au cas large bande. Il s'agit simplement de calculer la matrice de covariance $\Gamma_M(k_q)$ pour chaque bande de fréquence k_q , s'exprimant d'après (4.10) sous la forme

$$\Gamma_M(k_q) = E[\mathbf{M}^H(k_q)\mathbf{M}(k_q)] = \mathcal{V}(k_q)\Gamma_{S^0}(k_q)\mathcal{V}^H(k_q) + \sigma_b^2\mathcal{I}_N. \quad (4.12)$$

L'application de la méthode MUSIC à chacune de ces matrices conduit à Q pseudo-spectres qui permettent d'estimer les positions des sources à partir des maxima de leur moyenne. Néanmoins, une telle solution s'avère difficile à mettre en œuvre dans un contexte expérimental nécessitant un temps de réponse contraint. De plus, l'indépendance des décompositions en éléments propres conduit à des espaces signaux et bruits différents pour chaque fréquence. Le problème est donc de définir une combinaison cohérente de ces sous-espaces de façon à générer un sous-espace signal unique admettant des propriétés algébriques fonction du nombre et de la position des sources. La communauté du Traitement du Signal a proposé des extensions large bande prenant en compte ces considérations et permettant de diminuer drastiquement le nombre d'opérations nécessaires au traitement de chacune des fréquences k_q . Nous les présentons dans la suite.

4.3.2.1 Alignement et matrices de focalisation

Comme indiqué précédemment, le vecteur des observations $\mathbf{M}(k_q)$ correspondant à la $q^{\text{ième}}$ bande de fréquence k_q s'écrit

$$\mathbf{M}(k_q) = \mathcal{V}(k_q)\mathbf{S}^0(k_q) + \mathbf{B}(k_q) \quad (4.13)$$

$$= \left(\mathbf{V}(\theta_1, k_q) | \dots | \mathbf{V}(\theta_S, k_q) \right) \mathbf{S}^0(k_q) + \mathbf{B}(k_q), \quad q = 1, \dots, Q. \quad (4.14)$$

Les méthodes que nous allons présenter sont basées sur le principe de *focalisation*, exploitant le fait que les retards de phase $\phi_n(\theta_s, k_q) = k_q c \tau_n(\theta_s)$, $n = 1, \dots, N$, pour la fréquence k_q et associés à la $s^{\text{ième}}$ source en provenance de la direction θ_s , se transforment d'une fréquence à l'autre dans le rapport des fréquences. Dans ces conditions, nous sommes amenés à considérer une fréquence de référence k_0 et à appliquer à la matrice de covariance $\Gamma_M(k_q)$ une opération transformant les vecteurs d'antenne $\mathbf{V}(\theta_s, k_q)$ en un nouveau vecteur d'antenne $\mathbf{V}(\theta_s, k_0)$. Ainsi, les matrices de covariance $\Gamma_M(k_q)$, $q = 1, \dots, Q$, sont transformées en *matrices de covariance focalisées* pour lesquelles les vecteurs d'antenne $\mathbf{V}(\theta_s, k_0)$ sont indépendants de la fréquence k_q . Certaines méthodes se fondant sur ce principe appliquent aux matrices des vecteurs d'antenne $\mathbf{V}(\theta, k_q)$ la matrice de focalisation $\mathcal{T}(k_q)$ de sorte que

$$\mathcal{T}(k_q)\mathbf{V}(\theta, k_q) = \mathbf{V}(\theta, k_0). \quad (4.15)$$

En appliquant les Q matrices de focalisation $\mathcal{T}(k_q)$, $q = 1, \dots, Q$, à l'ensemble des vecteurs d'antenne, nous obtenons le vecteur signal focalisé

$$\mathcal{T}(k_q)\mathbf{M}(k_q) = \mathcal{V}(k_0)\mathbf{S}^0(k_q) + \mathcal{T}(k_q)\mathbf{B}(k_q), q = 1, \dots, Q. \quad (4.16)$$

Nous définissons alors la matrice de covariance focalisée Γ_M^0 des sorties des microphones par

$$\Gamma_M^0 = \sum_{q=1}^Q E \left[\mathcal{T}(k_q)\mathbf{M}(k_q)(\mathcal{T}(k_q)\mathbf{M}(k_q))^H \right] \quad (4.17)$$

$$= \sum_{q=1}^Q \mathcal{T}(k_q)\Gamma_M(k_q)\mathcal{T}^H(k_q). \quad (4.18)$$

En utilisant les équations (4.12) et (4.15), il vient

$$\Gamma_M^0 = \mathcal{V}(k_0)\Gamma_{S^0}^0\mathcal{V}^H(k_0) + \sigma_b^2\Gamma_B^0, \quad (4.19)$$

avec

$$\Gamma_{S^0}^0 = \sum_{q=1}^Q \Gamma_{S^0}(k_q), \text{ et } \Gamma_B^0 = \sum_{q=1}^Q \mathcal{T}(k_q)\mathcal{T}^H(k_q). \quad (4.20)$$

La méthode MUSIC basée sur ce principe de focalisation est fondé sur la décomposition en valeurs et vecteurs propres généralisés du faisceau matriciel (Γ_M^0, Γ_B^0) apparaissant dans (4.19) [Wang 1985]. Une telle décomposition permet alors de s'appuyer sur le même principe que la méthode MUSIC pour la détermination de sous-espaces signal et bruit. En comparaison avec l'extension naïve large bande de MUSIC consistant en un traitement indépendant de chacune des composantes fréquentielles, le principe de focalisation permet donc de traiter l'ensemble des fréquences à partir du seul faisceau matriciel (Γ_M^0, Γ_B^0) . Sa décomposition en valeurs propres et vecteurs propres généralisés permet, comme précédemment, de séparer le sous-espace signal $\mathbb{G}_S$ du sous-espace bruit $\mathbb{G}_B$, identiques pour chacune des fréquences selon la propriété d'alignement, mais nécessite la détermination de Q matrices $\mathcal{T}(k_q)$ satisfaisant la relation (4.15). Une estimation de ces matrices, basée sur une connaissance à priori des directions d'arrivées des sources, est proposée dans [Wang 1985]. Une alternative basée sur l'analyse modale détaillée dans le chapitre 3 est présentée dans [Abhayapala 2003] et ne nécessite aucune connaissance à priori sur la position des sources. Enfin, plus récemment, [Ward 2004] propose une prolongation de la propriété de focalisation dans l'espace des formations de voie. Cette approche permet de diminuer les dimensions des matrices de covariance focalisées et de réduire encore la charge de calcul nécessaire à la localisation. Nous allons la présenter maintenant.

4.3.2.2 MUSIC large bande dans l'espace des sorties des formations de voie

Généralités Dans cette partie, nous proposons d'exploiter les sorties de P formations de voie implémentées dans le domaine fréquentiel comme représenté figure 3.2, page 66, avec $S \leq P < N$. Comme précédemment, chacune des sorties des N microphones de l'antenne est décomposée en Q bandes étroites centrées sur les fréquences f_q ou k_q , $q = 1, \dots, Q$, de sorte que l'expression de

la sortie $Y_p(r, k_q)$ de la $p^{\text{ième}}$ formation de voie pour la $q^{\text{ième}}$ bande de fréquence à la distance r s'écrit ¹

$$Y_p(r, k_q) = \mathbf{W}_{k_q,p}^H(r) \mathbf{M}(k_q), p = 0, \dots, P-1, \quad (4.21)$$

avec $\mathbf{W}_{k_q,p}(r) = (w_{1,p}^{[k_q]}(r), \dots, w_{N,p}^{[k_q]}(r))^T$, $q = 1, \dots, Q$ le vecteur de poids de la $p^{\text{ième}}$ formation de voie pour la fréquence k_q et la distance r , et $\mathbf{M}(k_q)$ le vecteur des observations vérifiant

$$\mathbf{M}(k_q) = \mathcal{V}(k_q) \mathbf{S}^0(k_q) + \mathbf{B}(k_q) \quad (4.22)$$

$$= \left(\mathbf{V}(r_1, \theta_1, k_q) | \dots | \mathbf{V}(r_S, \theta_S, k_q) \right) \mathbf{S}^0(k_q) + \mathbf{B}(k_q). \quad (4.23)$$

Ainsi, le vecteur des sorties de P formations de voie $\mathbf{Y}(r, k_q) = (Y_0(r, k_q), \dots, Y_{P-1}(r, k_q))^T$ peut se mettre sous la forme

$$\mathbf{Y}(r, k_q) = \mathcal{W}_{k_q}^H(r) \mathbf{M}(k_q), q = 1, \dots, Q, \quad (4.24)$$

où $\mathcal{W}_{k_q}(r)$ désigne la matrice de pondération pour la distance r , de dimension $N \times P$

$$\mathcal{W}_{k_q}(r) = (\mathbf{W}_{k_q,0}(r), \dots, \mathbf{W}_{k_q,P-1}(r))^H, q = 1, \dots, Q. \quad (4.25)$$

La réponse d'antenne de la $p^{\text{ième}}$ formation de voie pour la fréquence k_q et la distance r s'écrit alors

$$D_{k_q,p}(\mathbf{r}) = D_{k_q,p}(r, \theta) = \mathbf{W}_{k_q,p}^H(r) \mathbf{V}(r, \theta, k_q). \quad (4.26)$$

Le cœur de la méthode réside dans l'utilisation de formations de voie invariantes en fréquences et synthétisées de telle sorte que

$$\forall r, D_{k_q,p}(r, \theta) = Y_p(\theta), q = 1, \dots, Q, \quad (4.27)$$

où $Y_p(\theta)$ désigne l'harmonique sphérique d'ordre p , définie par (3.55) page 94 indiquant que

$$Y_p(\theta) = \sqrt{\frac{2p+1}{4\pi}} P_p(\cos \theta), p = 1, \dots, P, \quad (4.28)$$

avec $P_p(\cdot)$ la fonction de Legendre de degré p . Or, comme indiqué dans la section 3.4.1, tout diagramme de directivité de la forme (4.26) peut aussi s'écrire

$$D(r, \theta, k) = \sum_{m=0}^M \alpha_m(k) R_m(r, k) Y_m(\theta), \quad (4.29)$$

où les termes $\alpha_m(k)$ et $R_m(r, k)$ désignent respectivement les coefficients modaux et les termes de normalisation modale définis par

$$\begin{cases} \alpha_m(k) &= -2jk \sqrt{\pi(2m+1)} \sum_{n=1}^N w_1^{[k]} j_m(kz_n) \\ R_m(r, k) &= r e^{jkr} h_m^{(2)}(kr), \end{cases}$$

où m désigne l'indice des modes, $h_m^{(2)}(\cdot)$ la fonction de Hankel sphérique de seconde espèce et $j_m(\cdot)$ la fonction de Bessel sphérique.

¹Remarquons que contrairement aux notations précédentes, la définition de $Y_p(r, k_q)$ fait intervenir la transposée hermitienne du vecteur $\mathbf{W}_{k_q,p}(r)$.

Invariance selon la distance Imaginons que nous ayons déterminé un vecteur de poids $\mathbf{W}_{k_q,p}^\infty$ pour lequel la réponse d'antenne en champ lointain $D_{k_q,p}^\infty(\theta) = \lim_{r \rightarrow \infty} D_{k_q,p}(r, \theta)$ donnée par (4.26) soit telle que $D_{k_q,p}^\infty(\theta) = Y_p(\theta)$. Nous pouvons montrer alors que la relation qui lie les ensembles de vecteurs de pondération en champ lointain et en champ proche pour les mêmes réponses d'antenne, s'écrit [Ward 2004]

$$\mathcal{W}_{k_q}(r) = \mathcal{W}_{k_q}^\infty \mathcal{R}_{k_q}^H(r), q = 1, \dots, Q, \quad (4.30)$$

avec $\mathcal{W}_{k_q}^\infty = (\mathbf{W}_{k_q,0}^\infty, \dots, \mathbf{W}_{k_q,P-1}^\infty)^H$ et $\mathcal{R}_{k_q}(r) = \text{diag}(1/R_0(r, k_q), \dots, 1/R_{P-1}(r, k_q))$. Plus clairement, la relation (4.30) indique comment ajuster les N coefficients des P formations de voie de la fréquence k_q pour obtenir, pour tout r , $D_{k_q,p}(r, \theta) = D_{k_q,p}^\infty(\theta) = Y_p(\theta)$.

Matrice de covariance focalisée Nous allons maintenant travailler avec la matrices de covariance $\Gamma_Y(r)$ des sorties des formations de voie, de dimension $P \times P$, et non plus des sorties des microphones, de dimension $N \times N$. Comme nous supposons que $P < N$, les dimensions du problème sont maintenant réduites et peuvent être à l'origine d'un gain notable en terme de complexité calculatoire. Cette matrice s'écrit

$$\Gamma_Y(r, k_q) = E[\mathbf{Y}(r, k_q) \mathbf{Y}^H(r, k_q)] \quad (4.31)$$

$$= \mathcal{W}_{k_q}^H(r) \mathcal{V}(k_q) \Gamma_{S^0}(k_q) \mathcal{V}^H(k_q) \mathcal{W}_{k_q}(r) + \sigma_b^2 \mathcal{W}_{k_q}^H(r) \mathcal{W}_{k_q}(r). \quad (4.32)$$

Nous définissons alors les matrices de covariance focalisées

$$\Gamma_Y^0(r) = \sum_{q=1}^Q \Gamma_Y(r, k_q), \text{ et } \Gamma_B^0(r) = \sum_{q=1}^Q \mathcal{W}_{k_q}^H(r) \mathcal{W}_{k_q}(r). \quad (4.33)$$

Comparativement aux matrices de covariance (4.20) dans l'espace des microphones, nous remarquons que la notion de matrice de focalisation est automatiquement introduite suite au passage dans l'espace des sorties des formations de voie. Cette propriété fondamentale permet de s'affranchir de l'estimation des Q matrices de focalisation $\mathcal{T}(k_q)$. Plus précisément, comme nous disposons de réponses d'antenne indépendantes de la fréquence quelle que soit la distance r , nous obtenons la propriété focale

$$\mathcal{W}_{k_q}^H(r) \mathbf{V}(r, \theta, k_q) = \mathcal{W}_{k_0}^H(r) \mathbf{V}(r, \theta, k_0) = \tilde{\mathbf{V}}_0(r, \theta), \forall q \in [1, \dots, Q], \forall \theta, \forall r, \quad (4.34)$$

où k_0 désigne par exemple la fréquence centrale du domaine de fréquence traité [Ward 2004].

Utilisation pour la localisation La méthode MUSIC basée sur le principe de focalisation est fondé sur la décomposition en valeurs et vecteurs propres généralisés du faisceau matriciel $(\Gamma_Y^0(r), \Gamma_B^0(r))$. Soient $l_n(r)$ et $\mathbf{U}_p(r)$, $p = 0, \dots, P-1$, respectivement les valeurs propres et vecteurs propres généralisés du faisceau matriciel $(\Gamma_Y^0(r), \Gamma_B^0(r))$, avec $l_p(r) \geq l_{p+1}(r)$. Nous avons les propriétés suivantes [Wang 1985] :

1. $l_S(r) = l_{S+2}(r) = \dots = l_{P-1}(r) = \sigma_b^2$;
2. l'ensemble des vecteurs $\mathbf{U}_S(r), \dots, \mathbf{U}_{P-1}(r)$ est orthogonal au vecteur $\mathbf{V}(\theta, k_0)$, c'est à dire $\mathbf{V}^H(r, \theta, k_0) \mathcal{U}_B(r) = \mathbf{0}$, avec $\mathcal{U}_B(r) = (\mathbf{U}_S(r) | \dots | \mathbf{U}_{P-1}(r))$.

De plus, l'ensemble des vecteurs colonne de $\mathcal{U}_S(r) = (\mathbf{U}_0(r) | \dots | \mathbf{U}_{S-1}(r))$, engendrant le sous-espace signal et l'ensemble des vecteurs colonne de $\mathcal{U}_B(r)$ engendrant le sous-espace bruit vérifient les propriétés suivantes [Wang 1985] :

$$\mathbf{U}_S(r)^H \Gamma_B^0(r) \mathbf{U}_B(r) = 0_{S \times (P-S)}, \quad (4.35)$$

$$\mathbf{U}_S(r)^H \Gamma_Y^0(r) \mathbf{U}_S(r) = \mathcal{I}_S, \quad (4.36)$$

$$\mathbf{U}_B(r)^H \Gamma_B^0(r) \mathbf{U}_B(r) = \mathcal{I}_{P-S}. \quad (4.37)$$

Il est proposé dans [Ward 2004] la définition suivante du pseudo-spectre MUSIC $P(r, \theta)$

$$P(r, \theta) = \frac{\tilde{\mathbf{V}}_0^H(r, \theta) \tilde{\mathbf{V}}_0(r, \theta)}{\tilde{\mathbf{V}}_0^H(r, \theta) \mathbf{U}_B(r) \mathbf{U}_B^H(r) \tilde{\mathbf{V}}_0(r, \theta)}, \quad (4.38)$$

présentant S pics en $(r, \theta) = (r_s, \theta_s)$, $s = 1, \dots, S$ permettant d'estimer l'azimut et la distance de la source. En pratique, le faisceau matriciel (Γ_Y^0, Γ_B^0) ne peut être qu'estimé par le faisceau $(\hat{\Gamma}_Y^0, \hat{\Gamma}_B^0)$. Si la matrice $\hat{\Gamma}_B^0$ est estimée directement selon (4.33), l'estimation de la matrice $\hat{\Gamma}_Y^0$ nécessite le moyennage sur L fenêtres temporelles des signaux issus des formations de voie, de telle sorte que

$$\hat{\Gamma}_Y(r, k_q) = \frac{1}{L} \sum_{l=1}^L \mathbf{Y}_l(r, k_q) \mathbf{Y}_l^H(r, k_q). \quad (4.39)$$

Finalement, la méthode MUSIC dans l'espace des formations de voie peut être synthétisée en les étapes suivantes :

0. Synthèse hors ligne de PQ formations de voie en champ lointain $D_{k_q, p}^\infty(\theta) = Y_p(\theta)$, $p = 0, \dots, P-1$ pour l'ensemble des fréquences k_q du domaine $[k_{\min} = k_1, \dots, k_{\max} = k_Q]$. Il en résulte Q matrices $\mathcal{W}_{k_q}^\infty$;
1. Acquisition des LQ N -uplets $\mathbf{M}_1(k_q), \dots, \mathbf{M}_L(k_q)$ issus de la décomposition en Q fréquences des signaux reçus par les N microphones, pour L fenêtres temporelles d'observation ;
2. Pour un ensemble de valeurs de r , calcul selon (4.24) des LQ P -uplets $\mathbf{Y}_1(r, k_q), \dots, \mathbf{Y}_L(r, k_q)$ issus des P formations de voie dont les coefficients sont donnés pour chaque valeur de r par l'équation (4.30) ;
3. Calcul des Q estimées $\hat{\Gamma}_Y(k_q)$ des matrices interspectrales des sorties des formations de voie, données par l'équation (4.39) ;
4. Calcul de l'estimé du faisceau matriciel focalisé $(\hat{\Gamma}_Y^0(r), \hat{\Gamma}_B^0(r))$ selon l'équation (4.33) ;
5. Décomposition du faisceau matriciel estimé en les valeurs propres généralisées $l_0 \geq l_1 \geq \dots \geq l_{P-1}$ et vecteurs propres généralisés $\hat{\mathbf{U}}_0, \dots, \hat{\mathbf{U}}_{P-1}$, pour chacune des distances r ;
6. Séparation des espaces signal $\hat{\mathcal{G}}_S$ et bruit $\hat{\mathcal{G}}_B$;
7. Définition du pseudo-spectre $P(r, \theta)$, sur l'ensemble des valeurs de r utilisé à l'étape 2, et pour différentes valeurs de θ ;
8. Et enfin, détection des positions en tant que les maxima de $P(r, \theta)$.

4.3.3 Synthèse modale des harmoniques sphériques

La première partie de l'algorithme décrit ci-dessus nécessite la synthèse en champ lointain de P formations de voie définies selon l'équation (4.27). Pour cela, nous proposons d'utiliser la méthode de synthèse modale définie dans le chapitre 3. Celle-ci a pour objectif de minimiser l'écart entre les coefficients modaux $\alpha_m^d(k_q)$ de la réponse d'antenne souhaitée $D_p^d(\theta)$ et ceux décrivant la formation de voie recherchée $D_{k_q,p}^\infty(\theta)$, pour l'ensemble des fréquences k_q , $q = 1, \dots, Q$.

Le comportement de référence $D_p^d(\theta)$ peut, comme toute réponse d'antenne, s'exprimer sous la forme (4.29). En champ lointain, cette même expression peut être ajustée pour obtenir

$$D_p^d(\theta) = \sum_{m=0}^{\infty} \alpha_m^d(k_q) \frac{j^{m+1}}{k} Y_m(\theta). \quad (4.40)$$

Comme nous souhaitons atteindre, d'après (4.27), $D_p^d(\theta) = Y_p(\theta)$, il en résulte directement

$$\alpha_m^d(k_q) = \begin{cases} 0 & \text{pour } m \neq p \\ \frac{k_q}{j^{m+1}} & \text{pour } m = p. \end{cases}$$

La représentation modale du comportement de référence est donc triviale. Nous voyons ici tout l'avantage de poser le problème de synthèse dans la base des harmoniques sphériques. Nous devons donc, pour chacune des fréquences k_q , définir P diagrammes d'antenne approximant au mieux les P premières harmoniques sphériques. Or, compte tenu du faible nombre de microphones constituant le capteur, il est évident que nous ne serons pas capables de déterminer les coefficients des formations de voie pour des harmoniques d'ordre élevé. De façon à évaluer la valeur maximale de P atteignable, nous avons synthétisé par résolution du problème d'optimisation (3.61) page 95 les 6 premières fonctions $Y_p(\theta)$, $p = 0, \dots, 5$ pour la fréquence $f = 400$ Hz. La puissance des coefficients est limitée à la valeur $\delta = 100$, et la minimisation est effectuée sur $M+1 = 14$ coefficients modaux. Les résultats de synthèse sont représentés figure 4.18, sur laquelle est tracée en rouge la fonction objectif $Y_p(\theta)$. Nous pouvons constater que le problème d'optimisation permet d'approximer sans erreur visible les trois premières harmoniques sphériques. La fonction $Y_4(\theta)$ est correctement décrite, à l'exception des deux directions périphériques $\theta = 0^\circ$ et $\theta = 180^\circ$. Enfin, les harmoniques de rang supérieures ne peuvent être correctement approchées : le nombre réduit de microphones et la contrainte δ importante conduit à des réponses d'antenne très éloignées du comportement recherché. Nous nous limiterons donc dans la suite à $P = 4$.

L'étape suivante consiste à synthétiser les quatre premières harmoniques sphériques pour l'ensemble des fréquences comprises entre 400 Hz et 3 kHz. Nous avons effectué cette synthèse dans les mêmes conditions que précédemment, avec un pas de 100 Hz entre chacune des fréquences f_q . Après résolution du problème d'optimisation, nous obtenons les diagrammes de directivité représentés figure 4.19 pour chacune des fréquences. Les réponses d'antenne possèdent toutes un comportement angulaire très voisin de celui défini par les fonction $Y_p(\theta)$. Nous pouvons néanmoins remarquer certaines variations selon la fréquence pour les deux directions périphériques. Étonnamment, ces déformations sont principalement localisées à des fréquences supérieures à environ 2.8 kHz, alors que nous avons vu dans le chapitre précédent que seules les basses fréquences posaient problème lors de la synthèse.

4.3.4 Résultats de localisation

Nous proposons dans cette partie d'exploiter selon l'algorithme décrit au §4.3.2.2 les formations de voie synthétisées précédemment et représentées figure 4.19. La méthode MUSIC dans

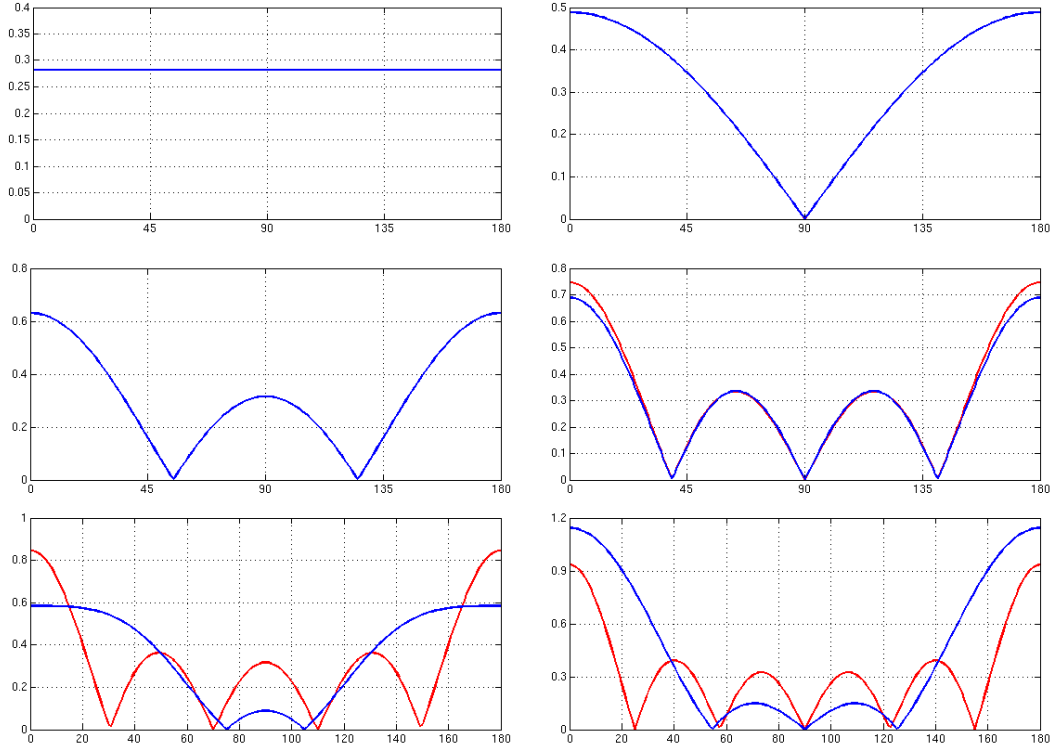


FIG. 4.18 – Synthèse modale en champ lointain des 6 premières harmoniques sphériques, pour $f = 400$ Hz, et une puissance des coefficients limitée à $\delta = 100$.

l'espace des formations de voie nécessite le calcul, pour chacune des distances r à tester, des valeurs propres et vecteurs propres généralisés du faisceau matriciel ($\hat{\Gamma}_Y^0(r), \hat{\Gamma}_B^0(r)$). Compte tenu de la puissance de calcul nécessaire à leur détermination, il nous semble opportun dans un premier temps de ne chercher à estimer que l'azimut de la source sonore. Cela est rendu possible par exploitation de la propriété selon laquelle le pseudo-spectre $P(r, \theta)$ admet des maxima en la même valeur θ quelle que soit la distance r à laquelle il est calculé [Ward 2004]. De cette façon, l'estimation de la position (r_s, θ_s) de la source peut se décomposer en deux étapes [Jeffers 2002] :

- Estimation de l'azimut θ_s , par recherche du maximum de la fonction $P^\infty(\theta)$, calculée à partir de l'unique faisceau matriciel estimé ($\hat{\Gamma}_Y^0(r), \hat{\Gamma}_B^0(r)$) pour $r \rightarrow \infty$. Cela revient à travailler uniquement avec les PQ formations de voie synthétisées en champ lointain ;
- Détection du maximum de la fonction $P(r, \hat{\theta}_s)$, permettant l'estimation de la distance r_s de la source au capteur.

Nous simulons cette première étape sous Matlab en considérant un environnement constitué de deux sources sonores, émettant deux signaux vocaux enregistrés de manière indépendante, simulées comme émettant depuis les positions $(45^\circ, 2\text{m})$ et $(130^\circ, 1.5\text{m})$. Le bruit ajouté à chacun des signaux issus des microphones est choisi indépendant, et sa puissance est ajustée de sorte que le rapport signal sur bruit soit de l'ordre de 10 dB. Nous procédons ensuite à l'échantillonnage des signaux à une fréquence $f_e = 15$ kHz, et à une décomposition fréquentielle sur 1024 points

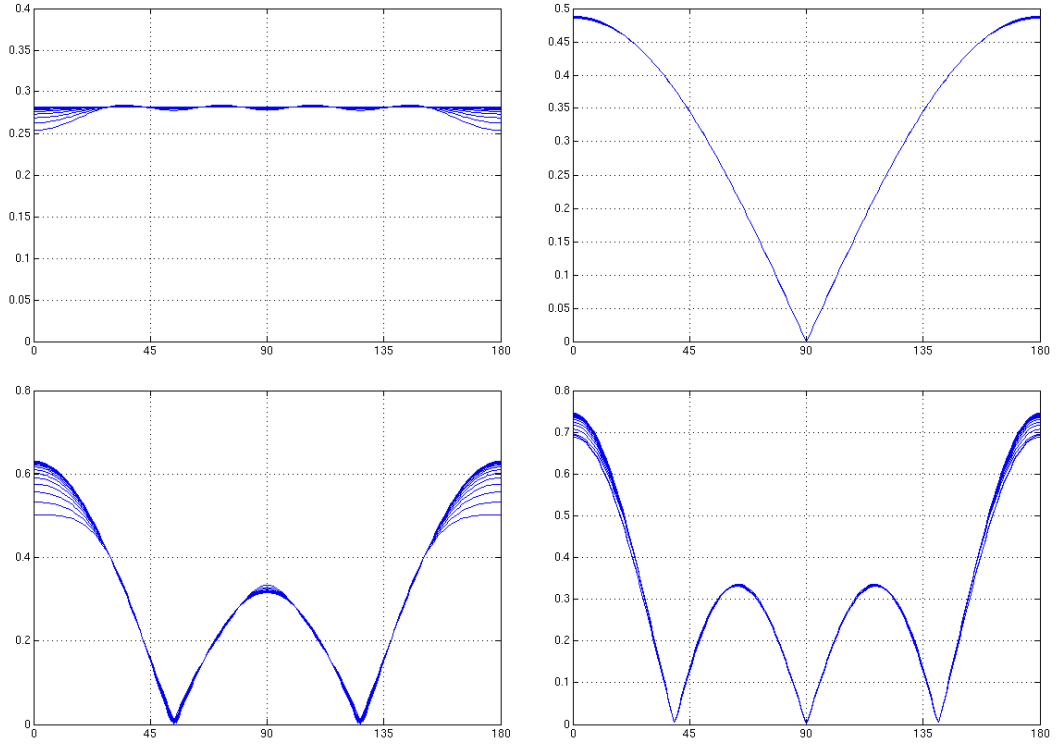


FIG. 4.19 – Résultats des formations synthétisées pour les 4 premières harmoniques sphériques, pour des fréquences comprises entre 400 Hz et 3 kHz.

par FFT tous les 1024 points. Cela revient à considérer 177 fréquences f_q dans l'intervalle 400 Hz – 3 kHz. Chacune des composantes fréquentielles est ensuite traitée par les formations de voie synthétisées précédemment en champ lointain jusqu'à construction des matrices $\mathbf{Y}_l(\infty, k_q)$. Les Q matrices de covariance $\hat{\Gamma}_Y(\infty, k_q)$ sont déterminées selon (4.39) avec $L = 16$. Enfin, le faisceau matriciel $(\hat{\Gamma}_Y^0(\infty), \hat{\Gamma}_B^0(\infty))$ est calculé par (4.33). Au final, compte tenu des durées nécessaires à l'observation des signaux pour la construction des estimées des covariances, la fréquence potentielle de localisation est de l'ordre de 2 Hz.

Estimation de l'azimut Nous avons simulé sur 5s de signal ce processus et calculé la fonction $P(\infty, \theta)$ au cours du temps. Nous obtenons après cette durée environ 20 pseudo-spectres MUSIC. Un de ces résultats est représenté sur la figure 4.20. Nous pouvons constater que l'azimut de la source est correctement estimé. La résolution de la localisation est évidemment sans commune mesure avec celle obtenue lors de l'utilisation des formations de voie pour le tracé d'une carte de puissance acoustique. De plus, bien que les sources soient situées à une distance respective de 1.5m et 2m, l'utilisation de formations de voie synthétisées en champ lointain permet d'aboutir tout de même une localisation très précise. Enfin, comparativement à l'extension naïve de MUSIC au cas large bande déjà proposée en robotique, cette approche reste tout à fait envisageable pour une utilisation en ligne sur une plateforme mobile.

Estimation de la distance La seconde étape consiste à calculer la fonction $P(r, \theta_s)$. Pour ce faire, nous discrétisons les distances comprises entre 1 m et 3 m par pas de 10 cm. Pour chacune

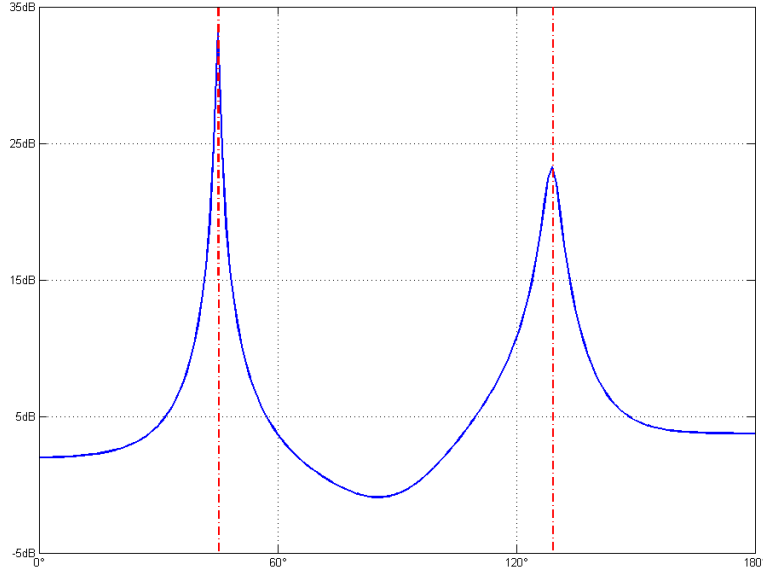


FIG. 4.20 – Pseudo-spectre MUSIC, dans sa version large bande dans l'espace des formations de voie, pour une valeur de r infinie. Les deux sources émettent depuis les direction indiquées en pointillé.

d'entres elles, nous procédons au calcul des matrices de covariances alignées et à leur décomposition en valeurs propres et vecteurs propres généralisés. Pour une fenêtre temporelle de 1024 points, nous aboutissons, après séparation des sous-espaces bruit et signal, à la courbe représentée figure 4.21. Nous pouvons constater que le pseudo-spectre admet deux valeurs maximales en $r = 1.6$ m et $r = 2$ m, tandis que l'émission des sources est simulée depuis la distance $r_s = 1.5$ m et 2 m. L'estimation de la distance semble donc correcte. D'ailleurs, si nous calculons ce même pseudo-spectre pour les différentes fenêtre d'observation, nous constatons que la position de ce maximum varie assez peu au cours du temps, de sorte qu'il est possible d'estimer la position des sources présentes dans l'environnement. Globalement, l'erreur maximale sur la distance estimée obtenue est de l'ordre de 30 cm, tandis que l'estimation des azimuts reste quasi-parfaite au cours du temps. Finalement, nous constatons que bien que les 4 premières harmonique sphériques n'aient pas été synthétisées parfaitement pour l'ensemble des fréquences d'intérêt, la propriété focale (4.34) nécessaire pour l'application de cette méthode reste suffisamment vraie. Dès lors, les espaces signaux et bruits de chacune des composantes fréquentielles alignées sont effectivement identiques, si bien que le moyennage fréquentiel défini par (4.33) prend véritablement un sens.

4.3.5 Conclusion

Nous avons évalué dans cette section une extension récente de l'algorithme haute-résolution MUSIC dans l'espace des formations de voie, appliqué à des signaux large bande. D'un point de vue algorithmique, cette méthode permet de réduire considérablement la puissance de calcul nécessaire à la détermination du pseudo-spectre. Plutôt que de traiter individuellement chacune des composantes fréquentielles, cette approche envisage l'utilisation de propriétés de focalisation pour ramener l'ensemble du problème en une unique fréquence k_0 . L'alignement fréquentiel nécessaire est de plus naturellement opéré par un choix judicieux de formations de voie invariantes en fréquence et en distance. De plus, il semble que cette méthode soit suffisamment robuste à l'er-

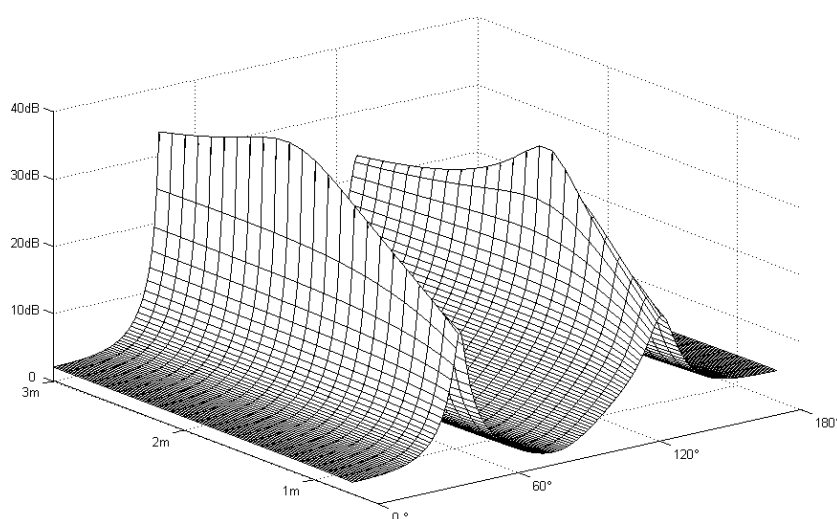


FIG. 4.21 – Estimation de la distance. Les maxima du pseudo-spectre correspondent aux véritables positions simulées d’émission des deux sources.

reur de synthèse des diagrammes de directivité, avec pour conséquence une estimation correcte de la distance. Il nous reste cependant à quantifier précisément l’influence des autres paramètres (nombre P de formations de voie synthétisées, rôle de l’estimateur de la covariance) avant de conclure sur la pertinence de cette méthode pour la détermination de la distance. Néanmoins, nous avons montré que cette technique permettait d’aboutir à une estimée correcte de l’azimut, à partir de formations de voie synthétisées en champ lointain, quelle que soit la distance entre la source et l’antenne. Finalement, bien que cette méthode soit finalisée sur le plan théorique, nous n’avons pas trouvé de publications relatives à sons implémentation pratique.

4.4 Conclusion

Nous avons présenté dans ce chapitre différentes méthodes de localisation basées sur l’exploitation de formations de voie large bande invariantes en fréquences. Nous avons dans un premier temps exploité leur propriété de filtrage spatial pour la construction d’une carte de puissance acoustique de l’environnement sonore de l’antenne. Cette utilisation demeure la plus courante en robotique, puisqu’elle permet d’envisager des cadences de localisation de l’ordre de la cadence vidéo. Nous avons montré que l’amélioration du comportement basse fréquence des diagrammes de directivité était à l’origine d’une meilleure résolution de la localisation. En contrepartie, l’invariance en fréquence est à l’origine de la nécessaire prise en compte des effets du champ proche lors de la localisation d’une source sonore à proximité de l’antenne. Cependant, sous réserve d’adapter la formation de voie à la distance d’émission, l’estimation de l’azimut de la source demeure correcte. Nous avons ensuite proposé l’évaluation d’une extension récente de l’algorithme MUSIC pour des signaux large bande. Celle-ci nécessite la définition d’un nombre important de formations de voie large bande pouvant être déterminées efficacement par la méthode de synthèse que nous avons proposée dans le chapitre 3. Cette évaluation constitue un travail préliminaire quant à l’applicabilité de la technique dans un contexte robotique en temps contraint. Basée sur la décomposition en valeurs propres et vecteurs propres généralisés d’un faisceau matriciel, elle semble constituer une alternative extrêmement intéressante en terme de temps de calcul pour

une utilisation en temps réel. Cependant, nos premières constatations tendent à prouver qu'une telle décomposition du faisceau matriciel est assez sensible numériquement, si bien qu'une implémentation sur un FPGA à virgule fixe semble difficile. De plus, celle-ci serait alors à implémenter en VHDL, et donc singulièrement délicate à programmer. De ce fait, cette dernière technique requièrera vraisemblablement de déporter une partie des calculs sur la CPU du robot, ce que nous cherchions précisément à éviter. Enfin, du fait de l'estimation des matrices de covariance sur des durées temporelles non négligeables, la cadence de localisation accessible demeure relativement faible et ne pourra probablement excéder 2 Hz. Nous avons aussi tenu à présenter en début de ce chapitre le travail important de conception du capteur sonore. Nous espérons que celui-ci sera fonctionnel très bientôt afin de disposer de mesures en environnement réel utiles à la validation expérimentale des méthodes de localisation évaluées ici en simulation semi-réalistes.

Conclusion

Le travail présenté dans cette thèse traite de la localisation de source sonore large bande dans le contexte de la robotique mobile. Cette thématique, peu traitée par la communauté Robotique, constitue la première étude en rapport avec l'acoustique au sein du groupe Robotique Action Perception du LAAS-CNRS. A ce titre, nous avons cherché tout au long de ce manuscrit à fournir au lecteur non initié l'ensemble des bases théoriques que nous avons jugées indispensables à la bonne compréhension de ce problème de traitement du signal.

Nous avons dans un premier temps parcouru une grande partie des méthodes de localisation proposées dans le cadre de la robotique mobile. Deux types d'approches s'affrontent. Les unes visent à reproduire géométriquement ou algorithmiquement le fonctionnement de notre système auditif. S'appuyant sur l'extraction difficile d'indices acoustiques, ces techniques bioinspirées aboutissent à des résultats de localisation d'une qualité parfois étonnante pour l'exploitation de seulement deux microphones. Toutefois, en l'état actuel, les algorithmes pèchent souvent par leur manque de robustesse au bruit et reposent pour la plupart sur des modèles de propagation validés uniquement en chambre anéchoïque. Parions cependant que la communauté Robotique saura améliorer à court terme ces premiers résultats. Dans un deuxième temps, nous avons introduit les approches basées sur des techniques éprouvées de Traitement du Signal. Des modèles précis de la propagation y sont employés afin de traiter analytiquement le problème de localisation. Les formalismes permettent de prendre en compte la présence de bruit dans l'environnement, et conduisent à des résultats de localisation très précis, mais souvent au prix d'une complexité algorithmique non compatible avec un contexte temps réel. A l'issue de cette étude, nous avons fait le choix de retenir les approches du type "formation de voie". Hélas, sous leur forme conventionnelle, elles ne permettent pas d'intégrer dans le processus de localisation les plus basses fréquences des sources sonores.

Nous avons alors cherché à améliorer le comportement basse fréquence de formations de voie large bande, implémentées dans le domaine temporel ou fréquentiel. En effet, des travaux récents de la communauté Acoustique nous ont indiqué qu'il est possible d'obtenir un diagramme d'antenne invariant en fréquence. Une telle propriété semble à priori extrêmement intéressante en vue de l'exploitation des grandes longueur d'onde pour la localisation. Or, de nombreux travaux requièrent un nombre élevé de microphones et conduisent souvent à des capteurs sonores d'une taille inenvisageable pour la robotique. Nous avons donc proposé la définition d'un problème d'optimisation et évalué l'extension immédiate large bande de méthodes convexes de la littérature. Ces premiers résultats nous ont montré toute la difficulté à travailler avec une antenne de petite taille, constituée de seulement huit microphones : nous avons obtenu des diagrammes de directivité quasiment invariants en fréquence, mais au prix d'une augmentation non négligeable de la sensibilité au bruit de la formation de voie. De plus, le nombre important de contraintes rend la résolution extrêmement lente et limite l'évolutivité par ajouts de contraintes convexes relatives à la robustification par rapport à d'éventuelles incertitudes. Nous avons alors proposé

une nouvelle méthode de synthèse, basée sur la représentation des diagrammes d'antenne dans la base des harmoniques sphériques. Cette solution permet une réduction drastique du nombre de contraintes et conduit à des temps de résolution beaucoup plus raisonnables. Nous avons alors bénéficié de ce formalisme pour évoquer et proposer des solutions aux problèmes liés à l'utilisation d'une formation de voie en champ proche.

Enfin, nous avons présenté deux méthodes de localisation exploitant les formations de voie large bande précédemment déterminées. La première est basée sur le tracé d'une carte de puissance acoustique de l'environnement. Elle demeure la méthode la plus employée en robotique pour la localisation. Nous avons illustré, à travers plusieurs simulations, l'amélioration de la résolution de la localisation apportée par l'invariance en fréquence des diagrammes d'antenne synthétisés. Néanmoins, la localisation s'avère possible uniquement pour des rapports signal sur bruit supérieurs à environ 20 dB. Cette limite peut toutefois être abaissée en étant moins exigeant sur la largeur du lobe principal de la réponse d'antenne. Puis nous avons évalué une extension récente de la méthode haute résolution MUSIC exploitant les sorties de formations de voie large bande. Cette étude préliminaire nous invite à envisager une implémentation en temps réel de l'estimation de l'azimut de la source.

▷ Perspectives à court terme

Nous avons tenu à présenter dans le dernier chapitre de ce manuscrit le travail de conception du capteur sonore à proprement dit. Nous avons défini un cahier des charges ambitieux visant à rendre l'ensemble des paramètres d'acquisition configurables logiciellement. De cette façon, la carte électronique conçue pendant cette thèse est suffisamment générique pour une exploitation dans d'autres gammes de fréquences ou pour d'autres cadences d'échantillonnage. Ce travail, comme tout travail expérimental, s'est avéré bien plus long que prévu, si bien que le capteur n'est à ce jour pas encore opérationnel. Nous devons auparavant finaliser le codage de la liaison USB sur le FPGA. Cette première étape est fondamentale puisqu'elle nous permettra de récupérer en temps réel les signaux des huit microphones et de les stocker afin de valider les différentes méthodes de localisation présentées dans ce manuscrit.

Bien entendu, certains développements théoriques évoqués ici sont loin d'être terminés. Nous avons par exemple proposé une nouvelle méthode d'optimisation convexe dans l'espace modal, visant à minimiser l'écart *en gain et phase* entre deux réponses d'antenne. Etendre la synthèse au gain uniquement pourrait permettre d'obtenir des diagrammes de directivité de meilleure qualité. Il en va de même pour les contraintes exprimant la robustesse vis à vis d'incertitudes sur les caractéristiques des microphones (défauts d'appariement, etc.). Toutes ces considérations ont déjà été prises en compte dans la littérature, mais aucune d'entre elles n'est directement transposable dans la base des harmoniques sphériques. Plus généralement encore, il serait envisageable de poser le problème de synthèse sur un continuum de valeurs angulaires et fréquentielles, afin de s'affranchir de la définition toujours délicate d'une grille de discrétisation des contraintes. Nous pensons que la convexification d'un tel problème peut être rendue possible par une extension du théorème de Kalman-Yakubovick-Popov, couramment employé pour la synthèse de filtres numériques RIF ou RII par programmation semi-définie en combinaison avec des arguments de factorisation spectrale [Wu 1996] [Alkire 2002] .

▷ Perspectives à moyen terme

La carte de puissance acoustique, déterminée à l'aide des formations de voie optimisées, peut

être exploitée dans la définition de fonctions de vraisemblance ou de fonctions d'importance dans un schéma de suivi auditif [Ward 2003] ou visio-auditif par filtrage particulière [Pérez 2004]. Cette nouvelle modalité extéroceptive serait à même de compléter efficacement les algorithmes de suivi de personne déjà implémentés au LAAS-CNRS [Brèthes 2006] pour des objectifs d'interaction homme-robot. Dans un autre domaine, nous pouvons envisager de développer des boucles de commande référencées audio, basées sur les mêmes techniques que pour les asservissements visuels. Cela revient à proposer une fonction de tâche, ou fonction d'erreur, directement définie à partir de données sensorielles [Samson 1991] dont la régulation à zéro assure le déplacement de la plateforme mobile en direction de la source [Espiau 1992]. Cependant, nous avons illustré dans ce manuscrit qu'il était très courant que la précision de la localisation soit dépendante de la position de la source vis à vis du capteur. Proposer une méthode de commande visant à déplacer le robot, et donc le capteur lui même, risque de dégrader sensiblement les performances de la localisation au cours du mouvement. Dès lors se pose la question de sa planification afin de garantir la qualité de traitement effectué par l'antenne tout le long de la trajectoire du robot.

▷ Perspectives à long terme

Au delà des simples boucles référencées audio évoquées précédemment, il peut être pertinent de proposer des objectifs de commande référencée multicapteurs, en s'intéressant tout particulièrement aux modalités visuelle et auditive. Cette question est centrale en neurosciences, où se pose également le problème de l'intégration multisensorielle pour la reconstruction de l'espace. En effet, nous disposons de deux référentiels distincts, dits respectivement oculocentré et craniocentré, liés aux modalités visuelle et auditive et dans lesquels chacune des informations prélevées par le capteur sont codées. La question est alors de savoir si nous devons conserver cette description indépendante, ou plutôt chercher à en représenter une dans le repère de l'autre. Cette étude faisait partie du projet ROBEA "Egocentre" dans lequel cette thèse s'inscrit, et dont l'objectif était d'étudier l'intérêt d'un référentiel dit d'"origine sensorielle" unique pour l'intégration multisensorielle et sensorimotrice. Nous avons montré dans [Argentieri 2006] que pour des systèmes constitués d'une chaîne cinématique simple, le choix d'un référentiel commun au sommet de la chaîne permet d'unifier la définition des différentes tâches référencées multicapteurs. Dans ce cadre, nous avons initialement prévu d'intégrer le capteur sonore présenté dans ce manuscrit au sein d'une "tête multicapteur" équipée de caméras. Un tel dispositif pourrait constituer une plateforme expérimentale permettant de valider à plus long terme l'ensemble de ces remarques dans un contexte robotique.

Bibliographie

- [Abhayapala 1999] P. Thushara D. Abhayapala. *Modal Analysis and Synthesis of Broadband Nearfield Beamforming Arrays*. Thèse de Doctorat, Australian National University, December 1999.
- [Abhayapala 2003] T.D. Abhayapala & H. Bhatta. *Coherent broadband source localization by modal space processing*. International Conference on Telecommunications, vol. 2, pages 1617– 1623, March 2003.
- [Alkire 2002] B Alkire & L. Vandenberghe. *Convex optimization problems involving finite autocorrelation sequences*. Mathematical Programming, vol. 93, n° 3, pages 351–359, June 2002.
- [Argentieri 2006] S. Argentieri, P. Barone, V. Cadenat, S. Celebrini, F. Chaumette, P Chiron, P Danès, J.Y Fourquet, B. Gao, N. Mansard, V Padois, P. Souères, M. Taïx & Y Trotter. *Projet EGOCENTRE : enchaînement dynamique de tâches référencées multi-capteurs pour l'exécution de mouvements complexes en robotique mobile*. 3èmes Journées du Programme Interdisciplinaire ROBEA,, 2006.
- [Asano 1999] Futoshi Asano, Hideki Asoh & Toshihiro Matsui. *Sound Source Localization and Signal Separation for Office Robot Jijo-2*. IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems, pages 243–248, Aug. 1999.
- [Asch 2006] Georges Asch. *Les capteurs en instrumentation industrielle*. Dunod, 6 edition, March 2006.
- [Aziz 1996] P.M. Aziz, H.V. Sorensen & J. van der Spiegel. *An Overview of Sigma-Delta Converters*. IEEE Signal Processing Magazine, vol. 13, n° 1, pages 61–84, January 1996.
- [Balakrishnan 1998] V. Balakrishnan & L. Vandenberghe. *Linear Matrix Inequalities for Signal Processing. An overview*. Annual Conference on Information Sciences and Systems, March 1998.
- [Berglund 2005] Erik Berglund & Joaquin Sitte. *Sound Source Localisation through Active Audition*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 509–514, Aug. 2005.
- [Boyd 1994] S. Boyd, L. El Ghaoui, E. Feron & V. Balakrishnan. *Linear matrix inequalities in system and control theory*, vol. 15. Studies in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), 1994.
- [Brèthes 2006] L Brèthes, F. Lerasle & P. Danès. *Particle filtering strategies for visual tracking dedicated to H/R interaction*. International Workshop on Vision Based Human-Robot Interaction, 2006.

- [Brooks 1999] R. Brooks, T.B. Senior & P.L.E. Uslenghi. *The Cog Project : Building a Humanoid Robot*. C.L. Nehaniv, editeur, Computations for Metaphors, Analogy, and Agents. Springer Verlag, 1999.
- [Brown 1993] Guy J. Brown. *Computational auditory scene analysis : A representational approach*. Journal of the Acoustical Society of America, vol. 94, n° 4, page 2454, October 1993.
- [Butler 1983] Butler & C. C. Helwig. *The Spatial Attributes of Stimulus Frequency in the Median Sagittal Plane and their Role in Sound Localization*. American Journal of Otolaryngology, vol. 4, n° 73, page 165, 1983.
- [Campbell 1999] A.T. Campbell, H.G. De Meer, M.E. Kounavis, K. Miki, J. Vicente & D. Vilella. *A Survey of Programmable Networks*. ACM SIGCOMM Computer Communications Review, vol. 29, n° 2, pages 7–23, April 1999.
- [Cavaco 1999] Sofia Cavaco & John Hallam. *A Biologically Plausible Acoustic Azimuth Estimation System*. Third International Workshop on Computational Auditory Scene Analysis of the Sixteenth International Joint Conference on Artificial Intelligence, pages 78–87, 1999.
- [Choi 2003] Changkyu Choi, Donggeon Kong, Jaywoo Kim & Seokwon Bang. *Speech enhancement and recognition using circular microphone array for service robots*. IEEE/RSJ Int. Conference on Intelligent Robots and Systems, vol. 3, pages 3516–3521, Oct. 2003.
- [Chou 1995] Thomas Chou. *Frequency-independent beamformer with low response error*. International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pages 2995–2998, May 1995.
- [Dolph 1946] C.L. Dolph. *A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level*. Proceedings of IRE, vol. 34, pages 335–348, June 1946.
- [Espiau 1992] B. Espiau, F. Chaumette & P. Rives. *A new approach to visual servoing in robotics*. IEEE Transactions on Robotics and Automation, vol. 8, n° 3, pages 313–326, June 1992.
- [Eyre 2000] J. Eyre & J. Bier. *The evolution of DSP processors*. IEEE Signal Processing Magazine, vol. 17, n° 2, pages 43–51, March 2000.
- [Gnansia 2005] Dan Gnansia. *Modèle auditif en temps réel*. Master’s thesis, Université Pierre et Marie Curie, July 2005.
- [Goodwin 1993] Michael M. Goodwin & Gary W. Elko. *Constant beamwidth beamforming*. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pages 169–172, April 1993.
- [GRAS 2004] GRAS. *Selection Guide for Microphones and Preamplifiers*. Rapport technique, G.R.A.S. Sound & Vibration A/S, 2004.
- [Hall 1966] Edward T. Hall. *The hidden dimension*. Doubleday, 1966.
- [Handzel 2002] Amir A. Handzel & P. S. Krishnaprasad. *Biomimetic Sound-Source Localization*. IEEE Sensors Journal, vol. 2, pages 607–616, 2002.
- [Handzel 2003] Amir A. Handzel, Sean B. Andersson, Martha Gebremichael & P.S. Krishnaprasad. *A Biomimetic Apparatus for Sound-source Localization*. IEEE Conference on Decision and Control, vol. 6, pages 5879 – 5884, Dec. 2003.

-
- [Hartmann 1999] William M. Hartmann. *How We Localize Sound*. Physics Today, vol. 52, n° 11, pages 24–29, 1999.
- [Hebrank 1974] Jack Hebrank & D. Wright. *Spectral Cues Used in The Localization of Sound Sources on The Median Plane*. Journal of the Acoustical Society of America, vol. 56, n° 6, pages 1829–1834, December 1974.
- [Heckmann 2005] Martin Heckmann, F. Joubin & E. Korner. *Sound Source Separation for a Robot Based on Pitch*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 203–208, Aug. 2005.
- [Huang 1995] Jie Huang, Noboru Ohnishi & Noboru Sugie. *A Biomimetic System for Localization and Separation of Multiple Sound Sources*. IEEE Transactions on Instrumentation and Measurement, vol. 44, n° 3, pages 733–738, June 1995.
- [Huang 1997a] J. Huang, N. Ohnishi & N. Sugie. Separation of multiple sound sources by using directional information of sound source, vol. 1, n° 4. Springer Tokyo, 1997.
- [Huang 1997b] Jie Huang, Noboru Ohnishi & Noboru Sugie. *Sound localization in reverberant environment based on the model of the precedence effect*. IEEE Transactions on Instrumentation and Measurement, vol. 46, n° 4, pages 842–846, August 1997.
- [Huang 1997c] Jie Huang, T. Supaongprapa, I. Terakura, N. Ohnishi & N. Sugie. *Mobile robot and sound localization*. IEEE/RSJ International Conference on Intelligent Robots and System, vol. 2, pages 683–689, September 1997.
- [Huang 1999] Jie Huang, Tadawute Supaongprapa, Ikutaka Terakura, Fuming Wang, Noboru Ohnishi & Noboru Sugie. *A Model-Based Sound Localization System and Its Application to Robot Navigation*. Robotics and Autonomous Systems, vol. 27, pages 199–209, 1999.
- [Huang 2002] Jie Huang, K. Kume, A. Saji, M. Nishihashi, T. Watanabe & W.L. Martens. *Robotic spatial sound localization and its 3D sound human interface*. International Symposium on Cyber Worlds, pages 191–197, Nov. 2002.
- [Humanski 1988] R. A. Humanski & R. A. Butler. *The Contribution of the Near and Far Ear Toward Localization of Sound in the Sagittal Plane*. Journal of the Acoustical Society of America, vol. 83, n° 10, page 2300, 1988.
- [Irino 1997] T. Irino & R. D. Patterson. *A time-domain, level-dependent auditory filter : The gammachirp*. Journal of the Acoustical Society of America, vol. 101, pages 412–419, 1997.
- [Jeffers 2002] Russell Jeffers, Kristine L. Bell & Harry L. Van Trees. *Broadband passive range estimation using MUSIC*. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pages 2921–2924, May 2002.
- [Johannesma 1972] P. I. M. Johannesma. *The pre-response stimulus ensemble of neurons in the cochlear nucleus*. IPO, editeur, Symposium on Hearing Theory, pages 58–69, 1972.
- [Jouhaneau 1999] Jacques Jouhaneau. Notions élémentaires d’acoustique, électroacoustique. Tech. & Doc./Lavoisier, December 1999.

- [Kennedy 1998] Rodney A. Kennedy, Thushara D. Abhayapala & Darren B. Ward. *Broad-band nearfield beamforming using a radial beampattern transformation*. IEEE Transactions on Signal Processing, vol. 46, n° 8, pages 2147–2156, August 1998.
- [Kennedy 1999] Rodney A. Kennedy, Darren B. Ward & Thushara D. Abhayapala. *Near-field Beamforming Using Radial Reciprocity*. IEEE Transactions on Signal Processing, vol. 47, n° 1, pages 33–40, January 1999.
- [Khalil 1994] F. Khalil, J.P. Jullien & A. Gilloire. *Microphone Array for Sound Pickup in Teleconference Systems*. Journal of the Acoustical Society of America, vol. 42, n° 9, pages 691–700, Sept. 1994.
- [Knapp 1976] C. Knapp & G. Carter. *The generalized correlation method for estimation of time delay*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 24, n° 4, pages 320–327, 1976.
- [Krim 1996] H. Krim & M. Viberg. *Two decades of array signal processing research : the parametric approach*. IEEE Signal Processing Magazine, vol. 13, n° 4, pages 67–94, July 1996.
- [Kumon 2005] Makoto Kumon, Tamoko Shimoda, Ryuichi Kohzawa & Zenta Iwai. *Audio Servo for Robotic Systems with Pinnae*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 885–890, Aug. 2005.
- [Kuratoro 2005] S. Kuratori, N. Suzuki, K. Nakadai, H.G. Okuno & H. Amano. *Implementation of Active Direction-Pass Filter on Dynamically Reconfigurable Processor*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 515–520, Aug 2005.
- [Lebret 1997] Hervé Lebret & Stephen Boyd. *Antenna array pattern synthesis via convex optimization*. IEEE Transactions on Signal Processing, vol. 45, n° 3, pages 526–531, March 1997.
- [Litovsky 1999] Ruth Y. Litovsky, H. Steven Colburn, William A. Yost & Sandra J. Guzman. *The Precedence Effect*. Journal of the Acoustical Society of America, vol. 4, pages 1633–1654, October 1999.
- [Löfberg 2004] J. Löfberg. *YALMIP : A Toolbox for Modeling and Optimization in MATLAB*. Taipei, Taiwan, 2004. Available from <http://control.ee.ethz.ch/~joloef/yalmip.php>.
- [Lopez-Poveda 1996] E.A Lopez-Poveda & R. Meddis. *A physical model of sound diffraction and reflections in the human concha*. Journal of the Acoustical Society of America, vol. 100, n° 5, pages 3248–3259, November 1996.
- [Luo 2003] Zhi-Quan Luo. *Mathematical programming : Applications of convex optimization in signal processing and digital communication*, vol. 97. Springer-Verlag GmbH, July 2003.
- [Ma 2004] Xiaohong Ma, Lili Liang & Fuliang Yin. *Generalized Crosspower-Spectrum Phase Method*. IEEE International Conference on Communications, Circuits and Systems, vol. 2, pages 826–829, June 2004.
- [Mahajan 2001] Ajay Mahajan & Maurice Walworth. *3-D Position Sensing Using the Differences in the Time-of-Flights from a Wave Source to Various Receivers*. IEEE Transactions on Robotics and Automation, vol. 17, n° 1, pages 91–94, February 2001.

-
- [Matsusaka 1999] Yosuke Matsusaka, Tsuyoshi Tojo, Sentaro Kubota, Kenji Furukawa, Daisuke Tamiya, Keisuke Hayata, Yuichiro Nakano & Tetsunori Kobayashi. *Multi-person Conversation via Multi-modal Interface - A Robot who Communicate with Multi-user -*. Eurospeech, pages 1723–1726, 1999.
- [Mattos 2004] Leonardo Mattos & Edward Grant. *Passive Sonar Applications : Target Tracking and Navigation of an Autonomous Robot*. IEEE International Conference on Robotics and Automation, vol. 5, pages 4265–4270, May 2004.
- [McCowan 2001] Iain A. McCowan. *Robust Speech Recognition using Microphone Arrays*. Thèse de Doctorat, Queensland University of Technology, Australia, 2001.
- [Middlebrooks 1991] John C. Middlebrooks & David M. Green. *Sound Localization by Human Listeners*. Annual Reviews on Psychology, vol. 42, pages 135–159, 1991.
- [Mills 1960] A. W. Mills. *Laterization of High-Frequency Tones*. Journal of the Acoustical Society of America, vol. 34, pages 32–132, 1960.
- [Mizoguchi 2004] H. Mizoguchi, Y. Tamai, K. Shinoda, S. Kagami & K. Nagashima. *Invisible messenger : visually steerable sound beam forming system based on face tracking and speaker array*. IEEE/RSJ International Conference on Intelligent Robots and System, vol. 3, pages 2417–2421, October 2004.
- [Murray 2005] John Murray, Stefan Wermter & Harry Erwin. *Auditory Robotic Tracking of Sound Sources using Hybrid Cross-Correlation and Recurrent Networks*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 891–896, 2005.
- [Nakadai 2000a] K. Nakadai, T. Lourens, H.G. Okuno & H. Kitano. *Active Audition for Humanoid*. 17th National Conference on Artificial Intelligence, 2000.
- [Nakadai 2000b] K. Nakadai, T. Matsui, H.G. Okuno & H. Kitano. *Active Audition System And Humanoid Exterior Design*. International Conference on Intelligent Robots and Systems, vol. 2, pages 1453–1461, Nov. 2000.
- [Nakadai 2001] K. Nakadai, H.G. Okuno & H. Kitano. *Epipolar geometry based sound localization and extraction for humanoid audition*. IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, pages 1395–1401, Nov. 2001.
- [Nakadai 2002a] Kazuhiro Nakadai, Hiroshi G. Okuno & Hiroaki Kitano. *Auditory Fovea Based Speech Separation and Its Application to Dialog System*. IEEE/RSJ International Conference on Intelligent Robots and System, vol. 2, pages 1320–1325, Sept. 2002.
- [Nakadai 2002b] Kazuhiro Nakadai, Hiroshi G. Okuno & Hiroaki Kitano. *Real-time Sound Source Localization and Separation for Robot Audition*. International Conference on Spoken Language Processing (ICSLP-2002), pages 193–196, September 2002.
- [Nakadai 2003] Kazuhiro Nakadai, Daisuke Matsuura, Hiroshi G. Okuno & Hiroaki Kitano. *Applying Scattering Theory to Robot Audition System : Robust Sound Source Localization and Extraction*. Proceedings of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems, vol. 2, n° 1147–1152, Oct. 2003.
- [Nakadai 2005] Kazuhiro Nakadai, Hirofumi Nakajima, Kentaro Yamada, Yuji Hasegawa, Takahiro Nakamura & Hiroshi Tsujino. *Sound Source Tracking with Directivity Pattern Estimation Using a 64ch Microphone Array*. IEEE/RSJ

- International Conference on Intelligent Robots and Systems, pages 196–202, Aug. 2005.
- [Ohashi 2005] Yasuaki Ohashi, Tsuyoki Nishikawa, Hiroshi Saruwatari, Akinobu Lee & Kiyohiro Shikano. *Noise-robust hands-free speech recognition based on spatial subtraction array and known noise superimposition*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 533–537, Aug. 2005.
- [Okuno 2004] Hiroshi G. Okuno, Tetsuya Ogata, Kazunori Komatani & Kazuhiro Nakadai. *Computational Auditory Scene Analysis and Its Application to Robot Audition*. International Conference on Informatics Research for Development of Knowledge Society Infrastructure, pages 73–80, March 2004.
- [Okuyama 2002] Fukutaro Okuyama, Jun ya Takayama, Shinji Ohyama & Akira Kobayashi. *A study on determination of a sound wave propagation direction for tracing a sound source*. Proceedings of the 41st SICE Annual Conference, vol. 2, pages 1102–1104, Aug. 2002.
- [Omologo 1994] Maurizio Omologo & Piergiorgio Svaizer. *Acoustic Event Localization using A Crosspower-Spectrum Phase Based Technique*. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pages 273–276, April 1994.
- [Omologo 1997] M. Omologo & P. Svaizer. *Use of the Crosspower-Spectrum Phase in Acoustic Event Location*. IEEE Transactions on Speech and Audio Processing, vol. 5, n° 3, pages 288–292, May 1997.
- [Park 1992] Sangil Park. *Principles of Sigma-Delta Modulation for Analog-to-Digital Converters*. Rapport technique, Motorola, 1992.
- [Patterson 1995] R. D. Patterson, M. Allerhand & C. Giguère. *Time domain modelling of peripheral auditory processing : a modular architecture and a software platform*. Journal of the Acoustical Society of America, vol. 98, pages 1890–1894, 1995.
- [Pérez 2004] P. Pérez, J. Vermaak & A. Blake. *Data fusion for visual tracking with particles*. Proceedings of the IEEE, vol. 92, n° 3, pages 495–513, March 2004.
- [Pujol 2006] R. Pujol. *Promenade autour de la cochlée <www.cochlee.info> INSERM - Université Montpellier 1*, 2006.
- [Purves 2004] Dale Purves, George J. Augustine, David Fitzpatrick, William C. Hall, Anthony-Samuel LaMantia, James O. McNamara & S. Mark Williams. *Neuroscience*. Sinauer Associates, 3 edition, 2004.
- [Rossignol 2001] L. Rossignol, G. Scorletti & V. Fromion. *Filter design under magnitude constraints is a finite dimensional convex optimization problem*. IEEE Conference on Decision and Control, vol. 4, pages 3575–3580, Dec. 2001.
- [Samson 1991] C. Samson & K. Ait-Abderrahim. *Feedback stabilization of a nonholonomic wheeled mobile robot*. IEEE International Workshop on Intelligent Robots and Systems, vol. 3, pages 1242–1247, Nov. 1991.
- [Schmidt 1979] R. O. Schmidt. *Multiple Emitter Location and Signal Parameter Estimation*. RADC Spectrum Estimation Workshop, Oct. 1979.
- [Scholnik 2001] Dan P. Scholnik & Jeffrey O. Coleman. *Superdirectivity and SNR constraints in wideband array-pattern design*. IEEE Radar Conference, pages 191–186, May 2001.

-
- [Silverman 1992] H. F. Silverman & S. E. Kirtman. *A Two-Stage Algorithm for Determining Talker Localisation from Linear Microphone Array Data*. Computer Speech and Language, vol. 6, pages 129–152, 1992.
- [Strybel 1984] T. Z. Strybel & D. R. Perrott. *Discrimination of Relative Distance in the Auditory Modality : the Success and Failure of the Loudness Discrimination Hypothesis*. Journal of the Acoustical Society of America, vol. 76, n° 20, page 318, 1984.
- [Tamai 2004] Yuki Tamai, Satoshi Kagami, Hiroshi Mizoguchi, Yutaka Amemiya, Koichi Nagashima & Tachio Takano. *Real-Time 2 Dimensional Sound Source Localization by 128-Channel Huge Microphone Array*. IEEE International Workshop on Robot and Human Interactive Communication, pages 65–70, Sept. 2004.
- [Tamai 2005] Yuki Tamai, Yoko Sasaki, Satoshi Kagami & Hiroshi Mizoguchi. *Three Ring Microphone Array for 3D Sound Localization and Separation for Mobile Robot Audition*. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 903–908, Aug. 2005.
- [Tseng 1992] C. Tseng & L.J. Griffiths. *A simple algorithm to achieve desired patterns for arbitrary arrays*. IEEE Transactions on Signal Processing, vol. 49, n° 11, pages 2737–2746, November 1992.
- [Tutuncu 2006] R.H. Tutuncu, K.C. Toh & M.J. Todd. *SDPT3 - A MATLAB Software Package for Semidefinite-Quadratic-Linear Programming*. <http://www.math.nus.edu.sg/~mattohc/sdpt3.html>, 2006.
- [Valin 2003] Jean-Marc Valin, Francois Michaud, Jean Rouat & Dominic L'etoumeau. *Robust Sound Source Localization Using a Microphone Array on a Mobile Robot*. Proceedings of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems, vol. 2, pages 1228–1233, Oct. 2003.
- [Valin 2004] Jean-Marc Valin, Francois Michaud, Brahim Hadjou & Jean Rouat. *Localization of Simultaneous Moving Sound Sources for Mobile Robot Using a Frequency-Domain Steered Beamformer Approach*. Proceedings of the IEEE International Conference on Robotics and Automation, vol. 1, pages 1033–1038, 2004.
- [Valin 2006] J.-M. Valin, F. Michaud & J. Rouat. *Robust 3D Localization and Tracking of Sound Sources Using Beamforming and Particle Filtering*. IEEE International Conference on Acoustics, Speech, and Signal Processing, April 2006.
- [Van Immerseel 2003] Luc Van Immerseel & Stefaan Peeters. *Digital implementation of linear gammatone filters : Comparison of design methods*. Acoustics Research Letters Online, vol. 4, n° 3, pages 59–64, July 2003.
- [Van Veen 1988] B.D. Van Veen & K.M. Buckley. *Beamforming : a versatile approach to spatial filtering*. IEEE ASSP Magazine, vol. 5, n° 2, pages 4–24, April 1988.
- [Wang 1985] H. Wang & M. Kaveh. *Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, pages 823–831, 1985.
- [Wang 2003] Fan Wang, Venkataramanan Balakrishnan, Philip Yuanping Zhou, Jason Jiangnan Chen, Roy Yang & Colin Franck. *Optimal Array Pattern*

- Synthesis Using Semidefinite Programming*. IEEE Transactions on Signal Processing, vol. 51, pages 1172–1183, 2003.
- [Wang 2004] Qing Hua Wang, Teodor Ivanov & Parham Aarabi. *Acoustic robot navigation using distributed microphone arrays*. Information Fusion, vol. 5, n° 2, pages 131–140, June 2004.
- [Ward 1996] Darren B. Ward, Rodney A. Kennedy & Robert C. Williamson. *FIR filter design for frequency invariant beamformers*. IEEE Signal Processing Letters, vol. 3, pages 69–71, March 1996.
- [Ward 1997] Darren B. Ward & Gary W Elko. *Mixed nearfield/farfield beamforming : a new technique for speech acquisition in a reverberant environment*. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 1997.
- [Ward 1998] Darren B. Ward, Zhi Ding & Rodney A. Kennedy. *Broadband DOA estimation using frequency invariant beamforming*. IEEE Transactions on Signal Processing, vol. 56, n° 5, pages 1463–1469, May 1998.
- [Ward 2003] D.B. Ward, E.A. Lehmann & R.C. Williamson. *Particle filtering algorithms for tracking an acoustic source in a reverberant environment*. IEEE Transactions on Speech and Audio Processing, vol. 11, n° 6, pages 826–836, Nov. 2003.
- [Ward 2004] Darren B. Ward & Thushara D. Abhayapala. *Range and bearing estimation of wideband sources using an orthogonal beamspace processing structure*. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pages 109–112, May 2004.
- [Wu 1996] Shao-Po Wu, Stephen Boyd & Lieven Vandenberghe. *FIR Filter Design via Semidefinite Programming and Spectral Factorization*. Proceedings of Decision and Control, vol. 1, pages 271–276, Dec. 1996.
- [Yamamoto 2004] Shunichi Yamamoto, Kazuhiro Nakadai, Hiroshi Tsujino, Toshio Yokoyama & Hiroshi G. Okuno. *Improvement of Robot Audition by Interfacing Sound Source Separation and Automatic Speech Recognition with Missing Feature Theory*. IEEE International Conference on Robotics and Automation, vol. 2, pages 1517–1523, May 2004.
- [Ziomek 1995] Lawrence J. Ziomek. *Acoustic field theory and space-time signal processing*. CRC Press, Inc., 1995.

Résumé

Le système auditif de l'homme fournit de nombreuses informations sur son environnement sonore. Nous sommes par exemple capables de localiser précisément l'origine d'un son et d'en interpréter sa signification, si bien qu'il paraît aujourd'hui extrêmement difficile de se passer de ces informations sonores dans un monde dynamique et imprévisible. Pourtant, la robotique mobile n'a que très peu intégré cette modalité auditive, bien qu'elle apparaisse indispensable pour compléter les informations délivrées par les autres capteurs extéroceptifs tels que les caméras, les télémètres lasers ou les détecteurs ultra-sonores.

Nous présentons dans cette thèse la conception d'un système auditif artificiel, composé de 8 microphones omnidirectionnels et d'une carte d'acquisition/traitement, pour la localisation de source sonore. Cette problématique a déjà été très largement traitée dans la littérature issue du Traitement du Signal et de l'Acoustique. Cependant, des contraintes inédites d'embarquabilité et de temps réel imposées par la robotique limitent l'applicabilité de ces méthodes pour des signaux large bande tels que la voix. Après une étude bibliographique approfondie sur les solutions de localisation déjà proposées en robotique, nous envisageons la définition de formations de voie invariantes en fréquence. Celles-ci sont synthétisées selon une nouvelle méthode d'optimisation convexe basée sur la représentation modale des diagrammes de directivité. Cette solution permet d'améliorer sensiblement la résolution des basses fréquences, aussi bien en champ proche qu'en champ lointain, malgré l'utilisation d'une antenne de microphones de petite taille. Nous exploitons ensuite ces nouvelles formations de voie optimisées pour le tracé d'une carte de puissance acoustique de l'environnement. Comparativement aux méthodes conventionnelles, la localisation s'avère plus précise et permet de traiter des signaux large bande sur la bande de fréquence 400Hz-3kHz. Nous évaluons enfin une extension récente de la méthode MUSIC dans l'espace des formations de voie pouvant être compatible avec les contraintes liées au contexte robotique.

Mots-clés: Formation de voie, optimisation convexe, localisation de source, robotique mobile.

Abstract

The auditory system provides the human with many informations on its acoustic environment. For instance, we are able to precisely localize the origin of a sound and interpret its meaning, so that it would be very difficult to do without these auditory cues in our dynamic and evolving world. Yet, mobile robotics has seldom integrated this auditory modality despite its compulsoriness in order to complete the informations delivered by other exteroceptive sensors such as cameras, laser range-finders or ultrasonic detectors.

This doctoral thesis deals with the design of an artificial auditory system for sound source localization, composed of an array of 8 omnidirectional microphones and an acquisition/processing board. This problematics has already been studied in Signal Processing and Acoustics. However, the unusual embeddability and real-time constraints imposed by the robotics context limit the applicability of such methods to wideband signals such as human voice. After a thorough bibliographical study of the approaches to localization proposed in Robotics, the synthesis of frequency-invariant beamformers is envisaged. An original solution is proposed, based on convex optimization and relying on the modal representation of antenna directivity patterns. Compared to classical methods, its resolution at low frequencies is enhanced, be the source in the farfield or in the nearfield, despite the small size of the array. These optimized beamformers are then exploited so as to compute an acoustic power map of the environment. The localization turns to be more precise than with conventional approaches, and can process wideband signals in the frequency range 400Hz-3kHz. Last, a recent beamspace extension of the high-resolution MUSIC method is assessed, which could cope with the robotics constraints.

Keywords: Beamforming, convex optimization, sound source localization, mobile robotics.

