

UNIVERSITE SAAD DAHLAB DE BLIDA

Faculté des Sciences de l'Ingénieur

Département d'Electronique

MEMOIRE DE MAGISTER

en Electronique

Spécialité : Signaux et Systèmes

REHAUSSEMENT DE LA PAROLE PAR LA TECHNIQUE DE LA FORMATION
DE VOIES DANS UN ENVIRONNEMENT REEL

Par

Fatima Zohra BOUCHIBANE

devant le jury composé de :

M. Djendi	Maitre de conférences, U.de Blida	Président
M. Lagha	Maitre de conférences, U. de Blida	Examineur
M. Ould Zmirli	Maitre de conférences, U. de Médéa	Examineur
M. Bensebti	Professeur, U. de Blida	Rapporteur

Blida, Juin 2011

RÉSUMÉ

Ce travail décrit la conception d'un système de rehaussement de la parole basé sur la séparation spatiale de sources, et ceci en utilisant un réseau de deux microphones. Les signaux sont acquis et traités via la carte son d'un micro ordinateur. L'algorithme de corrélation généralisé (PHAT-GCC : Phase Transform Generalized Cross Correlation) est utilisé pour déterminer les directions d'arrivées des sources, et le formateur de voies conventionnel est employé pour orienter le diagramme de directivité du réseau de microphones dans la direction souhaitée tout en essayant de minimiser l'énergie du signal interférent. Tous les traitements dans ce mémoire sont faits par ordinateur en utilisant un logiciel de calcul. L'algorithme PHAT-GCC est utile dans les applications temps réel et permet une bonne poursuite des sources mais, sa sensibilité aux multi-trajets et aux multiples sources dégrade ses performances. Le formateur de voies conventionnel permet la séparation spatiale d'un signal bande étroite du signal parole bande large et il travaille mieux dans le cas où les sources sont positionnées d'une manière orthogonale.

ABSTRACT

This work describes the design of a two channel microphone array system dedicated to speech enhancement based on spatial separation of sources. Microphones signals are acquired and processed by a computer sound card. The based correlation algorithm (PHAT-GCC: Phase Transform Generalized Cross Correlation) is used to determine the direction of arrival of the sources, and the conventional beamforming is used to steer the microphone array directivity pattern to the desired direction while trying to minimize the signal of interference source. All of the processing in this text is done using calculating software. The PHAT-GCC algorithm is useful for real time applications that allow good source tracking; however, its sensibility to multipath and multiples sources degrades its performance. The conventional beamforming is able to spatially separate a narrow band signal from the wideband speech signal and it performs better especially when the sources are orthogonally positioned.

ملخص

هذا العمل يصف تصميم نظام لتحسين الخطاب على أساس الفصل المكاني للمصادر وذلك باستخدام مجموعة من اثنين من الميكروفونات. يتم الحصول على الإشارات ومعالجتها باستخدام بطاقة الصوت من جهاز الكمبيوتر. يتم استخدام خوارزمية (PHAT-GCC) لتحديد جهات المصادر، وتقنية تكوين القنوات التقليدية في توجيه نمط التوجيه لمجموعة الميكروفونات نحو الاتجاه المطلوب مع محاولة التقليل من الطاقة القادمة من المصدر الغير مرغوب فيه. تتم جميع المعالجات في هذه الأطروحة عن طريق الكمبيوتر باستخدام برنامج حساب. الخوارزمية (PHAT-GCC) مفيدة في الوقت الحقيقي، تسمح بالتتبع المستمر لاماكن المصادر، ولكن حساسيتها للصدى وتعدد المصادر يحط من الأداء الجيد. المكون للقنوات يسمح بالفصل المكاني لإشارة ذات نطاق ترددي ضيق من إشارة الخطاب، أفضل أداء يكون في الحالة التي يتم فيها وضع المصادر بشكل متعامد.

REMERCIEMENTS

Je souhaiterais tout d'abord remercier les personnes qui ont voulu prendre part au jury de ce travail, en commençons par Monsieur M. Djendi, maitre de conférences au département d'électronique à l'université Saad Dahlab de Blida, pour m'avoir fait l'honneur de présider le jury. Je suis également très reconnaissante vers Monsieur Ould zmirli, maitre de conférences à l'université Yahia Fares de Médéa et Monsieur M. Lagha maitre de conférences au département d'aéronautique de l'université de Blida qui ont accepté d'être membres de jury et pour l'intérêt qu'ils ont porté à ce travail.

Je remercie infiniment M. Bensebti Messaoud, professeur à l'université de Blida pour sa qualité d'encadrement, son aspect pratique et ses précieux conseils.

Je remercie aussi ma famille, mes amis et les membres du laboratoire SET précisément l'équipe des télécommunications pour leurs supports.

TABLES DES MATIERES

RÉSUMÉ	2
REMERCIEMENTS	5
TABLES DES MATIERES	6
LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX	8
LISTE DES ABREVIATIONS	11
INTRODUCTION	12
LOCALISATION DE SOURCES	15
1.1. Introduction	15
1.2. Réseaux de microphones	15
1.3. Formulation du problème et modèles des signaux	17
1.4. Stratégies de localisation des sources acoustiques	22
1.5. Comment les être humains localisent le son	26
1.6. La performance de la localisation	28
1.7. Ambigüité spatiale	28
1.8. Conclusion	29
FORMATION DE VOIES	30
2.1. Introduction	30
2.2. Formation de voies (ou FdV)	30
2.3. Types de formateur de voies	31
2.4. Structures des formateurs de voies numériques	32
2.5. Classes des formateurs de voies numériques	34
2.6. Suppression du signal avec les formateurs de voies optimaux	42
2.7. Problème des signaux large bande	43
2.8. Concept de base de la formation de voies	44
2.9. Conclusion	46
REALISATIONS ET RESULTATS	47
3.1. Introduction	47
3.2. Description du système	47

3.3. Localisation de sources	48
3.4. Formation de voies	77
3.5. Combinaison des deux systèmes	87
3.6. Conclusion	91
CONCLUSION ET PERSPECTIVES	92
REFERENCES	96

LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX

Figure 1.1 : Différentes configurations géométriques des réseaux de microphones ; (a) linéaire, (b) circulaire, (c) planaire.	16
Figure 1.2 : Modèle de canal d'une seule source et espace libre.	19
Figure 1.3 : Modèle de canal d'une seule source et un environnement réverbérant.	21
Figure 1.4 : Les stratégies de la localisation.	23
Figure 1.5 : Deux paires de sinusoides avec une différence de phase différente semblent identiques.	29
Figure 2.1 : Utilisation d'un réseau de microphone pour le filtrage spatial.	31
Figure 2.2 : Structures des formateurs de voies: (a) modèle bande étroite, (b) modèle bande large.	33
Figure 2.3 : Principe du MSC; (a) structure, (b) les réponses du canal principal, canaux auxiliaires et du système complet.	37
Figure 2.4 : Formateur de voies bande étroite en présence de deux sources.	38
Figure 2.5 : Schéma descriptif du GSC.	39
Figure 2.6 : Formateur de voies large bande; (a) domaine fréquentiel, (b) domaine temporel.	44
Figure 2.7 : Suppression d'interférence par un réseau de deux microphones.	46
Figure 3.1 : Description globale du système.	48
Figure 3.2 : Principe de la localisation des sources par l'estimation du retard temporel.	49
Figure 3.3 : La résolution spatiale versus la fréquence d'échantillonnage.	51
Figure 3.4 : Chaîne de localisation du signal parole par l'algorithme PHAT-GCC dans un environnement de simulation idéal.	52
Figure 3.5 : Performance de l'algorithme PHAT-GCC en fonction du type de la fenêtre utilisée.	54
Figure 3.6 : Spectres d'amplitude des fenêtres (Hamming, Hanning, Blackman et Chebyshev).	55
Figure 3.7 : Résultats de la localisation du locuteur dans un environnement idéal.	56
Figure 3.8 : Configuration du matériel pour la localisation d'une source de parole dans un environnement acoustique réel.	57
Figure 3.9 : Résultat de la localisation de la parole dans un environnement réel.	57
Figure 3.10 : Deux échantillons de la fonction PHAT-GCC prise à des instants différents.	58
Figure 3.11 : Résultat de la localisation du locuteur dans un environnement réel.	58
Figure 3.12 : Fonction d'auto-corrélation d'un bruit blanc centré.	59
Figure 3.13 : Configuration utilisée pour la localisation d'une source de bruit blanc.	60
Figure 3.14 : Un échantillon de la fonction PHAT-GCC pour la localisation d'un bruit blanc.	60
Figure 3.15 : Résultat de la localisation d'une source de bruit blanc.	61

Figure 3.16 : Configuration utilisée pour la localisation d'une source de signal périodique.	61
Figure 3.17 : Un échantillon de la figure PHAT-GCC pour la localisation d'un signal périodique.	62
Figure 3.18 : Résultat de la localisation d'un signal périodique.	63
Figure 3.19 : Un échantillon de la fonction PHAT-GCC pris pendant une période du non activité vocale.	64
Figure 3.20 : Résultat de la localisation pendant les périodes du non activité vocale.	64
Figure 3.21 : Histogramme montre la distribution des directions d'arrivée dans le cas d'absence des sources audio.	65
Figure 3.22 : Configuration du matériel pour caractériser l'effet des multi-trajets.	66
Figure 3.23 : Un échantillon de la fonction PHAT-GCC pour la localisation d'une source de bruit blanc en présence des multi-trajets.	66
Figure 3.24 : Résultat de la localisation en présence des trajets multiples.	67
Figure 3.25 : Configuration utilisée pour caractériser l'effet de multiples sources sur le PHAT-GCC.	67
Figure 3.26 : Un échantillon de la fonction PHAT-GCC pour la localisation de multiples sources.	68
Figure 3.27 : Résultat de la localisation en présence des sources multiples.	68
Figure 3.28 : Effet du seuil.	69
Figure 3.29 : Histogramme montre l'amélioration de la performance du PHAT-GCC.	70
Figure 3.30 : Chaîne de la localisation du locuteur sous l'environnement simulink.	71
Figure 3.31 : Configuration du système pour la localisation d'une source de parole en présence d'un bruit blanc.	72
Figure 3.32 : Résultats de la localisation d'une source de la parole en présence d'un bruit blanc.	72
Figure 3.33 : Erreur de localisation en présence d'un bruit blanc gaussien.	73
Figure 3.34 : Configuration du système pour la localisation d'une source de la parole en présence d'un bruit harmonique.	74
Figure 3.35 : Résultats de la localisation de la parole en présence d'un bruit périodique.	74
Figure 3.36 : Erreur de localisation en présence d'un bruit périodique.	75
Figure 3.37 : Diagramme de directivité ;(a) un seul microphone, (b) un réseau de microphones.	77
Figure 3.38 : Configuration utilisée pour le tracé du diagramme de directivité d'un réseau de deux microphones.	78
Figure 3.39 : Diagramme de directivité d'un réseau à deux microphones, (a) $D=\lambda/4$, (b) $D=\lambda/2$, (c) $D=\lambda$, (d) $D=3\lambda$.	79
Figure 3.40 : Description de la configuration utilisée pour caractériser l'effet de la distance intra-microphones.	82
Figure 3.41 : La performance d'un FdV à bande étroite en fonction de la distance intra-microphones.	82
Figure 3.42 : La performance d'un FdV à bande étroite en fonction de la séparation angulaire.	83
Figure 3.43 : Configuration utilisée pour la séparation de deux sources à bande étroite.	84
Figure 3.44 : Résultats de la séparation des deux signaux sinusoïdaux.	85

Figure 3.45 : Configuration du système pour la séparation d'un signal large bande et un signal bande étroite.	86
Figure 3.46 : Résultats de la séparation de deux sources ; une source de parole et une source d'un signal périodique.	86
Figure 3.47 : Principe du rehaussement de la parole par la méthode de formation de voies.	87
Figure 3.48 : La conception complète du système sous Matlab.	88
Figure 3.49 : Spectrogrammes du signal original, bruité et du signal débruité.	89
Figure 3.50 : Diagramme de directivité d'un réseau de deux microphones.	90
Tableau 1.1 : Fonctions de pondération.	26
Tableau 3.1 : L'algorithme PHAT-GCC.	52
Tableau 3.2 : La résolution dans le cas : $D=20\text{cm}$, $F_s=22050\text{ Hz}$.	56
Tableau 3.3 : Paramètres du système de la localisation.	71
Tableau 3.4 : Paramètres choisis pour le système complet.	87

LISTE DES ABREVIATIONS

ANC	: Adaptive Noise Canceler.
CPSP	: Cross-Power Spectrum Phase.
DSP	: Digital Signal Processor.
DOA	: Direction Of Arrivals.
DTI	: Différence de Temps Interauriculaires.
DNI	: Différence de Niveaux Interauriculaires.
FdV	: Formation de Voies.
GCC	: Generalized Cross Correlation.
GSC	: Generalized Sidelobes Canceler.
HRTF	: Head Related Transfer Function.
IP	: Internet Protocol.
LCMV	: Linearly Constrained Minimum Variance.
MIMO	: Multiple Input Multiple Output.
MSC	: Multiple Sidelobes Canceler.
MVDR	: Minimum Variance Distortionless Response.
PHAT	: PHASE Transform.
RIF	: Réponse Impulsionnelle Finie.
RF	: Radio Frequency.
RSB	: Rapport Signal sur Bruit.
RSIB	: Rapport Signal sur Bruit plus Interférence.
SIMO	: Single Input Multiple Output.
SRP	: Steered Response Power.
SVD	: Singular Value Decomposition.
SCOT	: Smoothed COherence Transform.
TDOA	: Time Difference Of Arrivals.
ULA	: Uniform Linear Array.
UCA	: Uniform Circular Array.
VOIP	: Voice Over IP.

INTRODUCTION

Le bruit est partout et dans la plupart des applications qui sont liées à l'acoustique et à la parole, tel que les interfaces homme-machine, les communications mains libres, la voix à travers IP (VOIP), les prothèses auditives, les systèmes de téléconférence et télécollaboration, et tant d'autres. Le signal d'intérêt (habituellement la parole) qui est capté par un microphone est généralement contaminé par le bruit. Par conséquent, le signal issu du microphone doit être nettoyé par des outils du traitement numérique du signal avant qu'il soit stocké, analysé ou transmis. Ce processus de nettoyage s'appelle souvent le rehaussement de la parole. Ce domaine a et continue d'attiré l'attention des chercheurs pendant plusieurs décennies.

Le bruit peut être défini en tant que n'importe quel signal non désiré perturbant le signal portant l'information utile. Il est présent avec des degrés différents dans presque tous les environnements [1]. Il peut causer des erreurs de transmission et peut même perturber un processus de communication; par conséquent le traitement du bruit est une partie importante et intégrale dans les systèmes de télécommunications modernes et les systèmes de traitement du signal.

Le but d'un algorithme de rehaussement de la parole est d'estimer le signal de parole à partir de sa version corrompue par un bruit. Avec un seul microphone, les algorithmes de rehaussement de la parole se fondent seulement sur les informations temporelles ou spectrales contenues dans le signal d'entrée. Ils peuvent être efficaces si le contenu spectral du bruit change lentement relativement au signal mais ils rapportent généralement une déformation du signal désiré [2]-[3]. Avec un réseau de microphones le bruit peut être réduit sans l'affectation de l'intelligibilité de la parole [4].

Dans les communications mains libres, le couplage acoustique entre les microphones et les hauts parleurs produit des échos ce qui rend les

communications en temps réel difficiles. Les réseaux de microphones sont vus comme un moyen très efficace pour réduire le niveau des échos en dirigeant le lobe principal du réseau vers la source d'intérêt et les zéros vers les hauts parleurs [5].

Dans les salles, et dans le contexte de mains libres, le microphone ne capte pas uniquement le trajet direct du signal voulu mais aussi ses versions retardées et atténuées dues aux réflexions. Ce phénomène des multi-trajets introduisant des distorsions spectrales dans le signal observé est appelé réverbération. Des grands efforts ont été effectués durant les quatre dernières décennies en vue de trouver des solutions pratiques à ce problème en exploitant bien sûr les réseaux de microphones.

Dans des environnements acoustiques, l'information sur la position des sources joue un rôle très important dans des applications telles que la poursuite automatique du locuteur par une caméra pour la vidéoconférence et la formation de voies. L'estimation des positions des sources, souvent appelé localisation de sources a eu un considérable intérêt dans la recherche pendant des décennies. Un réseau de deux ou trois microphones est utilisé pour estimer les angles d'arrivées des sources. Pour l'estimation du nombre des sources et leurs localisations, plusieurs algorithmes intéressants existent pour le cas des sources à bande étroite où la recherche a déjà commencée pour résoudre le problème dans le cas des sources large bande.

La localisation d'une source acoustique est un procédé bien développé du système auditif humain. Utilisant seulement deux capteurs, ce système biologique a une remarquable précision en résolvant la position des locuteurs et d'autres sources acoustiques. Avec deux oreilles et un traitement fait par le cerveau, l'être humain a la capacité de résoudre le problème du « cocktail party ». Les chercheurs n'ont pas encore une idée claire de ce genre de traitement et comment exploiter les réseaux de microphones pour pallier à ce problème.

Malgré le nombre d'applications énorme des réseaux de microphones, le marché est encore en attente de bonnes solutions pratiques.

Dans ce mémoire, l'idée traitée consiste à employer un réseau de deux microphones et un processeur DSP par analogie avec le système auditif humain pour localiser des sources acoustiques. Ensuite exploiter cette information pour séparer spatialement la source d'intérêt parmi les sources interférentes. L'acquisition des signaux et leurs prétraitement est réalisé en temps réel grâce à la carte son d'un micro-ordinateur. Ceci est pour explorer les possibilités de ce moyen beaucoup moins coûteux que d'autres systèmes d'acquisition de données dans le marché.

Organisation du mémoire

Ce mémoire comporte trois chapitres :

Le premier chapitre est dédié à la localisation des sources audio; le principe et les différents modèles du canal acoustique sont décrits. Les stratégies de la localisation existantes sont traitées globalement tout en détaillant la méthode utilisée.

La technique de séparation de sources dite formation de voies est considérée dans le deuxième chapitre ; son principe, ses structures ainsi que les méthodes souvent utilisés sont détaillées.

Dans le troisième chapitre, les expériences réalisées avec les résultats obtenus suivis des interprétations sont présentées.

Enfin, ce mémoire sera conclut par une synthèse des résultats obtenus ainsi que par une revue des multiples perspectives futures.

CHAPITRE 1

LOCALISATION DE SOURCES

1.1. Introduction

La localisation d'une source acoustique est un procédé bien développé du système auditif humain. Utilisant seulement deux oreilles, ce système biologique a une remarquable précision en résolvant la position des locuteurs et d'autres sources acoustiques. Les oreilles humaines avec le cerveau peuvent exactement localiser et poursuivre les sources.

La localisation du locuteur et la poursuite automatique dans un environnement réverbérant sont des tâches provocantes et souvent nécessaires dans des applications audio telles que le téléphone mobile à mains libres, la reconnaissance de la parole et les systèmes de téléconférences.

Dans ce chapitre, des algorithmes du traitement de signal dédiés à la localisation des sources audio sont présentés. Des techniques de haute qualité basées sur l'estimation du retard temporel en utilisant que deux microphones sont discutées. Ces algorithmes peuvent être utilisés pour l'estimation de l'angle d'arrivée de l'onde sonore traversant un réseau de microphone unidimensionnel.

1.2. Réseaux de microphones

Un réseau de microphones consiste de deux ou plusieurs microphones positionnés de telle manière à profiter au maximum de l'information spatiale. Les réseaux de microphones peuvent avoir différentes géométries voir figure (1.1) [6].

1.2.1. Réseaux linéaires

La structure la plus étudiée et la plus souvent utilisée est le réseau linéaire. Il est constitué d'un certain nombre de microphones répartis sur une droite, si les éléments adjacents sont en plus équidistants on dit que le réseau est linéaire et

uniforme (ULA uniform linear array) voir figure (1.1.a). Le microphone de référence peut être placé à l'une des extrémités du réseau ou au centre du réseau.

1.2.2. Réseaux circulaires

Si les éléments sont arrangés d'une manière circulaire comme c'est illustré dans la figure (1.1.b), alors ce réseau est connu comme un réseau circulaire uniforme (UCA uniform circular array). La référence de phase est prise dans le centre de cercle et l'angle d'incidence est mesuré par rapport à l'axe qui réunit le centre du cercle avec le microphone de référence. Avec le même nombre d'éléments et la même séparation, le réseau circulaire produit un faisceau plus large que celui correspondant au réseau linéaire. Mais il est plus performant que le réseau linéaire en termes de diversité à la réception.

1.2.3. Réseaux planaires

Les réseaux présentés auparavant permettent de résoudre seulement la composante azimutale des directions d'arrivée. Un autre type de réseau, largement utilisé en pratique, permettant d'utiliser les deux composantes des directions d'arrivée est le réseau de type planaire ou rectangulaire.

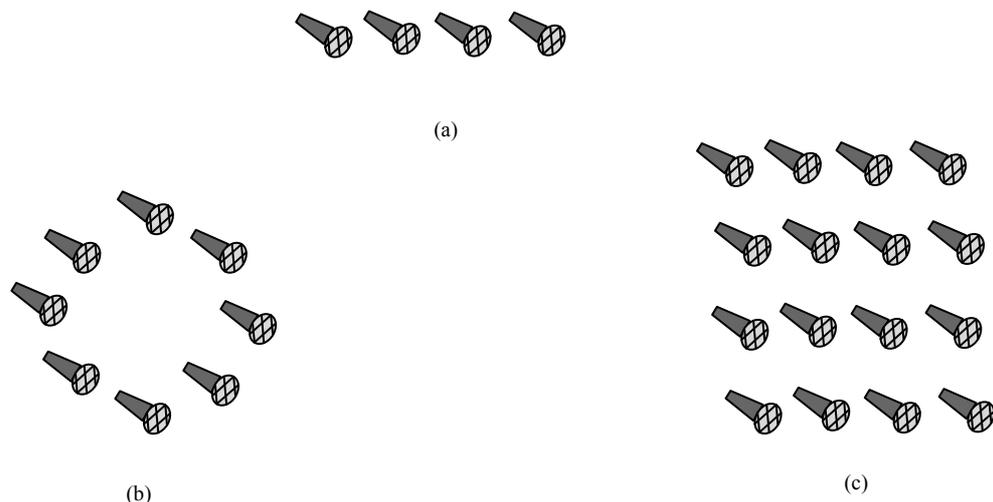


Figure 1.1 : Différentes configurations géométriques des réseaux de microphones ; (a) linéaire, (b) circulaire, (c) planaire.

1.2.4. Réseaux conformes

Les réseaux de microphones décrits auparavant sont basés sur un positionnement régulier et symétrique. Ceci n'est pas toujours convenable dans des situations où on aura besoin d'autres structures (elliptiques, triangulaires, cubiques...) conformes aux besoins.

1.3. Formulation du problème et modèles des signaux

Le problème d'estimation du DOA (Direction Of Arrivals) concerne la mesure de l'angle d'incidence du signal source au niveau des microphones. Selon l'environnement acoustique, deux situations peuvent être distinguées : l'environnement libre (anéchoïque) où les microphones reçoivent que le trajet direct et l'environnement réverbérant où chaque microphone reçoit plusieurs copies réfléchies et atténués en plus du trajet direct. Pour chaque situation, on différencie le cas d'une seule source du cas de multiples sources par ce que la complexité d'estimation n'est pas nécessairement la même dans les deux conditions. Alors, en total, quatre modèles de signaux peuvent être considérés [4].

On considère un réseau rectiligne composé de N microphones identiques, équidistants, omnidirectionnels et de gain unitaire. La distance entre les microphones est inférieure ou égale à $\lambda/2$, où λ est la longueur d'onde correspondante à la fréquence f des sources, le cas des sources à bande étroite est considéré. Ce réseau reçoit M ($M < N$) signaux émis par des sources lointaines auxquelles se superpose un bruit additif.

On notera:

$$s(k) = [s_1(k)s_2(k) \dots s_M(k)]^T, \quad (1.1)$$

Le vecteur des signaux émis par les M sources.

$$\text{et } b(k) = [b_1(k)b_2(k) \dots b_N(k)]^T, \quad (1.2)$$

Le vecteur des bruits reçus sur les N microphones.

Le vecteur des signaux reçus sur les N microphones s'écrit :

$$y(k) = [y_1(k)y_2(k) \dots y_N(k)]^T, \quad (1.3)$$

Ainsi, le signal reçu par le $i^{\text{ème}}$ microphone est donné par :

$$y_i(k) = \sum_{m=1}^M s_m(k - \tau_{m,k}) + b_i(k), \quad (1.4)$$

Où $\tau_{m,k} = (i - 1)\tau_k$ est le retard temporel entre le premier et le $i^{\text{ème}}$ microphone. Le premier microphone étant pris comme référence.

En utilisant la notation matricielle, le vecteur des signaux reçus s'écrit dans le domaine fréquentiel,

$$Y(f) = G(\theta, f)S(f) + B(f), \quad (1.5)$$

Où $S(f)$ est le vecteur signal, donné par :

$$S(f) = \begin{bmatrix} S_1(f) \\ S_2(f) \\ \vdots \\ S(f) \end{bmatrix}, \quad (1.6)$$

$B(f)$ est le vecteur bruit reçu.

$G(\theta, f)$ est appelée matrice de transfert sources-réseau et contient les informations sur les directions d'arrivée des M sources.

On définit la matrice inter-spectrale, des signaux reçus à la fréquence f , par :

$$R_{YY}(f) = E[Y(f)Y^T(f)] \quad (1.7)$$

Cette matrice est estimée par :

$$\hat{R}_{YY}(f) = \frac{1}{L_r} \sum_{l=1}^{L_r} Y(f)Y^T(f), \quad (1.8)$$

Où L_r représente le nombre de réalisations.

En présence d'un bruit non corrélé avec les signaux sources, la matrice $R_{YY}(f)$ s'écrit :

$$R_{YY}(f) = G(\theta, f)R_{SS}(f)G^T(\theta, f) + R_{BB}(f), \quad (1.9)$$

Où $R_{SS}(f)$ est la matrice interspectrale des signaux sources donnée par :

$$R_{SS}(f) = E[S(f)S^T(f)] \quad (1.10)$$

Et $R_{BB}(f)$ est la matrice inter-spectrale des bruits donnée par :

$$R_{BB}(f) = E[B(f)B^T(f)] \quad (1.11)$$

Si on suppose le bruit blanc et de variance σ^2 la matrice interspectrale du bruit devient :

$$R_{BB}(f) = \sigma^2 I_N \quad (1.12)$$

Avec , I_N la matrice identité de dimension $N \times N$.

1.3.1. Modèle d'une seule source et un espace libre

Supposons qu'il n'y a qu'une seule source dans un espace libre ou un réseau de N microphones est considéré. La puissance du signal parole $s(k)$ décroît en fonction de la distance séparant la source du réseau. Si le premier microphone est choisi comme un point de référence, le signal capté par le $n^{\text{ième}}$ microphone à l'instant k peut être exprimé ainsi :

$$\begin{aligned} y_n(k) &= \alpha_n s(k - t - \tau_{n1}) + b_n(k) \\ &= \alpha_n s[(k - t - \mathcal{F}_{n1}(\tau))] + b_n(k) \\ &= x_n(k) + b_n(k), n = 1, 2, \dots, N. \end{aligned} \quad (1.13)$$

Où α_n ($n = 1, 2, \dots, N$), est l'atténuation due à l'effet de la propagation, $s(k)$ est le signal source inconnu, t est le temps de propagation de la source vers le premier capteur, $b_n(k)$ est le bruit additif au capteur n , assumé d'être non corrélé avec le signal source et le bruit au niveau des autres capteurs, τ est le TDOA (appelé retard relatif) entre le capteur 1 et 2, et $\tau_{n1} = \mathcal{F}_n(\tau)$ est le TDOA entre le microphone 1 et le microphone n avec $\mathcal{F}_1(\tau) = 0$ et $\mathcal{F}_2(\tau) = \tau$. Pour $n = 3, \dots, N$, la fonction \mathcal{F}_n ne dépend pas uniquement de τ mais aussi de la géométrie du réseau. Par exemple pour un réseau linéaire et équidistant on a :

$$\mathcal{F}_n(\tau) = (n - 1)\tau, n = 2, \dots, N, \quad (1.14)$$

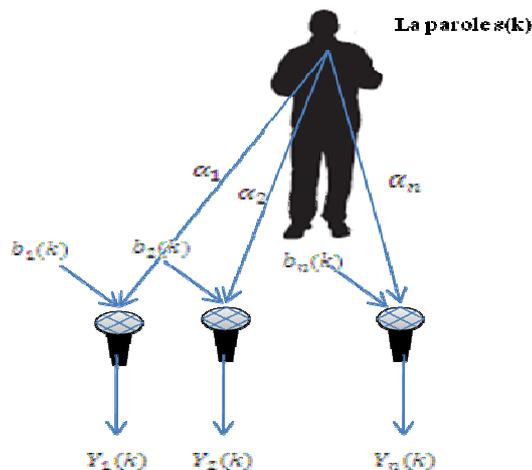


Figure 1.2 : Modèle de canal d'une seule source et espace libre.

Dans le cas du champ proche, \mathcal{F}_n dépend aussi de la position de la source. Notons que $\mathcal{F}_n(\tau)$ peut être une fonction non linéaire de τ pour un réseau non linéaire même dans le cas du champ lointain.

1.3.2. Modèle de sources multiples et espace libre

Encore dans l'espace libre, s'il ya plusieurs sources dans le champ acoustique, le signal reçu au niveau du n^{ieme} microphone devient :

$$y_n(k) = \sum_{m=1}^M \alpha_{nm} s_m[k - t_m - \mathcal{F}_n(\tau_m)] + b_n(k) = x_n(k) + b_n(k), n = 1, 2, \dots, N \quad (1.15)$$

Où M est le nombre de source, α_{nm} ($n = 1, 2, \dots, N, m = 1, 2, \dots, M$) sont les facteurs d'atténuations dues à l'effet de la propagation, $s_m(k)$ sont les signaux sources inconnus assumés être indépendants mutuellement, t_m est le temps de propagation de la source inconnue vers le microphone de référence, $b_n(k)$ est le bruit additif au niveau du n^{ieme} microphone qui est assumé être non corrélé non seulement avec les signaux sources mais aussi avec chaque bruit perçu au niveau des capteurs, τ_m est le TDOA entre le microphone 2 et le microphone 1 dû à la m^{ieme} source, et $\mathcal{F}_n(\tau_m)$ est le TDOA entre le microphone n et 1 pour la source m . Pour ce modèle, l'objectif de l'estimation du TDOA est de déterminer tous les paramètres τ_m en utilisant les observations au niveau des microphones.

1.3.3. Modèle d'une seule source et espace réverbérant

Malgré la simplicité du modèle espace libre, ce modèle ne prend pas en considération l'effet des multi trajets ce qui n'est pas adéquat pour caractériser le canal réel. On a besoin alors à un autre modèle qui décrit mieux ce phénomène de propagation. La réponse impulsionnelle acoustique est généralement modélisée par un filtre à réponse impulsionnelle finie (RIF). S'il ya qu'une seule source comme c'est illustrée dans la figure (1.3), le problème peut être modélisé comme un système (Single Input Multiple Output) SIMO et le signal reçu au niveau du n^{ieme} microphone peut être modélisé ainsi :

$$y_n(k) = g_n * s(k) + b_n(k) = x_n(k) + b_n(k), n = 1, 2, \dots, N \quad (1.16)$$

Avec g_n est la réponse impulsionnelle du canal acoustique entre la source et le microphone n . Sous forme vecteur/matrice l'équation précédente s'écrit comme suit :

$$Y_n(k) = G_n s(k) + b_n(k), n = 1, 2, \dots, N, \quad (1.17)$$

Où

$$Y_n(k) = [y_n(k) \dots y_n(k - L + 1)]^T, \quad (1.18)$$

$$G_n = \begin{bmatrix} g_{n,0} & \dots & g_{n,L-1} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & g_{n,0} & \dots & g_{n,L-1} \end{bmatrix} \quad (1.19)$$

$$s(k) = [s(k) s(k - 1) \dots s(k - L + 1) \dots s(k - 2L + 2)]^T, \quad (1.20)$$

$$b_n(k) = [b_n(k) \dots b_n(k - L + 1)]^T, \quad (1.21)$$

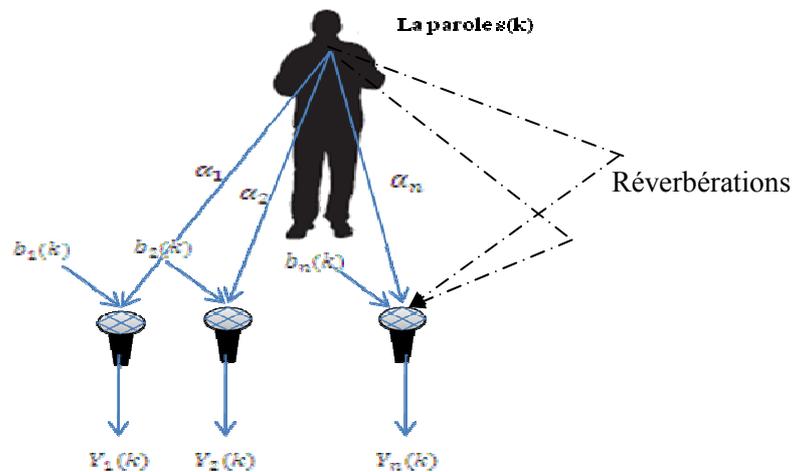


Figure 1.3 : Modèle de canal d'une seule source et un environnement réverbérant.

Où L est la longueur du canal acoustique le plus long dans le système SIMO. En comparaison avec le modèle d'espace libre, le TDOA τ peut être obtenu après l'identification aveugle du système SIMO, qui est vu comme un problème compliqué.

1.3.4. Modèle de sources multiples et environnement réverbérant

S'il ya plusieurs sources, le système peut être modélisé par un système multiple input multiple output (MIMO) avec M entrées et N sorties.

A un instant k on a :

$$y(k) = G s_{ML}(k) + b(k), \quad (1.22)$$

$$y(k) = [y_1(k) y_2(k) \dots y_N(k)]^T, \quad (1.23)$$

$$G = [G_1 G_2 \dots G_M], \quad (1.24)$$

$$G_m = \begin{bmatrix} g_{1m,0} & g_{1m,1} & \cdots & g_{1m,L-1} \\ g_{2m,0} & g_{2m,1} & \cdots & g_{2m,L-1} \\ \vdots & \vdots & \ddots & \vdots \\ g_{Nm,0} & g_{Nm,1} & & g_{Nm,L-1} \end{bmatrix}_{N \times L}, \quad (1.25)$$

$$m = 1, 2, \dots, M,$$

$$b(k) = [b_1(k) b_2(k) \dots b_N(k)]^T, \quad (1.27)$$

$$s_{ML}(k) = [s_1^T(k) s_2^T(k) \dots s_M^T(k)]^T, \quad (1.28)$$

$$s_m(k) = [s_m(k) s_m(k-1) \dots s_m(k-L+1)]^T \quad (1.29)$$

Où g_{nm} ($n = 1, 2, \dots, N, m = 1, 2, \dots, M$) est la réponse impulsionnelle de la source m vers le microphone n , tous les signaux sources sont mutuellement dépendants. Pour ce système et en vue d'estimer le TDOA, l'estimation aveugle du système MIMO doit être d'abord effectuée ce qui est un problème extrêmement difficile.

1.4. Stratégies de localisation des sources acoustiques

Les procédures de localisation des sources peuvent être classifiées en trois grandes catégories [7] voir figure (1.4) : celles basées sur la maximisation de la réponse d'un formateur de voies (SRP : Steered Response Power), techniques adoptants les concepts d'estimation à haute résolution spectrale et les approches exploitantes le retard temporel (TDOA). Ces grandes classes diffèrent selon l'environnement d'application et la méthode d'estimation. Dans la première classe, l'estimé de la position est obtenu directement après le formateur de voies retard et somme des données reçues au niveau des capteurs. La deuxième est basée sur le calcul de la matrice de corrélation et la dernière catégorie cherche à transformer le retard temporel entre les éléments du réseau en positions.

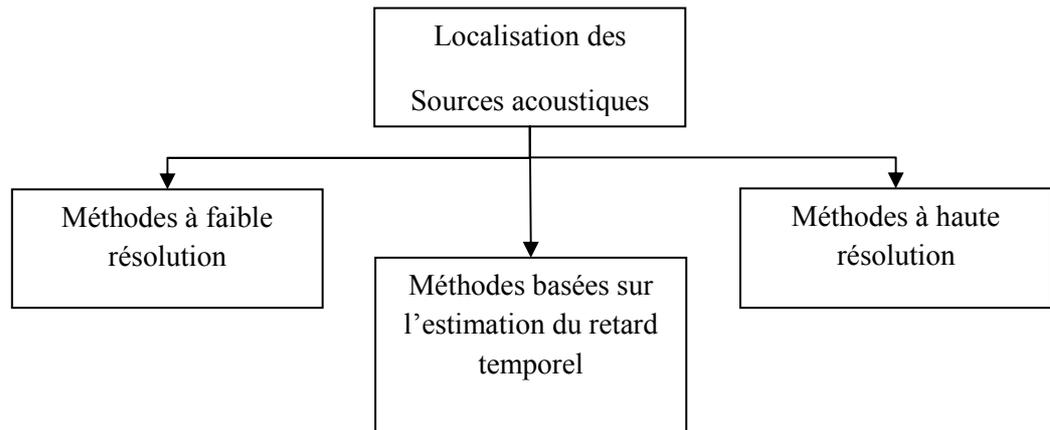


Figure 1.4 : Les stratégies de la localisation.

1.4.1. Méthode à faible résolution (formation de voies)

La formation de voies est parmi les premières méthodes développées pour la localisation des sources rayonnantes. Elle est basée sur le calcul de l'énergie reçue dans une direction privilégiée en réalisant des rotations virtuelles de l'antenne. Cette méthode est dite à faible résolution car sa résolution spatiale dépend de la largeur du diagramme de réception de l'antenne qui est liée à son tour à la longueur de l'antenne.

Le type le plus simple du SRP (Steered Response Power) est obtenu en utilisant la sortie d'un formateur de voies (retard et somme). Ce type de formateur consiste à appliquer des retards sur les signaux reçus de réseaux en vue de compenser le retard de la propagation. Les signaux alignés temporellement et sommés forment un seul signal à la sortie. Cette méthode est utilisée dans le cas d'un réseau de plus de deux microphones [8].

Le balayage de l'environnement acoustique se fait de $-\pi/2$ à $\pi/2$. La direction d'arrivée (DOA) de chaque source correspond alors aux directions montrant une puissance élevée à la sortie du formateur de voies. Cette méthode est simple à implémenter et moins coûteuse en terme de temps de calcul. La performance de cette méthode dépend aussi du contenu spectral du signal source qui est pratiquement inconnu [9].

2.4.2. Méthode à haute résolution (approches des sous-espaces)

Ces méthodes exploitent les statistiques d'ordre deux des signaux enregistrés. L'idée de base de ces méthodes est la décomposition de l'espace d'observation en deux sous espaces complémentaires : le sous-espace signal et le sous-espace bruit. C'est Pisarenko qui a eu l'idée en 1973 de faire une SVD (singular value decomposition) sur la matrice d'auto-corrélation du signal, et d'exploiter les données, en montrant que les plus petites valeurs propres correspondaient au bruit, ce qui permettait de faire une décomposition de l'espace et de déduire les angles en projetant une famille de vecteurs paramétrés par leur direction d'arrivée θ , pour θ variant de -90° à $+90^\circ$, sur le sous espace bruit, on ne retient que ceux dont la norme projetée constitue un minimum relatif.

Contrairement à la méthode de formation de voies, ces méthodes utilisent la propriété d'orthogonalité entre les deux sous espaces pour localiser les sources, ce qui les rend indépendantes des paramètres physiques associés à l'expérimentation. Avec ces méthodes, les sources très proches peuvent être distinguées grâce aux pics étroits correspondant à chaque source d'où le nom haute résolution [10].

1.4.3. Méthode basée sur des différences de temps d'arrivée (TDOA)

Dans une première étape un ou plusieurs délais entre différentes paires de microphones (c-à-d le TDOAs) sont estimés. En utilisant ces évaluations de TDOA est la géométrie du réseau de microphone, la position des sources peut être calculée dans une deuxième étape. Cette catégorie inclut les méthodes de l'inter-corrélation généralisées conceptuellement simples et largement répandues GCC (Generalized Cross Correlation).

1.4.3.1. Méthode d'inter-corrélation dite classique

La méthode classique consiste à calculer la fonction d'inter-corrélation $R_{y_1 y_2}(\tau)$ exprimée par :

$$R_{y_1 y_2}(\tau) = E[y_1(t)y_2(t - \tau)] \quad (1.30)$$

L'inter-corrélation représente, pour une différence de temps de propagation donnée, la similarité entre les signaux. En principe, la fonction d'inter-corrélation présente un maximum en τ , correspondant à la direction de l'onde incidente. En considérant les transformées de Fourier $Y_i(f)$ et $Y_j(f)$ des signaux $y_i(t)$ et $y_j(t)$, l'inter-corrélation peut être calculée en prenant la transformée de Fourier inverse de la densité inter-spectrale de puissance $G_{ij}(f)$. Cette relation est plus connue sous le nom de théorème de Wiener-Kinchine.

$$R_{y_1 y_2}(\tau) = \int_{-\infty}^{+\infty} G_{y_1 y_2}(f) e^{j2\pi f \tau} df \quad (1.31)$$

$$G_{y_1 y_2}(f) = Y_1(f) Y_2(f)^* \quad (1.32)$$

Le calcul de l'équation (1.30) nécessiterait l'intégration du signal sur un temps infiniment long. Cependant une estimation de l'inter-corrélation peut être calculée sur une fenêtre temporelle finie de longueur T . Le choix de la durée d'observation est un des paramètres importants de la localisation. La valeur optimale est déterminée en fonction des trajectoires de la source, des conditions d'expérimentation (rapport signal sur bruit, réverbération) et l'unité de calcul.

1.4.3.2. Méthodes d'inter-corrélation dites généralisées

Les méthodes généralisées se différencient de la méthode classique par le filtrage des signaux en amont de l'inter-corrélation (aussi appelé blanchiment). Ces filtres visent à amplifier les bandes de fréquences présentant un haut rapport Signal sur Bruit et diminuer celles où le bruit est dominant.

L'opération est effectuée via un filtre de pondération Ψ_g , appliqué à la densité inter spectrale.

$$R_j(\tau) = \int_{-\infty}^{+\infty} \Psi_g(f) G_j(f) e^{j2\pi f \tau} df \quad (1.33)$$

La méthode de la phase (PHAT Transform [11]) utilise la pondération:

$$\Psi_s(f) = \frac{1}{|G_j(f)|} \quad (1.34)$$

Le développement des équations (1.33) et (1.34), sous l'hypothèse que les bruits b_i sont non corrélés, montre que l'inter-corrélation se réduit à des diracs centrés sur les délais recherchés et ce, quelques soient les caractéristiques des signaux.

En pratique, la densité inter-spectrale est normalisée par les densités spectrales de puissance de chacun des signaux, où seule l'information de la phase est conservée. Une version modifiée de cet estimateur a été proposée par Rabinkin sous le nom de *CPSP modifié* (Modified Cross-Power Spectrum Phase)

$$\Psi_{CPSPm}(f) = \frac{1}{(|G_i(f)||G_j(f)|)^\rho} \quad (1.35)$$

Déterminé expérimentalement, le coefficient ρ permet de pondérer, de manière adéquate, la répartition de l'information entre l'amplitude et la phase. Lorsque ρ est nul, l'estimateur n'est autre que l'intercorrélacion classique, lorsqu'il vaut l'unité, celui de la phase.

Les différentes fonctions poids sont résumées dans le tableau (1.1).

Fonction de pondération	Approche
1	Intercorrélacion classique (CC)
$\frac{1}{\sqrt{G_{y_1y_1}(f)G_{y_2y_2}(f)}}$	Smoothed Coherence Transform (SCOT)
$\frac{1}{G_{y_1y_1}(f)}$	Roth (Wiener-Hopf weighting)
$\frac{1}{ G_{y_1y_2}(f) }$	PHAT

Tableau 1.1 : Fonctions de pondération.

La différence entre estimer le retard temporel par le calcul direct de l'intercorrélacion par la formule (1.30) ou l'estimer via la transformée de Fourier inverse de la densité interspectrale par la formule (1.33) réside essentiellement dans la complexité en terme de calcul ; dans le premier cas, la complexité est en $[N^2]$ par contre elle est en $[N \log_2 N]$ dans le deuxième cas [12].

1.5. Comment les être humains localisent le son

La perception des sons dans l'espace est essentiellement reliée à la perception binauriculaire, c'est-à-dire au fait d'avoir deux oreilles. La localisation des sons dans l'espace est complémentaire des autres sens, notamment la vision et la perception de son corps dans l'espace (proprio-perception). La localisation

comprend trois dimensions spatiales, soit l'azimut (gauche-droite), le plan vertical (haut-bas), et la distance.

1.5.1. Localisation à l'azimut (gauche - droite)

Les indices pertinents sont les différences de niveau interauriculaires (DNI) et les différences de temps interauriculaires (DTI). Ces deux indices sont indépendants.

1.5.1.1. Les Différences de Niveau Interauriculaires (DNI)

Il y a 120 ans, Lord Rayleigh avait déjà compris l'importance pour la localisation de la différence de niveaux d'intensité arrivant aux deux oreilles. À cause du phénomène de la diffraction, un son venant de la droite sera plus fort à l'oreille droite qu'à l'oreille gauche, et ce d'autant plus que la fréquence est élevée. En effet, la tête constitue un obstacle à franchir pour les ondes sonores et une "ombre acoustique" apparaît, d'autant plus marquée que la fréquence est élevée.

1.5.1.2. Les Différences de Temps Interauriculaires (DTI)

Pour les basses fréquences (grandes longueurs d'ondes $\lambda >$ diamètre de la tête), il semble que notre système auditif binaurculaire utilise la différence de phase ($\Delta \varphi$) ou différence de temps d'arrivée (Δt) entre les signaux arrivant aux deux oreilles.

Les être humains avec deux oreilles peuvent localiser la source du son en 3D. Deux microphones dans l'air peuvent détecter la direction en 1D dans la gamme $\pm 90^\circ$.

Pour localiser en 3D, les être humains utilisent les données suivantes [13]:

- Différences de Temps Interauriculaires (DTI) c'est la différence dans le temps d'arrivées du son au niveau des deux oreilles. Elle dépend de la position de la source.
- Différences de niveaux Interauriculaires (DNI) c'est la différence de l'intensité du son captés par les deux oreilles. Elle dépend aussi de la position de la source.
- La fonction de transfert de la tête (Head Related Transfert Function HRTF).
- Réverbérations.
- La connaissance à priori du son.

1.6. La performance de la localisation

Le but primaire d'un système de localisation de sources est l'exactitude du résultat. En général, la précision de l'estimé dépend de plusieurs facteurs : la qualité et la quantité des microphones employés, la géométrie du système, le bruit ambiant et les réverbérations, le nombre de sources et leurs contenus spectraux. Généralement, la performance d'une technique de localisation s'améliore en augmentant le nombre de microphones (ex : 512 éléments) particulièrement quand le canal acoustique est défavorable. Mais, quand les conditions acoustiques sont favorables et les microphones sont bien placés, le nombre de microphones sera réduit (ex : 4 éléments) [7]. En plus de la précision de la position, cette position doit être adaptée fréquemment pour l'utilisation dans la poursuite ou la formation de voies.

1.7. Ambiguïté spatiale

Il existe une relation entre les fréquences contenants dans le signal incident et la séparation maximale possible entre un pair de microphones. Considérons deux sinusoïdes ayant la même fréquence mais avec une différence de phase de ϕ entre eux. Cette différence de phase doit être entre π et $-\pi$. Un retard de phase ϕ qui est plus grand que π peut ne pas être distingué d'un retard de $2\pi - \phi$ et vice versa. Par exemple, considérer la sinusoïde de la figure (a) avec une autre sinusoïde retardée de $5\pi/4$. Dans la figure (b), on a la même sinusoïde avec une autre retardée de $2\pi - 5\pi/4 = 3\pi/4$. C'est clairement vu que ces deux situations sont identiques.

Ceci impose que la distance de séparation doit être bien choisie afin d'éviter l'ambiguïté spatiale. Dans le cas de signal à bande étroite, l'ambiguïté spatiale prend lieu quand le déphasage entre les signaux d'un pair de microphones dépasse π et la distance sera en fonction de cette fréquence f . Dans le cas d'un signal large bande ayant une fréquence maximale f_{max} , le scénario du pire des cas est considéré et la distance de séparation sera en fonction de la fréquence maximale [10].

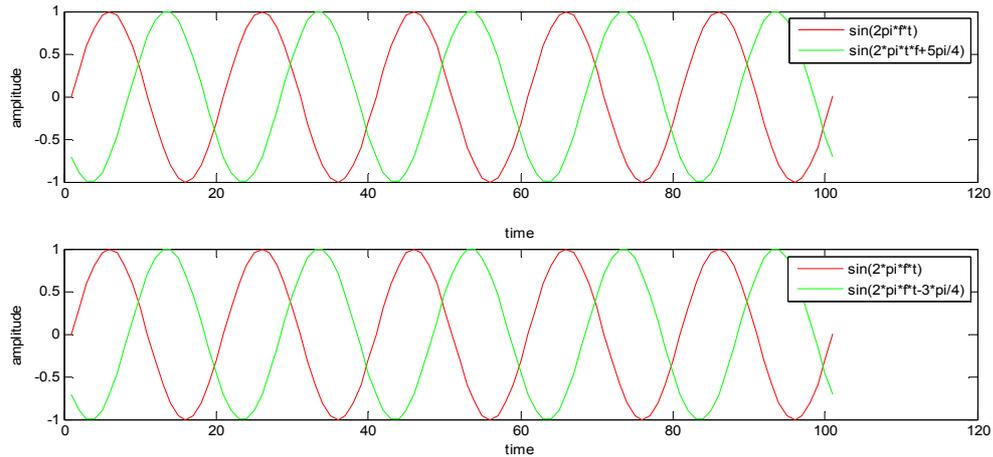


Figure 1.5 : Deux paires de sinusôides avec une différence de phase différente semblent identiques.

En résumé, il faut garantir qu'il y ait moins d'une longueur d'onde entre les deux microphones.

$$\forall f \in [-f_{max}, f_{max}], \forall \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \left|2\pi \frac{fD}{c} \sin\theta\right| < \pi,$$

Soit
$$D < \frac{c}{2f_{max}} \left(D < \frac{\lambda_{min}}{2}\right).$$

1.8. Conclusion

Ce chapitre a introduit le problème de la localisation des sources acoustiques. Les diverses configurations (sources-réseau) sont discutées ; le cas réel correspond souvent à un espace réverbérant avec la présence de multiples sources qui est un cas très difficile à traiter.

Les différentes stratégies de localisation sont décrites. Les méthodes à haute résolution semblent très efficaces dans le cas où les sources sont très proches et à bande étroite, les méthodes à faible résolution nécessitent un grand nombre d'éléments de réseau pour augmenter la résolution. La résolution dans les méthodes basées sur l'estimation du retard temporel dépend de la fréquence d'échantillonnage et la géométrie du réseau, ces méthodes sont moins complexes en termes de temps de calcul, donc beaucoup adéquats pour des applications en temps réel.

CHAPITRE 2

FORMATION DE VOIES

2.1. Introduction

Le traitement d'antenne est la discipline qui utilise un réseau de capteurs répartis dans l'espace pour recevoir des signaux provenant de différentes sources. Son principe est d'exploiter les caractéristiques spatiales des signaux reçus afin d'en extraire de l'information. Ainsi, le réseau de capteurs effectue un échantillonnage spatial d'un champ d'ondes (pouvant être de nature acoustique ou électromagnétique) et un traitement est appliqué sur les échantillons formés [14].

Généralement, des signaux perturbateurs (interférences) viennent s'ajouter au signal d'intérêt. Comme leurs directions d'arrivée sont la plupart du temps différentes, il est possible de les séparer par un filtrage spatial, voir figure (2.1).

Le but primaire de l'utilisation d'un réseau de microphones est de pouvoir commander son diagramme de directivité de sorte que le faisceau principal du gain global du réseau s'oriente vers une direction désirée et les zéros s'orientent vers les directions non désirés [6]. Ceci est réalisé en réglant les poids complexes associés à chaque élément.

2.2. Formation de voies (ou FdV)

La formation de voies ou le beamforming en anglais est une technique de traitement du signal. Elle peut être considérée comme un filtre spatial pour chaque direction d'intérêt. Ce filtre va permettre de récupérer ou d'émettre des signaux dans des directions particulières avec un gain maximum, c'est-à-dire de favoriser le gain pour les directions d'intérêt et d'atténuer les signaux dans les directions non souhaitées (interférences). A la sortie de ce filtre, on n'aura que les signaux dans la direction d'intérêt [3].

On mentionne que les systèmes d'antennes connus sous le vocable «d'antennes intelligentes» (Smart Antennas) utilisent la technique de formation de voies.

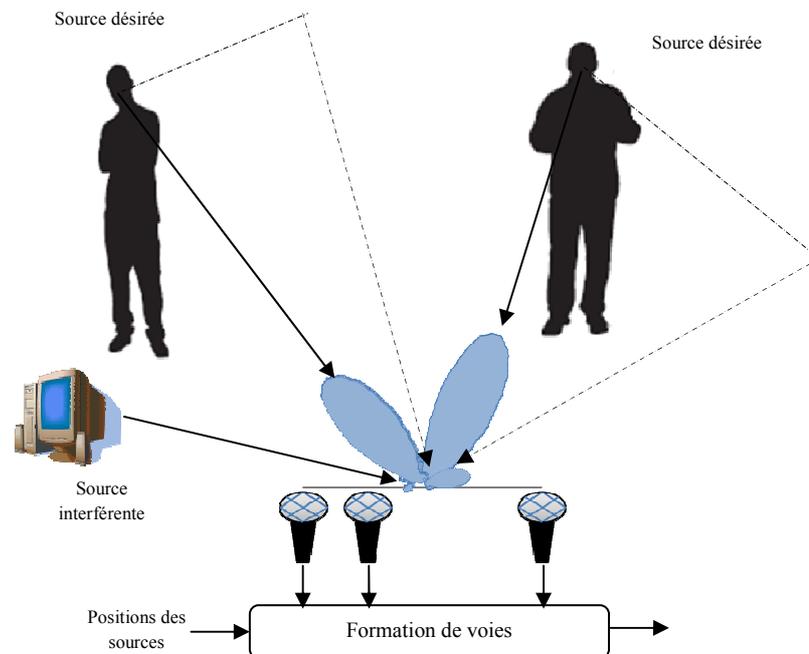


Figure 2.1 : Utilisation d'un réseau de microphone pour le filtrage spatial.

2.3. Types de formateur de voies

La formation de voies étant un outil très important dans le traitement d'antenne il est essentiel d'en connaître les différents types ainsi que les techniques mises en œuvre et leurs avantages respectifs.

Selon les types d'applications, on distingue deux types de réseau formateur de voies: le réseau de FdV analogique (Analog Beamforming) et le réseau de FdV par le calcul ou numérique (Digital Beamforming) [15]. Le premier type réalise la FdV sur un support RF (ligne de transmission, guide d'onde...), alors que le deuxième type la réalise sur un support numérique, beaucoup plus flexible. En fait, ce type de formateur est la somme de deux technologies : la technologie d'antenne et la technologie numérique. On s'intéresse lors de ce mémoire au deuxième type.

2.4. Structures des formateurs de voies numériques

Les algorithmes de la FdV varient en fonction de la distance entre les sources et le réseau d'antennes par rapport à la taille du réseau. Si les sources sont assez près du réseau on dit qu'on est dans le champ proche, la courbure du front d'onde doit être prise en considération. Au contraire si les sources sont assez loin du réseau on dit alors qu'on est dans le champ lointain et les fronts d'onde au niveau du réseau peuvent être vus comme des ondes planes [16]. Dans les problèmes de traitement d'antenne classiques comme le sonar, le radar et les communications numériques, le modèle du champ lointain est réel et pragmatique. Les algorithmes de la FdV ont aussi assumé que les sources sont localisées dans le champ lointain en vue de faciliter le traitement.

Une autre information qui doit être connue a priori pour la conception d'un FdV est la largeur de bande du signal source soit bande étroite ou bande large car ceci a un grand impact sur le choix de sa structure.

La figure (2.2) illustre deux formateurs de voies [17]. Le premier, échantillonne l'onde dans l'espace est typiquement utilisé pour des signaux à bande étroite. La sortie $z(k)$ à un instant k est donnée par la combinaison linéaire des données reçues au niveau de chaque élément i à l'instant k :

$$z(k) = \sum_{i=1}^N w_i^* y_i(k), \quad (2.1)$$

Où $*$ représente le conjugué complexe. Il est conventionnel de multiplier les données par les conjuguées des poids pour simplifier la notation.

Le deuxième FdV échantillonne l'onde dans l'espace et dans le temps, il est utilisé dans le cas des signaux ayant une large bande de fréquence. La sortie peut être exprimée ainsi :

$$z(k) = \sum_{i=1}^N \sum_{p=0}^{K-1} w_{i,p}^* y_i(k-p), \quad (2.2)$$

Où $K-1$ est le nombre de retards pour chaque canal de microphone i .

On note que les deux équations peuvent s'écrire comme suit pour la simplification :

$$z(k) = w^H y(k), \quad (2.3)$$

Avec w est le vecteur poids et $y(k)$ est le vecteur des données reçues, le H représente le hermitien (le transposé du conjugué complexe).

La réponse fréquentielle d'un filtre RIF avec des poids w_p^* , $1 \leq p \leq N$ et des retards de T secondes est donnée par :

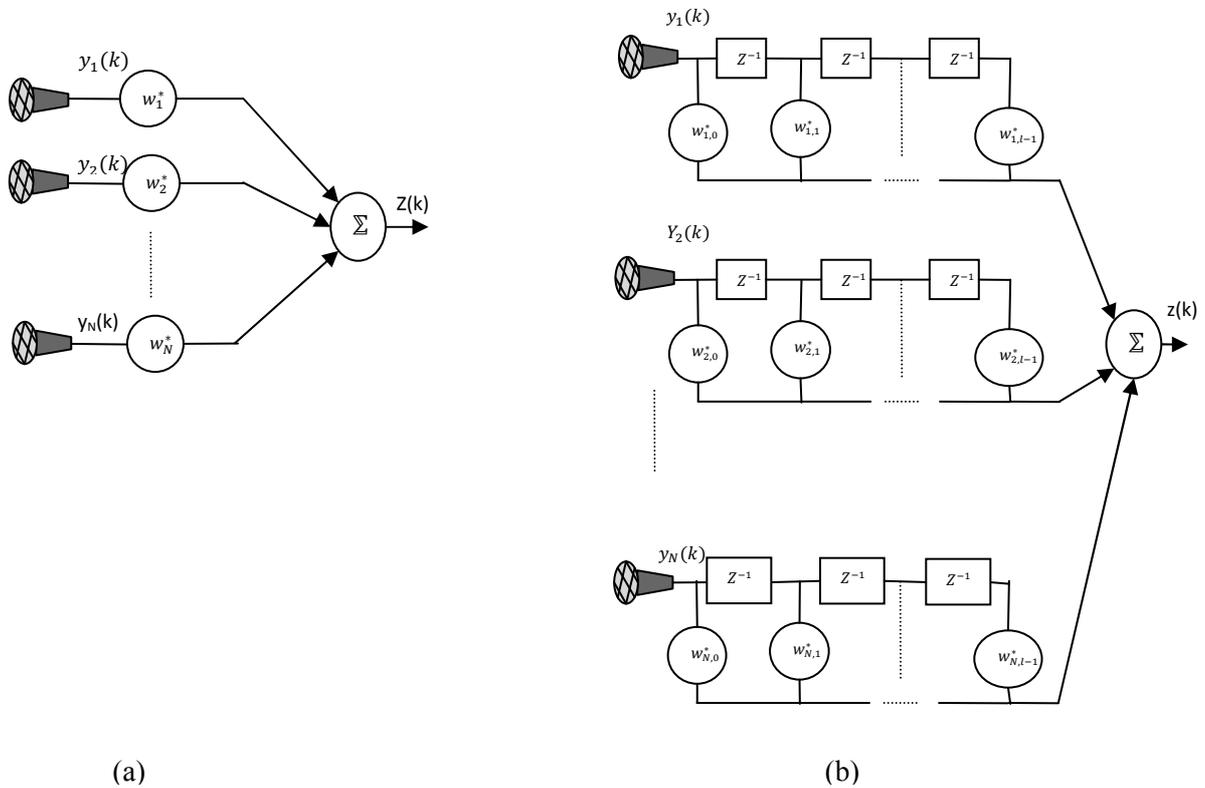


Figure 2.2 : Structures des formateurs de voies: (a) modèle bande étroite, (b) modèle bande large.

$$r(\omega) = \sum_{p=1}^N w_p^* e^{-j\omega T(p-1)} \quad (2.4)$$

$$r(\omega) = w^H d(\omega) \quad (2.5)$$

Où

$$w^H = [w_1^* \ w_2^* \ \dots \ w_N^*] \quad (2.6)$$

Et

$$d(\omega) = [1 \ e^{j\omega T} \ e^{j\omega 2T} \ \dots \ e^{j\omega(N-1)T}]^H \quad (2.7)$$

$r(\omega)$ est la réponse du filtre à une sinusoïde complexe de fréquence ω .

$d(\omega)$ est un vecteur décrivant les phases de la sinusoïde complexe à chaque poids de filtre RIF.

Similairement, la réponse d'un formateur de voies est définie dans le plan complexe en fonction de la position et la fréquence. La position est généralement une quantité de 3 dimensions, mais souvent on s'intéresse qu'à une ou deux dimensions, dans ce mémoire on s'intéresse uniquement à l'azimut. On assume que le signal est une onde plane complexe ayant une DOA θ et une fréquence ω . Par convenance, la phase au niveau du microphone 1 est zéro, ce qui implique $y_1(k) = e^{j\omega k}$ et $y_i(k) = e^{j\omega[k-\Delta_i(\theta)]}$, $2 \leq i \leq N$. $\Delta_i(\theta)$ représente le retard temporel due à la propagation du premier microphone vers le $N^{ième}$ microphone. La sortie du formateur de voies :

$$z(k) = e^{j\omega k} \sum_{i=1}^N \sum_{p=0}^{k-1} w_{i,p}^* e^{-j\omega[\Delta_i(\theta)+p]} = e^{j\omega k} r(\theta, \omega), \quad (2.8)$$

Où $\Delta_1(\theta) = 0$.

$r(\theta, \omega)$ est la réponse du formateur de voies et peut être exprimée sous forme de vecteur :

$$r(\theta, \omega) = w^H d(\theta, \omega) \quad (2.9)$$

$$d(\theta, \omega) = [1 e^{j\omega\tau_2(\theta)} e^{j\omega\tau_3(\theta)} \dots e^{j\omega\tau_N(\theta)}]^H \quad (2.10)$$

Où $\tau_i(\theta)$, $2 \leq i \leq N$ sont les retards temporels dus à la propagation. $d(\theta, \omega)$ est connu sous le nom du vecteur directionnel. Les caractéristiques non idéales des microphones peuvent être incorporées dans $d(\theta, \omega)$ en multipliant chaque décalage en phase par $a_i(\theta, \omega)$, qui décrit la réponse associée à chaque microphone en fonction de la fréquence et la direction.

2.5. Classes des formateurs de voies numériques

Les formateurs de voies peuvent être classifiés soit en des formateurs indépendants des données ou des formateurs optimaux (statistically optimum beamforming) selon la méthode de calcul des poids. Les poids dans les formateurs indépendants des données ne sont pas choisis en fonction des données reçues, ils sont choisis pour vérifier une certaine réponse spécifique pour tous les scénarios signal/interférence. Les poids dans l'autre type sont choisis en se basant sur les statistiques des données pour optimiser sa réponse [18].

2.5.1. Formation de voies indépendante des données

Les poids dans ces FdVs sont calculés pour approximer une certaine réponse désirée indépendamment des statistiques des données. On discute la formation de voies dans le sens classique c'est-à-dire mettre une puissance unité dans la direction d'intérêt et mettre des zéros partout ailleurs.

Pour ce formateur, la pondération est choisie de façon à sommer en phase les signaux provenant d'une direction θ_0 [19] :

$$w = d(\theta_0, \omega) \quad (2.11)$$

Le réseau se focalise alors dans cette direction et une amplitude maximale est obtenue dans cette direction.

Le signal en sortie du FdV s'écrit :

$$z(k) = d^H(\theta_0, \omega)y(k) \quad (2.12)$$

Où $d^H(\theta_0, \omega)$ représente le hermitien du vecteur directionnel de la direction visée et $y(k)$ est le vecteur des signaux reçus au niveau du réseau d'antenne.

On a alors :

$$z(k) = d(\theta_0, \omega)d^H(\theta_0, \omega)s(k) + d^H(\theta_0, \omega)b(k) \quad (2.13)$$

$s(k)$ étant le signal émis et $b(k)$ le vecteur des bruits.

Cela s'écrit encore :

$$z(k) = \|d^H(\theta_0, \omega)\|^2 s(k) + d^H(\theta_0, \omega)b(k) \quad (2.14)$$

Supposons que le vecteur directionnel dans la direction θ_0 , $d(\theta_0, \omega)$ s'écrit :

$$d(\theta_0, \omega) = \begin{bmatrix} a_0(\theta_0) \exp(j\omega\tau_1(\theta_0)) \\ \vdots \\ a_N(\theta_0) \exp(j\omega\tau_N(\theta_0)) \end{bmatrix} \quad (2.15)$$

Lorsqu'on applique une pondération $w = a(\theta_0)$ à un signal $v(t)$ émis dans une direction θ_1 différente de θ_0 , le signal $v(t)$ émis sera alors multiplié par un facteur conduisant à une atténuation non maîtrisée en dehors du lobe principal (les

positions des maxima locaux et les zéros du diagramme en dehors du lobe principal sont aléatoires).

Ce dispositif réalise donc à la fois un pointage électronique de l'antenne dans la direction θ_0 et un filtrage spatial permettant d'atténuer les signaux arrivant des autres directions.

En termes de complexité, ce formateur est simple à mettre en œuvre, seule la connaissance des vecteurs directionnels et les directions d'arrivées sont requises pour son implémentation.

Néanmoins, cette méthode de FdV ne permet pas de contrôler le niveau d'atténuation du diagramme dans des directions particulières. Elle ne permet donc pas d'atténuer la contribution de brouilleurs éventuels qui peuvent perturber la communication. Cela conduit à l'étude du second type de formateur de voies.

2.5.2. Formation de voies optimales (statistically optimum beamforming)

Dans ce type de formateurs de voies, les poids sont choisis en se basant sur les statistiques des données reçues au niveau du réseau. Le but est d'optimiser la réponse du FdV de telle manière à avoir moins de contribution du bruit et des interférences. Plusieurs critères d'optimisation des poids sont discutés. Au cours de cette partie, on assume que les données sont stationnaires au sens large et que les statistiques d'ordre deux sont connus.

2.5.2.1. Méthode MSC (Multiple Sidelobe Canceler)

Le MSC est probablement le premier formateur de voies optimal. Il consiste d'un canal principal et un ou plusieurs canaux auxiliaires comme c'est illustré dans la figure 2.3 [20]. Le canal principal peut être soit un seul capteur haut gain ou un formateur de voies indépendant des données ayant une réponse très directive qui pointe dans la direction du signal désiré. Le but est de choisir des poids pour les canaux auxiliaires qui annulent les interférences captées par les lobes secondaires du canal principal, ceci implique que la combinaison linéaire des réponses des canaux auxiliaires est identique à la réponse aux interférences du canal principal.

En général, ce n'est pas toujours possible d'annuler tous les interférents, ce qui donne naissance à un bruit blanc .

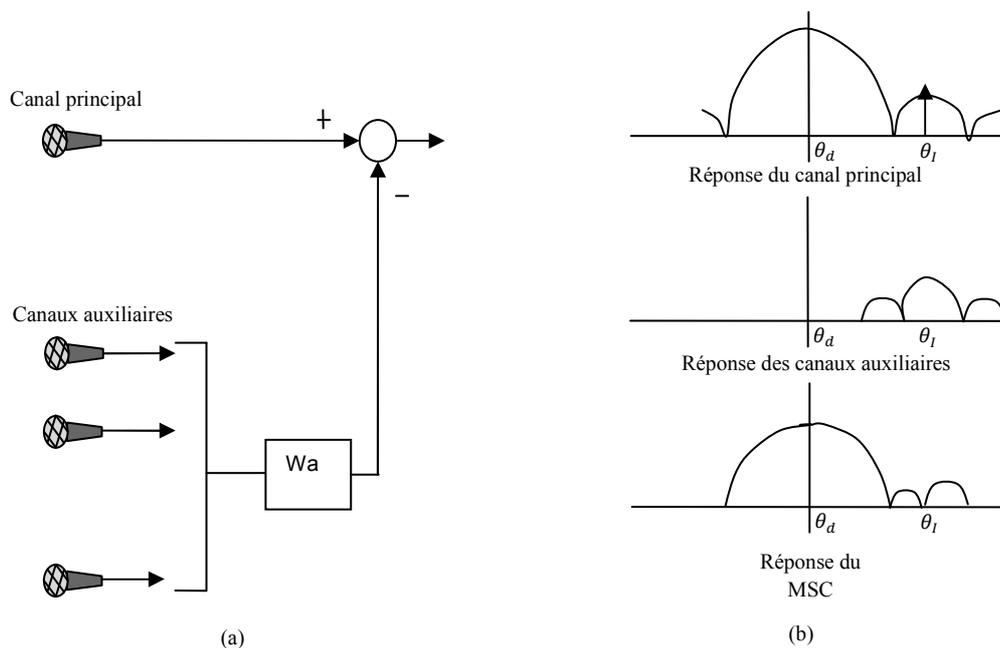


Figure 2.3 : Principe du MSC; (a) structure, (b) les réponses du canal principal, canaux auxiliaires et du système complet.

2.5.2.3. Formation de voies à variance minimum LCMV(Frost)

Dans cette section, la manière de contrôler la réponse du formateur de voies par l'utilisation des contraintes linéaires est expliquée. L'idée de base derrière la FdV LCMV est de poser une contrainte sur la réponse du FdV de façon à laisser passer le signal incident de la direction souhaitée avec un gain et une phase spécifiés. Les poids sont choisis en minimisant la variance ou la puissance à la sortie tout en respectant la contrainte posée. Ceci à l'effet de préserver le signal désiré et minimiser la contribution due aux interférences et au bruit à la sortie du FdV.

La réponse d'un FdV est donnée par $w^H d(\theta, \omega)$, la contrainte linéaire $w^H d(\theta, \omega) = g$ est posée, ou g est une composante complexe, ceci dit que n'importe quel signal venant de la direction θ ayant la fréquence ω est passé vers la sortie avec un gain g . la minimisation de la contribution de tout signal venant d'autres directions est accomplie en choisissant les poids qui minimisent la

puissance à la sortie $E\{|z|^2\} = w^H R_y w$ le problème du LCMV peut être écrit comme suit :

$$\min w^H R_y w \quad (2.16)$$

Sous la contrainte

$$d^H(\theta, \omega)w = g^* \quad (2.17)$$

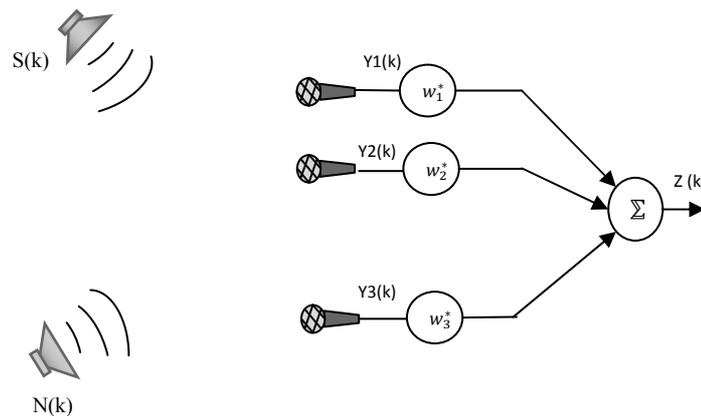


Figure 2.4 : Formateur de voies bande étroite en présence de deux sources.

La méthode de Lagrange est utilisée pour résoudre ce problème, le résultat suivant est obtenu :

$$w = g^* \frac{R_y^{-1} d(\theta, \omega)}{d^H(\theta, \omega) R_y^{-1} d(\theta, \omega)} \quad (2.18)$$

R_y est la matrice de covariance du signal reçu par le réseau, $y(t)$, qui comprend le signal utile, le bruit et les interférences éventuelles. Notez, qu'en pratique, la présence d'un bruit non corrélé assure que : R_y est invertible. Si $g = 1$, alors ce formateur de voies est appelé MVDR (minimum variance distortionless response) ou méthode de CAPON. L'unique contrainte linéaire dans l'équation peut être facilement généralisée en un multiple de contraintes linéaires pour plus contrôler le diagramme de rayonnement. Par exemple, s'il existe une source d'interférence fixe connue dans la direction φ , un gain de zéro sera forcé dans cette direction en plus, un gain g sera maintenu pour la direction désirée. Ceci est exprimé comme suit :

$$\begin{bmatrix} d^H(\theta, \omega) \\ d^H(\varphi, \omega) \end{bmatrix} w = \begin{bmatrix} g^* \\ 0 \end{bmatrix} \quad (2.19)$$

$$C^H w = f \quad (2.20)$$

Où C est la matrice des contraintes et f le vecteur des réponses.

La complexité de ce formateur dépend alors d'une part du nombre d'échantillons requis pour l'estimation de la matrice de covariance mais aussi de la complexité associée à l'inversion de cette dernière.

2.5.2.3.1. L'annulateur de lobe secondaire généralisé Méthode GSC (Generalized Sidelobe Canceler)[21]

Le GSC représente une formulation alternative du problème LCMV. Le GSC essentiellement est un mécanisme qui change un problème de minimisation avec contraintes en un problème de minimisation sans contraintes [22].

Le GSC comprend trois blocs ; un FdV fixe qui alignent les composantes du signal désiré, matrice bloquante (blocking matrix) qui bloque les composantes du signal désirée contenant dans les signaux de bruit de références et un ANC (Adaptive Noise Canceler) multi canal qui élimine les composantes du bruit qui fuit dans les lobes secondaires du FdV fixe.

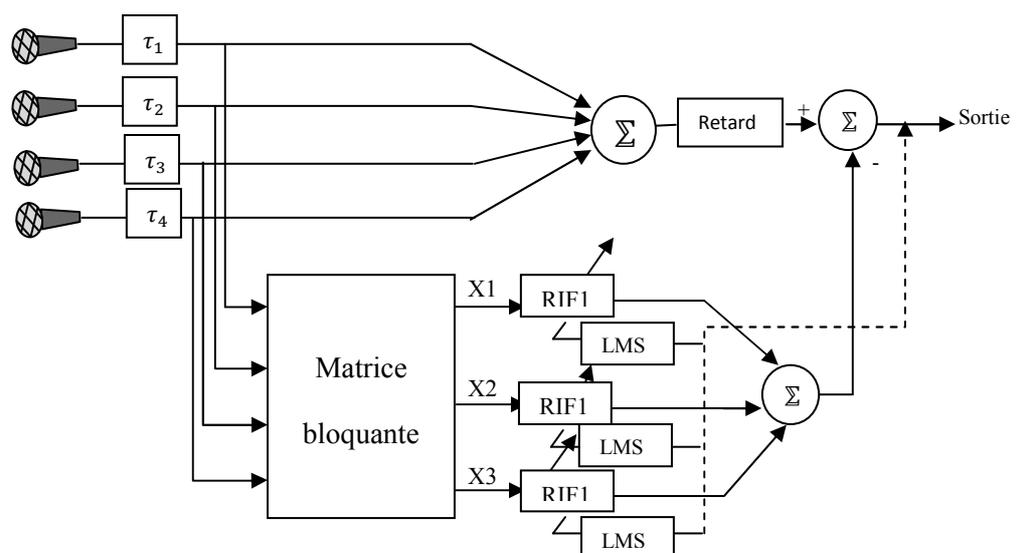


Figure 2.5 : Schéma descriptif du GSC.

2.5.2.4. Maximisation du rapport signal sur bruit

Ici les poids sont choisis directement pour maximiser le rapport signal sur bruit (RSB) où le signal à bruit plus interférence (RSBI) en sortie du FdV . Une solution générale pour le calcul des poids nécessite la connaissance des matrices de covariance du signal désiré R_{ss} et du bruit R_{bb} . L'obtention de ces deux connaissances dépend de l'application. Par exemple, dans le radar, R_{bb} peut être estimé pendant le temps où aucun signal n'est transmis et R_{ss} peut être connue à partir de l'impulsion transmise et la direction d'intérêt. Si le signal est à bande étroite de fréquence ω et une direction θ , alors

$$R_{ss} = \sigma^2 d(\theta, \omega) d^H(\theta, \omega) \quad (2.21)$$

Le récepteur estime les puissances des signaux utiles et interférents à travers l'estimation des matrices de corrélation des signaux. Les poids sont alors calculés pour maximiser le rapport de ces puissances et éventuellement annuler les interférences.

En présence d'interférences, le signal reçu au niveau du réseau de microphones s'écrit :

$$y(k) = a(\theta_0)s(k) + x_I(k) + b(k) \quad (2.22)$$

$y_I(k)$ étant le vecteur des signaux interférents. En sortie du FdV, le signal s'écrit :

$$z(k) = w^H a(\theta)s(k) + w^H [y_I(k) + b(k)] \quad (2.23)$$

On a alors la puissance du signal en sortie du FdV, P_{signal} qui s'écrit :

$$P_{signal} = E[\{w^H a(\theta)s(k)\}\{w^H a(\theta)s(k)\}^H] \quad (2.24)$$

Soit :

$$P_{signal} = E[\|w^H a(\theta)s(k)\|^2] \quad (2.25)$$

En notant P_s la puissance du signal

$$P_{signal} = P_s \|w^H a(\theta)\|^2 \quad (2.26)$$

Soit

$$n_{bi}(k) = y_I(k) + b(k) \quad (2.27)$$

Le vecteur représentant la contribution du bruit et des signaux interférents. La puissance des bruits (bruit et interférences) en sortie s'écrit :

$$P_{bruit} = E[\{w^H n_{bi}(k)\}\{w^H n_{bi}(k)\}^H] \quad (2.28)$$

Soit :

$$P_{bruit} = w^H E[n_{bi}(k)n_{bi}^H(k)]w \quad (2.29)$$

Soit

$$C = E[n_{bi}(k)n_{bi}^H(k)] \quad (2.30)$$

la matrice de covariance des bruits et des interférences, la puissance des bruits s'écrit :

$$P_{bruit} = w^H C w \quad (2.31)$$

Le *RSBI* étant le rapport de la puissance du signal utile sur celle des interférences et du bruit,

$$RSBI = \frac{P_s \|w^H a(\theta)\|^2}{w^H C w} \quad (2.32)$$

La maximisation du RSBI est donc équivalente à la minimisation de la puissance du bruit et des interférences en sortie de FdV.

Calculer la pondération MVDR revient donc à résoudre :

$$\min_w w^H C w \quad (2.33)$$

Sous la contrainte :

$$w^H a(\theta_0) = 1 \quad (2.34)$$

La résolution de ce problème conduit à la pondération suivante :

$$w_{mvd} = \frac{C^{-1} a(\theta_0)}{a^H(\theta_0) C^{-1} a(\theta_0)} \quad (2.35)$$

Le signal utile n'étant pas présent dans l'équation (2.35), cette méthode est robuste et donne de bons résultats. Elle reste cependant délicate à implémenter car elle nécessite la connaissance du bruit et des interférences indépendamment du signal utile.

La complexité de ce formateur est la même que celle du formateur MVDR.

2.5.2.2. Emploi d'un signal de référence

Si le signal désiré est connu, les poids peuvent être calculés en minimisant l'erreur quadratique moyenne entre la sortie du formateur de voies et le signal désiré. Bien sûr la connaissance du signal désiré élimine l'utilisation de la formation de voies, mais dans quelques applications, un signal de référence est indispensable.

Le vecteur des poids dépend de l'intercovariance entre le signal indésiré inconnu et le signal de référence. Une performance acceptable est obtenue lorsque le signal de référence est incorrélié avec les signaux d'interférence [19].

$$\begin{aligned} E[\|w^H y(k) - s(k)\|^2] \\ = w^H R_{yy} w - w^H E[y(k)s^H(k)] - E[y(k)s^H(k)]^H w + P_s \end{aligned} \quad (2.36)$$

Où

$$E[y(k)s^H(k)] \equiv R_{ys} \quad (2.37)$$

représente l'intercorrélation entre le vecteur signal reçu au niveau du réseau et le signal pilote (de référence), P_s est la puissance du signal pilote.

On cherche la pondération w solution de :

$$\min_w E[\|w^H y(k) - s(k)\|^2] \quad (2.38)$$

L'avantage de cette méthode de FdV est que l'on n'a pas besoin de connaître les directions d'arrivée du signal utile contrairement à la méthode MVDR. Par contre, la connaissance du signal utile est requise. Le vecteur de pondération solution du problème est :

$$w = R_{yy}^{-1} E[y(k)s^H(k)] = R_{yy}^{-1} R_{ys} \quad (2.39)$$

La complexité de ce formateur est alors moindre que pour les deux autres formateurs adaptatifs.

2.6. Suppression du signal avec les formateurs de voies optimaux

Les FdVs optimaux nécessitent la connaissance de quelques caractéristiques du signal désiré soit ses statistiques (méthode du maximum SNR

et méthode basée sur le signal de référence), soit sa direction (pour le MSC), soit son vecteur directionnel $d(\theta, \omega)$ (pour le formateur LCMV). Si ces connaissances sont imprécises, le FdV optimal atténuera le signal désiré comme s'il était un signal interférent. La réduction du signal désiré est parfois importante. Une deuxième cause pour la réduction du signal désiré est la corrélation entre ce dernier et l'un des interférents. Ceci peut être dû aux multi-trajets. Quand le signal désiré et les interférents sont non corrélés, le FdV atténue les interférents pour minimiser la puissance de sortie. Hors, si le signal désiré est corrélé avec les signaux interférents, le signal désiré sera aussi atténué.

2.7. Problème des signaux large bande

Les techniques de formation de voies utilisées pour les signaux à bande étroite s'avèrent non intéressante pour les applications acoustiques puisque la parole est typiquement un signal large bande. La longueur d'onde varie en fonction de la fréquence alors le diagramme de rayonnement ne sera pas le même pour toutes les fréquences constituant le signal large bande. Pour cela, si on utilise un formateur de voies bande étroite pour des signaux à large bande, le bruit et les interférents venant des directions différentes ne vont pas être atténués uniformément.

Une manière pour remédier à ce problème est d'utiliser des sous réseau [16]. Chaque sous réseau (linéaire et équidistant) est conçu pour fonctionner à une seule fréquence. Mais cette solution nécessite un grand réseau avec beaucoup de microphones. Une autre méthode pour concevoir un formateur de voies large bande en tirant profit des techniques appliquées aux signaux bande étroite est de décomposer le signal large bande en des signaux bande étroite et réaliser un formateur de voies bande étroite pour chaque fréquence comme le montre la figure (2.6).

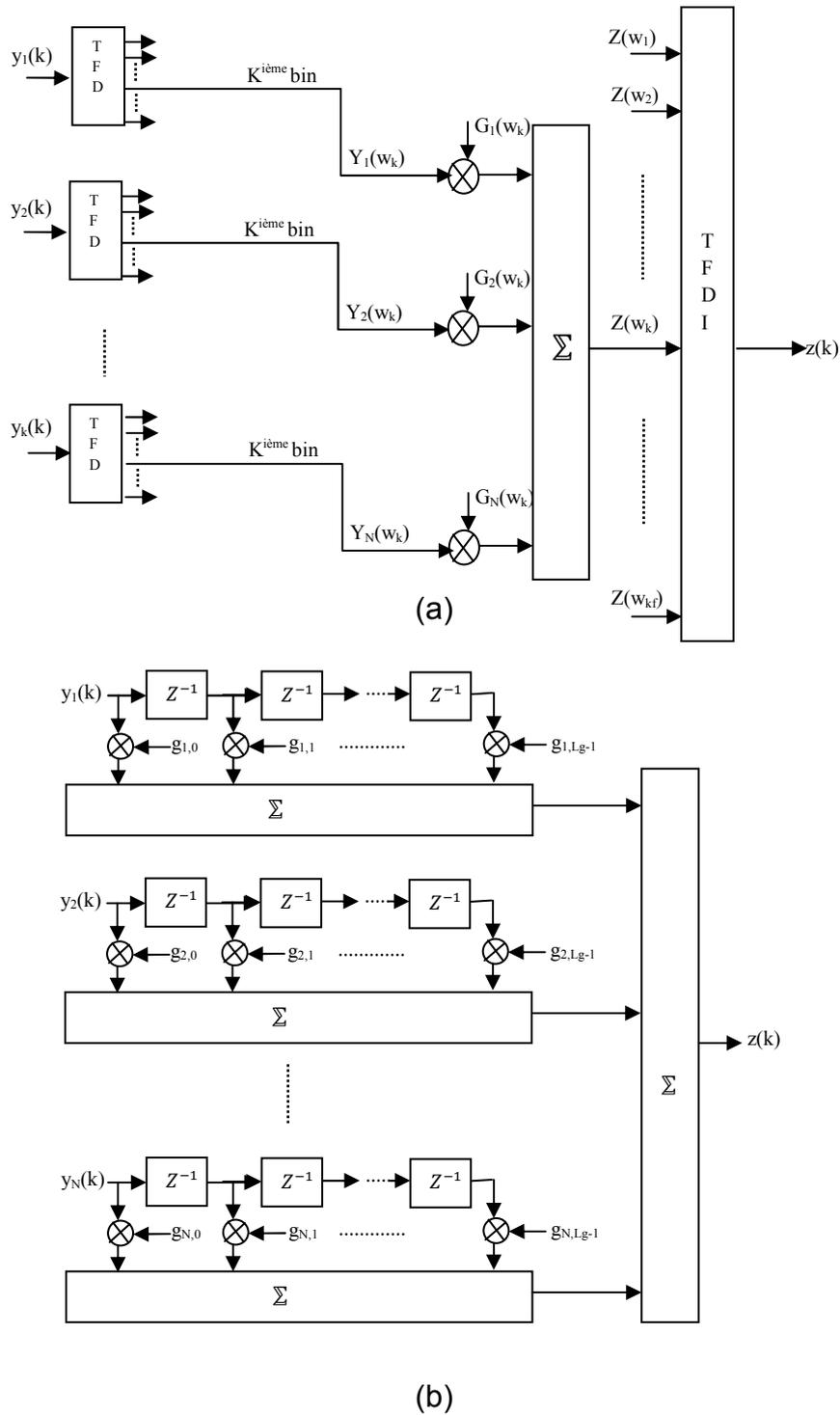


Figure 2.6 : Formateur de voies large bande; (a) domaine fréquentiel, (b) domaine temporel.

2.8. Concept de base de la formation de voies

La procédure utilisée pour modifier le diagramme de rayonnement d'un réseau d'antenne pour favoriser la réception d'un signal désiré en même temps

supprimer des signaux interférents à travers le calcul des poids complexes est illustrée dans l'exemple suivant [15]:

Considérons le réseau d'antenne montré dans la figure qui consiste de deux microphones omnidirectionnels espacés de $\lambda/2$. Le signal désiré $s(t)$ provient d'une direction formant un angle $\theta_s = 0$ avec la normale au réseau et le signal interférent $i(t)$ provenant de la direction $\theta_i = \frac{\pi}{6}$. Les deux signaux ayant la même fréquence f_0 . Le signal de chaque élément est multiplié par un poids complexe, et les signaux pondérés sont sommés pour former la sortie du réseau. La sortie du réseau due au signal désiré est :

$$z_d(t) = Ae^{j2\pi f_0 t}(w_1 + w_2) \quad (2.40)$$

La sortie du réseau due au signal interférent est :

$$z_i(t) = Ne^{j2\pi f_0 t}w_1 + Ne^{j(2f_0 t + \frac{\pi}{2})}w_2 \quad (2.41)$$

La sortie totale est :

$$z = z_d + z_i = s(t) \quad (2.42)$$

Ce qui implique :

$$w_1 + w_2 = 1 \quad (2.43)$$

$$w_1 + e^{j\pi/2}w_2 = 0 \quad (2.44)$$

La solution est donnée

$$\text{par : } w_1 = \frac{1}{2} - j\frac{1}{2}, w_2 = \frac{1}{2} + j\frac{1}{2} \quad (2.45)$$

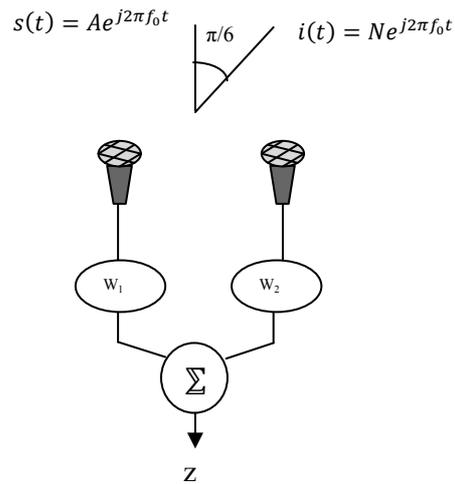


Figure 2.7 : Suppression d'interférence par un réseau de deux microphones.

Avec ces poids, le réseau va accepter le signal désiré et simultanément supprimer le signal interférent. La méthode utilisée dans cet exemple exploite le fait qu'il n'y a qu'un seul signal interférent et utilise des informations a priori concernant la fréquence et la direction d'arrivée des deux sources or, dans le cas pratique les détails concernant le nombre de sources, leurs directions et la nature des source ne sont pas connus à priori.

2.9. Conclusion

La technique de formation de voies était bien décrite dans ce chapitre tout en citant les différentes classes et structures. Dans plusieurs cas des réseaux adaptatifs, la fonction de transfert acoustique doit être connue à priori ou au moins peut être estimée à partir des données reçues (il faut noter que dans le cas de propagation avec retard uniquement, la fonction de transfert acoustique se réduit en un vecteur directionnel constitué uniquement de composantes donnant information sur la phase), et ceci est un moyen qui facilite le calcul des poids associés aux éléments de réseau. Le problème de la formation de voies large bande a été introduit ; deux techniques peuvent être utilisées pour le résoudre sont discutées.

CHAPITRE 3

REALISATIONS ET RESULTATS

3.1. Introduction

Les réseaux de microphones ou (d'antennes acoustiques) peuvent être un moyen efficace pour combattre les dégradations de la parole dues au bruit et aux réverbérations dans les terminaux de communication mains-libres. L'objectif dans ce mémoire est de réaliser un système de rehaussement de la parole basé sur la séparation spatiale de sources par la technique de formation de voies. Pour le réaliser, l'étape de la localisation de sources est primordiale, deux sources audio sont considérées (source désirée et source interférente) qui doivent être localisées par leurs angles d'azimut, ensuite la technique de formation de voies sera employée pour modifier le diagramme de directivité du réseau de microphone en vue de le maximiser dans la direction de la source désirée et le minimiser dans la direction de la source interférente. Au cours de ce chapitre, la conception de ce système sera réalisée et testée sous l'environnement Matlab qui permet l'acquisition de données et leurs prétraitements via la carte son du micro-ordinateur.

3.2. Description du système

Le schéma dans la figure (3.1), décrit d'une manière globale les principales fonctions du système de rehaussement de la parole qu'on désire réaliser. Ce système est constitué de deux micro-ordinateurs, un destiné à l'émission des signaux de test tandis que l'autre a pour rôle l'acquisition des signaux et l'exécution des deux algorithmes. Les transducteurs utilisés sont deux hauts parleurs à l'émission et deux microphones omnidirectionnels à la réception ; leurs diagrammes de directivité sont considérées idéals.

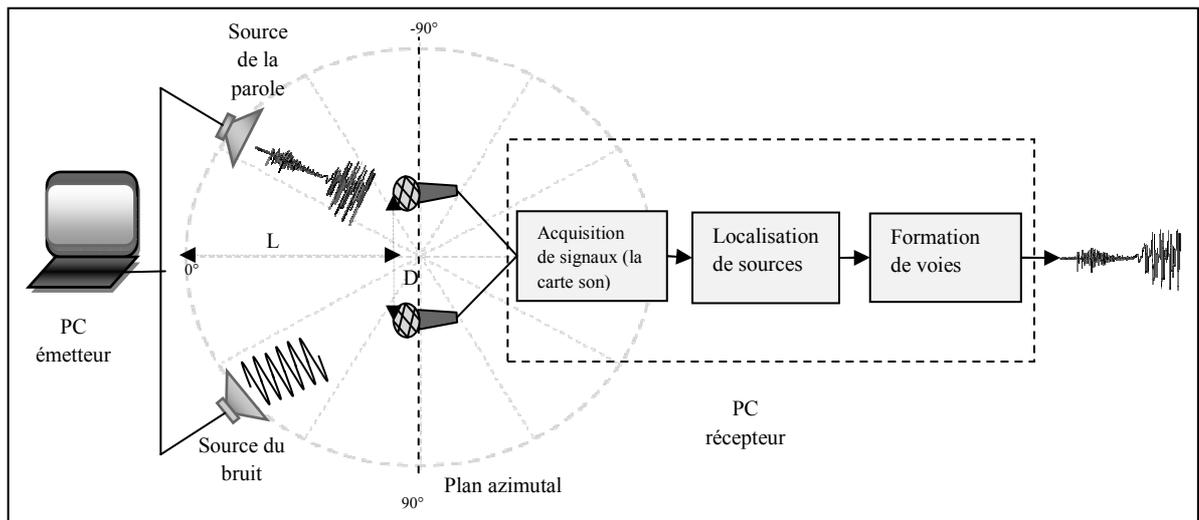


Figure 3.1 : Description globale du système.

Les expériences sont réalisées dans un laboratoire d'où un espace dégagé est préservé. Les distances intra-microphones et sources-réseau seront justifiées au cours de ce chapitre.

3.3. Localisation de sources

3.3.1. Introduction

Dans un environnement acoustique, la localisation de sources joue un rôle très important dans des applications diverses telles que la poursuite automatique du locuteur par une camera dans une salle de vidéoconférence, la suppression du bruit acoustique par la formation de voies...

La localisation comprend trois dimensions spatiales, soit l'azimut (gauche-droite), le plan vertical (haut-bas), et la distance. Ici, uniquement le plan azimutal qui est pris en considération.

D'après les méthodes de localisation citées dans le premier chapitre, les méthodes d'estimation du retard temporel semblent les plus simples et possibles à employer.

3.3.2. Principe

Soit, selon la figure (3.2), deux microphones omnidirectionnels soumis à une onde acoustique plane progressive sous un angle d'incidence tel qu'un front d'onde atteint d'abord le premier, puis le second après un temps de propagation τ ; le second microphone capte la même forme d'onde que le premier, mais avec

un retard. Par un raisonnement géométrique, on constate que τ a une relation avec l'angle d'incidence θ_1 et la distance D entre les microphones.

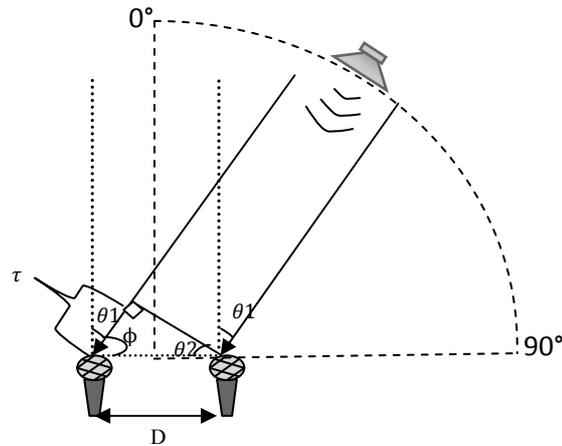


Figure 3.2 : Principe de la localisation des sources par l'estimation du retard temporel.

$$\text{On a : } \theta_1 + \phi = \pi/2, \quad (3.1)$$

$$\text{Et } \theta_2 + \phi = \pi/2, \quad (3.2)$$

$$\text{Alors } \theta_1 = \theta_2, \quad (3.3)$$

$$\text{On a en plus } \sin(\theta_2) = \tau * \frac{c}{D} \quad (3.4)$$

$$\text{Alors } \theta_1 = \arcsin\left(\tau * \frac{c}{D}\right), \quad (3.5)$$

Avec c est la célérité du son, τ est le retard temporel.

La mesure du retard temporel τ permet de déterminer l'angle d'incidence θ_1 .

3.3.3. La résolution angulaire

La résolution angulaire peut être définie comme étant la capacité d'un système de discriminer deux sources légèrement espacées. Elle est déterminée par le nombre de différentes mesures de directions d'arrivées peuvent être effectuées entre 0 et π .

Considérons le même réseau de microphones précédent. Pour une distance intra-microphone D , on a le temps de parcours de cette distance τ :

$$\tau = \frac{D}{c} \quad (3.6)$$

Ce temps de parcours τ équivaut à un nombre d'échantillons N , tel que :

$$N = \tau * F_s \quad (3.7)$$

Le nombre N correspond aussi au nombre des angles pouvant être détectés.

$$N = \frac{D}{c} * F_s \quad (3.8)$$

D'après l'équation (3.8), l'amélioration de la résolution angulaire peut être achevée par les deux méthodes suivantes:

-Augmenter la distance intra-microphone D ; Ceci implique une taille importante du réseau, ce qui n'est pas toujours adéquat surtout dans des situations où l'espace est limité. Aussi, une grande distance D peut causer le problème de l'ambiguïté spatiale qui a une grande influence dans le contexte de la formation de voies. En plus, cela nécessite l'utilisation d'une large fenêtre d'analyse puisque le retard serait important.

-Augmenter la fréquence d'échantillonnage; Ceci augmente le nombre d'échantillons constituant le retard τ , d'où chaque échantillon correspond à un angle d'arrivée, mais ceci influence sur la qualité d'estimation et la rapidité du système.

Dans la figure (3.3), la variation de la résolution angulaire en fonction de deux fréquences d'échantillonnage est illustrée. Pour une distance $D = 20cm$, $F_s = 8KHz$, on obtient $N = 5$ échantillons et pour une $F_s = 22KHz$, $N = 12$ échantillons.

- La transformation (retard-angle) est non linéaire, une haute résolution est observée lorsque la source est en face le réseau du microphone (broad-side), par contre elle commence à se diminuer en se déplaçant vers les endroits où la source se trouve colinéaire avec le réseau (end-fire).

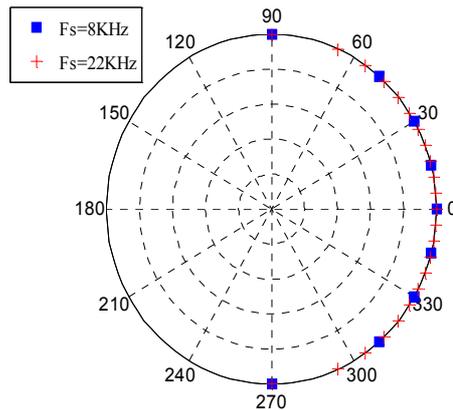


Figure 3.3 : La résolution spatiale versus la fréquence d'échantillonnage.

3.3.4. Localisation du locuteur

Dans cette section, on s'intéresse à la localisation d'une source de parole. L'algorithme PHAT-GCC (PHASE Transform Generalized Cross Correlation) [11] est utilisé et testé dans un environnement de simulation puis dans un environnement réel. Cet algorithme est basé sur le calcul de l'inter-corrélation via l'estimation de la densité inter-spectrale. Le filtre de blanchiment (PHAT) est employé pour rendre le pic indiquant le retard plus étroit donc facile à détecter. Ci-dessous, la description de cet algorithme.

Algorithme PHAT-GCC [23]

Les étapes de l'algorithme PHAT-GCC

Obtenir une trame d'observation à un instant k :

$$\mathbf{x}_n(\mathbf{k}) = [x_n(\mathbf{k}), x_n(\mathbf{k} + 1), \dots, x_n(\mathbf{k} + N - 1)]^T$$

Estimer le spectre de $\mathbf{x}_0(\mathbf{k})$:

$$X_0(\mathbf{k}') = \sum_{k=0}^{N-1} x_0(\mathbf{k}) e^{-\frac{j2\pi k k'}{N}} = FFT_N\{x_0(\mathbf{k})\}$$

Estimer le spectre de $\mathbf{x}_1(\mathbf{k})$:

$$X_1(\mathbf{k}') = \sum_{k=0}^{N-1} x_1(\mathbf{k}) e^{-\frac{j2\pi k k'}{N}} = FFT_N\{x_1(\mathbf{k})\}$$

Calcul de la densité interspectrale :

$$\frac{s_{x_0 x_1}(\mathbf{k}')}{|s_{x_0 x_1}(\mathbf{k}')|} = \frac{\mathbf{E}\{X_0(\mathbf{k}') X_1^*(\mathbf{k}')\}}{|\mathbf{E}\{X_0(\mathbf{k}') X_1^*(\mathbf{k}')\}|}$$

Estimer la fonction PHAT :

$$\hat{\Psi}_{PHAT}(\mathbf{m}) = \sum_{k'=0}^{K'-1} \frac{S_{x_0x_1}(k')}{|S_{x_0x_1}(k')|} e^{j2\pi mk'/K'} = \mathbf{FFT}_N^{-1} \left\{ \frac{S_{x_0x_1}(k')}{|S_{x_0x_1}(k')|} \right\}$$

Maximiser la fonction $\hat{\Psi}_{PHAT}$:

$$\max \hat{\Psi}_{PHAT}(\mathbf{m})$$

Tableau 3.1 : L'algorithme PHAT-GCC.

3.3.4.1. Localisation du locuteur dans un environnement de simulation idéal

L'idée ici est d'évaluer la performance de l'algorithme PHAT-GCC en fonction des caractéristiques spectrales du signal parole indépendamment de l'effet du canal acoustique, cela en utilisant un canal de propagation de simulation dont uniquement le trajet direct est considéré. Ses performances seront testées en fonction de quelques paramètres à identifier.

L'échantillon de la parole utilisé dans le test est un fichier déjà existant dans la bibliothèque du Matlab « speech_dft.wav », quantifié sur 16 bits et échantillonné à 22.05 KHz. Ici, le canal acoustique est supposé idéal c'est-à-dire qu'uniquement le retard entre les éléments du réseau est considéré. C'est ce retard (TDOA1) qui doit être retrouvé pour la localisation. Un autre retard (TDOA2) est ajouté après le PHAT-GCC, essentiel pour la localisation dans l'intervalle $[-\frac{\pi}{2}, +\frac{\pi}{2}]$, à condition qu'il soit retranché après. Ce TDOA2 doit être supérieur ou égale au retard maximal qui peut être atteint (source est colinéaire avec le réseau). L'échantillon en plus revient au fait que pendant l'acquisition, un retard d'un échantillon se crée entre les deux canaux.

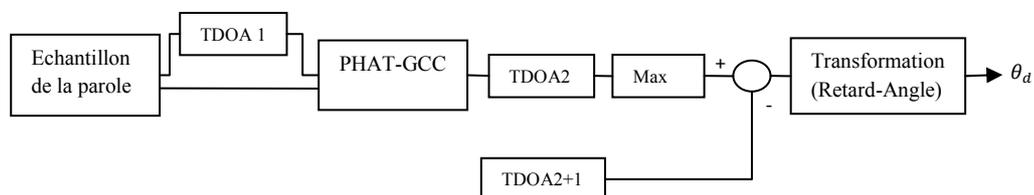


Figure 3.4 : Chaîne de localisation du signal parole par l'algorithme PHAT-GCC dans un environnement de simulation idéal.

a. Influence de la taille de la fenêtre d'analyse

Généralement, la plupart des algorithmes audio fonctionnera plus efficacement sur de plus grands blocs de données. Donc il est normal d'aller vers l'emploi de

fenêtre d'analyse de grande taille [24]. Malheureusement ceci pose un problème critique dans des applications telles que la téléphonie mobile où la contrainte du temps réel est indispensable. Une raison principale de limiter la taille de la fenêtre d'analyse survient quand les caractéristiques du signal changent pendant cette fenêtre d'analyse. Il y a deux points importants à expliquer ici. Le premier est la stationnarité du signal, le deuxième est la résolution.

La stationnarité du signal

En pratique, beaucoup de signaux audio ne tendent pas à rester stationnaires pour longtemps. Pour cela, des fenêtres d'analyse de petite taille sont nécessaires pour capturer les changements brefs du signal. Dans l'analyse de la parole, plusieurs mouvements de muscle qui produisent la parole changent lentement ayant pour résultat un changement lent des caractéristiques spectrales de la parole. Un principe de base utile est que la parole peut être supposée stationnaire ou pseudo-stationnaire, pendant environ 20-30ms.

La résolution temps-fréquence

Dans le domaine fréquentiel, la *FFT* du signal audio échantillonné à une fréquence F_s contient $N/2 + 1$ intervalle de fréquence positif (bîne). Chaque bîne donne une information sur l'énergie du signal d'origine dans un intervalle très petit. La largeur de cet intervalle est liée directement à la fréquence d'échantillonnage et au nombre d'échantillons à analyser, F_s/N . pour avoir une meilleure résolution fréquentielle, une longue durée d'échantillons doit être accumulée.

La longueur de la fenêtre doit d'une part être suffisante pour assurer une bonne résolution fréquentielle; d'autre part elle doit être limitée si l'on veut suivre fidèlement l'évolution dans le temps du spectre vocal. Ces deux exigences sont contradictoires.

Au cours de ce chapitre, la taille de la fenêtre d'analyse est fixée à 23 ms.

b. Influence du type de la fenêtre d'analyse

Les calculs numériques devront s'effectuer sur des séquences de durées limitées, il faut tronquer le signal. Ceci revient à le multiplier par une fenêtre d'observation rectangulaire [25]. Dans le domaine fréquentiel, ceci équivaut à une

convolution par une fonction *sinc*. Les ondulations qui ont résultent sont responsables d'une distorsion qui porte le nom leakage. La réduction du leakage conduit à l'utilisation de fenêtres mieux conditionnées dans le domaine fréquentiel.

Simulation 1 : Influence du type de la fenêtre d'analyse

Pour choisir le type de fenêtre approprié, l'algorithme est testé pour quatre types de fenêtres courantes et prédéfinies de Matlab, plusieurs réalisations ont été effectuées et les histogrammes de la figure (3.8) ont été obtenus.

Ces histogrammes représentent la fréquence d'occurrence de chaque direction d'arrivée en pourcentage. L'algorithme est exécuté environ 8500 fois et à chaque fois la direction d'arrivée est estimée. La vraie direction d'arrivée est de 18.13°.

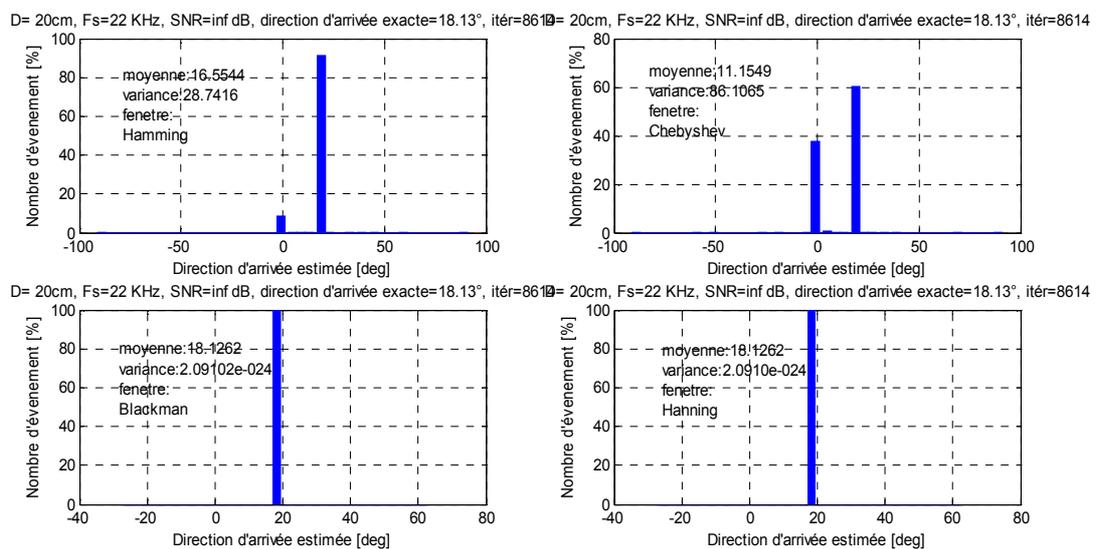


Figure 3.5 : Performance de l'algorithme PHAT-GCC en fonction du type de la fenêtre utilisée.

Observation

Meilleures performances ont été observées lors de l'utilisation des fenêtres Hanning et Blackman (100%) par rapport à celles obtenus par l'utilisation des autres fenêtres, Hamming (90%) et Chebyshev (60%).

Afin de justifier la différence entre les résultats obtenus, les spectres d'amplitudes des quatre fenêtres employées sont tracés, voir figure (3.6).

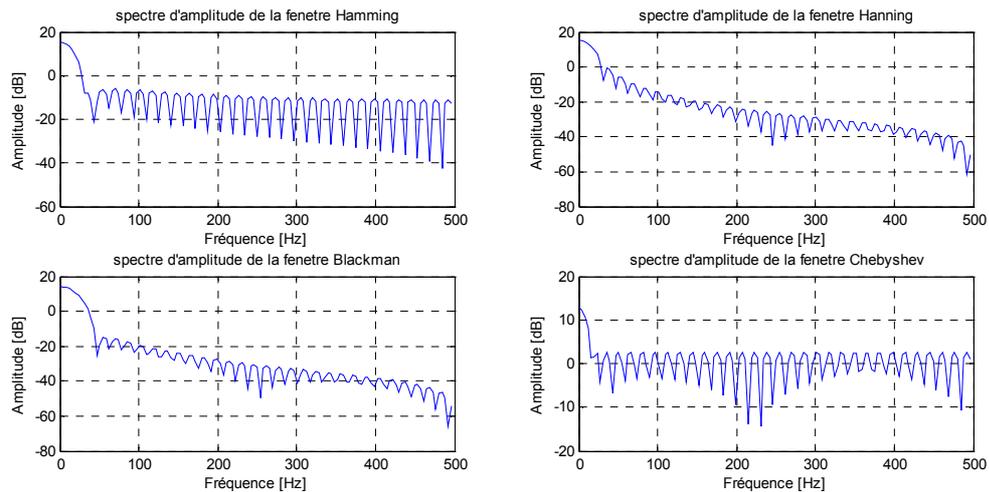


Figure 3.6 : Spectres d'amplitude des fenêtres (Hamming, Hanning, Blackman et Chebyshev).

Les spectres d'amplitude des fenêtres montrent une différence au niveau de la pente de décroissance des lobes secondaires ; dans Hanning et Blackman environ -18dB/oct [26], mais, par contre un comportement (equiripple) est observé dans le cas de la fenêtre Chebyshev[20] et une pente de -6dB/oct dans Hamming [26].

Résultat

L'amplitude des lobes secondaires de la fenêtre utilisée dans l'algorithme PHAT-GCC influe considérablement sur la qualité de la localisation. Les fenêtres de Hanning et Blackman présentent une meilleure performance par rapport aux autres fenêtres.

Une fenêtre de Hanning sera utilisée par la suite dans l'algorithme PHAT-GCC. L'exécution de cet algorithme a mené aux résultats illustrés dans la figure ci-dessous :

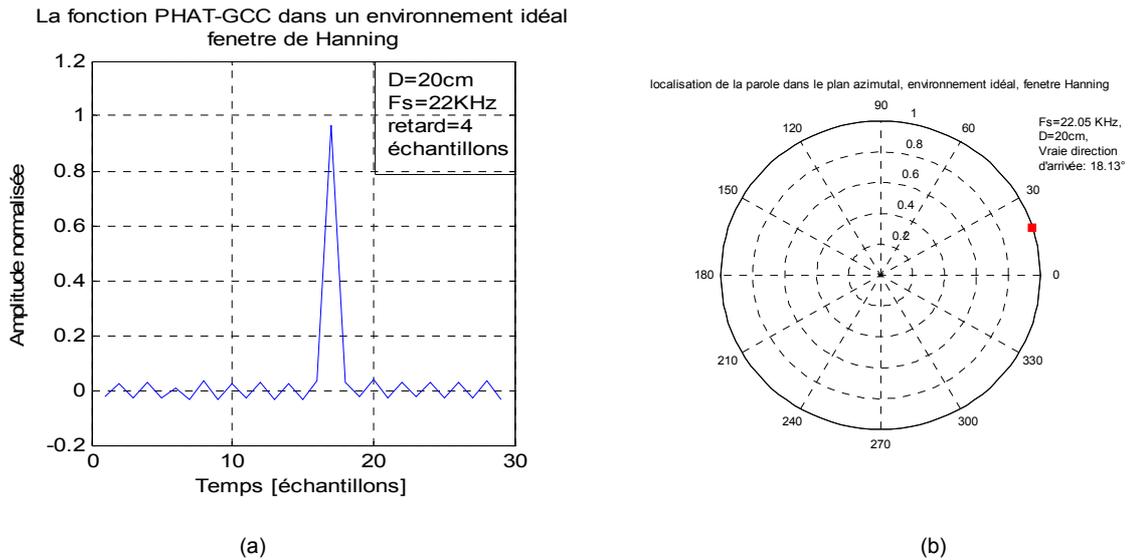


Figure 3.7 : Résultats de la localisation du locuteur dans un environnement idéal.

La figure (3.9.a) représente la fonction PHAT-GCC; Le grand pic correspond à un retard $TDOA_1 + TDOA_2 + 1$, ici $TDOA_2 = 13$ échantillons (fixé par l'utilisateur), alors $TDOA_1 = 17$ (indiqué par le pic) - 13 ($TDOA_2 + 1$) = 4 échantillons. Selon le Tableau 3.2, 4 échantillons correspond à une $DOA = 18.13^\circ$. $\theta = \text{asin} \left(\left(\frac{u}{F_s} \right) * \left(\frac{c}{D} \right) \right)$.

- u est l'index correspondant au pic.
- F_s est la fréquence d'échantillonnage.
- c est la vitesse du son (343m/s).
- D est la distance intra capteurs.

Echantillon	1	2	3	4	5	6	7	8	9	10	11	12	13
Angle	4.46°	8.94°	13.49°	18.13°	22.88°	27.81°	32.98°	38.48°	44.42°	51.05°	58.82°	68.96°	90°

Tableau 3.2 : La résolution dans le cas : $D=20\text{cm}$, $F_s=22050\text{ Hz}$.

La figure (3.9.b) représente la position de la source dans le plan azimuthal, le 0° représente le cas broad-side et le 90° et 270° représentent les cas end-fire. L'atténuation en puissance n'est pas considérée dans ce cas.

3.3.4.2. Localisation du locuteur dans un environnement réel

Le PHAT-GCC est un algorithme d'estimation de la direction d'arrivée des sources audio qui a prouvé ses performances vis-à-vis les changements spectraux d'un signal parole. Est-ce qu'il peut performer de la même manière dans un environnement acoustique réel ?, cet environnement réel est un espace réservé dans un laboratoire, dégagé de toute source de multi trajets.

Expérience 1 : Localisation d'une source de parole

L'expérience décrite dans la figure (3.8) est réalisée afin de tester les performances de l'algorithme PHAT-GCC vis-à-vis le canal acoustique réel. Le même signal de parole utilisé précédemment est généré par l'ordinateur émetteur, les distances intra-microphones et source-réseau sont spécifiées selon [9].

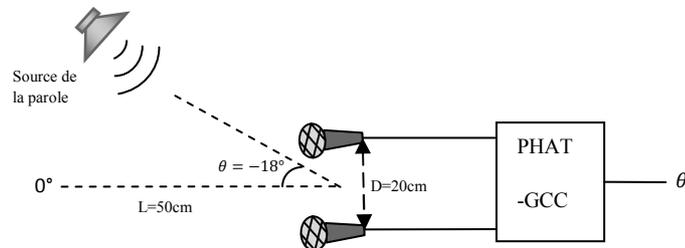


Figure 3.8 : Configuration du matériel pour la localisation d'une source de parole dans un environnement acoustique réel.

Observation

Initialement, l'algorithme donne des résultats erronés, le pic indiquant le retard est complètement noyé dans le bruit. En faisant augmenter la puissance du signal de test et déplacer la source, le pic devient identifiable parmi les autres pics grâce à ses mouvements, voir figure (3.9).

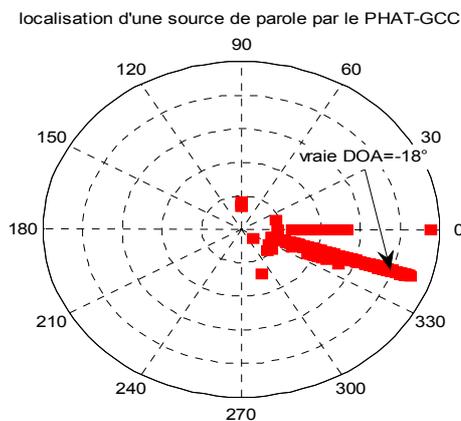


Figure 3.9 : Résultat de la localisation de la parole dans un environnement réel.

Une variation de la puissance du pic indiquant la direction d'arrivée est observée au cours de l'exécution de l'algorithme. L'atténuation de ce pic provoque une mauvaise décision.

La figure (3.10) représente deux échantillons de la fonction PHAT-GCC prise à deux instants différents. Dans le premier échantillon, un pic important d'amplitude

égale à 0.55 est observé par contre, son amplitude diminue jusqu'à 0.21 dans le deuxième échantillon.

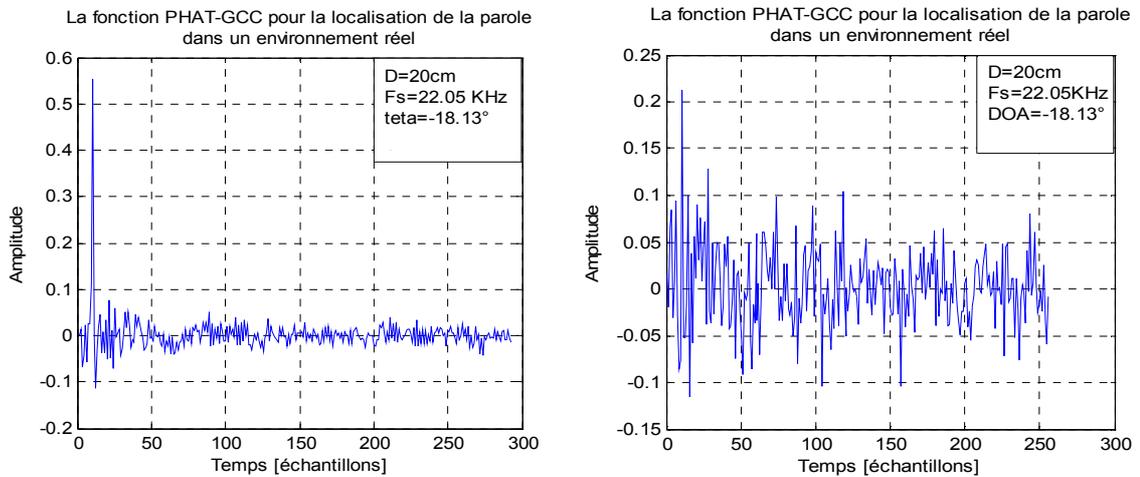


Figure 3.10 : Deux échantillons de la fonction PHAT-GCC prise à des instants différents.

Environ 220 réalisations successives de l'algorithme PHAT-GCC ont été effectuées et le résultat obtenu est présenté dans la figure (3.11).

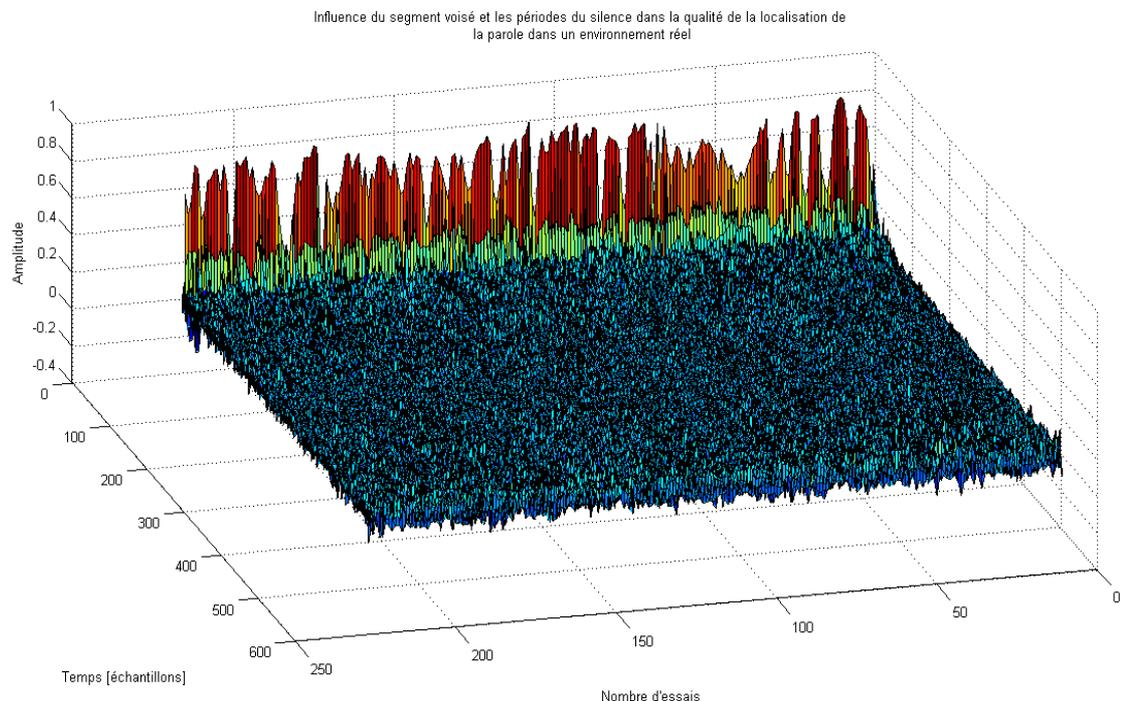


Figure 2.11 : Résultat de la localisation du locuteur dans un environnement réel.

Une variation de la puissance du pic concerné par la localisation est observée le long de ces réalisations.

Hypothèse

Etant donné que le signal parole est composé de trois segments de caractéristiques spectrales différentes, soit le segment du signal voisé, le segment du signal non voisé, et le segment du silence. L'algorithme se manifeste différemment dans ces trois régions.

Puisque la reproduction du signal parole peut être modélisée par deux générateurs de signaux ; un générateur d'un signal périodique pour modéliser les signaux voisés et un générateur d'un signal aléatoire pour modéliser les signaux non voisés [27]. Alors, on vérifiera le comportement de l'algorithme vis-à-vis un bruit aléatoire, vis-à-vis un signal périodique, et en absence de source audio.

Expérience 2 : Cas du signal non voisé

Un bruit blanc gaussien de moyenne $\mu = 0$, et d'une variance $\sigma^2 = 0.1$, est généré par l'ordinateur émetteur. Sa fonction d'auto-corrélation est illustrée dans la figure (3.12).

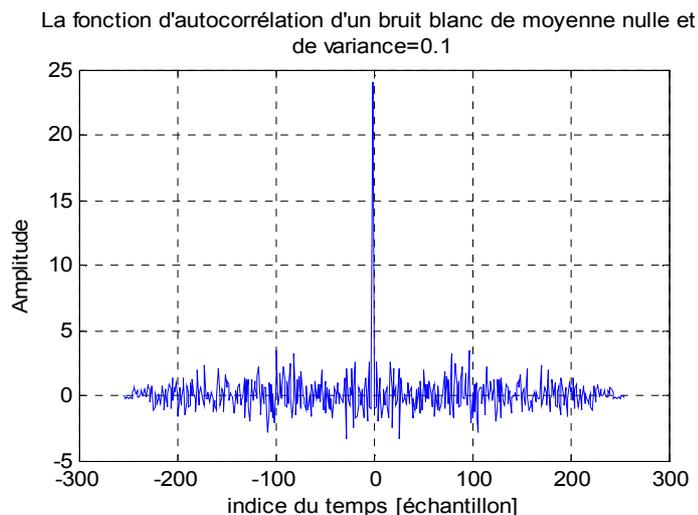


Figure 3.12 : Fonction d'auto-corrélation d'un bruit blanc centré.

Par définition, le bruit blanc est un signal idéalisé caractérisé par une densité spectrale de puissance constante sur une étendue infinie [26]. Sa fonction d'autocorrélation est une impulsion de Dirac à l'origine. Plus la nature d'un signal est aléatoire plus sa fonction d'autocorrélation se rapproche d'une impulsion. Cette propriété est mise à profit pour déterminer le retard entre un bruit et une version retardée de celui-ci. La position du maximum de leur fonction d'intercorrélation est une très bonne estimation du retard entre les deux signaux.

La disposition du réseau et la source est montrée dans la figure (3.13).

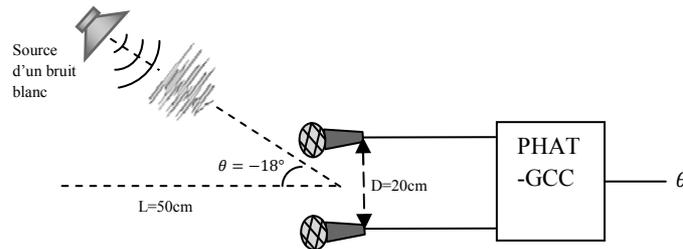


Figure 3.13 : Configuration utilisée pour la localisation d'une source de bruit blanc.

Observation

L'échantillon de la fonction PHAT-GCC illustré dans la figure (3.14) montre un pic de puissance importante dans le retard correspondant à la vraie direction d'arrivée.

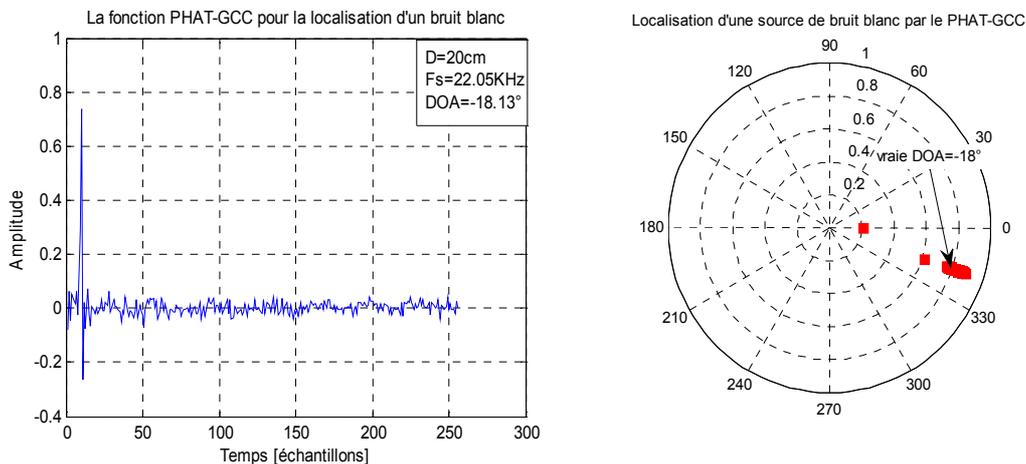


Figure 3.14 : Un échantillon de la fonction PHAT-GCC pour la localisation d'un bruit blanc.

Plusieurs réalisations successives de l'algorithme dans les mêmes conditions ont menées au résultat illustré dans la figure (3.15).

Tous les pics indiquant le retard correspondant à l'angle de provenance du bruit ayant une puissance constante en fonction du temps et qui est très importante par rapport aux puissances des autres pics. Donc, il n'ya pas lieu de confusion entre le trajet direct et les autres trajets.

Résultat

La nature aléatoire du bruit blanc cause l'apparition d'un pic d'intercorrélacion très étroit et plus important en terme de puissance, ce qui rendre

la détection du maximum beaucoup plus exacte, donc une bonne estimation de la direction d'arrivée de la source.

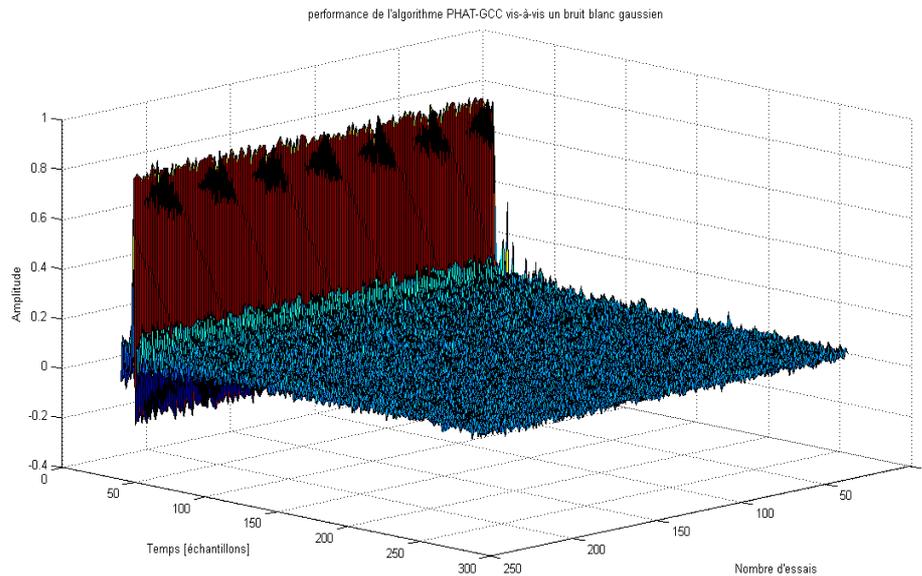


Figure 3.15 : Résultat de la localisation d'une source de bruit blanc.

Expérience 3 : Cas du signal visé

Soit un signal sinusoïdal de fréquence $f = 500 \text{ Hz}$ généré par l'ordinateur émetteur. A la réception, le PHAT-GCC est exécuté. La figure (3.16) illustre l'arrangement du matériel.

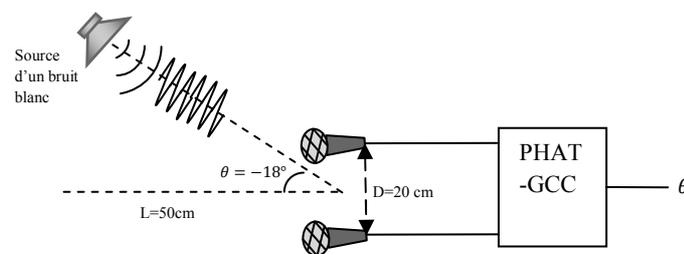


Figure 3.16 : Configuration utilisée pour la localisation d'une source de signal périodique.

Théoriquement, pour un signal périodique réel $x(t)$ de période T_0 , on définit la corrélation de la manière suivante :

$$R_{xx}(t) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(\tau)x(\tau - t) d\tau \quad (3.9)$$

Tous les signaux périodiques peuvent s'exprimer sous la forme d'un développement en série de Fourier

$$x(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos 2\pi n F_0 t + b_n \sin 2\pi n F_0 t) \quad (3.10)$$

En appliquant la relation précédente dans la définition de la corrélation, nous obtenons la fonction d'autocorrélation suivante :

$$R_{xx}(t) = a_0^2 + \sum_{n=1}^{\infty} (a_n^2 + b_n^2) \cdot \cos(2\pi n F_0 t) \quad (3.11)$$

Nous pouvons effectivement conclure que la fonction d'autocorrélation conserve l'information fréquence, mais pas l'information phase [28] [29]. Le signal d'autocorrélation possède toutes les fréquences comprises dans le signal initial et uniquement ces fréquences. Par contre la représentation de la fonction d'autocorrélation est d'autant plus distordue que le signal est riche en harmoniques étant donné que l'addition de ces signaux périodiques se fait sans cohérence de phase.

L'exécution du PHAT-GCC a mené au résultat présenté dans la figure ci-dessous :

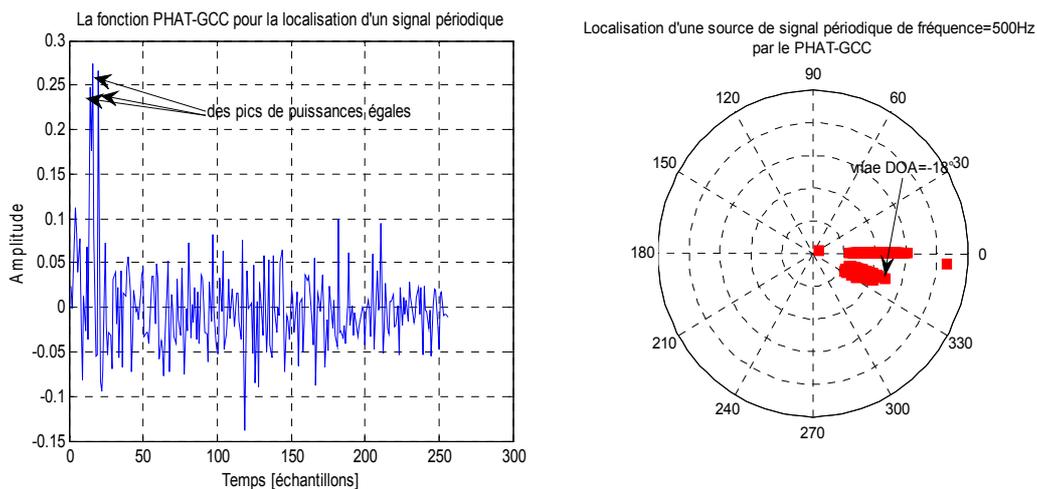


Figure 3.17 : Un échantillon de la figure PHAT-GCC pour la localisation d'un signal périodique.

Observation

Des pics très proches de puissances presque égales et moins importantes (environ 0.25) que celles observées dans le cas d'une source de bruit blanc sont aperçus, une mauvaise localisation aura lieu.

La fonction PHAT-GCC est tracée pour environ 220 réalisations successives et la figure (3.18) montre le résultat obtenu.

Le pic qui indique la direction d'arrivée de la source n'est pas identifiable parmi d'autres pics ayant presque la même puissance, une mauvaise localisation est alors obtenue dans toutes les réalisations.

Résultat

Le signal utilisé est un signal à bande étroite, sa fonction d'autocorrélation présente un pic large est de faible puissance. Le filtre de blanchiment PHAT s'efforce de rendre ce pic étroit et détectable, mais malgré ça une confusion aura lieu. Par conséquent, une mauvaise localisation est obtenue, ce qui est conforme au résultat théorique prouvé précédemment.

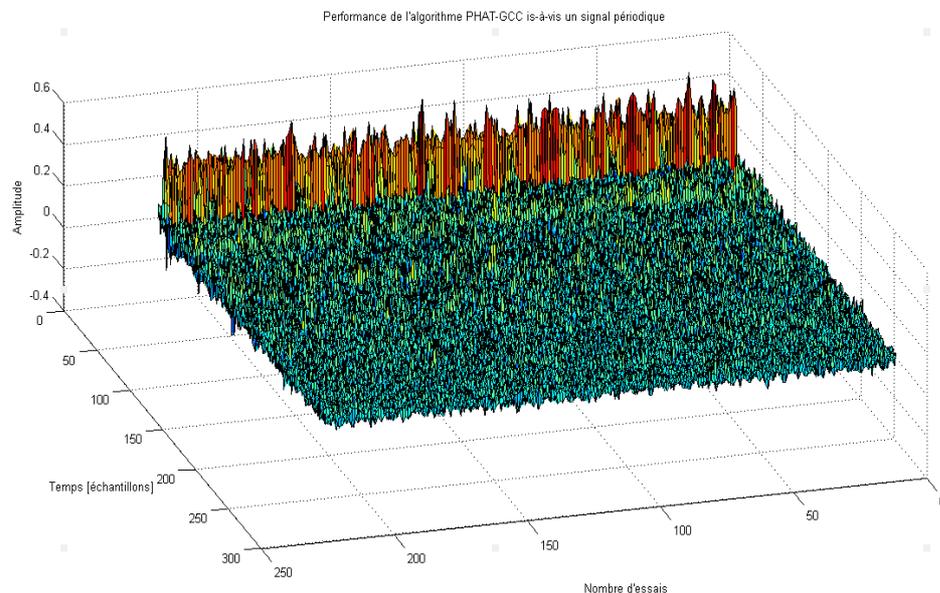


Figure 3.18 : Résultat de la localisation d'un signal périodique.

Expérience 4 : Cas du non activité vocale

Pendant les périodes de silence, en absence de toute source acoustique l'algorithme est exécuté. La figure (3.19) représente un échantillon de la fonction PHAT-GCC.

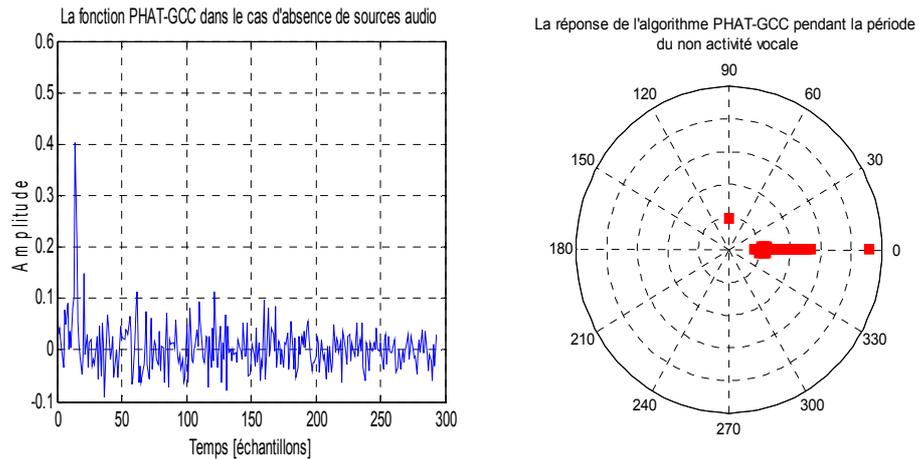


Figure 3.19 : Un échantillon de la fonction PHAT-GCC pris pendant une période du non activité vocale.

Observation

Des pics de puissance moyenne indiquant des angles d'arrivées aléatoires sont aperçus, Ces pics peuvent être dus au bruit ambiant, le bruit thermique interne, les réverbérations et les lobes secondaires des fenêtres et ceci peut être une des causes qui perturbe le processus de la localisation.

Plusieurs réalisations successives de l'algorithme dans les mêmes conditions ont menées au résultat présenté dans la figure ci-dessous.

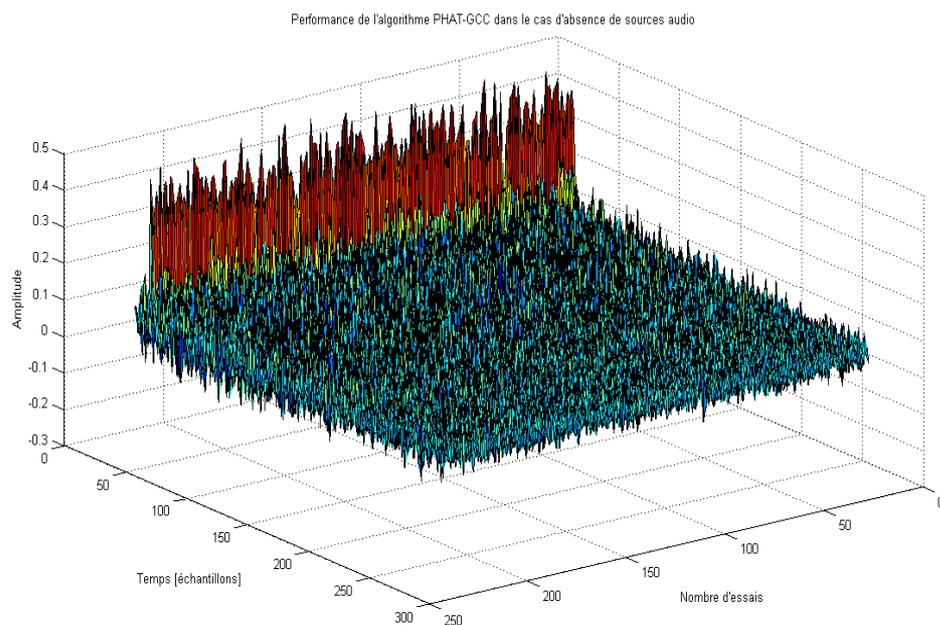


Figure 3.20 : Résultat de la localisation pendant les périodes du non activité vocale.

Pour toutes les réalisations, un pic plus ou moins important indiquant une direction d'arrivée 0° est observé avec un grand pourcentage comme le prouve l'histogramme de la figure (3.21).

Résultat

L'hypothèse supposée dans le paragraphe (3.3.4.2) a été vérifiée. L'algorithme PHAT-GCC donne un bon estimé de la direction d'arrivée d'une source de la parole pendant le segment du signal non voisé dû à la nature aléatoire de celui-ci et un mauvais estimé pendant la période du segment voisé présentant une périodicité ainsi que pendant les périodes du non activité vocale.

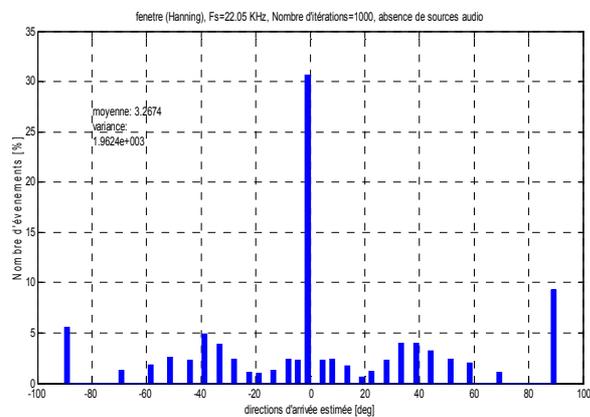


Figure 3.21 : Histogramme montre la distribution des directions d'arrivée dans le cas d'absence des sources audio.

Expérience 5 : Influence des multi-trajets

L'expérience réalisée ici consiste à évaluer le comportement de l'algorithme PHAT-GCC dans le cas de présence d'un obstacle créant des trajets multiples. La configuration illustrée dans la figure (3.22) indique la disposition de l'obstacle par rapport à la source et par rapport aux microphones. Une source de bruit blanc est employée car ce signal a une densité de puissance constante, donc le signal obtenu par corrélation caractérise uniquement l'environnement de propagation.

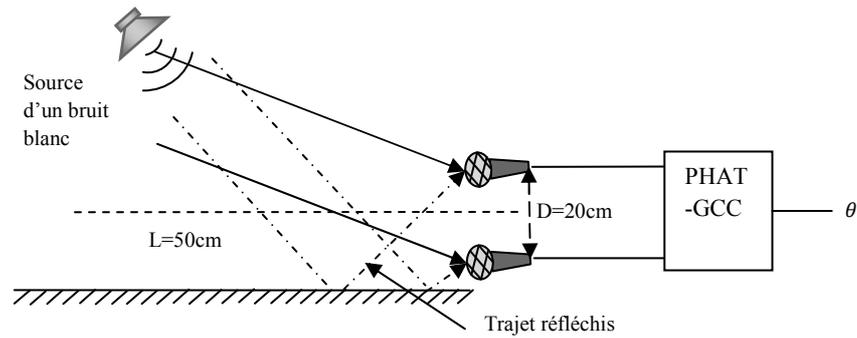


Figure 3.22 : Configuration du matériel pour caractériser l'effet des multi-trajets.

La figure (3.23) illustre un échantillon de la fonction PHAT-GCC après l'exécution de l'algorithme.

Observation

Un pic correspondant au trajet direct et d'autres pics correspondants aux trajets réfléchis sont aperçus. Leurs puissances sont très proches ce qui génère une source d'ambiguïté dans la localisation.

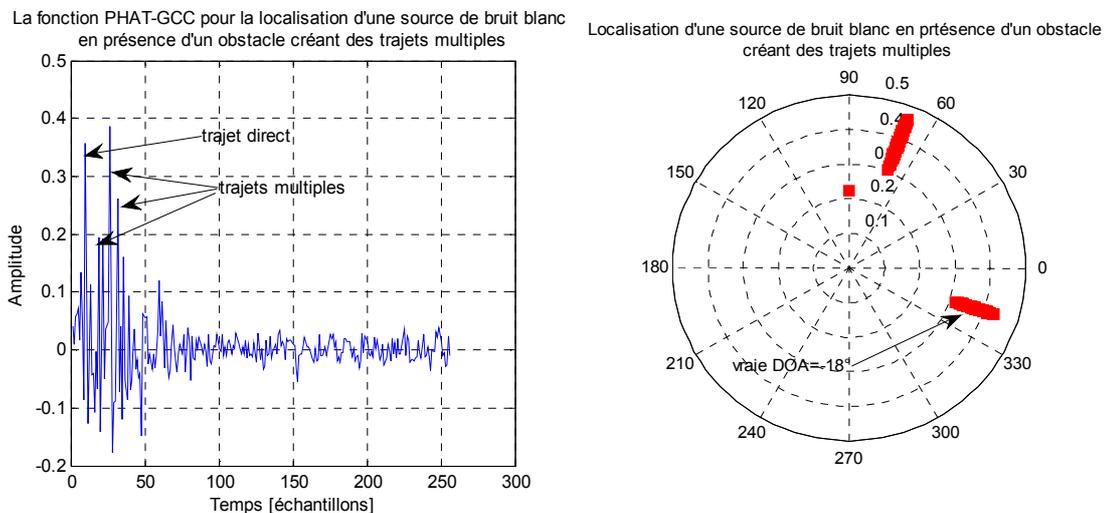


Figure 3.23 : Un échantillon de la fonction PHAT-GCC pour la localisation d'une source de bruit blanc en présence des multi-trajets.

Pour plusieurs réalisations successives de l'algorithme, la figure (3.24) montre le résultat obtenu.

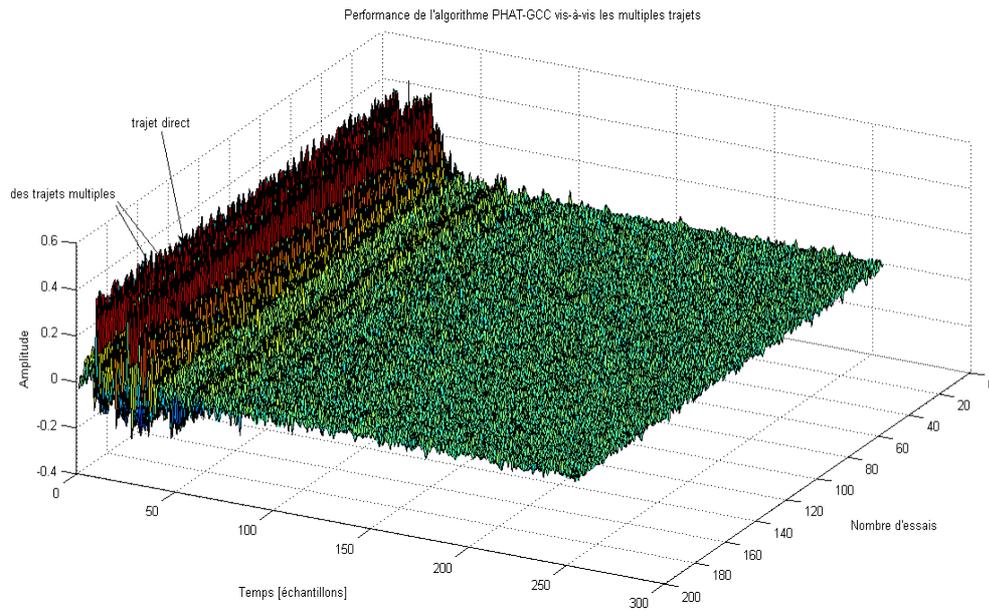


Figure 3.24 : Résultat de la localisation en présence des trajets multiples.

Le PHAT-GCC se manifeste également dans toutes les réalisations.

Résultat

La présence de plusieurs pics de puissances très proches dans la fonction PHAT-GCC pose un problème dans la détection du pic correspondant à l'angle d'arrivée de la source. Le phénomène des multi-trajets crée une ambiguïté dans la localisation.

Expérience 6 : Influence des multiples sources

Dans cette expérience, l'objectif est d'évaluer la capacité de l'algorithme PHAT-GCC à localiser une source désirée en présence d'une autre source identique. Pour cela la configuration illustrée dans la figure (3.25) est utilisée.

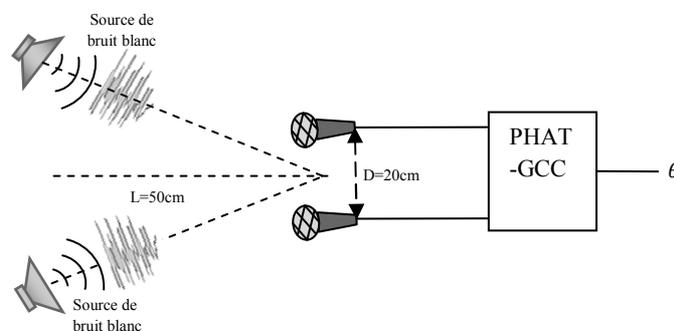


Figure 3.25 : Configuration utilisée pour caractériser l'effet de multiples sources sur le PHAT-GCC.

Après l'exécution de l'algorithme, le résultat présenté dans la figure (3.26) est obtenu.

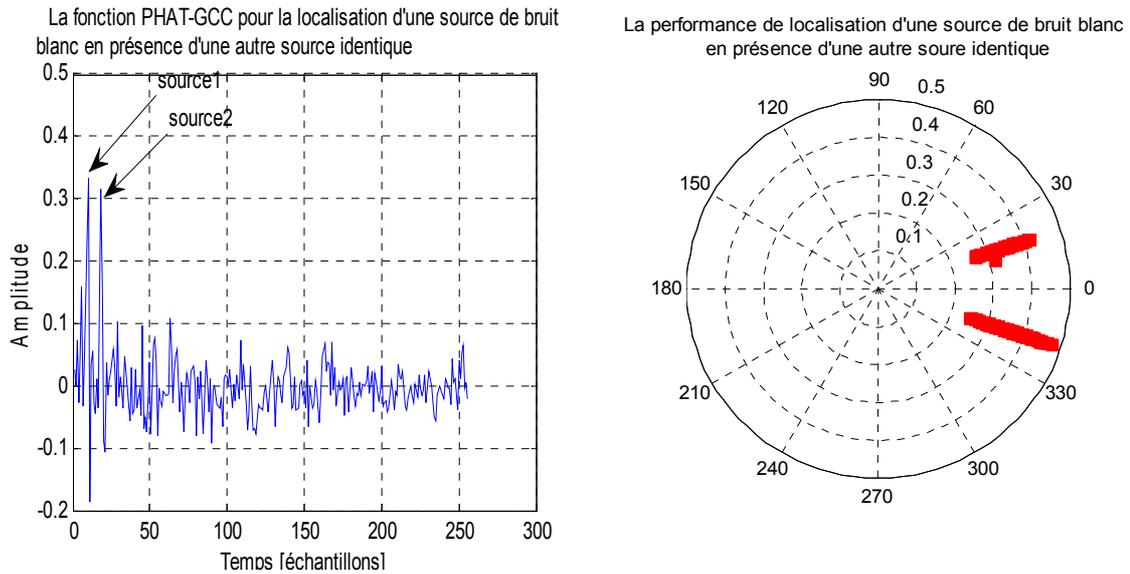


Figure 3.26 : Un échantillon de la fonction PHAT-GCC pour la localisation de multiples sources.

Observation

Deux pics de puissances importantes sont observés, chacun d'eux indique la direction d'arrivée d'une source.

L'exécution successive de l'algorithme a mené au résultat montré dans la figure (3.27):

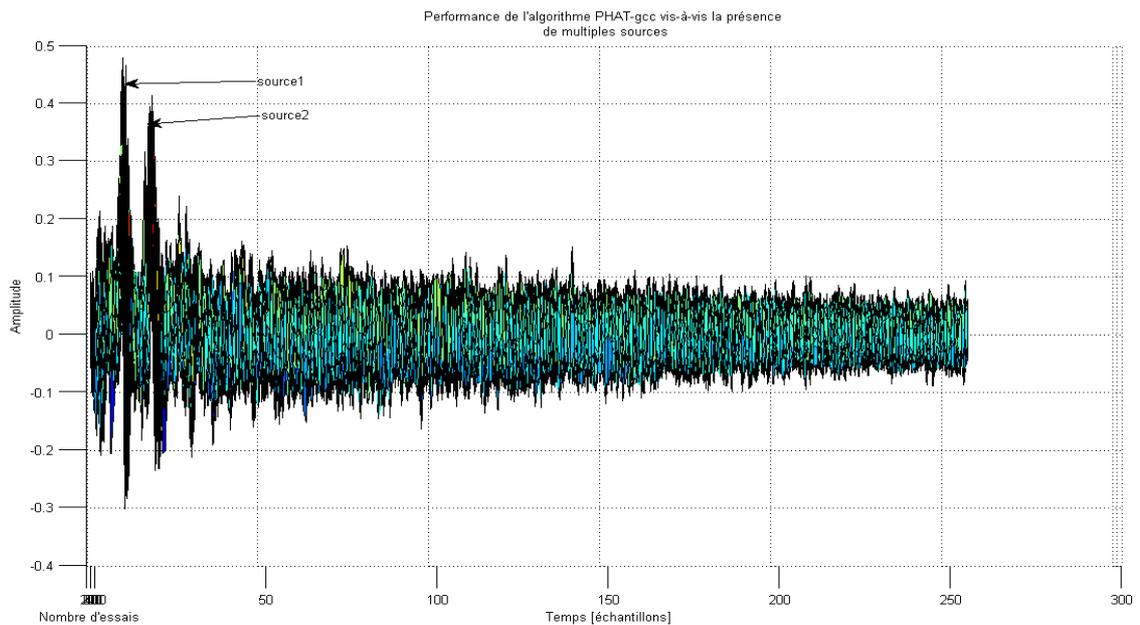


Figure 3.27 : Résultat de la localisation en présence des sources multiples.

Les deux sources peuvent être localisées à la fois, mais l'attribution de chaque pic à sa propre source pose une ambiguïté.

Résultat

La présence de plusieurs sources audio dans le même champ de localisation présente un problème pour l'algorithme PHAT-GCC. Chacune des sources cause l'apparition d'un pic, et si les sources sont de même nature et ayant des puissances égales ces pics vont avoir la même puissance, donc la détection du maximum sera ambiguë.

Expérience 7 : Seuillage

En vue de remédier au problème de la mauvaise localisation pendant les périodes du signal voisé et du silence, un seuil de détection peut être fixé comme l'illustre la figure (3.28). Si la puissance du pic dépasse le seuil fixé, le retard correspondant est considéré, sinon l'ancienne valeur est préservée en supposant que le locuteur n'a pas changé d'emplacement.

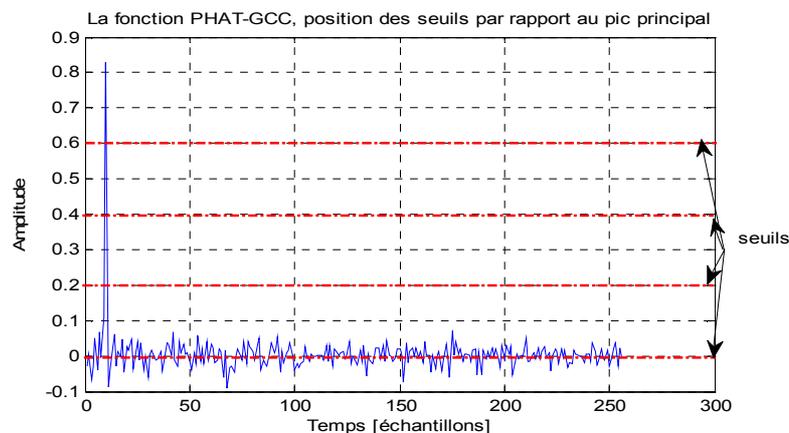


Figure 3.28 : Effet du seuil.

Alors, pour fixer cette valeur de seuil, l'histogramme de la figure (3.29) est tracé pour les quatre seuils suivants : 0, 0.2, 0.4 et 0.6.

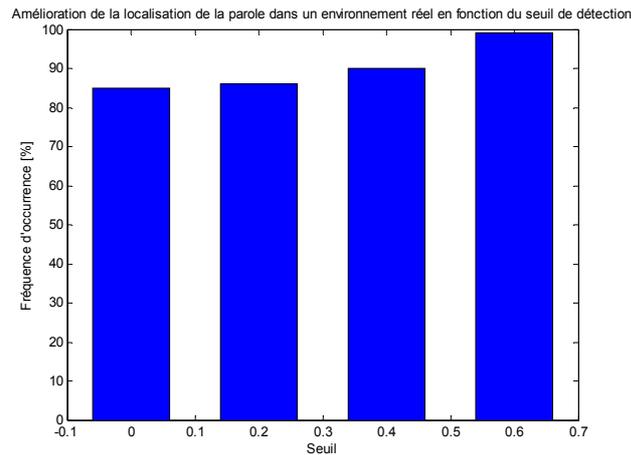


Figure 3.29 : Histogramme montre l'amélioration de la performance du PHAT-GCC.

Observation

Pour le même nombre de réalisations, la fréquence d'occurrence de la vraie direction d'arrivée est calculée en pourcentage et comme c'est attendu le seuil de 0.6 donne le maximum de performance.

Résultat

La performance de la localisation s'améliore en fonction du niveau de la détection du pic. Mais malheureusement ça peut causer des erreurs de localisation si le locuteur change sa position et baisse sa voix.

3.3.4.3. Choix des paramètres et amélioration de la localisation

D'après les expériences précédentes, la fenêtre de Hanning sera utilisée, la longueur de la fenêtre d'analyse sera fixée selon le déplacement du locuteur, ici on suppose qu'il se déplace tout les 23 ms. La taille de la *FFT* est fixée selon [30] $N_{FFT} = 512$ pour éviter le zero padding. Le niveau de la détection du pic sera pris à 0.6. Pendant les périodes de silence et les segments correspondant aux sons voisés, le locuteur est supposé être dans le même emplacement précédent et une nouvelle valeur sera prise dès que le pic dépasse le seuil de détection précisé.

Un espace d'expérimentation dégagé de toute source de multi-trajets doit être réservé.

Le schéma de la figure (3.30) illustre l'algorithme PHAT-GCC comme c'est conçu sous l'outil simulink du MATLAB.

LOCALISATION D'UNE SOURCE AUDIO DANS UN ENVIRONNEMENT REEL PAR L'ALGORITHME PHAT-GCC

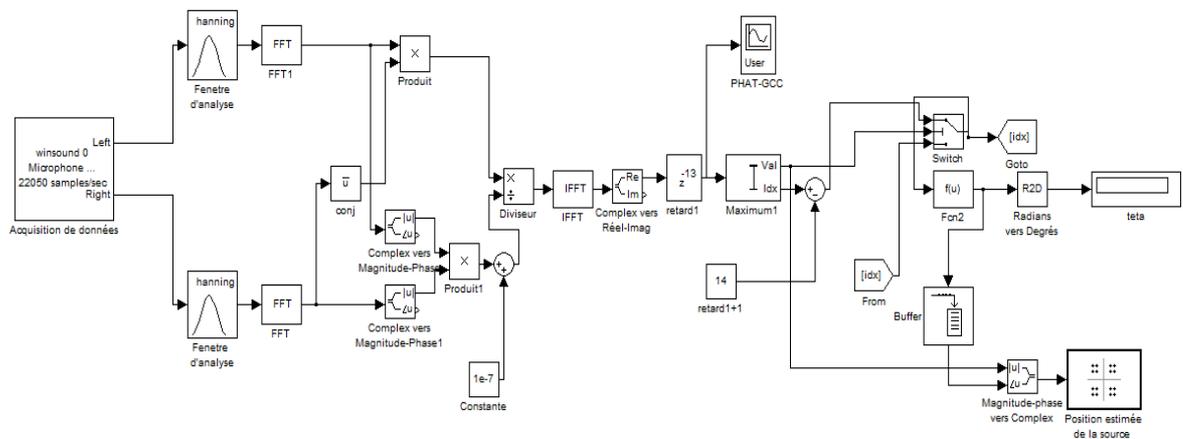


Figure 3.30 : Chaîne de la localisation du locuteur sous l'environnement simulink.

Et le tableau (3.3) résume les valeurs des paramètres choisis :

Durée de la trame d'analyse	23ms
NFFT	512
Type de la fenêtre d'analyse	Hanning
Distance intra-microphones	20cm
Distance sources-réseau	50cm
Fréquence d'échantillonnage	22050Hz
Seuil de détection	0.6

Tableau 3.3 : Paramètres du système de la localisation.

Rappel

L'objectif visé consiste à localiser deux sources audio, dont une est désirée et l'autre est interférente, puis essayer d'éliminer spatialement la source interférente en tirant profit uniquement de leurs directions d'arrivée. Donc, il faut trouver une source de bruit dont sa présence n'influe pas sur la qualité de la localisation.

Expérience 8 : Localisation d'une source de la parole en présence d'un bruit blanc

Cette expérience vise à vérifier la performance de l'algorithme PHAT-GCC à localiser une source de la parole en présence d'une autre source qui génère un bruit blanc gaussien dans le même champ de localisation.

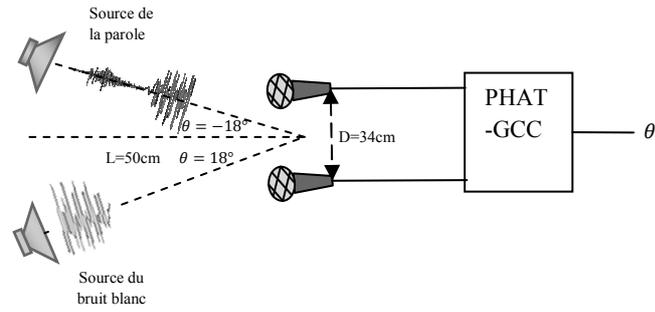


Figure 3.31 : Configuration du système pour la localisation d'une source de parole en présence d'un bruit blanc.

Les résultats obtenus pour un $SNR = 10 \text{ dB}$ sont illustrés dans la figure (3.32) :

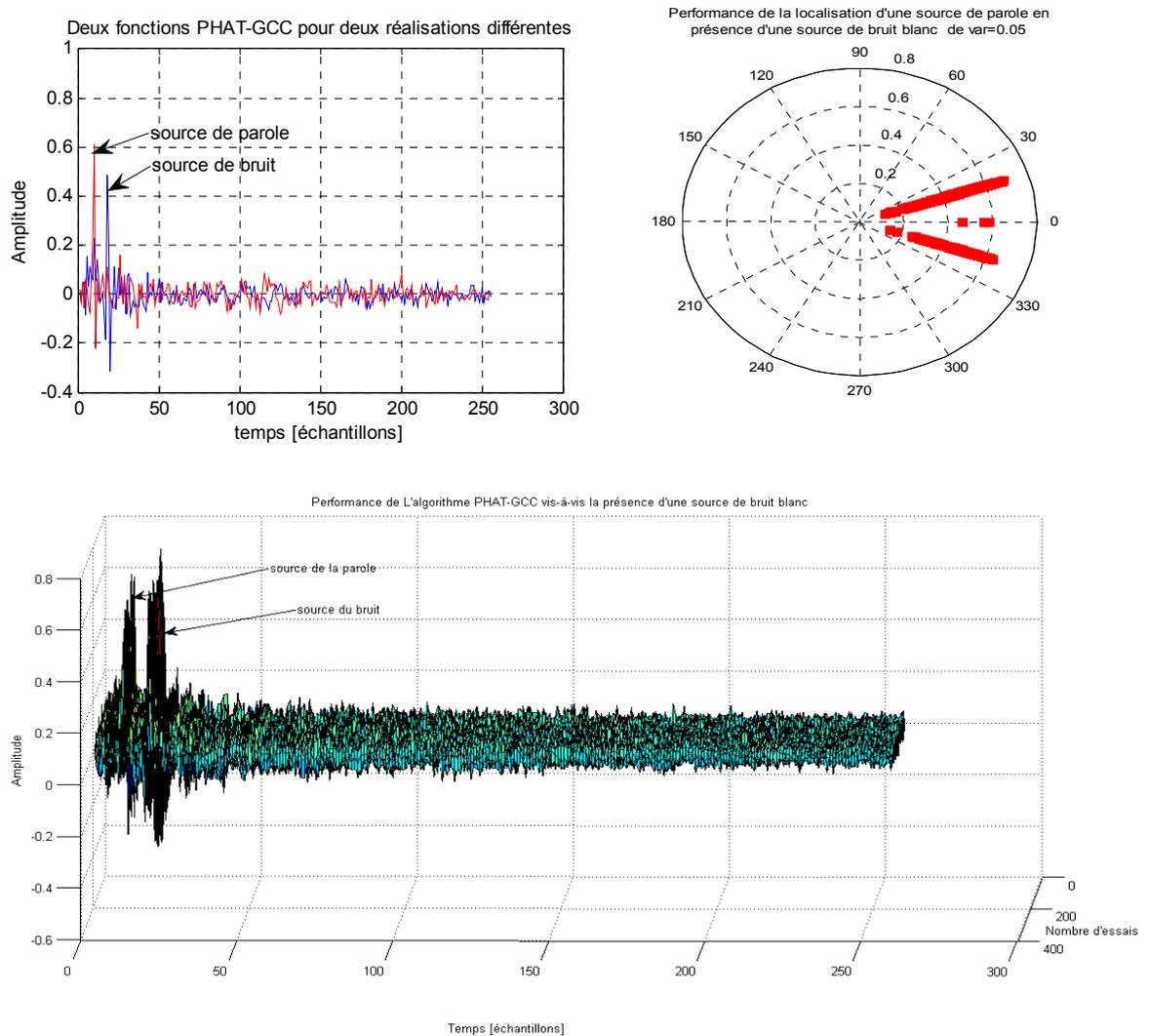


Figure 3.32 : Résultats de la localisation d'une source de la parole en présence d'un bruit blanc.

L'erreur de localisation RMSE : $\delta(\theta) = \sqrt{\frac{1}{K} \sum_{i=1}^K |\hat{\theta}_i - \theta|^2}$ est mesurée pour différentes valeurs de SNR, et la courbe présentée dans la figure (3.33) illustre l'influence de la puissance du bruit sur l'exactitude du résultat de la localisation.

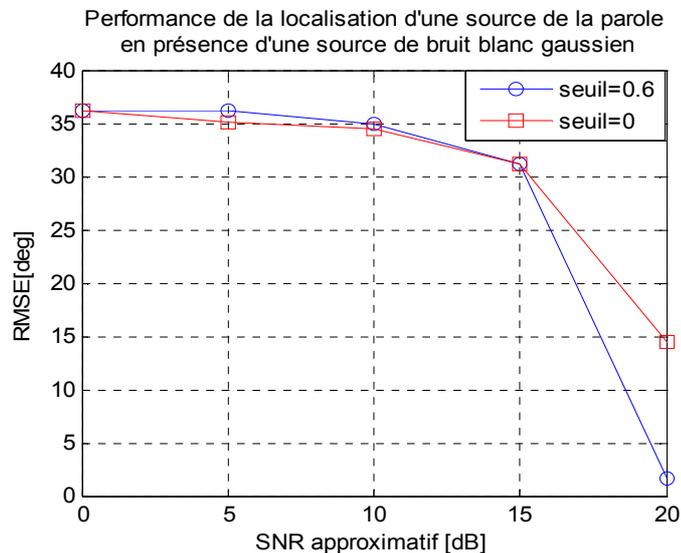


Figure 3.33 : Erreur de localisation en présence d'un bruit blanc gaussien.

Observation

Les deux sources sont localisées, car les deux pics ayant presque la même amplitude qui dépasse le seuil fixé.

Pour un SNR important (20dB), une erreur de 15° est observée dans le cas où aucun seuil n'est fixé, et une erreur de 2° lorsque le seuil est fixé à 0.6. L'erreur reste très importante pour des SNR inférieur à 20dB.

Résultat

La performance de l'algorithme se détériore en présence d'un bruit blanc gaussien et ceci même en fixant le seuil de détection à 0.6.

Expérience 9 : Localisation d'une source de la parole en présence d'un bruit périodique

La même expérience est refaite sauf que cette fois-ci le bruit blanc est remplacé par un bruit périodique, le SNR pris dans ce cas est de 0dB. La figure (3.34) décrit l'arrangement du matériel.

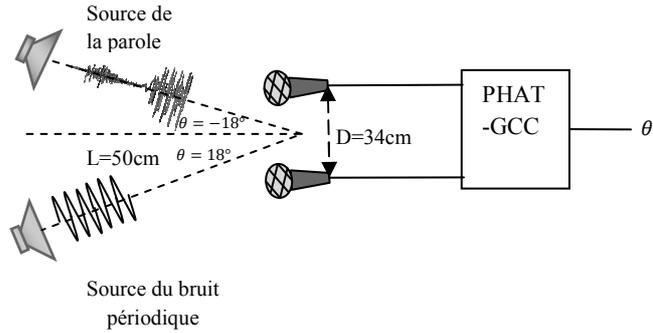


Figure 3.34 : Configuration du système pour la localisation d'une source de la parole en présence d'un bruit harmonique.

L'exécution de l'algorithme a menée aux résultats suivants :

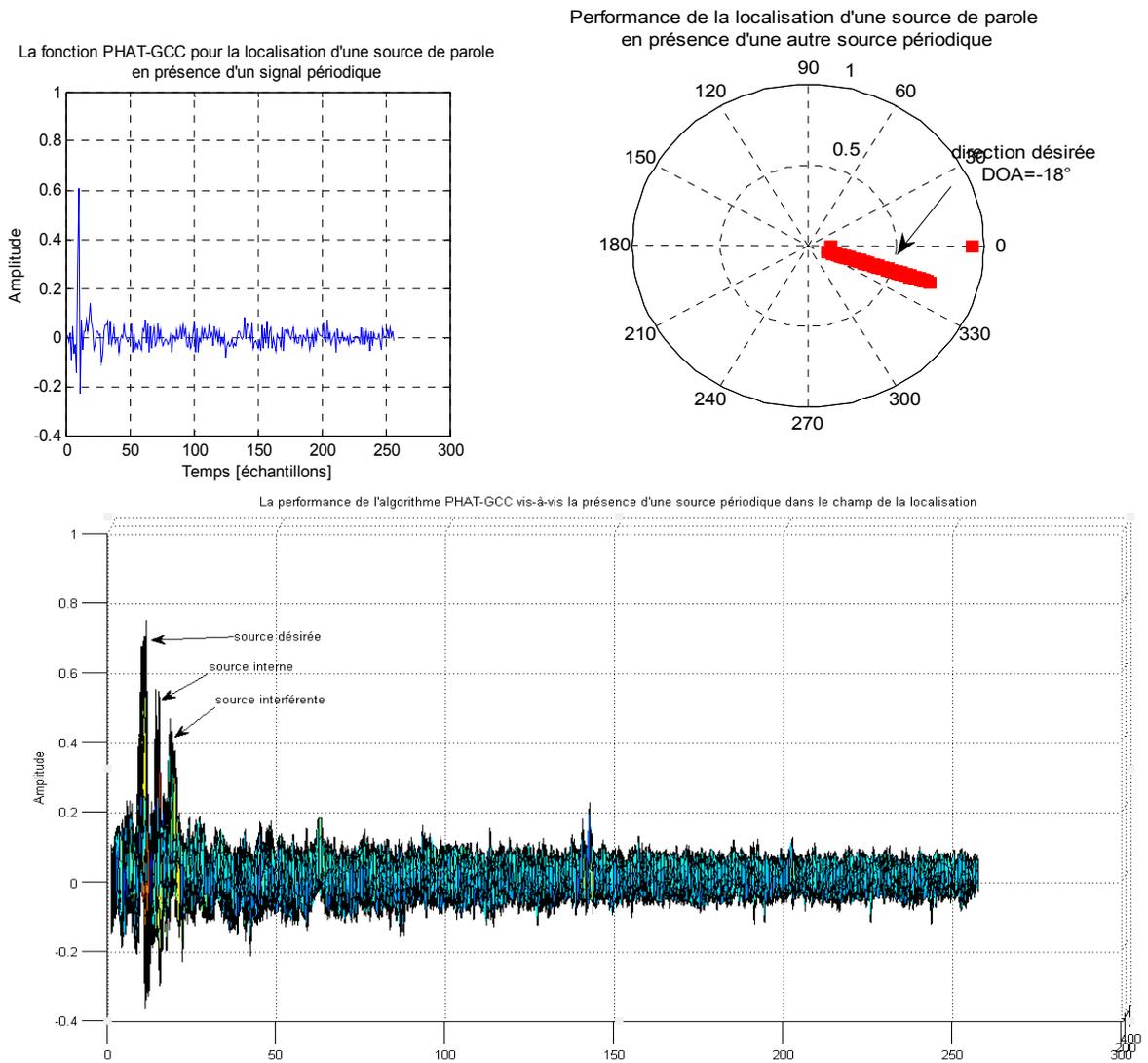


Figure 3.35 : Résultats de la localisation de la parole en présence d'un bruit périodique.

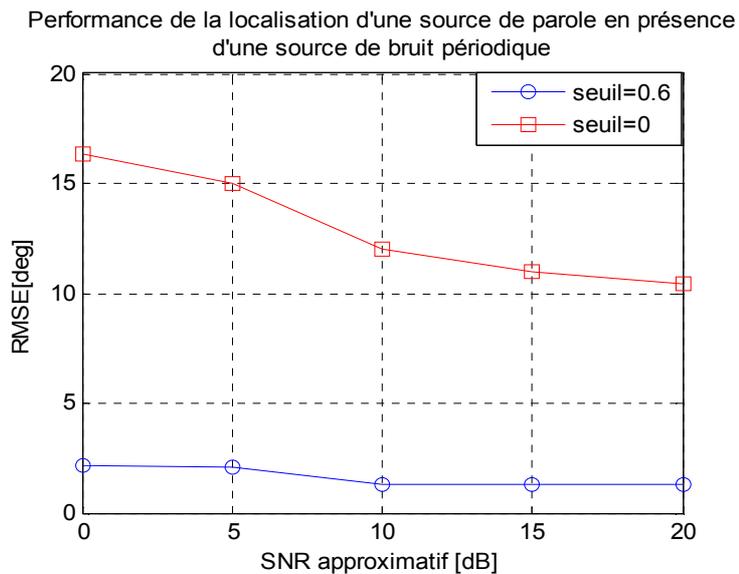


Figure 3.36 : Erreur de localisation en présence d'un bruit périodique.

Observation

On remarque la présence de trois pics indiquant successivement la source désirée, le bruit interne et la source interférente, mais de puissances différentes. Le pic du signal désiré est souvent supérieur au seuil fixé, par contre la puissance des autres pics est inférieure.

L'erreur observée dans ce cas est moins importante que celle observée dans le cas précédent. Elle est minimale de 2° et presque constante quelque ce soit le SNR dans le cas où le seuil est fixé à 0.6.

Résultat

La présence d'un bruit périodique n'influe pas sur la qualité de localisation. Ceci est dû au fait que le pic indiquant le retard n'est pas assez important pour être détecté.

3.3.5. Conclusion

Dans cette partie, l'algorithme de localisation PHAT-GCC basé sur l'estimation du retard temporel TDOA est choisi parce qu'il est simple en termes d'implémentation et temps de calcul, convenable pour les implémentations en temps réel et applicable dans le cas des signaux large bande. Les résultats de la simulation et des tests pratiques ont montrés que cet algorithme souffre de trois sources d'ambiguïté :

- La partie périodique constituant le segment voisé du signal parole.

- Les trajets multiples.
- Et la présence de multiples sources actives simultanément dans le champ de la localisation.

3.4. Formation de voies

3.4.1. Introduction

L'intérêt de l'utilisation d'un réseau de microphones, plutôt qu'un seul est que l'élément isolé a un diagramme de directivité figé dans l'espace (sauf si on le fait tourner mécaniquement). En disposant de plusieurs éléments dans l'espace, et en pondérant chacun d'eux, on peut jouer sur les retards entre les signaux reçus par les différents capteurs pour obtenir une répartition spatiale du gain, variable avec la loi de la pondération.

On peut même, en contrôlant la loi de la pondération du réseau, orienter le maximum du gain vers un angle choisi. En changeant l'orientation de ce maximum, on effectue un balayage de l'espace sans faire tourner le réseau mécaniquement. La figure (3.37) illustre la différence entre l'utilisation d'un seul ou plusieurs microphones.

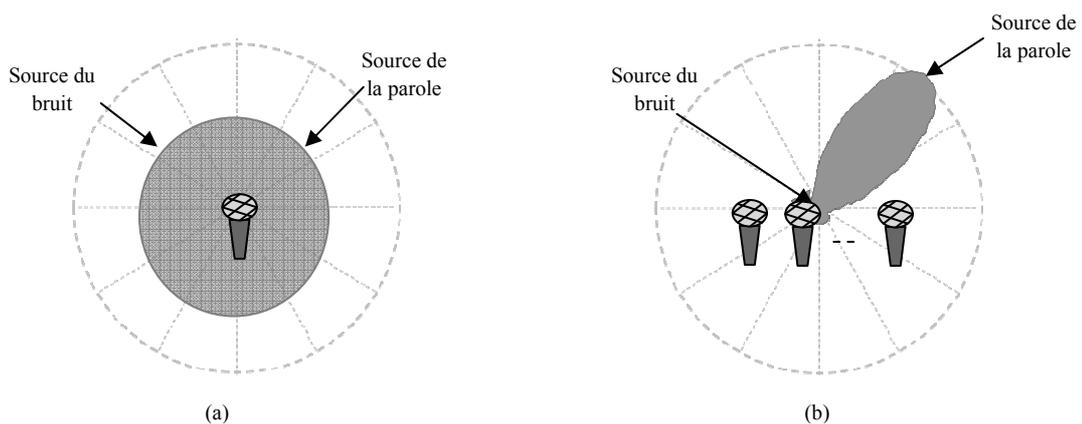


Figure 3.37 : Diagramme de directivité ;(a) un seul microphone, (b) un réseau de microphones.

3.4.2. Formulation de la réponse spatiale du réseau

Soit un réseau de deux microphones de pas D (spécifié ultérieurement). La distance entre la source et les microphones est de $L > 2D^2/\lambda$. On se place dans l'hypothèse d'une propagation libre en champ lointain. Les signaux captés par les microphones supposés être à bande étroite, séparément accessibles, sont pondérés puis sommés pour fournir le signal résultant. Ce signal est sensé être le plus proche possible du signal envoyé par la source utile notée $s(t)$. Le signal $s(t)$ émis par la source est transformé par le milieu de propagation avant d'être capté.

Cette transformation, selon les hypothèses faites précédemment, s'écrit suivant le vecteur directionnel $d(\theta, \omega) = \begin{bmatrix} 1 & e^{j\frac{2\pi}{\lambda}D \sin(\theta)} \end{bmatrix}$.

La dimension de ce vecteur est égale au nombre de microphones. Le signal reçu au niveau du premier microphone est considéré arrivé sans retard et sans atténuation. Le signal reçu au niveau du deuxième microphone est considéré reçu sans atténuation et avec un retard qui est fonction de la distance intra-microphones et la direction d'arrivée.

3.4.3. L'ambiguïté spatiale

En vue de spécifier la distance de séparation optimale entre les éléments du réseau, le diagramme de directivité de ce réseau est tracé pour différentes distances D .

Simulation 2 : Ambiguïté spatiale

Pour le tracé du diagramme de directivité on considère le réseau de deux microphones de la figure suivante :

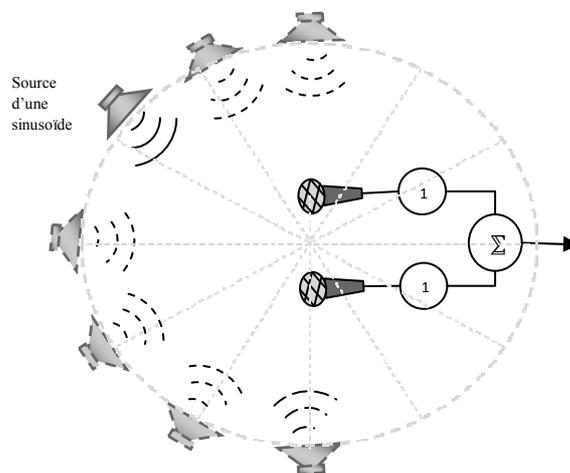


Figure 3.38 : Configuration utilisée pour le tracé du diagramme de directivité d'un réseau de deux microphones.

On génère un signal sinusoïdal par une source située dans le champ lointain. Cette source sera déplacée autour du réseau en balayant toutes les directions.

En faisant varier la distance D entre les deux microphones, les diagrammes de directivité suivants sont obtenus par simulation:

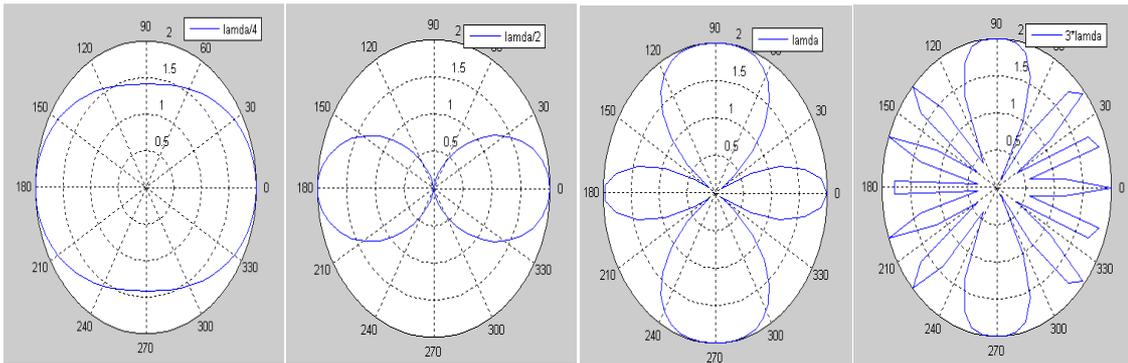


Figure 3.39 : Diagramme de directivité d'un réseau à deux microphones, (a) $D=\lambda/4$, (b) $D=\lambda/2$, (c) $D=\lambda$, (d) $D=3\lambda$.

Observation

Lorsque la distance de séparation entre les éléments du réseau augmente, la largeur du lobe principal dans le diagramme de directivité diminue et le phénomène des lobes images (gratings lobes) apparaît.

Résultat

Le rétrécissement de la largeur du lobe principal est avantageux en termes de SIR car le signal interférent sera mieux rejeté, mais l'apparition des lobes images crée un autre problème d'ambiguïté ; les signaux perturbateurs peuvent être captés par ces lobes.

L'espacement entre les éléments du réseau doit satisfaire $D < \lambda/2 = c/2f$. Par analogie avec le théorème d'échantillonnage de Nyquist, ce résultat peut être interprété comme le théorème d'échantillonnage spatial.

3.4.4. Calcul des poids

L'étape principale d'un FdV est le calcul des poids associés à chaque élément du réseau. On considère un réseau de deux microphones espacés d'un pas D , Deux sources audio situées dans le champ lointain, dont chacune peut se retrouver dans un quart du plan à part.

Le déphasage $\Delta\varphi$ créée au niveau de chaque microphone est donnée par :

$$\Delta\varphi = \frac{2\pi D}{\lambda} \sin(\vartheta) \quad (3.8)$$

La matrice de transfert (sources-réseau) est générée par l'emploi des vecteurs directionnels comme c'est démontré ci-dessous.

$$G = \begin{pmatrix} 1 & e^{j\frac{2\pi D}{\lambda}\sin\vartheta_1} \\ e^{j\frac{2\pi D}{\lambda}\sin\vartheta_2} & 1 \end{pmatrix} \quad (3.9)$$

Les colonnes correspondent aux éléments du réseau, et les lignes correspondent aux sources.

Les signaux reçus par le réseau y_i sont donné par :

$$y_1 = se^{j2\pi ft} + ne^{j(2\pi ft + \frac{2\pi D}{\lambda}\sin\vartheta_2)} \quad (3.10)$$

$$y_2 = se^{j(2\pi ft + \frac{2\pi D}{\lambda}\sin\vartheta_1)} + ne^{j(2\pi ft)} \quad (3.11)$$

Le signal résultant à la sortie du FdV est donné par :

$$\begin{aligned} z &= y_1 w_1 + y_2 w_2 \\ &= se^{j2\pi ft} w_1 + se^{j(2\pi ft + \frac{2\pi D}{\lambda}\sin\vartheta_1)} w_2 + ne^{j(2\pi ft + \frac{2\pi D}{\lambda}\sin\vartheta_2)} w_1 + ne^{j(2\pi ft)} w_2 \end{aligned} \quad (3.12)$$

Pour avoir une sortie égale à l'entrée désirée :

$$\begin{aligned} z &= s(t) \Rightarrow \\ \begin{cases} w_1 + w_2 e^{j\frac{2\pi D}{\lambda}\sin\vartheta_1} = 1 \\ w_1 e^{j\frac{2\pi D}{\lambda}\sin\vartheta_2} + w_2 = 0 \end{cases} \end{aligned} \quad (3.13)$$

$$\begin{pmatrix} 1 & e^{j\frac{2\pi D}{\lambda}\sin\vartheta_1} \\ e^{j\frac{2\pi D}{\lambda}\sin\vartheta_2} & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (3.14)$$

Il faut que :

$$w = U' G^{-1} \quad (3.15)$$

3.4.5. Mesure de performance

Cette section introduit une mesure de performance appelée SIR_{imp} (Signal to Interference Ratio improvement [3]), pour réaliser cette mesure il faut différencier entre les deux types de réseau de microphones, les réseaux compacts et les réseaux distribués. Dans le cas des réseaux compacts, toutes les entrées ont le même SIR , donc la référence est prise en faisant la moyenne des SIR mesurés au niveau de tous les microphones. Dans le cas des réseaux distribués, le microphone de référence est connu à priori, c'est le microphone qui présente le SIR le plus important.

Le cas traité dans ce mémoire est bien le réseau compact dont le calcul du SIR_{imp} est le suivant :

Le signal observé au niveau de chacun des microphones est la somme de la contribution du signal désiré et le signal interférent :

$$x_m(k) = x_{sig}(k) + x_{int}(k), \quad (3.16)$$

Similairement, le signal à la sortie peut être écrit ainsi :

$$y_m(k) = y_{sig}(k) + y_{int}(k), \quad (3.17)$$

Avec :

$$y_{sig}(k) = \sum_{m=1}^M (w_m * x_{sig})(k), \quad (3.18)$$

Et :

$$y_{int}(k) = \sum_{m=1}^M (w_m * x_{int})(k). \quad (3.19)$$

On introduit une mesure de la réduction du signal désiré (S_R : Signal Reduction) défini par le rapport de la moyenne des puissances du signal désiré au niveau de tous les microphones et la puissance du signal désiré à la sortie.

$$SR(k) \triangleq \frac{\sum_{m=1}^M \frac{E\{x_{sig}^2(k)\}}{M}}{E\{y_{sig}^2\}}, \quad (3.20)$$

On introduit aussi une mesure de la réduction du signal interférent (IR : Interference Reduction) défini par le rapport de la moyenne des puissances du signal interférent à travers tous les microphones et la puissance du signal interférent à la sortie.

$$IR(k) \triangleq \frac{\sum_{m=1}^M \frac{E\{x_{int}^2(k)\}}{M}}{E\{y_{int}^2\}}, \quad (3.21)$$

L'amélioration en termes de rapport signal sur interférence est défini par :

$$SIR_{imp}(k) = \frac{IR(k)}{SR(k)}. \quad (3.22)$$

3.4.6. Séparation de sources

Expérience 10 : Influence de la distance intra-microphones

Il s'agit de spécifier la distance intra-microphone qui donne le maximum de performance en termes de SIR_{imp} . Le schéma de la figure (3.40) décrit l'expérience réalisée.

On fait varier l'espacement intra-microphones et à chaque fois le SIR_{imp} est mesuré. Pour mesurer le S_R on élimine la source du signal interférent et pour mesurer l' I_R on élimine le signal désiré.

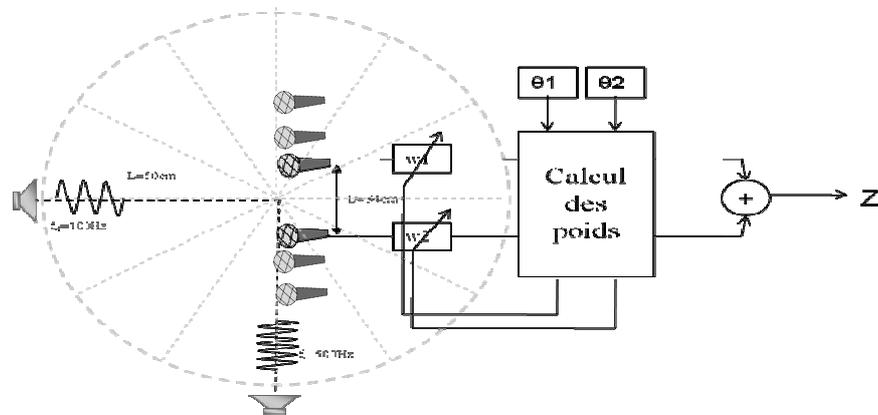


Figure 3.40 : Description de la configuration utilisée pour caractériser l'effet de la distance intra-microphones.

La courbe de la figure (3.41) montre le résultat obtenu.

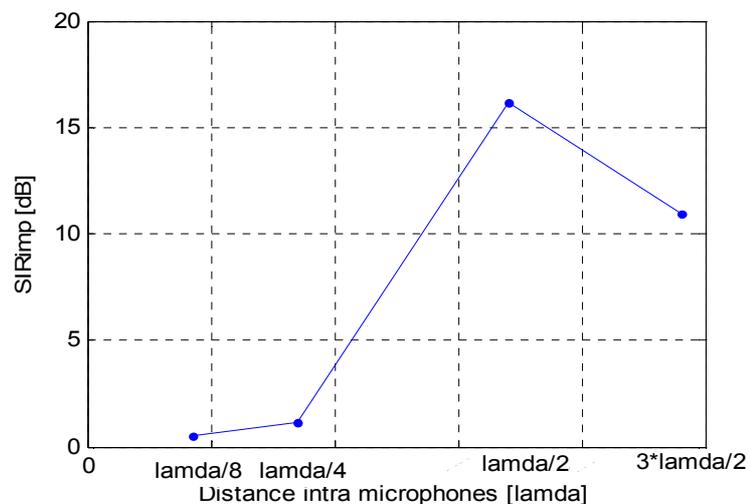


Figure 3.41 : La performance d'un FdV à bande étroite en fonction de la distance intra-microphones.

Observation

Le maximum d'amélioration en termes de SIR_{imp} (16 dB) est obtenu lorsque l'espacement intra-microphone est égale à $\lambda/2$. ce λ correspond à la longueur d'onde du signal interférent.

Résultat

Un signal interférent situant colinéaire avec le réseau peut être atténué de 16 dB lorsque la distance intra-capteurs est égale à la demi-longueur d'onde du signal interférent.

Expérience 11 : Influence de la séparation angulaire

Cette expérience vise à caractériser la performance du système vis-à-vis la séparation angulaire. Pour cela, la source désirée se trouve en face le réseau de microphones et la source interférente change d'emplacement suivant les valeurs de l'intervalle $[\pi/6, \pi/4, \pi/3, \pi/2]$.

La figure (3.42) illustre le résultat obtenu :

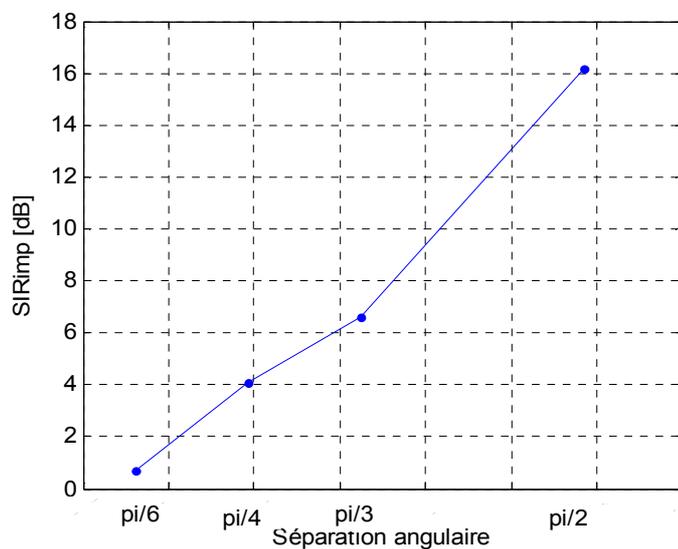


Figure 3.42 : La performance d'un FdV à bande étroite en fonction de la séparation angulaire.

Observation

D'après le résultat observé dans la figure (3.42), l'emplacement de la source du signal interférent a un grand impact sur la performance du système. La

meilleure position est quand le signal interférent se trouve colinéaire avec le réseau de microphones.

Résultat

Les diverses configurations sources-réseau créent des diagrammes de directivité différents, dont l'apparition des lobes secondaires est inévitable dans la plupart des cas. Le signal interférent est généralement capté par ces lobes. Le seul cas qui ne présente pas ce problème est celui qui correspond à un positionnement orthogonal des sources.

Expérience 12 : Séparation de deux sources à bande étroite

Dans cette expérience, on veut séparer spatialement un mélange de deux signaux à bande étroite. Pour cela, on génère par l'ordinateur émetteur ; dans le canal gauche de l'haut parleur une sinusoïde de fréquence $f_d = 100\text{Hz}$ et dans le canal droit une sinusoïde de fréquence $f_i = 500\text{Hz}$ (le signal à annuler). La distance de séparation entre les deux microphones à la réception est $D = c/2f = 34\text{ cm}$. La figure (3.43) décrit l'expérience réalisée. Les positions des sources sont supposées connues a priori ; le signal désiré est en face les microphones, le signal non désiré se trouve colinéaire avec le réseau des microphones.

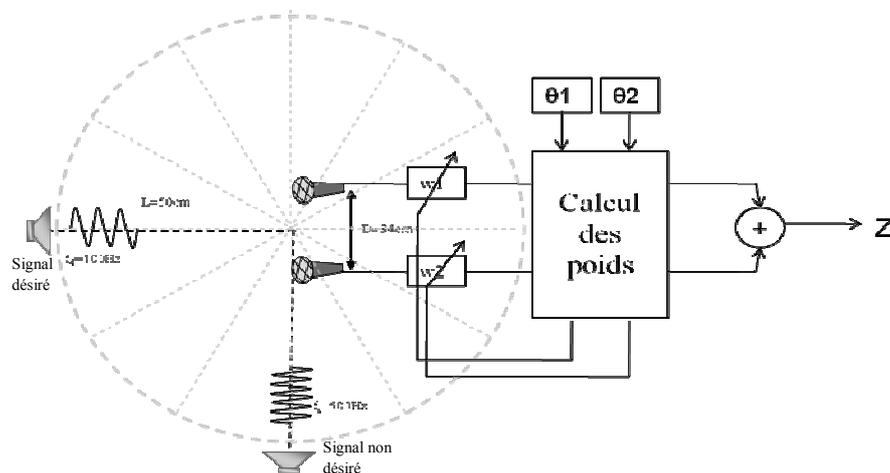


Figure 3.43 : Configuration utilisée pour la séparation de deux sources à bande étroite.

Les résultats présentés dans la figure (3.44) illustrent les signaux recueillis par les deux microphones et le signal à la sortie du FdV.

Observation

D'après la forme du signal capté par chacun des microphones, il semble que le signal de fréquence 100 Hz est modulé par le signal de fréquence 500 Hz, à la sortie du FdV uniquement le signal désiré est récupéré.

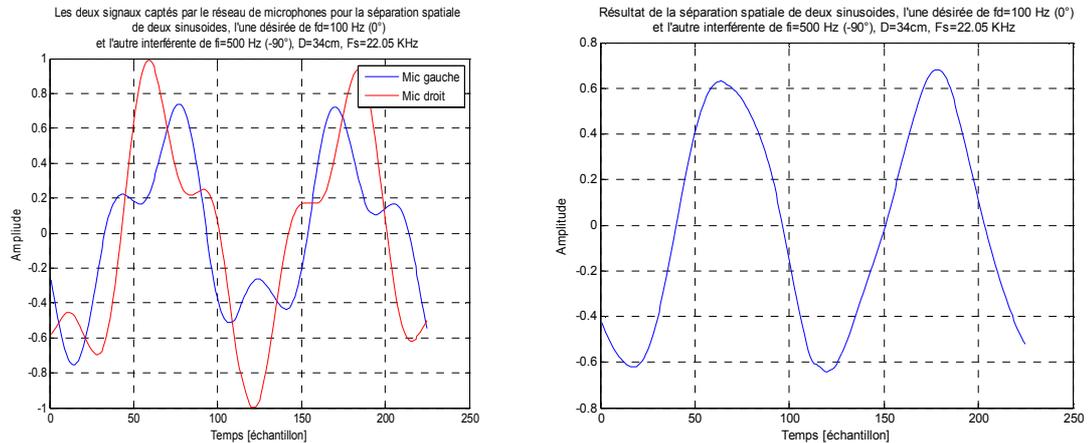


Figure 3.44 : Résultats de la séparation des deux signaux sinusoïdaux.

Résultat

La distance intra-microphone spécifiée à priori a créé un déphasage de π entre le signal de fréquence 500 Hz et sa version retardée, leurs somme a impliqué son annulation.

Expérience 13 : Séparation de deux sources une à bande large et l'autre à bande étroite

Le but dans cette section, est de séparer un signal périodique (bande étroite) d'un signal parole (large bande). Pour cela, la même configuration de la figure précédente est utilisée sauf que le signal désiré maintenant est un signal parole. Les directions d'arrivées des sources sont toujours supposés connus. Les résultats de la séparation sont présentés dans la figure (3.46).

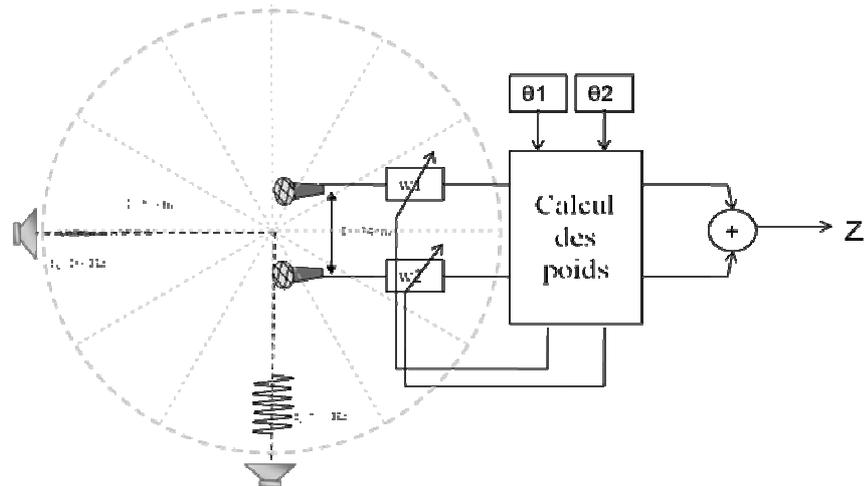


Figure 3.45 : Configuration du système pour la séparation d'un signal large bande et un signal bande étroite.

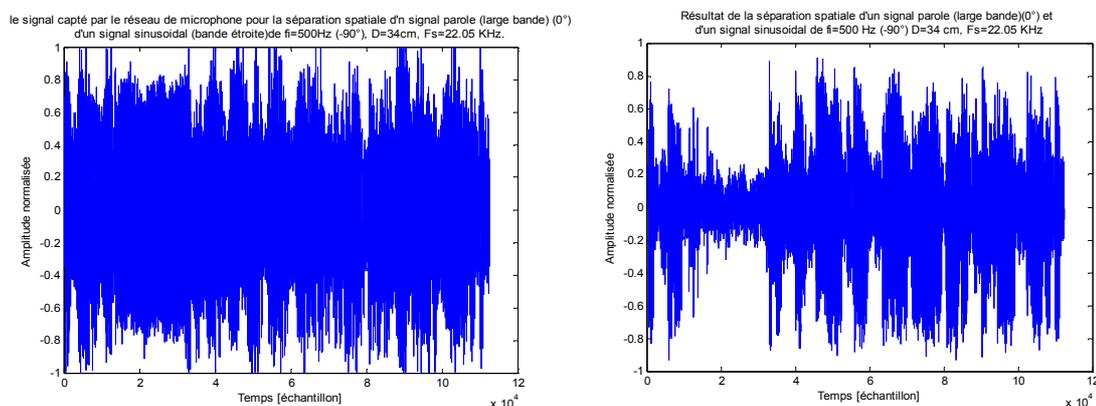


Figure 3.46 : Résultats de la séparation de deux sources ; une source de parole et une source d'un signal périodique.

Observation

Le signal bruité capté par les microphones est illustré en forme d'onde ainsi que le signal obtenu après la FdV. L'atténuation du signal interférent peut être vue dans les périodes du non activité vocale.

Résultat

Le signal interférent ne peut pas être annulé complètement à cause du non prise en compte de l'atténuation en puissance entre les deux signaux reçus par le réseau de microphones (Rayleigh) ainsi qu'aux caractéristiques non idéales du matériel utilisé.

3.5. Combinaison des deux systèmes

Expérience 14 : Rehaussement de la parole

L'objectif de cette section est expliqué dans le schéma de la figure (3.47). Puisque l'algorithme de la localisation utilisé n'est capable de localiser qu'une seule source, on a assumé que la DOA du signal perturbateur est connue à priori et la DOA du signal désiré sera estimée par le PHAT-GCC [31].

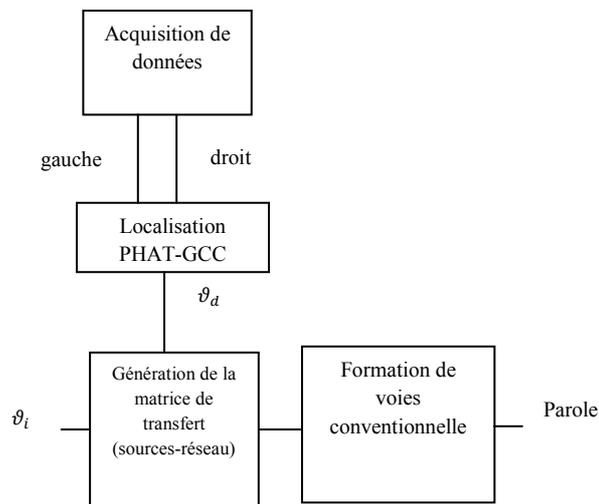


Figure 3.47 : Principe du rehaussement de la parole par la méthode de formation de voies.

A la réception, la chaîne de la figure (3.48) est simulée. Les paramètres du système sont résumés dans le tableau (3.4).

Taille de la fenêtre d'analyse	23ms
Type de la fenêtré d'analyse	Hanning
NFFT	512
Distance intra-microphones	34cm
Distance sources-réseau	50cm
Fréquence d'échantillonnage	22050Hz
Seuil de détection	0.6
Position de la source désirée	0°
Position de la source interférente	-90°

Tableau 3.4 : Paramètres choisis pour le système complet.

Après l'exécution de l'algorithme, les histogrammes présentés dans la figure (3.49) sont tracés.

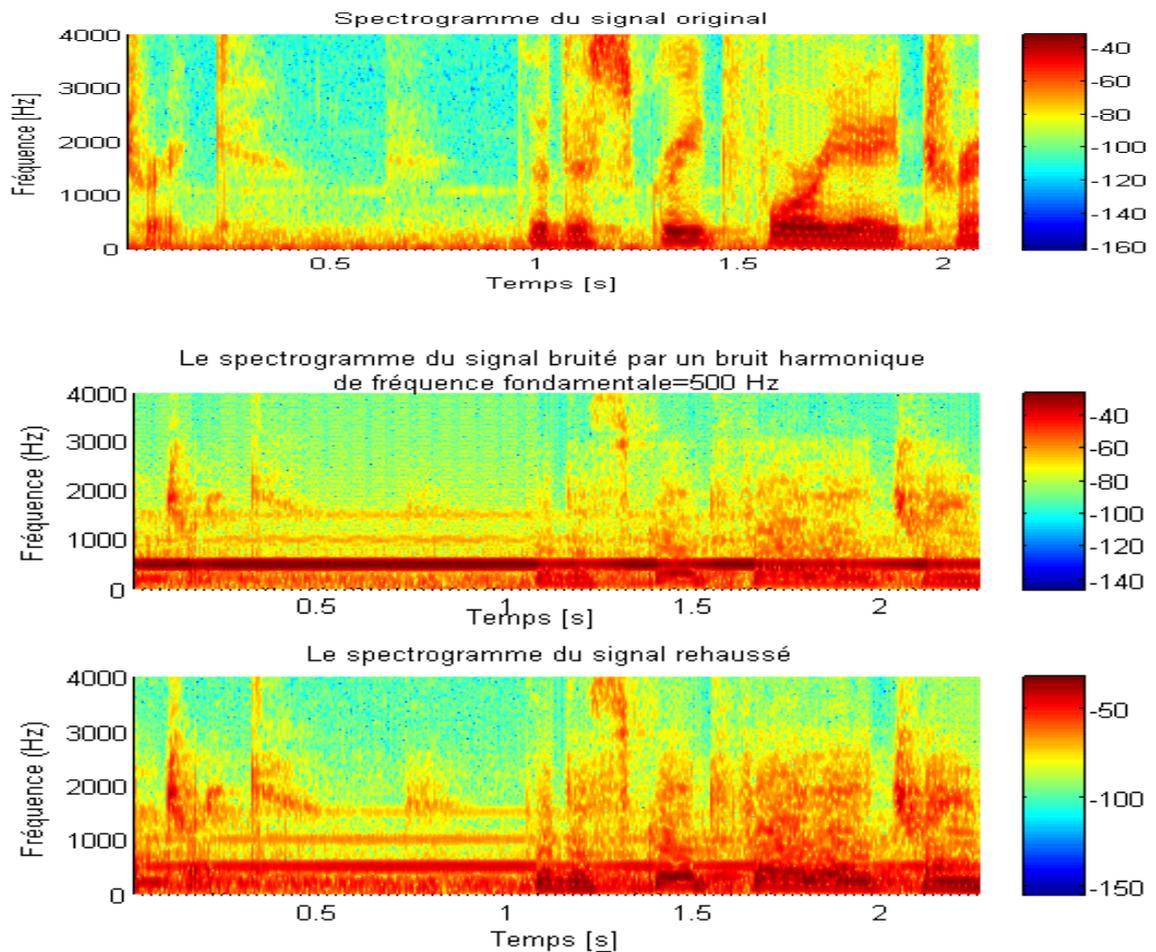


Figure 3.49 : Spectrogrammes du signal original, bruité et du signal débruité.

Observation

Le signal périodique interférent de fréquence égale à 500Hz se voit avec une puissance très importante dans le spectrogramme du signal capté par les microphones, tandis que sa puissance a diminué dans le deuxième spectrogramme du signal rehaussé par la FdV conventionnelle.

Résultat

Le signal interférent n'est pas annulé complètement à cause du non prise en compte de la contribution du canal acoustique en termes d'atténuation en puissance, et à cause des caractéristiques non idéales des microphones utilisés.

Expérience 15 : Tracé du diagramme de directivité du réseau

Cette fois ci, on veut tracer le diagramme de directivité d'un réseau de deux microphones. Le scénario choisi correspond à une source désirée en face le réseau et une source interférente colinéaire avec le réseau. Une fois les poids sont calculés on utilise une source de signal sinusoïdal de fréquence égale à 500Hz et on la fait déplacer de -90° à $+90^\circ$, les signaux recueillis à la sortie de chacun des microphones sont affectés d'un gain, puis sommés. La puissance est ensuite estimée par quadrature puis intégration temporelle. Et le résultat est présenté dans la figure (3.50).

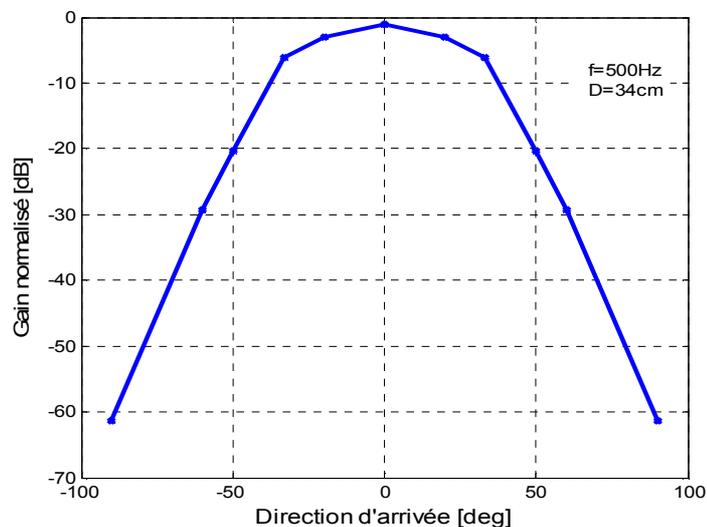


Figure 3.50 : Diagramme de directivité d'un réseau de deux microphones.

Observation

Ce diagramme de directivité indique la répartition du gain en fonction des directions des sources. Ici, le maximum du gain est orienté dans la direction du signal désiré (0°) et le minimum est orienté dans la direction du signal interférent (90°).

Résultat

Une atténuation importante du signal interférent est observée lorsque les sources sont orthogonales (60dB) vient du fait que la distance de séparation intra-microphone correspond à $\lambda/2$ crée ainsi un déphasage de π entre les deux versions du signal interférent, alors leur somme est nulle.

3.6. Conclusion

Dans ce chapitre, un système de séparation de sources a été réalisé. Les tests ont été faits dans un environnement idéal puis réel.

L'algorithme PHAT-GCC utilisé pour localiser les sources acoustiques est simple, rapide grâce à l'emploi de la *FFT* donc convenable pour la poursuite des sources en temps réel, mais il présente trois sources d'ambiguïté, une causée par le segment voisé du signal parole, une causée par les multiples sources et une autre causée par les multi-trajets.

Le formateur de voies employé est un FdV à bande étroite a permis l'annulation d'un signal bande étroite après un bon choix de la distance de séparation des microphones. Ses performances peuvent être améliorés si le matériel utilisé est bien caractérisé et le canal acoustique est estimé afin d'égaliser à la réception. Le système a présenté de meilleurs résultats dans le cas où les sources acoustiques sont orthogonalement positionnées ($60dB$).

CONCLUSION ET PERSPECTIVES

L'objectif dans ce travail était de réaliser un système de rehaussement de la parole basé sur la séparation spatiale des sources, connue sous le vocable beamforming ou formation de voies. L'idée principale de cette technique consiste à modifier le diagramme de directivité d'un réseau de microphones afin de maximiser son gain dans la direction de la source désirée et le minimiser dans la direction des sources interférentes. Pour réaliser cela, la connaissance des directions d'arrivées des sources est primordiale d'où la nécessité d'un algorithme de localisation des sources acoustiques. Pour tester la performance du système en temps réel, la carte son d'un ordinateur était utilisée comme une carte d'acquisition responsable au prétraitement des signaux et leurs restitutions.

La réalisation de ce système en temps réel et dans un environnement acoustique réel a rencontré des limites que ce soit dans l'acquisition de données, la localisation de sources ou la formation de voies.

L'acquisition et le traitement des signaux réels se fait généralement par un matériel (soft et hard) d'un coût très élevé (NI). Cependant, lors de ce mémoire, l'idée a été réalisée par un matériel faible coût disponible dans tous les établissements basé uniquement sur deux micro-ordinateurs, deux microphones omnidirectionnels et deux hauts parleurs grand public. La carte d'acquisition utilisée n'est rien d'autre que la carte son fournie avec tous les micro-ordinateurs; elle semble être un moyen très intéressant pour le test des algorithmes dédiés au traitement de signal (dans la gamme audio). La carte son intégrée permet d'utiliser uniquement deux microphones, ce qui empêche l'utilisation de plusieurs algorithmes que ce soit dans la localisation ou dans la formation de voies qui nécessitent plus de deux microphones pour leur fonctionnement.

L'algorithme de localisation utilisé est basé sur l'estimation de la différence en retard temporel (TDOA). Ça implique deux étapes; dans la première étape le TDOA entre un pair de microphones est estimé, et dans la deuxième étape la

position de la source est estimée à partir de ce TDOA et la géométrie du réseau. Cette procédure a l'avantage d'être applicable dans le cas des signaux large bande. La résolution de la localisation peut être ajustée d'une manière flexible en faisant varier la fréquence d'échantillonnage et ou la dimension du réseau de microphones. L'effet des multi-trajets peut perturber le processus de la localisation uniquement dans la première étape en plus qu'ils sont moins complexes.

Les méthodes d'inter-corrélation dite généralisés GCC sont très efficaces en terme de temps de calcul (grâce à l'emploi de la FFT, $N \log_2 N$) par rapport à l'intercorrélacion classique (N^2), en plus qu'ils n'ont pas besoin de faire une recherche exhaustive sur toutes les angles possibles. Elles prennent un temps de décision très court, donc elles ont la capacité d'une bonne poursuite de la source; un estimé est produit presque instantanément. Mais malheureusement ces méthodes ne performant pas bien dans le cas des multi-trajets; le PHAT-GCC présente un grand nombre d'écho en plus du trajet direct, or, que le TDOA obtenu à partir du trajet direct donne l'information utile pour la localisation des sources.

La structure périodique du segment voisé du signal parole présente une ambiguïté dans l'estimation du TDOA ; ceci peut être résolu en faisant une décomposition du signal parole en des segments contenant la partie voisée à part et la partie non voisée d'autre part et exploiter ensuite que la partie non voisée pour la localisation.

Une autre source d'ambiguïté dans la localisation apparait dans le cas de multiples sources actives simultanément ; le PHAT-GCC contient le trajet direct et les multi-trajets de toutes les sources en même temps. Dans ce cas, même si la localisation est parfaitement faite en exploitant que les trajets directs, l'assignement de ces TDOAs aux sources correspondantes est un autre problème.

Le rehaussement de la parole par la technique de formation de voies nécessite le calcul des poids associés aux éléments du réseau afin de sommer en phase les signaux venant de la direction souhaitée et créer un déphasage destructif entre les signaux venant de la direction non souhaitée. Avec un réseau composé uniquement de deux microphones, la largeur du lobe principal ne permet pas un bon pouvoir de séparation. Une atténuation maximale de 60 dB est achevée lorsque les sources sont positionnées d'une manière orthogonale et cette

performance diminue si la source interférente devient de plus en plus proche de la source désirée.

L'annulation d'un bruit à bande étroite a été réalisée sans problème grâce au choix approprié de la distance de séparation intra-microphones celle-ci correspond à la demi longueur d'onde de la source non désirée.

Pour améliorer les performances de ce système, la caractérisation du matériel utilisé ainsi que l'estimation adaptative du canal acoustique semble très nécessaire. Le cas d'un bruit large bande paraît impossible à résoudre avec uniquement deux microphones parce qu'il faut trouver la distance de séparation optimale de ces deux microphones permettant la rejection de toutes les fréquences contenants dans ce bruit. L'emploi de plusieurs microphones n'est pas un moyen pratique, mais l'utilisation de la technique des sous bandes peut être un bon moyen.

Le sujet traité dans ce mémoire est la combinaison de trois grands domaines de recherche qui sont l'acquisition de données par ordinateur, la localisation des sources acoustiques et la formation de voies. Il faut signaler que jusqu'au temps récent ces domaines sont en cours de la recherche et peu de travaux ont été effectués dans des environnements réel. La combinaison de ces trois domaines dans un seul travail est rare à cause de l'absence de solution parfaite pour chacun d'eux.

D'après les études de la psycho acoustique, la localisation chez l'être humain exploite tous les indices interauraux en plus de l'image et quelques autres paramètres. Tirer profit de cette technique humaine peut donner des résultats satisfaisants surtout par l'emploi de l'image comme un moyen de précision. Ils ont trouvés aussi qu'une certaine espèce d'hiboux (barn owl) a la capacité d'une bonne poursuite et ceci dans l'obscurité totale c'est-à-dire la localisation chez cet oiseau se fait sans le besoin de l'information image, la localisation donc sera plus rapide. Alors, une recherche sur la technique employée par cet oiseau peut mener à un bon algorithme de localisation.

Les cartes sons professionnelles externes existantes au marché présentent de bonnes caractéristiques et leur utilisation permet l'emploi de plus que deux

microphones. Cela permet le test des algorithmes nécessitant un grand nombre de microphones et l'emploi de différentes géométries.

Le cas traité dans ce mémoire représente une configuration simple, plusieurs simplifications sont assumées. Il faut viser à refaire le travail de telle manière à localiser les sources en trois dimensions et chercher à annuler les différents sources interférentes large bande venants de toutes les directions possibles.

REFERENCES

1. Vaseghi, S., "Advanced Digital Signal Processing and Noise Reduction", 4^{ème} edition, John Wiley & Sons Ltd, Chichester, (2006), 532 p.
2. Boll, S. F., "Suppression of Acoustic Noise in Speech Using Spectral Substraction", IEEE TRANSACTION ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol 27, p 113-120, (april 1979).
3. Bourgeois, J. et Minker, W., "Time Domain Beamforming and Blind Source Separation", Springer Science + Business Media, New York, (2009), 228 p.
4. Benesty, J. Chen, J. et Huang, Y., "Microphone Array Signal Processing", Springer-Verlag, Berlin, Germany, (2008), 245 p.
5. Kellermann, W., "Integrating Acoustic Echo Cancellation with Adaptive Beamforming Microphone Arrays", 25th German Annual conference on Acoustic (DAGA), Berlin, March p 14-19, (1999).
6. Allen, B. et Ghavami, M., "Adaptive Array Systems Fundamentals and Applications", John Wiley & Sons, Chichester, (2005), 290 p.
7. Brandestein, M. et Ward, D., "Microphone Arrays Signal Processing Techniques and Applications", Springer-Verlag, Berlin, (2001), 389 p, [en ligne].
8. Cobos, M. Lopez, J.J. et Martinez, D., "Two-Microphone Multi-speaker Localisation based on a Laplacian Mixture Model", Digital Signal Processing (Elsevier), Vol 21, Issue 1, p 66-67, (2011).
9. Yermeche, Z., "Soft-Constrained Subband Beamforming for speech Enhancement", these de doctorat, Sweden, (2007).
10. Varma, K., "Time-Delay-Estimate Based Direction-of-Arrival Estimation for Speech in Reverberant Environment", these de Magister, (2002).
11. Knapp, C.H. et Carter, G.C., "The Generalized Correlation Method For Estimation of Time Delay", IEEE Transactions on Acoustics, Speech, and Signal processing, V. 24, n°4, (Août 1976).
12. Valin, J.M, Michaud.F, Rouat. J. et Létourneau, D., "Robust Sound Source Localization Using a Microphone Array on a Mobile Robot", Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, p. 1228-1233.

13. Tashev, I., "Sound Capture and Processing: Practical Approaches", John Wiley & Sons, Chichester, (2009), 341 p, [en ligne].

14. Oudin, M., "Etude d'algorithme de Traitement d'Antenne sur signaux large bande et signaux Radar bande étroite à antenne tournante", thèse de doctorat, (2008).

15. Litva, J. et Kwok-Yeng Lo, T., "Digital beamforming in Wireless Communications", Artech House, Norwood, (1996), 297 p.

16. Huang, Y. Benesty, J. et Chen, J., "Acoustic MIMO Signal Processing Handbook", Springer-Verlag, Berlin, (2006), 391 p.

17. Madiseti, V. et Williams, D.B., "Digital Signal Processing Handbook", CRC/IEEE Press, (1999), 1690 p.

18. Van Veen, B.D. et Buckley, K.M., "Beamforming : A versatile Approach to Spatial Filtering", IEEE Acoustics, Speech, and Signal Processing (ASSP) Magazine, p 4-24, (April 1988).

19. Sow, G., "Méthodes d'Accès Multiple à Répartition Spatiale Pour Communications Par Satellite", thèse de doctorat, (2009).

20. Manolakis, D.G. Ingle, V.K. et Kogon, S.M., "Statistical and Adaptive Signal Processing", Artech House, Boston, (2005), 787 p.

21. Griffiths. L et Jim.C., "An Alternative Approach to Linearly Constrained Adaptive Beamforming", IEEE Trans. Antennas Propagat. AP-30, 27-34, (1982).

22. Benesty, J. Mohan Sondhi, M. et Huang, Y., "Springer Handbook of Speech Processing", Springer-Verlag, Berlin, (2007), 1188 p.

23. Chen, J. Benesty, J and Huang, Y., "Time Delay Estimation in Room Acoustic Environments: An Overview", EURASIP journal on Applied Signal Processing, vol 2006, p 1-19.

24. Macloughlin, I., "Applied Speech and Audio Processing", Cambridge University Press, (2009), 218 p.

25. Preumont, A., "Vibrations aléatoires et analyse spectrale", Presse Polytechnique et universitaires romandes, Lausanne, 1990, 331 p, [en ligne].

26. Tisserand, E., Pautex, J.F., et Schweitzer, P., "Analyse et Traitement des Signaux", 2^{ème} édition, Dunod, Paris, (2008), 330 p.

27. Rabiner, L.R. et Schafer, R.W., "Digital Processing of Speech Signals", Prentice-Hall, New Jersey, (1978), 509 p.

28. Cottet, F., "Aide Mémoire Traitement du signal", Dunod, Paris, (2000), 237 p.
29. Mari, J.L. Glanceaud, F et Coppens, F., "Traitement du signal pour Geologues et Geophysiciens" Technip, Paris, (1997), 249 p, [en ligne].
30. Hansler, E et Schmidt, G., "Topics in Acoustic Echo and Noise Control", Springer-Verlag, Heidelberg, (2006), 648 p.
31. Bouchibane, FZ et Bensebti, M., "Harmonic Noise Reduction using Conventional Beamforming in Real Environment", Image and Signal Processing and their Applications (ISPA) conference, Biskra, (December 2010).