

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

École Nationale Polytechnique



Département d'Electronique
Laboratoire de Communication et de Conversion Photovoltaïque

Mémoire de Master en Electronique

THÈME :

Localisation de sources par un réseau de microphones.

DJEMAI Mohamed

Sous la direction de

Pr. Adel BELOUHRANI

PhD. Mourad ADNANE

Présenté et soutenu publiquement le 17 Juin devant le jury composé de :

Président	Pr. M. MEHENNI	(ENP)
Examineur	Pr. R. AKSAS	(ENP)
Promoteur	Pr. A. BELOUHRANI	(ENP)
Co-Promoteur	Dr. M. ADNANE	(ENP)

ENP 2017

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

École Nationale Polytechnique



Département d'Electronique
Laboratoire de Communication et de Conversion Photovoltaïque

Mémoire de Master en Electronique

THÈME :

Localisation de sources par un réseau de microphones.

DJEMAI Mohamed

Sous la direction de

Pr. Adel BELOUHRANI

PhD. Mourad ADNANE

Présenté et soutenu publiquement le 17 Juin devant le jury composé de :

Président	Pr. M. MEHENNI	(ENP)
Examineur	Pr. R. AKSAS	(ENP)
Promoteur	Pr. A. BELOUHRANI	(ENP)
Co-Promoteur	Dr. M. ADNANE	(ENP)

ENP 2017

Dédicace

Je dédie ce travail à mes très chères mère et grande mère, dont le sacrifice, la tendresse, l'amour, la patience, le soutien, l'aide et l'encouragement sont l'essentiel de ma réussite. Sans elles je ne serai pas à ce stade aujourd'hui.

A mes oncles Mustapha et Foudil pour leur soutien continue durant mon parcours.

A Brahim, Mounir, Oussama, et Kamel.

A Rahim, Mehdi et Madjid pour leur présence.

M. DJEMAI

Remerciements

Je remercie Dieu le tout puissant de m'avoir donné la force et le courage pour réaliser ce travail.

Les travaux présentés dans ce mémoire ont été effectués au sein de l'École Nationale Polytechnique.

Ce travail que je présente a été effectué sous la direction de Mr *A. BELOUHRANI*, Professeur à l'École Nationale Polytechnique, et Mr *M. ADNANE*, Docteur et chef de département de l'électronique à l'École Nationale Polytechnique, qui ont suivi ce travail, pour leur orientation pédagogique dans l'élaboration de ce mémoire.

Je tiens aussi à remercier Mr *C. LARBES*, Docteur à l'École Nationale Polytechnique et Mr *Samir*, dirigeant de Orbit, pour m'avoir aidé avec le matériel nécessaire.

Je tiens à remercier Mr *M. MEHENNI*, Professeur à l'École Nationale Polytechnique, pour l'honneur qu'il me fait de présider le jury de ma soutenance.

Que Mr *R. AKSAS*, Professeur à l'École Nationale Polytechnique, soit convaincu de mon sincère reconnaissance pour avoir accepté d'examiner et de critiquer ce mémoire.

Enfin, je tiens à remercier tous les gens qui ont contribué à ma réussite tout au long de mon parcours d'étude.

ملخص

تلقي تحديد مواقع الاشارات بشبكة الميكروفونات مؤخرا اهتماما كبيرا باعتبارها وسيلة آلية لتتبع الأفراد. النهج الأكثر استخداما هو تقدير الفرق في وقت الوصول من اثنين من الميكروفونات. في الأدب، هناك العديد من الخوارزميات لتقدير الفرق في وقت الوصول. ونقدم ثلاثة خوارزميات مختلفة، يستند الاول على عبر الارتباط ويستخدم حيث لا يوجد سوى اثنين من الميكروفونات لتقدير الفرق في وقت الوصول، والثاني يقوم على التنبؤ الخطي المكاني، وهذا الأسلوب يسمح له باستخدام عدة ميكروفونات (أكثر من اثنين) لتقدير الفرق في وقت الوصول، الخوارزمية الثالثة التي يمكن أن ينظر إليها على أنه تعميم الطريقة الأولى في حالة عدة ميكروفونات.

الكلمات المفتاحية: معالجة شبكة الميكروفونات، الفرق في وقت الوصول، عبر الارتباط

Abstract

The sources localization with a microphone array has recently received significant attention as a means of automated tracking of individuals. The most widely used approach is to estimate the time difference of arrival (TDOA) of the signal between two microphones. In the literature, there are several algorithms for estimating TDOA. Three different algorithms were used, one based on cross-correlation, where only two microphones were used to estimate the TDOA, and the second was based on spatial linear prediction, using several microphones (More than two) to estimate the TDOA and have more resolution, the third algorithm which can be seen as a generalization of the first method in the case of several microphones.

Key words: microphone array signal processing, source localization, cross-correlation.

Résumé

La localisation de sources avec un réseau de microphones a récemment reçu une attention significative comme un moyen de suivi automatisé des individus. l'approche la plus utilisé est d'estimer la différence de temps d'arrivée (TDOA) du signal entre deux microphones. Dans la littérature, il existe plusieurs algorithmes pour l'estimation du TDOA. On a présenté trois algorithmes différents, la premier est basé sur la corrélation croisée (cross-correlation) où on utilise seulement 2 microphones pour estimer le TDOA, le deuxième est basé sur la prédiction linéaire spatiale, cette méthode permet d'utiliser plusieurs microphones (plus de deux) pour estimer le TDOA et avoir plus de résolution, le troisième algorithme qui peut être vu comme une généralisation de la première méthode dans le cas de plusieurs microphones.

Mots clés : réseau de microphones, formation de voies, localisation de sources, la corrélation croisée.

Table des matières

Table des figures

1	Introduction générale	7
1.1	Choix de la méthode de calcul de la direction d'arrivée	8
1.1.1	ILD	8
1.1.2	TDOA	9
2	Réseau de microphones	10
2.1	Les microphones	10
2.2	Directivité et réponse fréquentielle	10
2.3	Le réseau de microphones	11
2.4	L'espacement entre microphones	13
3	Localisation de sources	15
3.1	Introduction	15
3.2	Le modèle de signal	16
3.2.1	Le modèle du signal d'une seule source dans l'environnement libre . . .	17
3.2.2	Le modèle du signal de plusieurs sources dans l'environnement libre . .	17
3.2.3	Le modèle du signal d'une seule source dans l'environnement réverbérant	17
3.2.4	Le modèle du signal de plusieurs sources dans l'environnement réver- bérant	19
3.3	La méthode de la corrélation croisée (Cross-Correlation)	20
3.4	La méthode de la prédiction linéaire spatiale	22
3.5	L'algorithme de la corrélation croisée à multi canaux	24
	Conclusion générale	28
	Bibliographie	29

Table des figures

1.1	Méthodes de calcul de la direction d'une source sonore.	8
1.2	Exemple d'un réseau de deux microphones.	9
2.1	Diagramme de directivité.	11
2.2	réseau de microphones linéaires.	12
2.3	Recouvrement spatial : diagramme polaire du diagramme de directivité d'un réseau de microphone linéaire avec quatre éléments, en fonction de la direction horizontale θ ; Avec un échantillonnage spatial critique, $d = \lambda/2$ (gauche), et avec des effets d'alias pour λ (à droite).	14
3.1	Illustration de la direction d'arrivée avec 2 microphones, la source est dans le champ lointain, l'angle d'incidence est θ , et l'espacement entre 2 capteurs est d . .	16
3.2	Illustration de l'espace ouvert avec seulement une seule source.	18
3.3	Illustration de l'environnement réverbérant avec seulement une seule source. . .	19
3.4	Illustration de l'effet de l'aliasing.	22
3.5	Comparaison de $J_{1,min}(p)$ pour différentes nombre de microphones. (a) SNR=10, (b) SNR=-5.	25
3.6	Comparaison de $\det [\mathbf{R}_a(p)]$ pour différentes nombre de microphones. (a) SNR=10, (b) SNR=-5.	27

Chapitre 1

Introduction générale

La localisation de sources est déjà présente dans le domaine électromagnétique depuis longtemps, par exemple par l'utilisation d'antennes comme capteurs. Plusieurs techniques emploient, entre autres, des réseaux de capteurs, en combinaison avec les algorithmes de traitement de signal, pour déterminer l'origine des signaux reçus. De la même manière, il est possible de localiser des sources sonores à l'aide d'un réseau de microphones. L'onde sonore se propage alors dans un milieu, du point d'émission jusqu'au réseau de microphones, et les signaux enregistrés permettent de déterminer la position de la source par rapport au réseau. Plusieurs applications pratiques à cette localisation sont possibles, parmi lesquelles il y a la localisation d'un tireur embusqué ou bien l'identification de sources de pollution sonore dans un environnement complexe.

Plusieurs algorithmes existent déjà pour la localisation de sources à l'aide d'un réseau de microphones, par la détermination des directions d'arrivée (DOA) des signaux par rapport au réseau. Plusieurs de ces algorithmes fonctionnent pour des sources monochromatiques ou à bande étroite et exploitent les différences de phase entre les capteurs pour l'estimation des positions d'origines.

L'oreille est l'un des organes les plus complexes dans le corps humain [1]. En effet, de part sa nature, l'oreille est composée d'une multitude de cellules spécialisées permettant l'analyse du spectre sonore dans une gamme de fréquences allant de 22Hz à 22kHz. Cette analyse assure une perception assez précise de l'espace sonore qui nous entoure en permettant non seulement d'analyser les sons qu'elle capte mais aussi de localiser la source émettrice de ces sons, même si elle ne se trouve pas dans notre champ de vision.

Localiser une source sonore revient à déterminer sa direction d'arrivée ainsi que sa distance. La précision de cette localisation peut atteindre 1 degré pour certains mammifères.

La modélisation du fonctionnement de l'oreille a suscité, ces dernières années, un intérêt particulier notamment dans le domaine de la robotique. En effet, plusieurs équipes se sont attachées, durant cette dernière décennie, à produire des robots pouvant assurer les mêmes fonctions élémentaires qu'un être humain [2]

Dans la littérature, il existe plusieurs méthodes pour la détermination de la direction d'arrivée d'une onde sonore. Les deux les plus utilisées sont l'ILD (Interaural Level Differences) et le TDOA (Time Difference of Arrival) connue aussi sous le nom d'ITD (Interaural Time Difference)(Figure 1.1).

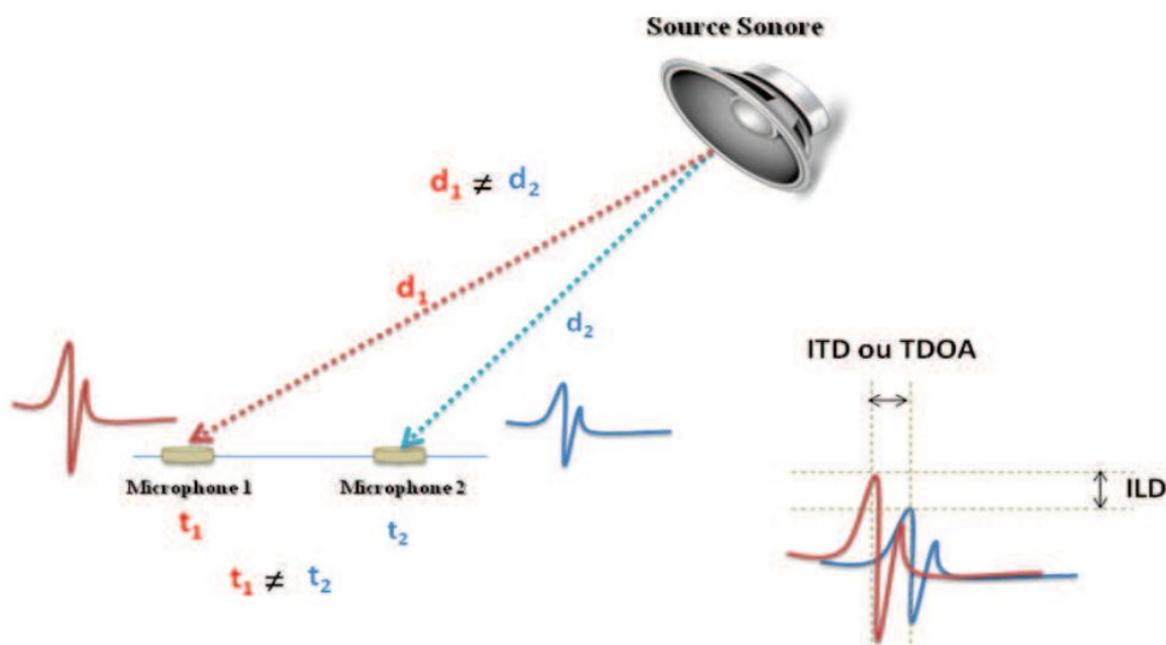


FIGURE 1.1 – Méthodes de calcul de la direction d'une source sonore.

1.1 Choix de la méthode de calcul de la direction d'arrivée

1.1.1 ILD

L'ILD (Interaural Level Differences) mesure la différence d'intensité sonore entre les signaux reçus par deux récepteurs (microphones ou oreilles). En fait, l'intensité d'un signal est géné-

ralement considérée comme l'un des indices principaux de perception de l'emplacement d'une source sonore. En effet, l'appareil auditif humain est très sensible à toute variation de cet indice.

L'intensité d'un son dans un champ libre est inversement proportionnelle au carré de la distance entre la source et le récepteur. Cependant, cette loi ne s'applique que pour une source relativement proche. En effet, la différence d'intensité devient négligeable en fonction de la distance réduite du récepteur. En outre, la connaissance à priori du niveau de la source sonore à l'émission est indispensable pour sa localisation du fait que le niveau du son à la réception est relatif à celui du son émis.

1.1.2 TDOA

L'une des techniques les plus utilisées est la technique de la différence du temps d'arrivée (TDOA) qui permet d'obtenir de bons résultats de localisation. Si l'on considère le cas d'un réseau de deux capteurs (microphones) (Figure 1.2), le principe de l'estimation de la direction d'arrivée d'une onde sonore repose sur le calcul de ses temps de capture par chaque microphone puis en déduire le retard (τ) ou la différence du temps d'arrivée (TDOA) enregistrée entre ces deux microphones. Ainsi, à l'instar du système auditif humain, l'exploitation du retard que met un son pour arriver à deux capteurs (microphones) pourrait nous conduire à déterminer la direction d'arrivée de ce son.

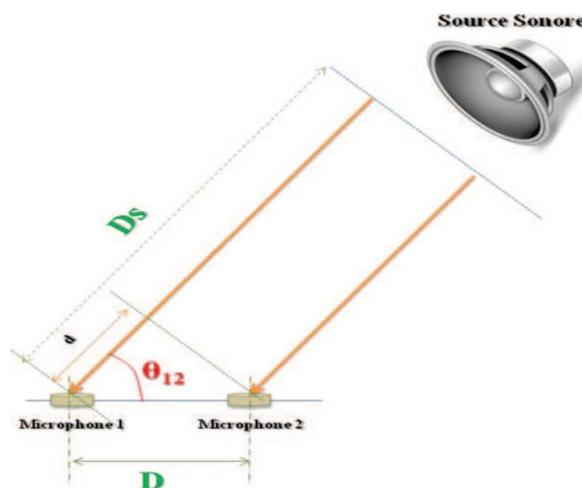


FIGURE 1.2 – Exemple d'un réseau de deux microphones.

Cette méthode est beaucoup plus précise et robuste dans le cas d'une source sonore assez éloignée (cas des ondes planes). C'est pour cette raison que la méthode TDOA a été retenue.

Chapitre 2

Réseau de microphones

2.1 Les microphones

Dans le traitement du signal d'un réseau de microphones, la géométrie et la taille du réseau ainsi que le nombre de microphones peuvent avoir un impact sur les performances du réseau. En outre, le type de microphones utilisés est un facteur important. Dans ce travail, deux types différents de microphones peuvent être utilisés, omnidirectionnels et unidirectionnels, pour lesquels les différentes caractéristiques seront discutées dans cette section.

2.2 Directivité et réponse fréquentielle

La directivité ou la réponse directionnelle d'un microphone décrit comment le son issu de différentes directions est capté par le microphone. Il est communément présenté dans un diagramme polaire du gain par rapport à l'angle, et les différents types de microphones sont appelés suivant la forme de leur directivité. Les deux types mentionnés dans ce travail sont présentés dans la figure 2.1. Un microphone omnidirectionnel répond également au son provenant de toutes les directions. La directivité d'un microphone dépend de la fréquence, mais un microphone omnidirectionnel optimal devrait avoir le même gain sur toutes les fréquences. La réponse unidirectionnelle montrée à la figure au dessus est un microphone cardioïde typique, nommé d'après son motif de gain en forme de cœur. Les microphones unidirectionnels répondent au son provenant de l'avant du microphone, tandis que la lecture des côtés et de l'arrière est réduite. La réponse des microphones unidirectionnels varie avec la fréquence et est généralement étroite aux fréquences élevées alors que ces microphones sont pratiquement omnidirectionnels aux basses fréquences. En outre, la réponse en fréquence des microphones unidirectionnels varie également avec la distance entre le microphone et la source sonore ; plus la distance est

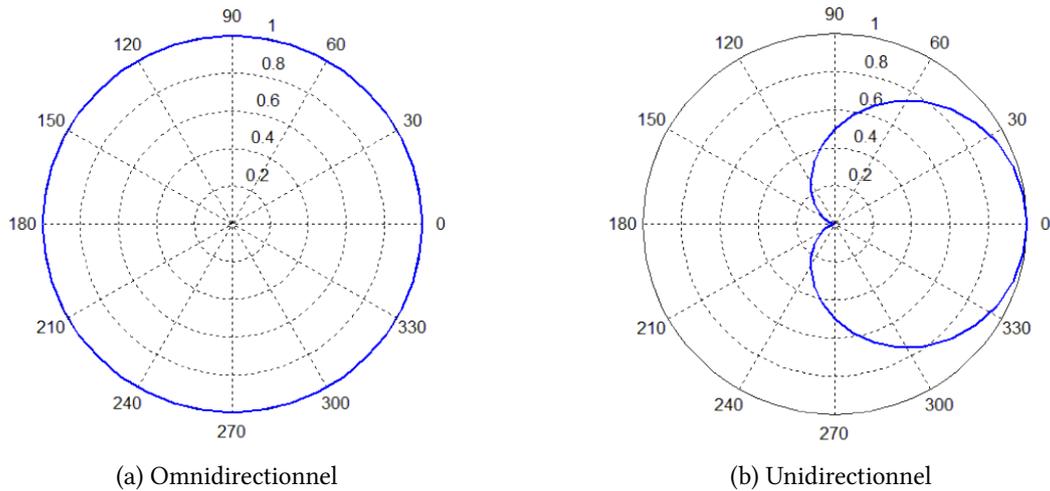


FIGURE 2.1 – Diagramme de directivité.

longue, plus la lecture des basses fréquences est une caractéristique connue sous le nom d'effet de proximité.

2.3 Le réseau de microphones

L'étude et la mise en œuvre des réseaux de microphones sont survenues il y a plus de 30 ans. Grâce à la recherche et aux développements expérimentaux poursuivis au cours du présent jour, le champ a mûri au point que la technologie basée sur le réseau a maintenant une applicabilité immédiate à un certain nombre de systèmes actuels et un vaste potentiel pour l'amélioration des produits existants et la création de futurs appareils.

Le réseau de capteurs acoustiques se compose d'un ensemble de capteurs acoustiques placés à différents endroits afin de recevoir un signal porté par des ondes de propagation. La figure 2.2 montre la disposition d'un réseau de microphone linéaire composé de M microphones avec K signaux incidents provenant des angles d'arrivée θ_K .

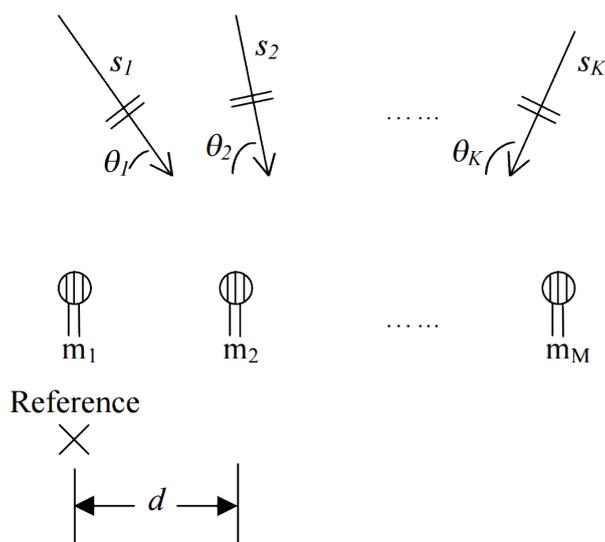


FIGURE 2.2 – réseau de microphones linéaires.

Dans cette analyse, on suppose que les ondes incidentes sont des ondes planes. Chacun des microphones recevra ainsi chacun des signaux, s_k , mais ceux-ci seront des versions retardées ou avancées dans le temps. Il faut donc un point de référence pour notre système de coordonnées afin que nous puissions calculer la phase de chaque signal, nous choisissons le microphone 1 comme origine. La distance entre deux microphones adjacents est d (voir la figure 2.2).

Le retard correspondant τ_i à chaque microphone est alors :

$$\tau_i = \frac{d_i \cos \theta_k}{v} \quad (2.1)$$

ou v est la vitesse du son ($343m/s$) et d_i est la distance entre la référence (l'origine) et l' i ème microphone.

L'entrée à l' i ème microphone du signal s_k est donc :

$$x_i(t) = s_k(t - \tau_i) \quad (2.2)$$

dans un scénario à bande étroite, petites variations autour de λ , le retard peut être exprimé par un simple déphasage :

$$x_i(t) = s_k(t) e^{-j\omega_0 \tau_i} \quad (2.3)$$

qui est équivalent à

$$x_i(t) = s_k(t)e^{-jkd \cos \theta_k} \quad (2.4)$$

où $k = \frac{2\pi}{\lambda}$ est le nombre d'onde.

la représentation de (2.3) est plus simple à utiliser que la représentation (2.2). Maintenant si on veut voir une représentation générale on doit introduire tous les signaux, et inclure aussi le bruit à chaque microphone :

$$x_i(t) = \sum_{k=1}^K s_k(t)e^{-jkd \cos \theta_k} + n_i(t) \quad (2.5)$$

2.4 L'espace entre microphones

Une des propriétés d'un réseau c'est que la largeur de faisceau diminue à mesure que l'espace d augmente. Donc, si nous voulons un faisceau plus précis, nous pouvons simplement augmenter l'espace d , ce qui conduit à une ouverture de réseau plus grande, mais on doit toujours éviter le recouvrement spatial (l'aliasing) causé par ce dernier. Le recouvrement spatial entraîne l'apparition de lobes parasites dans le diagramme de directivité, appelés lobes de grille, comme illustré dans la figure 2.3. Pour éviter cela dans la bande des fréquences qui nous intéresse, on doit avoir un déphasage entre deux microphones adjacents se trouvant dans l'intervalle $[-\pi, \pi]$. On peut calculer le déphasage de l'équation (2.4) qui est :

$$\phi_k = \frac{2\pi d \cos \theta_k}{\lambda} \quad (2.6)$$

on veut ϕ_k entre $-\pi$ et π , donc $d < \frac{\lambda_{min}}{2}$ où λ_{min} est le minimum des longueurs d'onde des signaux reçus, sinon on va avoir plus d'un lobe principal. Par conséquent, la distance d'espace critique requise pour traiter les signaux dans la bande passante de la parole B_s (300-3400 Hz) est $d = 5 \text{ cm}$. Par analogie à la théorème de Nyquist, ce résultat peut être exprimé comme un théorème d'échantillonnage spatial.

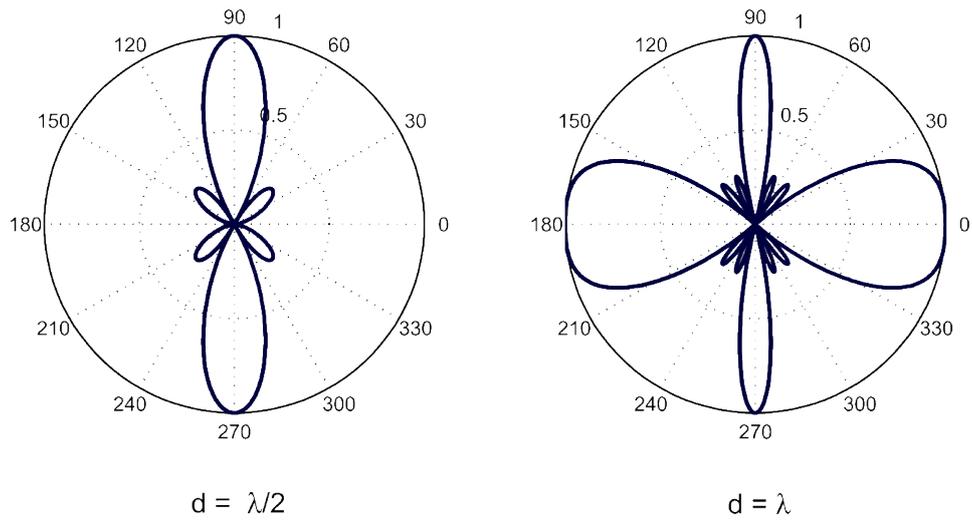


FIGURE 2.3 – Recouvrement spatial : diagramme polaire du diagramme de directivité d'un réseau de microphone linéaire avec quatre éléments, en fonction de la direction horizontale θ ; Avec un échantillonnage spatial critique, $d = \lambda/2$ (gauche), et avec des effets d'alias pour λ (à droite).

Chapitre 3

Localisation de sources

3.1 Introduction

L'estimation de la direction d'arrivée DOA (Direction-Of-Arrival) traite le cas où la source se trouve dans le champ lointain du réseau, comme l'illustre la Figure 3.1. Dans cette situation, la source rayonne une onde plane $s(k)$ qui se propage à travers l'air. La normale au front d'onde fait un angle θ avec la ligne rejoignant les capteurs dans le réseau linéaire, et le signal reçu à chaque microphone est une version retardée/avancée du signal à un capteur de référence. Pour voir cela, choisissons le premier capteur de la Figure 3.1 comme point de référence et notons d l'espacement entre les deux capteurs. Le signal au deuxième capteur est retardé par le temps requis pour que l'onde plane se propage dans $d \cos \theta$. Par conséquent, la différence de temps (retard) entre les deux capteurs est donnée par $\tau_{12} = d \cos \theta / c$, où c est la vitesse du son dans l'air. Si l'angle est compris entre 0 degrés et 180 degrés et si τ_{12} est connu alors θ est déterminé de manière unique, et vice versa. Par conséquent, l'estimation de l'angle d'incidence θ est essentiellement identique à l'estimation de la différence de temps τ_{12} . En d'autres termes, le problème d'estimation de l'angle d'arrivée (DOA) est le même que le problème d'estimation de la différence de temps d'arrivée (TDOA, Time-Difference-Of-Arrival) dans le cas du champ lointain.

Bien que l'angle d'incidence puisse être estimé à l'aide de deux capteurs ou plus, la distance entre la source sonore et le réseau de microphones est difficile, voir impossible si la source est dans le champ lointain du réseau. Cependant, si la source est située dans le champ proche, il est possible d'estimer non seulement l'angle à partir duquel l'onde atteint chaque capteur, mais aussi la distance entre la source et chaque microphone.

Indépendamment du fait que la source est située dans le champ lointain ou le champ proche, l'étape la plus fondamentale est d'estimer le TDOA entre différents microphones. Ce problème d'estimation serait une tâche facile si les signaux reçus n'étaient qu'une version retardée l'une de l'autre. En réalité, le signal de la source est généralement immergé dans le bruit ambiant, puisque nous vivons dans un

environnement naturel où l'existence du bruit est inévitable. En outre, chaque signal d'observation peut contenir plusieurs répliques atténuées et retardées du signal source en raison de réflexions à partir des murs et des objets. Cet effet de propagation multi-voies introduit des échos et des distorsions spectrales dans le signal d'observation, appelé réverbération, qui détériore sévèrement le signal de la source. La source peut également se déplacer de temps en temps. Tous ces facteurs font que l'estimation de TDOA est un problème complexe et difficile.

Ce document traite des idées de base de l'estimation du TDOA. Nous commencerons notre discussion avec des scénarios où il n'y a qu'une seule source. Nous allons ensuite explorer quelles approches peuvent être utilisées pour améliorer la robustesse de l'estimation du TDOA en ce qui concerne le bruit et la réverbération. Nombreuses idées fondamentales développées pour l'estimation du TDOA à une seule source peuvent être étendues à la situation à plusieurs sources. Pour illustrer cela, nous discuterons de la philosophie derrière l'estimation du TDOA à sources multiples.

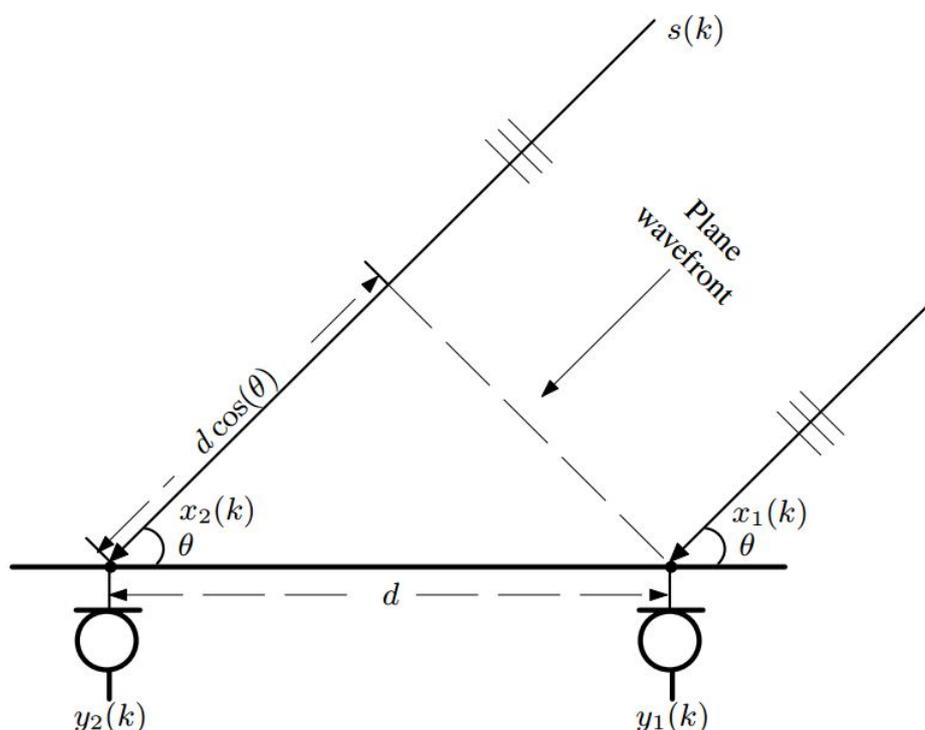


FIGURE 3.1 – Illustration de la direction d'arrivée avec 2 microphones, la source est dans le champ lointain, l'angle d'incidence est θ , et l'espacement entre 2 capteurs est d .

3.2 Le modèle de signal

Le problème d'estimation du TDOA concerne la mesure de la différence de temps entre les signaux reçus à différents microphones. Selon l'environnement acoustique, nous considérons deux situations :

l'environnement libre où chaque capteur reçoit uniquement le signal du chemin direct, et les environnements réverbérants où chaque capteur peut recevoir un grand nombre de signaux réfléchis en plus du chemin direct. Pour chaque situation, nous distinguons le cas source unique du scénario à source multiple, puisque les principes d'estimation et la complexité dans ces deux conditions peuvent ne pas être nécessairement les mêmes.

3.2.1 Le modèle du signal d'une seule source dans l'environnement libre

Supposons qu'il n'y ait qu'une seule source et que nous utilisons N microphones. Dans un espace ouvert, comme le montre la Figure 3.2, le signal de source $s(k)$ se propage et la puissance du signal diminue en fonction de la distance de la source. Si nous choisissons le premier microphone comme point de référence, le signal capté par le n -ième microphone au moment k peut être exprimé comme suit :

$$y_n(k) = \alpha_n s(k - t - \tau_{n1}) + v_n(k) \quad (3.1)$$

$$= \alpha_n s(k - t - \mathcal{F}_n(\tau)) + v_n(k) \quad (3.2)$$

$$= x_n(k) + v_n(k), n = 1, \dots, N. \quad (3.3)$$

Où $\alpha_n(k)$ ($n = 1, \dots, N$), qui prend des valeurs entre 0 et 1, est le facteur d'atténuation des effets de propagation, $s(k)$ est le signal de source, t est le temps de propagation de la source vers le premier microphone, $v_n(k)$ est le bruit additif à l' n ème microphone, ce dernier est considéré dé-corrélé avec le signal de la source et les signaux des autres microphones, τ est le TDOA entre le microphone 1 et 2, et $\tau_{n1} = \mathcal{F}_n(\tau)$ est le TDOA entre le microphone 1 et n avec $\mathcal{F}_1(\tau) = 0$ et $\mathcal{F}_2(\tau) = \tau$. De $n = 3, \dots, N$, $\mathcal{F}_n(\tau)$ dépend non seulement de τ mais aussi de la géométrie du réseau. Pour un réseau linéaire

$$\mathcal{F}_n(\tau) = (n - 1)\tau \quad n = 2, \dots, N. \quad (3.4)$$

3.2.2 Le modèle du signal de plusieurs sources dans l'environnement libre

S'il y a plusieurs sources, le signal reçu au n ème capteur devient

$$y_n(k) = \sum_{m=1}^M \alpha_{nm} s_m(k - t_m - \mathcal{F}_n(\tau_m)) + v_n(k) \quad (3.5)$$

$$= x_n(k) + v_n(k), \quad n = 1, \dots, N \quad (3.6)$$

Où M est le nombre de sources.

3.2.3 Le modèle du signal d'une seule source dans l'environnement réverbérant

Alors que les modèles idéaux de l'environnement libre ont le mérite d'être simples, ils ne tiennent pas compte de l'effet multi-voies. Par conséquent, ces modèles sont inadéquats pour décrire un véritable

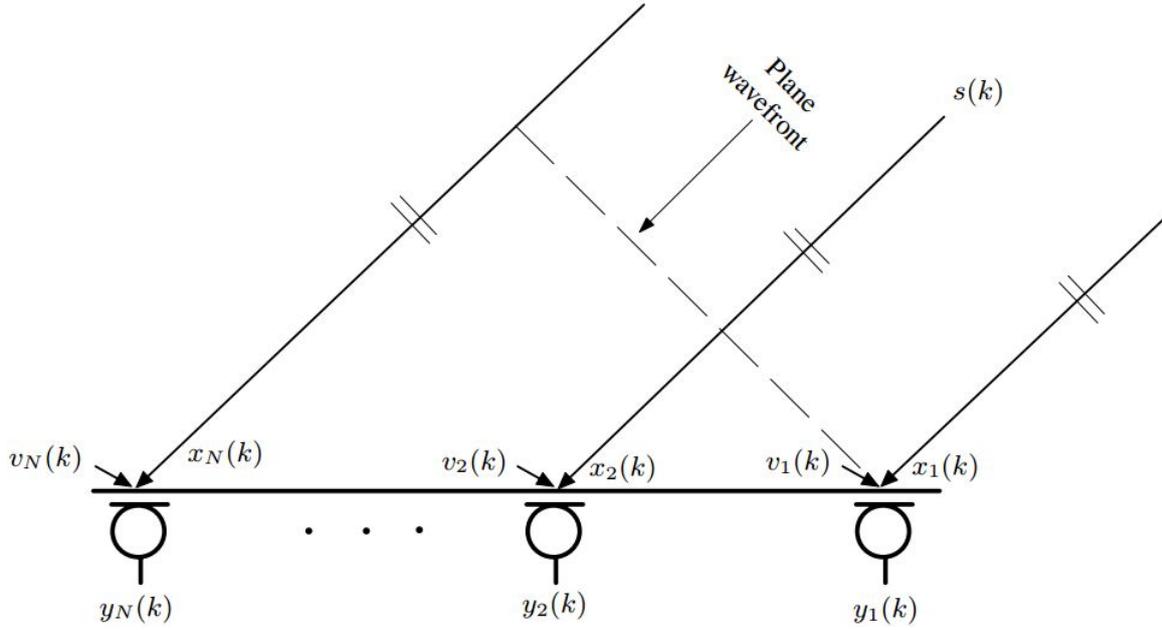


FIGURE 3.2 – Illustration de l'espace ouvert avec seulement une seule source.

environnement réverbérant et nous avons besoin d'une alternative plus complète et plus informative pour modéliser l'effet de la propagation multi-voies, ce qui conduit aux modèles dits réverbérants qui traitent la réponse impulsionnelle acoustique avec un filtre FIR. S'il n'y a qu'une seule source comme illustré à la Figure 3.3, le problème peut être modélisé en tant que système SIMO (Single-Input-Multiple-Output) et le signal du n ème microphone est donné par

$$y_n(k) = g_n * s(k - t - \mathcal{F}_n(\tau)) + v_n(k) \quad (3.7)$$

$$= x_n(k) + v_n(k), \quad n = 1, \dots, N \quad (3.8)$$

Où g_n est la réponse impulsionnelle de la source au n ème microphone. On réécrit 3.7 sous forme vecteur/matrice,

$$\mathbf{y}_n(k) = \mathbf{G}_n \mathbf{s}(k) + v_n(k), n = 1, \dots, N. \quad (3.9)$$

Où

$$\mathbf{y}_n(k) = \begin{bmatrix} y_n(k) & \cdots & y_n(k - L + 1) \end{bmatrix}^T$$

$$\mathbf{G}_n = \begin{bmatrix} g_{n,0} & \cdots & g_{n,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & g_{n,0} & \cdots & g_{n,L-1} \end{bmatrix}$$

$$\mathbf{s}(k) = \begin{bmatrix} s(k) & \cdots & s(k - L - 1) \end{bmatrix}^T$$

$$\mathbf{v}_n(k) = \begin{bmatrix} v_n(k) & \cdots & v_n(k - L + 1) \end{bmatrix}^T$$

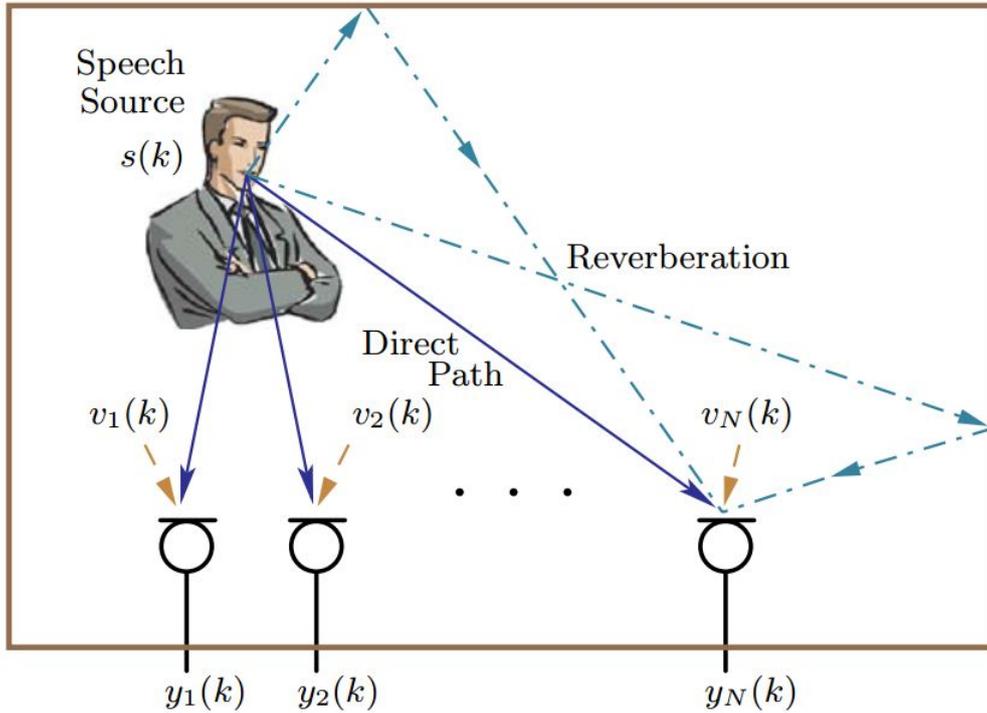


FIGURE 3.3 – Illustration de l’environnement réverbérant avec seulement une seule source.

Et L est la longueur de la réponse impulsionnelle du canal. Encore une fois, on suppose que $v_n(k)$ n’est pas corrélé avec le signal source et le bruit observé aux autres capteurs.

En comparant avec le modèle de l’environnement libre, TDOA τ dans ce modèle réverbérant est un paramètre implicite ou caché. Avec un tel modèle, l’estimation du TDOA ressemble à un problème plus difficile mais n’est heureusement pas insurmontable.

3.2.4 Le modèle du signal de plusieurs sources dans l’environnement réverbérant

S’il y a plusieurs sources, le réseau peut être modélisé en tant que système MIMO (Multiple-Input Multiple-Output) avec M entrées et N sorties. Au temps k , nous avons

$$\mathbf{y}(k) = \mathbf{G}\mathbf{s}_{ML}(k) + \mathbf{v}(k) \quad (3.10)$$

Où

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) & \cdots & y_N(k) \end{bmatrix}^T$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \cdots & \mathbf{G}_M \end{bmatrix}^T$$

$$\mathbf{G}_m = \begin{bmatrix} g_{1m,0} & g_{1m,1} & \cdots & g_{1m,L-1} \\ g_{2m,0} & g_{2m,1} & \cdots & g_{2m,L-1} \\ \vdots & \vdots & \ddots & \vdots \\ g_{Nm,0} & g_{Nm,1} & \cdots & g_{Nm,L-1} \end{bmatrix}_{N \times L}$$

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

$$\mathbf{v}(k) = \begin{bmatrix} v_1(k) & \cdots & v_N(k) \end{bmatrix}^T$$

$$\mathbf{s}_{ML}(k) = \begin{bmatrix} \mathbf{s}_1^T(k) & \mathbf{s}_2^T(k) & \cdots & \mathbf{s}_M^T(k) \end{bmatrix}^T$$

$$\mathbf{s}_m(k) = \begin{bmatrix} \mathbf{s}_m(k) & \mathbf{s}_m(k+1) & \cdots & \mathbf{s}_m(k-L+1) \end{bmatrix}^T$$

et g_{nm} ($n = 1, \dots, N$ $m = 1, \dots, M$) est la réponse impulsionnelle de la source m au microphone n . Comme pour le modèle de l'environnement libre à sources multiples, nous supposons que tous les signaux de sources sont indépendants, et $v_n(k)$ est non corrélé avec les signaux de sources et avec les autres signaux de bruits d'autres microphones.

3.3 La méthode de la corrélation croisée (Cross-Correlation)

Nous sommes maintenant prêts à étudier les algorithmes pour l'estimation du TDOA. Commençons par la méthode la plus simple : la corrélation croisée (CC). Considérons le modèle de signal dans l'environnement libre avec seulement deux capteurs, c'est-à-dire $N = 2$. La fonction de la corrélation croisée (CCF) entre les deux signaux d'observation $y_1(k)$ et $y_2(k)$ est définie comme

$$r_{y_1 y_2}(p) = E [y_1(k) y_2(k+p)] \quad (3.11)$$

Où $E[\cdot]$ est l'espérance.

En utilisant l'équation (3.1)

$$r_{y_1 y_2}(p) = \alpha_1 \alpha_2 r_{ss}(p - \tau) + \alpha_1 r_{ss}(p + t) + \alpha_2 r_{sv_1}(p - t - \tau) + r_{v_1 v_2}(p) \quad (3.12)$$

Si on suppose que le bruit $v_n(k)$ est non corrélé avec le signal de la source et aussi avec les autres signaux de bruits des autres microphones, $r_{y_1 y_2}(p)$ atteint le maximum pour $p = \tau$, donc on peut estimer le TDOA entre $y_1(k)$ et $y_2(k)$ par,

$$\hat{\tau} = \underset{p}{\operatorname{argmax}} r_{y_1 y_2}(p) \quad (3.13)$$

Dans l'implémentation numérique de (3.13), quelques approximations doivent être prises parce que $r_{y_1 y_2}(p)$ est inconnue et on doit l'estimer. Une estimation consiste à remplacer la CCF défini en (3.11) par son estimation de la moyenne. Supposons qu'à l'instant k , nous avons un ensemble d'échantillons d'observation

de $x_n, \{x_n(k), x_n(k+1), \dots, x_n(k+K-1)\}, n = 1, 2$, la CCF correspondante est estimée par

$$r_{y_1 y_2}(p) = \begin{cases} \frac{1}{K} \sum_{i=0}^{K-p-1} y_1(k+i) y_2(k+i+p), & p \geq 0 \\ r_{y_2 y_1}(-p) & p < 0 \end{cases} \quad (3.14)$$

Ou

$$r_{y_1 y_2}(p) = \begin{cases} \frac{1}{K-p} \sum_{i=0}^{K-p-1} y_1(k+i) y_2(k+i+p), & p \geq 0 \\ r_{y_2 y_1}(-p) & p < 0 \end{cases} \quad (3.15)$$

Où K est la taille du bloc. La différence entre (3.14) et (3.15) est que le premier conduit à un estimateur biaisé. Cependant, étant donné qu'il présente une variance d'estimation plus faible et qui est asymptotiquement impartiale, le premier a été largement adopté dans de nombreuses applications.

La méthode de la CCF est simple à mettre en œuvre. Cependant, sa performance est souvent affectée par de nombreux facteurs tels que l'auto-corrélation du signal, la réverbération, etc. De nombreux efforts ont été consacrés à l'amélioration de cette méthode, qui sera discuté dans la section suivante. Mais avant de terminer cette section, nous voudrions souligner un problème potentiel qui est souvent négligé dans l'estimation de TDOA : l'aliasing spatial. L'aliasing spatial peut provoquer une ambiguïté pour que le réseau distingue les signaux qui se propagent à partir de directions différentes. De même, ce problème affectera également l'estimation de TDOA. Pour voir cela, considérons un exemple simple où la source est un signal à bande étroite sous la forme de

$$s(k) = \cos(2\pi f k) \quad (3.16)$$

Si on néglige l'effet de l'atténuation et l'effet du bruit, on obtient,

$$y_1(k) = \cos(2\pi f(k-t)) \quad (3.17)$$

$$y_2(k) = \cos(2\pi f(k-t-\tau)) \quad (3.18)$$

En utilisant (3.11), on obtient

$$r_{y_1 y_2}(p) = E[y_1(k) y_2(k+p)] = \frac{1}{2} \cos(2\pi f(p-\tau)) \quad (3.19)$$

La Figure 3.4 illustre la CCF pour différentes fréquences. L'espacement entre les deux microphones est de 8 cm. En supposant que la vitesse du son est de $320m/s$, on peut facilement vérifier que lorsque $f > 2kHz$, il y aura un aliasing spatial. De la Figure 3.4, nous voyons que lorsque $f > 2kHz$, la CCF connaît plusieurs pics dans l'intervalle $[-\tau_{max}, \tau_{max}]$ (τ_{max} est le TDOA maximum possible et peut être déterminé à partir de l'espacement entre les deux microphones), ce qui rend difficile Pour rechercher le TDOA correct. Dans les applications de microphones, la source est habituellement la parole, qui se compose de plusieurs fréquences. Afin d'éviter le problème d'aliasing et d'améliorer l'estimation de TDOA, il faut passer les signaux par un filtre passe-bas avant de les passer aux algorithmes d'estimation. La fréquence de coupure peut être calculée en utilisant l'espacement des capteurs $f_c = c/2d$.

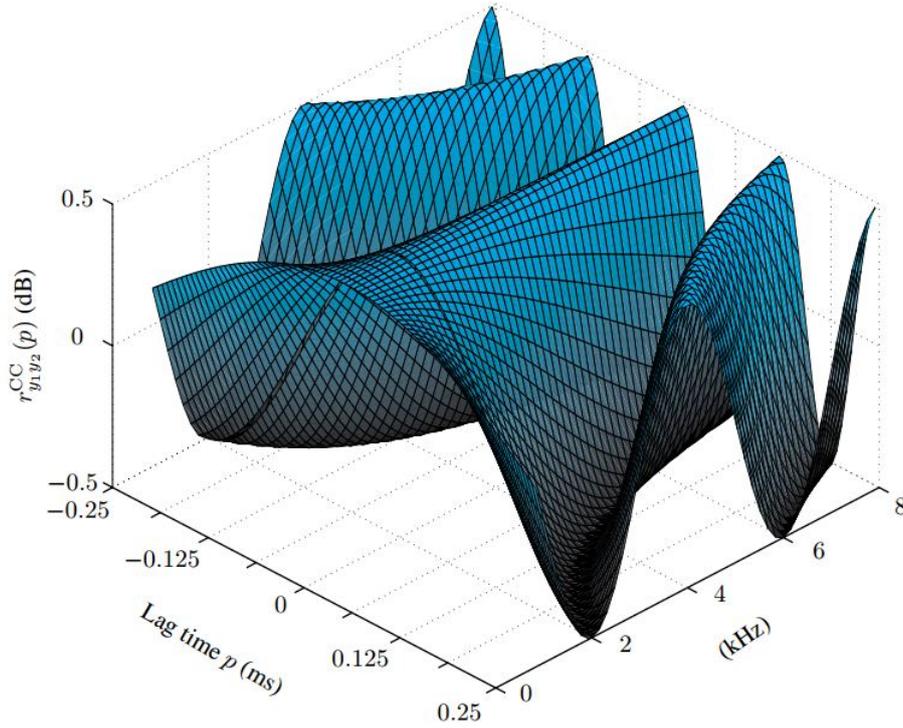


FIGURE 3.4 – Illustration de l’effet de l’aliasing.

3.4 La méthode de la prédiction linéaire spatiale

Dans cette section, nous explorons la possibilité d’utiliser plusieurs microphones (plus de 2) pour améliorer l’estimation de TDOA dans les environnements acoustiques défavorables. L’idée fondamentale est de profiter de l’information redondante fournie par plusieurs capteurs. Pour illustrer la redondance, considérons un système à trois microphones. Dans un tel système, il existe trois TDOA, τ_{12} , τ_{13} et τ_{23} . Apparemment, ces trois TDOA ne sont pas indépendants mais sont liés comme suit : $\tau_{13} = \tau_{12} + \tau_{23}$. Une telle relation, a été utilisée dans plusieurs algorithmes [3]. Dans ce qui suit, nous présentons un algorithme d’estimation du TDOA en utilisant la prédiction linéaire spatiale [4], qui profite de la redondance du TDOA parmi les multiples microphones.

Considérons le modèle de l’environnement libre dans (3.1) avec un réseau linéaire de N microphones ($N \geq 2$). Si la source est dans le champ lointain et nous négligeons les termes du bruit, on peut facilement vérifier que

$$y_n(k + \mathcal{F}_n(\tau)) = \alpha_n s(k - t), \quad n = 1, \dots, N \quad (3.20)$$

$y_1(k)$ est aligné avec $y_n(k + \mathcal{F}_n(\tau))$, à partir de cette relation, on peut définir,

$$e_1(k, p) = y_1(k) - \mathbf{y}_{a,2:N}^T(k, p) \mathbf{a}_{2:N}(p), \quad (3.21)$$

Où p est un variable pour déterminer τ

$$\mathbf{y}_{a,2:N}(k, p) = \begin{bmatrix} y_2(k + \mathcal{F}_2(p)) & \cdots & y_N(k + \mathcal{F}_N(p)) \end{bmatrix} \quad (3.22)$$

est le vecteur de signal aligné, et

$$\mathbf{a}_{2:N}(p) = \begin{bmatrix} a_2(p) & \cdots & a_N(p) \end{bmatrix}^T \quad (3.23)$$

Contient les coefficients de la prédiction linéaire spatiale. Minimiser la valeur du carré moyen du signal d'erreur de prédiction

$$J_1(p) = E \left[e_1^2(k, p) \right] \quad (3.24)$$

Conduit au système linéaire

$$\mathbf{R}_{a,2:N}(p) \mathbf{a}_{2:N}(p) = \mathbf{r}_{a,2:N}(p) \quad (3.25)$$

Où

$$\mathbf{R}_{a,2:N}(p) = E \left[\mathbf{y}_{a,2:N}(k, p) \mathbf{y}_{a,2:N}^T(k, p) \right] \quad (3.26)$$

$$= \begin{bmatrix} \sigma_{y_2}^2 & r_{a,y_2y_3}(p) & \cdots & r_{a,y_2y_N}(p) \\ r_{a,y_3y_2}(p) & \sigma_{y_3}^2 & \cdots & r_{a,y_3y_N}(p) \\ \vdots & \vdots & \ddots & \vdots \\ r_{a,y_Ny_2}(p) & r_{a,y_Ny_3}(p) & \cdots & \sigma_{y_N}^2 \end{bmatrix} \quad (3.27)$$

et la matrice de corrélation spatiale des signaux alignés avec

$$\sigma_{y_n}^2 = E \left[y_n^2(k) \right], n = 1, \dots, N \quad (3.28)$$

$$r_{a,y_iy_j}(p) = E \{ y_i(k + \mathcal{F}_i(p)) y_j(k + \mathcal{F}_j(p)) \}, i, j = 2, 3, \dots, N \quad (3.29)$$

Et

$$\mathbf{r}_{a,2:N}(p) = \begin{bmatrix} r_{a,y_1y_2}(p) & r_{a,y_1y_3}(p) & \cdots & r_{a,y_1y_N}(p) \end{bmatrix}$$

La relation (3.25) nous donne

$$\mathbf{a}_{2:N}(p) = \mathbf{R}_{a,2:N}^{-1}(p) \mathbf{r}_{a,2:N}(p)$$

Et avec la relation (3.21) on trouve l'erreur de prédiction minimale

$$e_1(k, p) = y_1(k) - \mathbf{y}_{a,2:N}^T(k, p) \mathbf{R}_{a,2:N}^{-1}(p) \mathbf{r}_{a,2:N}(p), \quad (3.30)$$

donc

$$J_1(p) = E \left[e_{1,min}^2(k, p) \right] = \sigma_{y_1}^2 - \mathbf{r}_{a,2:N}^T(p) \mathbf{R}_{a,2:N}^{-1}(p) \mathbf{r}_{a,2:N}(p) \quad (3.31)$$

Le temps p qui minimise $J_1(p)$ nous donne le TDOA entre le premier microphone et le deuxième, donc

$$\hat{\tau} = \operatorname{argmin} J_{1,min}(p) \quad (3.32)$$

S'il n'y a que deux microphones, c'est-à-dire $N = 2$, on peut facilement vérifier que cet algorithme est identique à la méthode de la CC. Cependant, au fur et à mesure que le nombre de microphones augmente, cet approche peut profiter des informations redondantes fournies par les autres microphones pour améliorer l'estimation de TDOA.

Pour illustrer cela, on considère un simple exemple de simulation [5], où on a un réseau linéaire composé de 10 microphones omnidirectionnels. L'espacement entre deux capteurs voisins est de $8cm$. Une source sonore située dans le champ lointain rayonne un signal de parole (femelle) sur le réseau, avec un angle d'incidence de $\theta = 75,5degree$. À chaque microphone, le signal est corrompu par un bruit blanc gaussien. Les signaux du microphone sont numérisés avec une fréquence d'échantillonnage de $16kHz$. La Figure 3.5 trace la fonction de coût $J_{1,min}(p)$ calculée à partir d'une trame (128 ms de longueur) de données dans deux conditions de le SNR . Lorsque $SNR = 10dB$, on voit que le système peut réaliser une estimation correcte du TDOA vrai avec seulement deux microphones. Toutefois, au fur et à mesure que le nombre de microphones augmente, la vallée de la fonction de coût devient mieux définie, ce qui permettra une recherche plus simple du minimum. Lorsque le SNR baisse à $-5dB$, cette fois, l'estimation avec deux microphones est incorrecte. Mais lorsque 4 ou plus de microphones sont utilisés, le système produit une estimation correcte. Les deux situations indiquent clairement que la performance d'estimation TDOA augmente avec le nombre de microphones.

3.5 L'algorithme de la corrélation croisée à multi canaux

On a vu dans la section précédente que la clé des techniques basées sur la prédiction spatiale est l'utilisation de la matrice de corrélation spatiale. Une méthode plus naturelle d'utilisation de la matrice de corrélation spatiale dans l'estimation TDOA est par le biais du coefficient de corrélation croisée multi canaux (MCC, Multichannel-Cross-Correlation) [4], qui mesure la corrélation entre les sorties du réseau et peut être visualisé comme une généralisation du coefficient de corrélation croisée classique dans le cas multi canaux et où il existe plusieurs processus aléatoires.

En utilisant (3.22) on définit

$$\mathbf{y}_a(k, p) = \begin{bmatrix} y_1(k) & y_2(k + \mathcal{F}_2(p)) & \cdots & y_N(k + \mathcal{F}_N(p)) \end{bmatrix}^T \quad (3.33)$$

On écrit la matrice de corrélation correspondante

$$\mathbf{R}_a(p) = E \left[\mathbf{y}_a(k, p) \mathbf{y}_a^T(k, p) \right] \quad (3.34)$$

$$= \begin{bmatrix} \sigma_{y_2}^2 & r_{a,y_2y_3}(p) & \cdots & r_{a,y_2y_N}(p) \\ r_{a,y_3y_2}(p) & \sigma_{y_3}^2 & \cdots & r_{a,y_3y_N}(p) \\ \vdots & \vdots & \ddots & \vdots \\ r_{a,y_Ny_2}(p) & r_{a,y_Ny_3}(p) & \cdots & \sigma_{y_N}^2 \end{bmatrix} \quad (3.35)$$

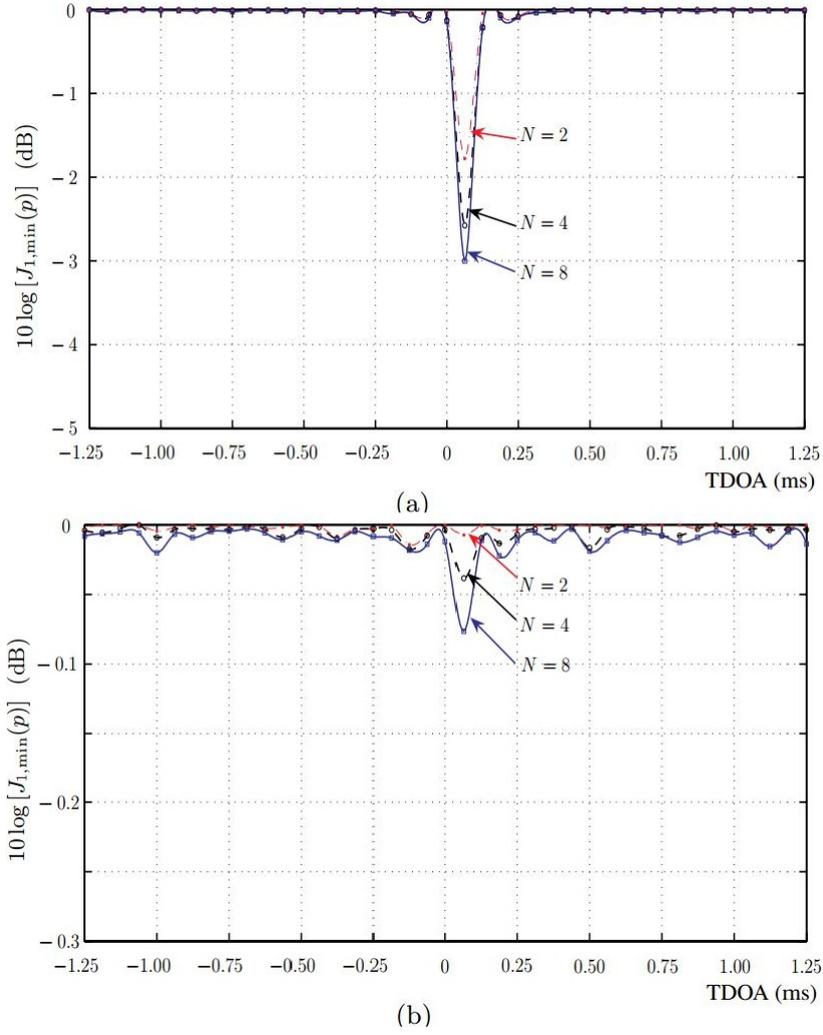


FIGURE 3.5 – Comparaison de $J_{1,\min}(p)$ pour différents nombre de microphones. (a) SNR=10, (b) SNR=-5.

La matrice $\mathbf{R}_a(p)$ peut s'écrire comme

$$\mathbf{R}_a(p) = \mathbf{\Sigma} \hat{\mathbf{R}}_a(p) \mathbf{\Sigma} \quad (3.36)$$

Où

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{y_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{y_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{y_N}^2 \end{bmatrix}$$

est une matrice diagonale, et

$$\hat{\mathbf{R}}_a(p) = \begin{bmatrix} 1 & \rho_{a,y_1y_2}(p) & \cdots & \rho_{a,y_1y_N}(p) \\ \rho_{a,y_2y_1}(p) & 1 & \cdots & \rho_{a,y_2y_N}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{a,y_Ny_1}(p) & \rho_{a,y_Ny_2}(p) & \cdots & 1 \end{bmatrix}$$

est une matrice symétrique, et

$$\rho_{a,y_Ny_2}(p) = \frac{r_{a,y_iy_j}(p)}{\sigma_{y_i}\sigma_{y_j}}, \quad i, j = 1, 2, \dots, N$$

est le coefficient de corrélation entre le i ème et le j ème signaux alignés.

On a $\hat{\mathbf{R}}_a(p)$ est symétrique et semi-définie positive, et tous les éléments diagonaux égalent 1, d'après [4], [6], on a

$$0 \leq \det[\hat{\mathbf{R}}_a(p)] \leq 1 \quad (3.37)$$

Où $\det[\]$ est le déterminant. S'il n'y a que deux microphones, c'est-à-dire $N = 2$, on peut facilement vérifier que le coefficient de corrélation au carré est lié à la matrice de corrélation spatiale normalisée par

$$\rho_{a,y_1y_2}^2(p) = 1 - \det[\hat{\mathbf{R}}_a(p)] \quad (3.38)$$

et par analogie le carré des coefficients de la (MCC) des N signaux alignés $y_n(k + \mathcal{F}_n(p))$ est

$$\rho_{a,y_1y_2}^2(p) = 1 - \det[\hat{\mathbf{R}}_a(p)] \quad (3.39)$$

$$= 1 - \frac{\det[\hat{\mathbf{R}}_a(p)]}{\prod_{i=1}^N \sigma_{y_n}^2} \quad (3.40)$$

les coefficients de la MCC a des propriétés qui sont acceptées sans démonstration [4],[6] :

1. $0 \leq \rho_{a,y_1y_2}^2(p) \leq 1$
2. Si 2 signaux sont totalement corrélés alors, $\rho_{a,y_1y_2}^2(p) = 1$.
3. Si tous les signaux ne sont pas complètement corrélés entre eux, alors $\rho_{a,y_1y_2}^2(p) = 0$.
4. Si l'un des signaux est complètement non corrélé avec les autres $N - 1$ signaux, alors la MCC mesurera la corrélation entre ces $N-1$ signaux restants.

En utilisant la définition de la MCC, on déduit le TDOA entre deux microphones par

$$\tau = \operatorname{argmax}_p \rho_{a,y_1y_2}^2(p) \quad (3.41)$$

qui est équivalente à

$$\tau = \operatorname{argmax}_p \{1 - \det[\hat{\mathbf{R}}_a(p)]\} \quad (3.42)$$

$$= \operatorname{argmax}_p \left\{1 - \frac{\det[\hat{\mathbf{R}}_a(p)]}{\prod_{i=1}^N \sigma_{y_n}^2}\right\} \quad (3.43)$$

$$= \operatorname{argmin}_p \det[\hat{\mathbf{R}}_a(p)] \quad (3.44)$$

$$= \operatorname{argmin}_p \det[\mathbf{R}_a(p)] \quad (3.45)$$

Pour illustrer l'estimation de TDOA avec l'algorithme MCC, nous étudions le même exemple qui a été utilisé dans la section précédente. La fonction de coût $\det [\mathbf{R}_a(p)]$ calculée pour les mêmes données utilisées dans la Figure 3.5 est tracée dans la Figure 3.6. Il est clair que l'algorithme permet d'obtenir une meilleure performance d'estimation.

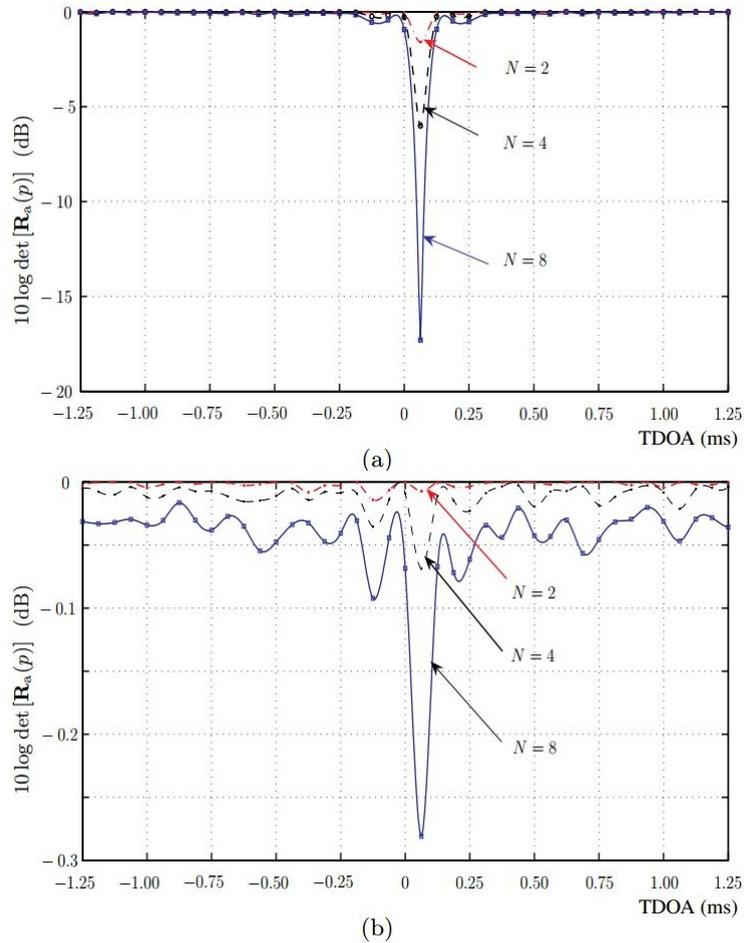


FIGURE 3.6 – Comparaison de $\det [\mathbf{R}_a(p)]$ pour différents nombre de microphones. (a) SNR=10, (b) SNR=-5.

Conclusion générale

Dans ce document, on a présenté le problème de l'estimation de la direction d'arrivée d'un signal (DOA). D'abord, on a comparé les deux méthodes d'estimation de la direction d'arrivée, TDOA et ILD, et on a constaté que l'estimation du TDOA est meilleure que l'estimation du ILD, et elle donne la meilleure résolution.

Nous avons choisi de nous concentrer exclusivement sur les principes de l'estimation TDOA puisque le problème de l'estimation du DOA est essentiellement identique à l'estimation du TDOA. Nous avons discuté l'idée fondamentale de l'estimation du TDOA basée sur le critère de la corrélation croisée qui permet d'utiliser seulement deux microphones. En pratique, le problème d'estimation peut être compliqué par le bruit et la réverbération. Afin d'améliorer la robustesse de l'estimation TDOA par rapport aux distorsions, nous avons discuté deux approches fondamentales : exploiter le fait que nous pouvons avoir plusieurs microphones, et utiliser un modèle de signal réverbérant plus pratique, ce qui a abouti à une large gamme d'algorithmes tels que la prédiction spatiale, la corrélation croisée multi canaux.

Bibliographie

- [1] Adel Thaljaoui, Damien Brulin, Thierry Val, and Nejah Nasri. Localisation d'une source sonore par un réseau de microphones. *Journées Nationales des Communications Terrestres - JNCT*, May 2014.
- [2] J.-M. ; Michaud F. ; Hadjou B. ; Rouat J VALIN et al., Valin. Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach. *IEEE International Conference*, volume 1, April 2004.
- [3] H. Silverman J. DiBiase and M. Brandstein. Robust localization in reverberant rooms. in *Microphone Arrays : Signal Processing Techniques and Applications*, 2001.
- [4] J. Chen J. Benesty and Y. Huang. Time-delay estimation via linear interpolation and cross-correlation. *IEEE Trans. Speech Audio Process*, volume 12 :pp. 509–519, 2004.
- [5] J. Benesty, J. Chen and Y Huang. *Microphone array signal processing*, volume 1. Springer, Montréal, QC/ Murray Hill, NJ/ Bridgewater, NJ, 2008.
- [6] J. Benesty J. Chen and Y. Huang. Robust time delay estimation exploiting redundancy among multiple microphones. *IEEE Trans. Speech Audio Process*, volume 11 :pp. 549–557, Nov 2003.