

Master Sciences de l'Ingénieur
Ingénierie des Systèmes Intelligents
(ISI)

Sorbonne Université

Traitement du son avancé
Compte-Rendu TP : Localisation de source
sonore avec un réseau de microphones :
Approches de beamforming

Étudiants (Groupe 1) :

Edouard Aboudaoud (28712992)
Lyes Mokhbi (21200513)



Année académique 2024-2025

1 Introduction

Ce rapport détaille l'implémentation et l'analyse de la localisation de source sonore à l'aide d'approches de beamforming avec un réseau de microphones de 8 capteurs omnidirectionnels MEMS. Lors du TP précédent, la propagation du son a été caractérisée, et ces informations seront maintenant utilisées pour déterminer la position d'une source sonore par rapport à un réseau linéaire de microphones. Nous travaillerons ici avec une fréquence d'échantillonnage de $F_s = 20$ kHz et une taille de buffer $BLK = 2048$.

2 Acquisition du buffer audio

Pour débuter le processus, il faut commencer l'acquisition du système audio et capturer un buffer audio. Les signaux résultants sont ensuite tracés en fonction du temps pour visualiser les données capturées.

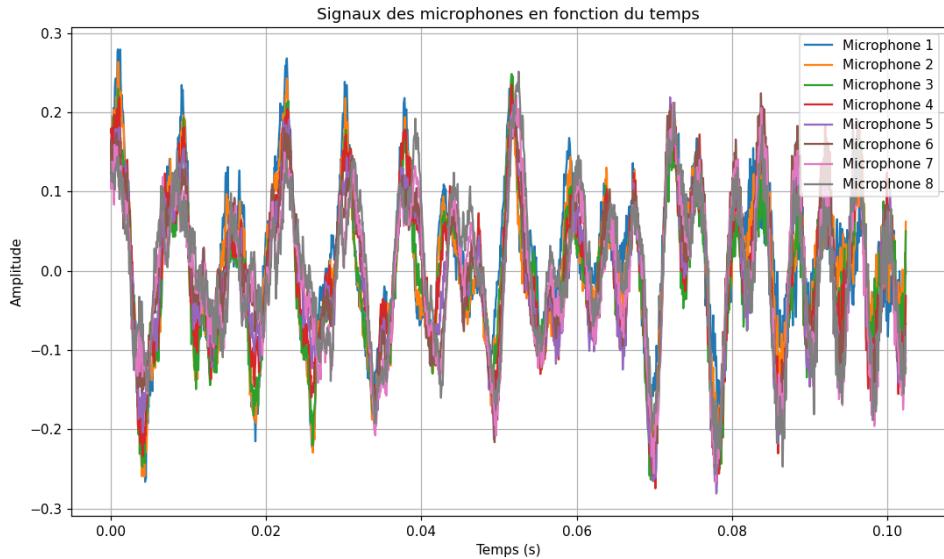


FIGURE 1 – Signal temporel du buffer audio capturé à partir du réseau de microphones.

Commentaire : Dans la Figure 1, nous observons le signal temporel du buffer audio. Ce signal correspond aux sorties brutes des microphones, qui seront ultérieurement traitées à l'aide de techniques de beamforming.

3 Filtres de Beamforming

Le beamforming consiste à appliquer un filtre sur chaque signal des microphones et à sommer les signaux filtrés pour former la sortie du beamformer. La réponse en fréquence du filtre est donnée par l'équation suivante :

$$W_n(f) = \mathcal{F}\{w_n(t)\} = e^{j2\pi \frac{f}{c}x_n \cos(\theta_0)} \quad (1)$$

où :

- $W_n(f)$ est la réponse en fréquence du filtre pour le n -ème microphone,
- f est la fréquence,

- c est la vitesse du son dans l'air,
- x_n est la position du n -ème microphone,
- θ_0 est la direction angulaire de focalisation du beamformer.

3.1 Position des microphones

La position x_n des microphones peut être exprimée en fonction du numéro du microphone n et de l'espacement d entre les microphones. Le premier microphone est numéroté 0, et l'origine est placée au centre du réseau. La position est donnée par l'équation :

$$x_n = \left(n - \frac{N}{2} \right) \cdot d \quad (2)$$

3.2 Fonction de filtre du beamformer

Pour calculer la réponse en fréquence du filtre pour chaque microphone, nous définissons une fonction Python `beam_filter`. Cette fonction prend en entrée le réseau de microphones, un vecteur de fréquences, la direction de focalisation θ_0 , et le numéro du microphone n , puis renvoie la réponse en fréquence correspondante du filtre.

L'implémentation de cette fonction peut être trouvée dans le jupyter notebook joint à ce rapport.

3.3 Comparaison des réponses en fréquence des filtres pour deux microphones à $\theta_0 = 0^\circ$

L'effet de la direction de focalisation θ_0 sur les réponses des filtres est étudié en comparant les filtres associés à deux microphones différents lorsque $\theta_0 = 0^\circ$, pour des fréquences allant de 0 à 5 kHz. Les résultats sont présentés ci-dessous.

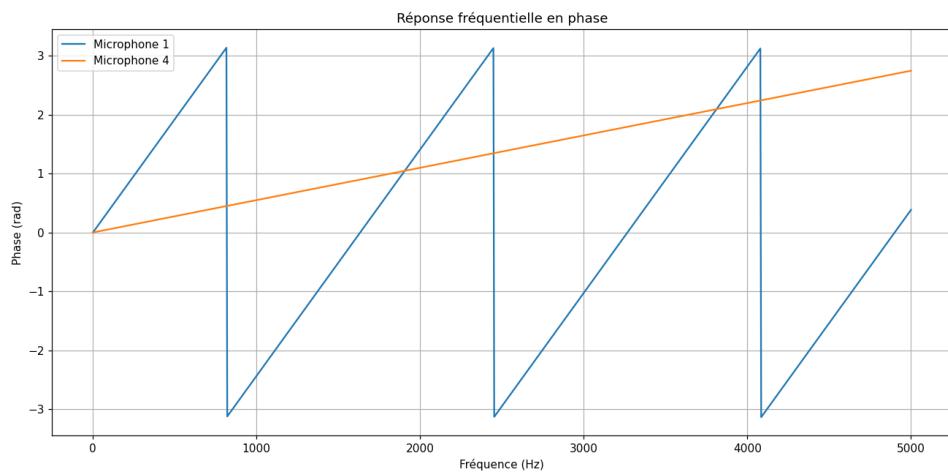


FIGURE 2 – Réponse en fréquence des filtres appliqués au Microphone 1 (bleu) et au Microphone 4 (orange).

Commentaire : Dans la Figure 2, La courbe de phase du Microphone 4 (en orange) montre une progression linéaire sans discontinuités apparentes. Cela correspond à une réponse en phase typique d'un filtre appliqué sur un microphone situé à une position

spécifique dans l'espace. Cette phase linéaire indique que le signal est traité de manière uniforme en termes de retard, ce qui est souvent le cas pour un microphone placé à un angle particulier par rapport à la source sonore. La phase du Microphone 1 (en bleu) présente une progression linéaire, mais elle inclut des discontinuités à intervalles réguliers. Ces discontinuités correspondent à des "sauts de phase" de $-\pi$ à π , ce qui est une propriété normale des représentations angulaires en phase. Cela peut être dû à une interaction complexe entre le filtre et la manière dont le signal est perçu par ce microphone, ou à un effet de traitement numérique des signaux.

3.4 Comparaison des filtres obtenus pour $\theta_0 = 90^\circ$

À $\theta_0 = 90^\circ$, la source sonore est positionnée de manière perpendiculaire à l'axe de l'antenne. Dans cette configuration, tous les microphones reçoivent théoriquement le signal sans différence de temps de propagation. Cela signifie qu'il n'y a pas de retard relatif entre les microphones, et donc, les filtres appliquent peu ou pas de correction de phase.

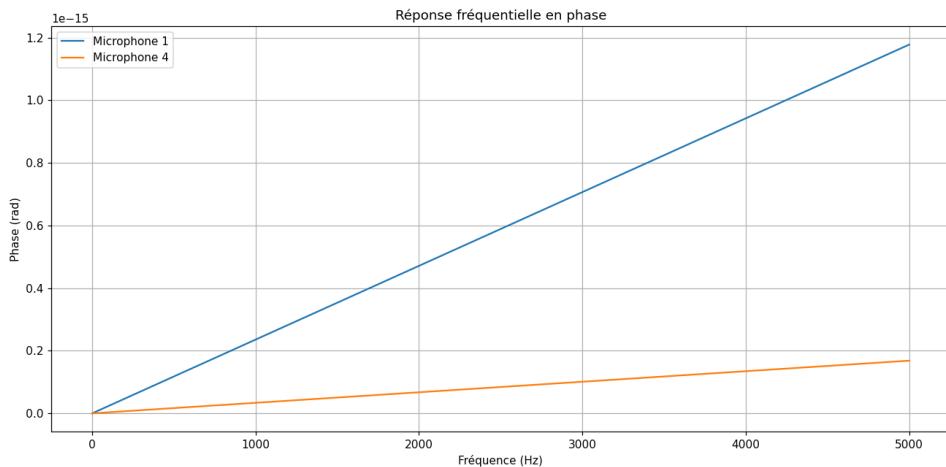


FIGURE 3 – Réponse en fréquence pour $\theta_0 = 90^\circ$.

Commentaire : Dans la Figure 3, À $\theta_0 = 90^\circ$, la phase des filtres est presque nulle, ce qui indique qu'il n'y a pas de besoin de correction importante. Tous les microphones étant positionnés de façon symétrique par rapport à la source, le signal est perçu de manière similaire par chaque microphone, et les ajustements de phase sont minimes.

4 Algorithme de Beamforming

Le processus de beamforming consiste en plusieurs étapes pour localiser une source sonore à une fréquence spécifique. Ces étapes incluent l'acquisition d'un frame audio, la réalisation d'une FFT, et le calcul des sorties du beamformer pour différentes directions.

4.1 Acquisition du buffer audio et calcul de la FFT

La première étape consiste à acquérir un buffer audio et à calculer sa FFT. La FFT du buffer permet de révéler les composantes fréquentielles des signaux en provenance de tous les microphones. La transformation de Fourier est donnée par :

$$M_{\text{fft}}(k) = \sum_{t=0}^{BLK-1} m(t) \cdot e^{-j2\pi kt/BLK} \quad (3)$$

où $M_{\text{fft}}(k)$ est la composante fréquentielle à l'indice k , $m(t)$ est le signal du microphone à l'instant t , et BLK est la taille du buffer.

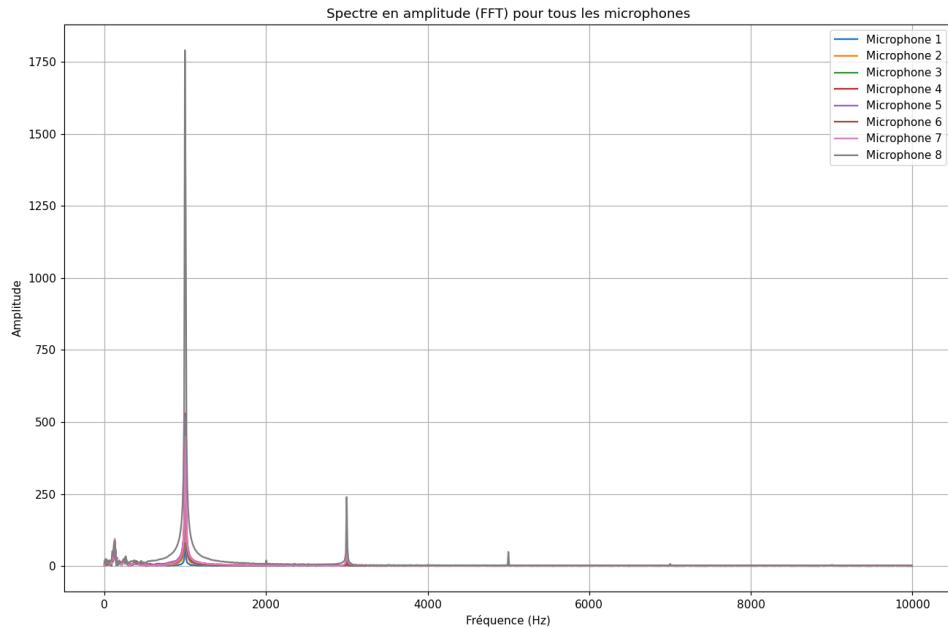


FIGURE 4 – FFT d'un buffer audio lors de l'émission d'un ton sinusoïdal à 1 kHz.

Commentaire : Dans la Figure 4, - Un pic clair et dominant est visible à la fréquence $F_0 = 1\text{kHz}$ comme attendu, car un signal pur sinusoïdal à cette fréquence a été émis. Cela confirme que le signal a bien été capté par tous les microphones. Les amplitudes des différents microphones sont très proches pour la fréquence principale (1kHz). On observe également des pics à des fréquences multiples de F_0 (2kHz, 3kHz), ces pics sont probablement résultant du bruits présent dans la salle

4.2 Sélection de la fréquence et application des filtres

Une fois les composantes fréquentielles obtenues après l'application de la transformée de Fourier, il est nécessaire de sélectionner la fréquence d'intérêt, dans ce cas, la fréquence $F_0 = 1\text{kHz}$, et d'extraire les valeurs FFT associées à cette fréquence pour chaque microphone. Cette étape repose sur l'identification de l'index k_0 correspondant à la fréquence la plus proche de F_0 dans le vecteur de fréquences positives obtenu lors de la FFT.

L'index k_0 est déterminé en trouvant la position du pic le plus proche de la fréquence F_0 dans le vecteur des fréquences positives positive_freqs, à l'aide de la formule suivante :

$$k_0 = \arg \min_k |\text{positive_freqs}[k] - F_0| \quad (4)$$

Une fois cet index k_0 identifié, la fréquence exacte associée à cet index est récupérée comme suit :

$$\text{exact_freq} = \text{positive_freqs}[k_0] \quad (5)$$

À partir de cette fréquence exacte, les valeurs FFT associées à tous les microphones à cette fréquence sont extraites. Ces valeurs sont représentées par le vecteur M , qui contient les amplitudes et phases des signaux de chaque microphone pour la fréquence F_0 . En notation :

$$M[k_0] = [M_1[k_0], M_2[k_0], \dots, M_N[k_0]] \quad (6)$$

où chaque $M_n[k_0]$ représente la composante FFT pour le n -ième microphone à la fréquence F_0 .

Les résultats obtenus pour $k_0 = 102$, correspondant à une fréquence exacte de 996,09 Hz, sont les suivants :

Amplitudes et Phases FFT pour chaque microphone

Microphone	Amplitude FFT	Phase FFT
1	78.62042	-2.0243607
2	239.8363	-1.3679484
3	419.63046	-1.113776
4	529.75995	-0.9095199
5	516.78784	-0.5971969
6	338.89374	0.00567186
7	448.90628	1.723821
8	1790.6444	2.604943

Interprétations :

Index $k_0 = 102$ et fréquence exacte :

- L'index $k_0 = 102$ correspond à une fréquence de 996.09375 Hz, ce qui est très proche de la fréquence source attendue (1 kHz).
- Cette petite différence est due à la résolution en fréquence limitée par la taille du tampon FFT (BLK) et la fréquence d'échantillonnage (Fs).

Amplitudes des microphones :

- Les amplitudes FFT varient entre les microphones, allant de 78.62 (Microphone 1) à 1790.64 (Microphone 8). Cela peut refléter plusieurs facteurs :
- Les microphones plus proches de la source peuvent capter un signal avec une amplitude plus élevée.
- Les amplitudes varient en fonction de l'orientation du réseau par rapport à la source sonore.

Phases des microphones :

- Les phases varient entre les microphones, allant de -2.02 rad (Microphone 1) à 2.60 rad (Microphone 8).
- Chaque microphone capte le signal à un moment légèrement différent, entraînant un déphasage.

Observations supplémentaires :

- L'amplitude est proportionnelle à la distance à la source. Les microphones plus proches captent des amplitudes plus élevées.
- La phase est directement liée au retard de propagation du signal sonore pour atteindre chaque microphone (sa position).

4.3 Calcul et application des filtres pour chaque microphone

Une fois la fréquence d'intérêt f_0 et la direction principale θ_0 définies, il est nécessaire de calculer les filtres associés à chaque microphone du réseau. Ces filtres sont obtenus à partir de la fonction `beam_filter` qui, pour chaque microphone, calcule un facteur de filtrage en fonction de la position θ_0 et de la fréquence f_0 obtenue précédemment.

Les filtres sont ensuite appliqués à la composante FFT de chaque microphone, ce qui permet d'obtenir un vecteur M_{filtered} contenant les résultats après filtrage.

Le calcul du filtre (Cf. Le notebook jupyter joint au rapport pour davantage de détails) pour le n -ième microphone se fait de la manière suivante :

$$\text{filter_value} = \text{beam_filter}(\text{antenne}, [f_0], \theta_0 = 0, \text{mic_nb} = n)[0]$$

Puis, le filtre est appliqué au vecteur M_n de chaque microphone :

$$M_{\text{filtered}}[n] = M[n] \times \text{filter_value}$$

Les résultats de l'application des filtres à chaque microphone montrent les contributions filtrées des signaux de chaque microphone, dont les valeurs de phase et d'amplitude sont ajustées en fonction du filtre appliqué.

Exemple des résultats après application des filtres :

Microphone	Composantes de M_{filtered}
1	$-18.43344065 + 76.42891507j$
2	$48.05455807 + 234.9727802j$
3	$362.39204926 + 211.56965028j$
4	$495.40447064 - 187.66990246j$
5	$213.64548864 - 470.55848847j$
6	$-22.26316136 - 338.16166154j$
7	$237.54011986 - 380.90880869j$
8	$603.76639869 - 1685.78564487j$

Interprétations :

Vecteur M_{filtered} après application des filtres :

- Le vecteur M_{filtered} contient les contributions des microphones après filtrage, exprimées sous forme complexe.
- Chaque élément complexe est composé de deux parties : l'amplitude (la magnitude) et la phase.

Amplitude et phase des microphones :

- Les amplitudes des microphones varient significativement, ce qui peut être dû à des facteurs comme la distance des microphones par rapport à la source sonore ou des réflexions acoustiques dans l'environnement. (**voir Figure 5 pour un histogramme des amplitudes**).

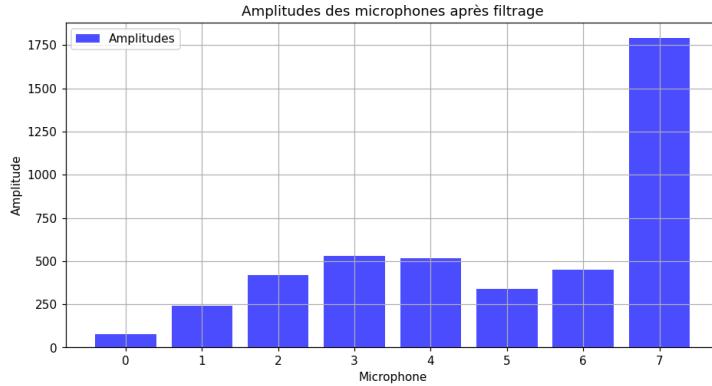


FIGURE 5 – Histogramme des amplitudes des microphones.

- Les phases varient également d'un microphone à l'autre, ce qui reflète le déphasage dû à la position relative de chaque microphone par rapport à la source sonore. (**voir Figure 6 pour l'évolution de la phase**).

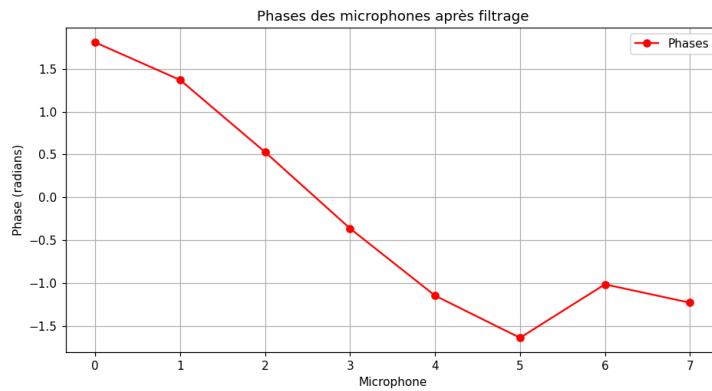


FIGURE 6 – Évolution de la phase des microphones.

4.4 Calcul de la puissance $P(\theta_0)$ pour différentes directions

Dans cette section, nous avons calculé la puissance $P(\theta_0)$ pour une source émettant dans deux directions différentes : une proche de $\theta_0 = 0^\circ$ et une autre éloignée, à $\theta_0 = 45^\circ$. Les résultats sont présentés dans le tableau ci-dessous.

$$P(\theta_0) = \left| \sum_{n=0}^{N-1} W_n(f) \cdot x_n \right|^2 \quad (7)$$

Où $P(\theta_0)$ est la puissance du beamformer pour la direction θ_0 , et $W_n(f)$ est la réponse en fréquence du filtre appliquée au n -ème microphone.

Direction θ_0	Puissance $P(\theta_0)$
Proche ($\theta_0 = 0^\circ$)	10,138,983.77
Éloignée ($\theta_0 = 45^\circ$)	15,500,366.36

TABLE 1 – Puissance $P(\theta_0)$ pour différentes directions de la source

Interprétations :

- Contrairement à ce que l'on pourrait attendre, la puissance pour la direction éloignée ($\theta_0 = 45^\circ$) est plus élevée que pour la direction proche ($\theta_0 = 0^\circ$).

Analyse des causes possibles :

- **Réflexions ou interférences acoustiques** : Des réflexions dans l'environnement pourraient introduire des contributions significatives dans une direction différente de $\theta_0 = 0^\circ$, comme $\theta_0 = 45^\circ$, ce qui pourrait augmenter la puissance observée.
- **Mauvaise estimation de la source réelle** : Si la source sonore est réellement plus proche de $\theta_0 = 45^\circ$, les phases seraient mieux alignées dans cette direction, ce qui expliquerait la puissance plus élevée observée dans cette direction.
- **Calibration des filtres ou des microphones** : Une calibration imparfaite des microphones ou des filtres pourrait provoquer des erreurs dans l'alignement des phases, affectant les résultats attendus et contribuant à une puissance plus élevée dans une direction éloignée.
- **Bruit ou sources parasites** : Des bruits ou des interférences provenant d'autres directions pourraient contribuer de manière inattendue à la puissance mesurée dans la direction éloignée ($\theta_0 = 45^\circ$).

Observations supplémentaires :

- La puissance observée dépend de la manière dont les signaux sont filtrés et du déphasage entre les différents microphones.
- Des facteurs externes comme les réflexions ou les erreurs de calibration peuvent influencer les résultats de manière significative, et leur impact doit être pris en compte dans l'analyse des résultats.

4.5 Calcul de la puissance et localisation directionnelle

Pour chaque direction θ_0 , la puissance de la sortie du beamformer est calculée et tracée en fonction de θ_0 afin de détecter la direction de la puissance maximale. La puissance est donnée par la formule 7

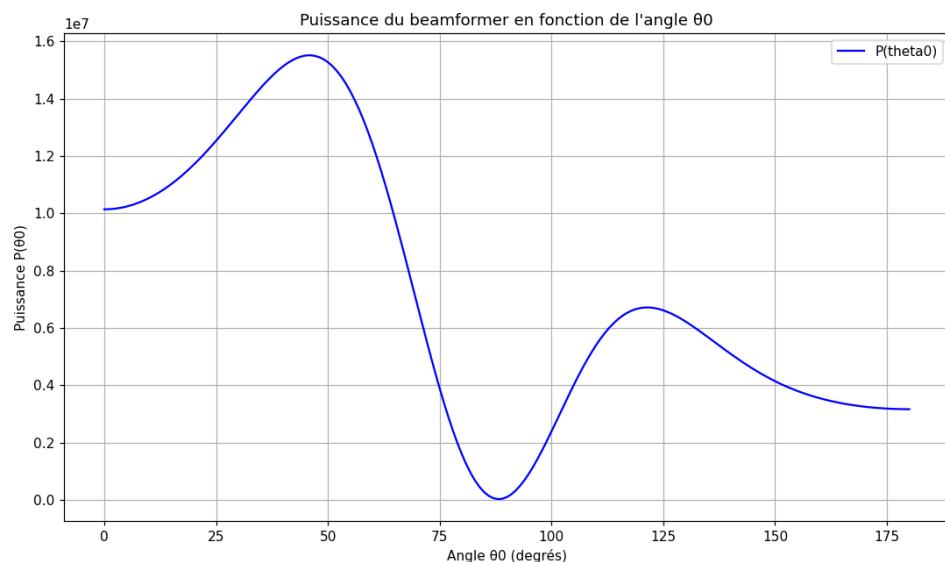


FIGURE 7 – Puissance de la sortie du beamformer en fonction de θ_0 .

Interprétations de la figure 7 :

- Un pic de puissance est clairement visible autour de $\theta_0 = 45^\circ$
- Une diminution de la puissance est observée pour des angles éloignés de cette direction, avec un minimum autour de $\theta_0 = 90^\circ$
- Cela suggère que la source sonore est située autour de $\theta_0 = 45^\circ$
- Dans un environnement non anéchoïque, des réflexions sur les murs ou objets environnants peuvent créer des sources virtuelles. Ces sources réfléchies apparaissent comme si elles venaient d'une direction différente, par exemple 45°
- Si l'intensité des réflexions est significative, le beamformer peut les interpréter comme la source principale.

5 Analyse des performances du beamformer

Dans cette section, nous analysons les performances du beamformer en examinant les cartes d'énergie pour différentes fréquences de la source et en évaluant la précision de l'estimation de la position dans le cas d'une source mobile.

5.1 Cartes d'énergie pour des fréquences fixes

Nous avons généré des cartes d'énergie pour des fréquences fixes $F_0 = 400$ Hz, 1 kHz, 2 kHz, et 4 kHz, émettant à partir d'une position fixe et arbitraire. Ces cartes permettent de visualiser l'intensité relative du signal en fonction de la position dans le champ d'écoute.

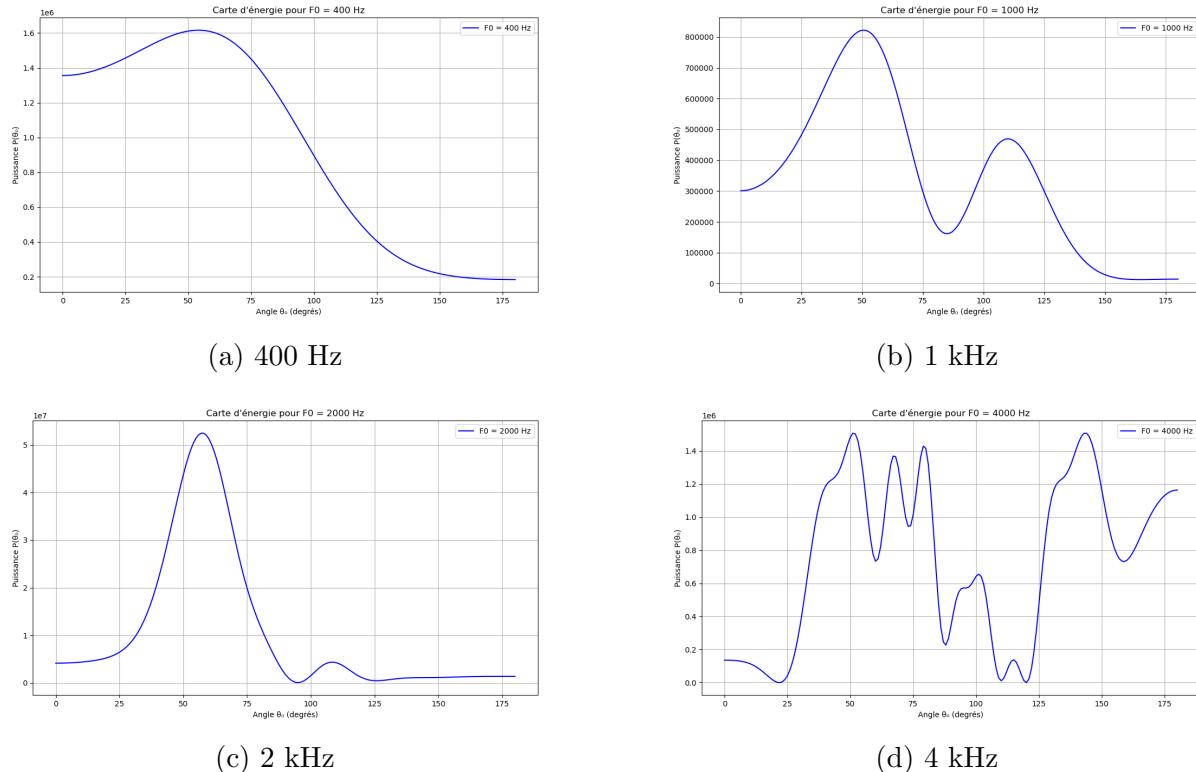


FIGURE 8 – Cartes d'énergie pour $F_0 = 400$ Hz, 1 kHz, 2 kHz, et 4 kHz pour une source émettant d'un angle de $\theta_0 = 0^\circ$.

Interprétation :

- Pour la fréquence $F_0 = 400 \text{ Hz}$ (voir Figure 8a), la courbe montre une intensité maximale autour de $\theta_0 = 50^\circ$, ce qui est conforme aux attentes, étant donné que le beamformer est focalisé sur la direction $\theta_0 = 0^\circ$. Ce décalage peut être attribué à une légère imprécision dans l'orientation du téléphone. L'intensité diminue progressivement à mesure que θ_0 s'éloigne de 50° , ce qui indique que la sensibilité du microphone diminue en dehors de son axe principal. L'atténuation en dehors de $\theta_0 = 50^\circ$ n'est pas très abrupte, ce qui est cohérent avec une faible directivité à cette fréquence.
- Pour la fréquence $F_0 = 1 \text{ kHz}$ (voir Figure 8b), l'intensité maximale est également observée près de $\theta_0 = 50^\circ$, mais l'atténuation est plus marquée à mesure que l'on s'éloigne de cette direction. De plus, un deuxième lobe apparaît autour de $\theta_0 = 120^\circ$. Ce phénomène peut être dû à une imprécision dans la localisation de la source, résultant probablement du fait que la fréquence de l'onde émise est supérieure à la fréquence limite du système. En effet, le système est limité en fréquence et fonctionne uniquement en dessous de f_{lim} . Ce comportement est dû au fait que la distance entre les deux microphones devient trop grande par rapport à la longueur d'onde λ , entraînant la capture de plusieurs périodes par le premier microphone avant que l'onde n'atteigne le second. Il devient alors impossible de déterminer correctement la vitesse du son en se basant sur les extréums ou les zéro-crossing. La fréquence limite f_{lim} peut être calculée à partir des caractéristiques du montage, avec la condition suivante :

$$\lambda > d$$

Or, $\lambda = \frac{c}{f}$, donc la fréquence doit respecter la relation :

$$f < \frac{c}{d}$$

Dans notre cas, $f_{lim} = 810 \text{ Hz}$.

- Pour la fréquence $F_0 = 2 \text{ kHz}$ (voir Figure 8c), l'intensité continue de diminuer à mesure que l'angle θ_0 s'éloigne de 0° , mais avec une chute plus nette. Nous observons deux lobes similaires à ceux du cas précédent, mais la largeur du lobe principal est plus petite et il est centré autour de $\theta_0 = 60^\circ$. De plus, l'amplitude du second lobe est plus faible comparée à celle observée pour $F_0 = 1 \text{ kHz}$. Cela indique une amélioration de la directivité, ce qui est attendu.
- Enfin, pour la fréquence $F_0 = 4 \text{ kHz}$ (voir Figure 8d), nous constatons des oscillations caractéristiques d'une perte de directivité, ce qui est prévu pour des fréquences élevées en raison du phénomène de repliement spatial.
- En résumé, à mesure que la fréquence augmente, la directivité du faisceau s'améliore, avec une atténuation plus rapide en dehors de la direction principale. Pour une fréquence élevée (4 kHz), la directivité devient plus pointue, mais le rayon de microphone continue de détecter la source dans toutes les directions.

5.2 Estimation de la position pour une source mobile

La méthode d'estimation repose sur le calcul de la puissance maximale P_{\max} à partir des signaux reçus par plusieurs microphones disposés en réseau. La procédure est décrite mathématiquement comme suit.

Soit **audio** une matrice représentant les signaux reçus, où chaque colonne correspond à un microphone. Le premier pas consiste à calculer la fréquence dominante k_0 pour chaque signal :

$$k_0 = \arg \max (|\mathcal{F}(\text{audio}(:, 1))|)$$

où \mathcal{F} désigne la transformation de Fourier discrète (FFT). Ensuite, pour chaque angle j allant de 0 à 360, nous calculons la somme des puissances associées à chaque direction :

$$P_j = \left| \sum_{i=1}^8 (\text{beam_filter}(\text{antenne}, f_{k_0}, j, i) \cdot \mathcal{F}(\text{audio}(:, i))_{k_0}) \right|^2$$

Ici, $\text{beam_filter}(\text{antenne}, f_{k_0}, j, i)$ est un filtre de faisceau qui dépend de l'antenne, de la fréquence f_{k_0} et de l'angle j , et $\mathcal{F}(\text{audio}(:, i))_{k_0}$ est la valeur de la FFT à la fréquence dominante k_0 pour le i -ème microphone.

Enfin, la position estimée de la source correspond à l'angle j_{\max} où la puissance est maximale :

$$P_{\max} = \arg \max(P)$$

Cela permet de déterminer la direction de la source mobile à chaque instant.

Pour une fréquence fixe de $F_0 = 1 \text{ kHz}$, nous avons déplacé une source autour du microphone. La position estimée en fonction du temps est représentée dans la figure suivante.

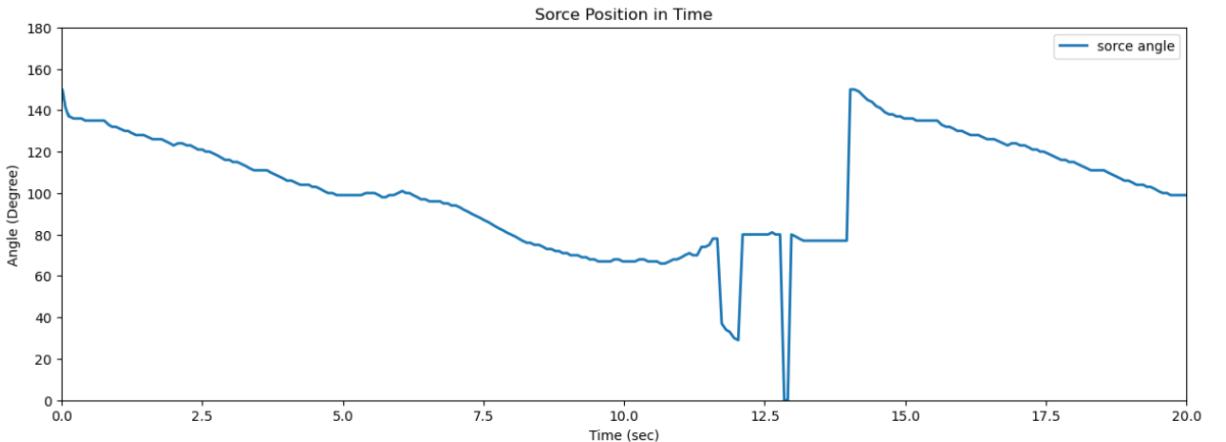


FIGURE 9 – Position estimée de la source mobile ($F_0 = 1 \text{ kHz}$) en fonction du temps.

Interprétation : Globalement, la méthode d'estimation semble robuste avec un suivi continu de l'angle sur la majeure partie de la période observée. La source mobile se déplace principalement dans une direction circulaire, ce qui se traduit par des variations progressives de l'angle au cours du temps.

Cependant, aux environs de $t = 12.5$ s, une chute rapide de l'angle est suivie d'une augmentation brusque. Après ces perturbations, la courbe reprend une tendance régulière et décroissante, qui se poursuit jusqu'à la fin de l'intervalle de 20 secondes.

Cette anomalie peut être expliquée par l'utilisation de **buffers** lors de la récupération du signal. En effet, une observation plus attentive de la courbe dans l'intervalle [15 s; 20 s] révèle que son évolution est identique à celle observée dans l'intervalle initial [0 s; 5 s]. Cela suggère que la dernière partie du signal correspond au début du signal, affiché de nouveau.

Conclusion

Ce rapport a exploré les performances d'un système de microphones et de beamforming pour la localisation d'une source sonore. Les résultats montrent que les variations d'amplitude et de phase entre les microphones sont influencées par la position de la source, les réflexions acoustiques et la calibration des équipements. L'analyse de la puissance directionnelle a révélé des résultats inattendus, suggérant l'impact des réflexions et des erreurs d'estimation de la source. Enfin, l'amélioration de la directivité avec l'augmentation de la fréquence a été observée, bien que des artefacts liés aux limites du système soient apparus à des fréquences plus élevées.