

UNIVERSITÉ DE BLIDA 1

Faculté des sciences

Département d'informatique



MÉMOIRE DE MASTER

En Informatique

Option : Ingénierie Des Logiciels

THÈME :

Détection et localisation en temps réel des événements sonores en vue de la conception d'un système de surveillance audio

Réalisé par :

HATTALI Ahmed & MOUSSELMAL Abdelaziz

Devant un jury composé de :

BENYAHIA Mohammed, Maitre Assistant A, Blida 1

Président

MIDOUN Khadidja, Maitre Assistant B, Blida 1

Examinatrice

YKHLEF Fayçal, Maître de Recherche, CDTA

Promoteur

YKHLEF Hadjer, Maitre de Conférence B, Blida1

Co-Promoteur

2021/2022

REMERCIEMENTS

Avant tout, nous remercions Allah le tout puissant de nous avoir donné le courage et la volonté pour mener à bien ce modeste travail.

Nos premiers remerciements sont adressés à notre promoteur, Dr. Fayçal YKHLEF, Maitre de Recherche au CDTA, qui nous a proposé ce thème et nous a fait confiance malgré nos connaissances plutôt légères dans le domaine de traitement de signal. Nous le remercions pour son aide précieuse et ses judicieux conseils et pour le temps qu'il nous a consacré.

Nous remercions les chercheurs du Centre de Développement des Technologies Avancées (C.D.T.A) et en particulier ceux de la Division Architecture des Systèmes et Multimédias (ASM) de nous avoir accueilli et d'avoir mis à notre disposition les conditions favorables pour la réalisation de ce présent travail.

Nous tenons à remercier notre co-encadreur Dr. Hadjer YKHLEF, Maitre de Conférence au niveau de l'Université SAAD DAHLEB, pour sa disponibilité ses conseils et son sens d'écoute et d'échange.

Nos remerciements vont également aux membres du jury pour avoir accepté d'examiner notre travail et de l'enrichir par leurs propositions.

Nous souhaitons aussi adresser nos remerciements au corps professoral et administratif de l'Université de BLIDA qui a contribué à la réussite de nos études universitaires.

DEDICACE

Je dédie ce modeste travail à :

A mes chers parents, que nulle dédicace ne puisse exprimer mes sincères sentiments,

Pour leur patience illimitée, leur encouragement continu, leur aide, en témoignage de mon profond amour et respect pour leurs grands sacrifices.

A ma sœur, à mon frère et à tous mes cousins qui m'encouragent toujours à me surpasser

A tous mes amis qui ont été toujours à mes côtés surtout dans les moments difficiles, A mon binôme Abdelaziz et à tous ceux qui ont contribué de près ou de loin pour que ce travail soit possible, je vous dis merci.

Ahmed

DEDICACE

Je dédie ce modeste travail à :

A ma très chère mère

Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se doit. Ton affection me couvre, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force pour affronter les différents obstacles.

A mon très cher père

Tu as toujours été à mes côtés pour me soutenir et m'encourager. Que ce travail traduit ma gratitude et mon affection.

A mon très cher frère, mes sœurs et mes oncles qui m'ont toujours encouragé

Puisse Dieu vous donne santé, bonheur, courage et surtout réussite

A mon meilleur ami Ahmed et à tous mes collègues du club scientifique ITC.

Abdelaziz

RESUME

L'objectif principal de ce projet de fin d'étude est la conception d'un système de surveillance audio capable de détecter les événements acoustiques impulsifs et de localiser leur source en temps réel. Nous utilisons l'analyse de la variance des énergies à court terme du signal pour détecter ces événements. Quant à la localisation, nous exploitons le *generalized cross-correlation phase transform* pour estimer l'angle d'incidence de la source. L'acquisition du signal audio est accomplie en utilisant le réseau de microphones incorporé dans la Microsoft Xbox Kinect v1. Les métriques d'évaluation utilisées pour mesurer les performances de la détection et la localisation sont respectivement : (i) Le F1-score et (ii) l'erreur absolue de l'écart entre l'angle mesuré et sa valeur exacte. Nous avons implémenté un prototype sous python. Les tests ont été effectués dans des environnements fermés. Les résultats obtenus sont prometteurs.

Mots clés : Détection, Localisation de la source sonore, Différence entre les temps d'arrivée, *Generalized cross-correlation phase transform*, microphones MEMS.

Abstract

The main objective of this graduation project is to design an audio monitoring system capable of detecting impulsive acoustic events and localizing their source in real-time. We use the variance analysis of short-term signal energies to detect these events. As for localization, we exploit the generalized cross-correlation phase transform to estimate the angle of incidence of the source. The acquisition of the audio signal is accomplished using the microphone array embedded in the Microsoft Xbox Kinect v1. The evaluation metrics used to measure the detection and localization performances are respectively: (i) the F1-score and (ii) the absolute error of the deviation between the estimated angle and its exact value. We have implemented a prototype under python. The tests were conducted in closed environments. The obtained results are promising.

Keywords: Detection, localization of the sound source, Time difference of arrival, Generalized cross-correlation phase transform, MEMS Microphones.

ملخص

الهدف الرئيسي من مشروع التخرج هو تصميم نظام مراقبة الصوت قادر على اكتشاف الأحداث الصوتية المندفعة وتحديد مصدرها في الوقت الفعلي. تعتمد مرحلة الكشف عن هاته الأحداث على تحليل تباين الطاقات قصيرة المدى للإشارة. أما بالنسبة لتحديد مصدر الصوت، فتتم باستغلال خوارزمية generalized cross-correlation phase transform لتقدير الزاوية. يتم إلتقاط الإشارة عن طريق استخدام مصفوفة لواقط صوتية مدمجة في جهاز Microsoft Xbox Kinect v1 . مقاييس التقييم المستخدمة لقياس أداء الكشف و تحديد المصدر هي على التوالي: F1-Score والخطأ المطلق لحساب الفرق بين الزاوية الحقيقية و الزاوية التي تم قياسها. تم القيام بتطبيق النموذج باستعمال لغة البايثون. تم اختبار النموذج المقترح داخل قاعة مغلقة، والنتائج المتحصل عليها واعدة ومحفزة.

الكلمات الدالة: الكشف، تحديد مصدر الصوت، الفاصل الزمني للإشارة الصوتية، -Generalized cross-correlation phase transform، ميكروفون MEMS.

LISTE DES ACRONYMES ET ABREVIATIONS

CPU : *Central processing unit*

EA : *L'erreur absolue*

Fe : *Fréquence d'échantillonnage*

GCC : *Generalized Cross Correlation*

GCC-PHAT : *Generalized Cross Correlation with Phase Transform*

GMM : *Gaussian Mixture Model*

PHAT : *Phase Transform*

L : *Taille de la séquence d'énergie*

MEMS : *Micro-Electro-Mechanical Systems*

N : *Taille du bloc*

RAM : *Random-Access Memory*

TDOA : *Time difference of arrival*

Th : *Threshold*

TABLE DES MATIERES

Remerciements	(i)
Dédicaces	(ii)
Résumé	(iii)
Liste d'abréviations et acronymes	(iv)
Table des matières	(v)
Liste des figures	(vi)
Liste des tableaux	(vii)

CHAPITRE 1 : INTRODUCTION.....7

1.1 Motivations	7
1.2 Contributions	7
1.3 Impact du projet	8
1.4 Organisation du mémoire	8

CHAPITRE 2 : SYSTEMES DE SURVEILLANCE AUDIO.....9

2.1 Introduction	9
2.2 Généralités sur les technologies de surveillance	9
2.3 Avantages de l'audio pour le monitoring	10
2.4 Applications de la surveillance audio	11
2.5 Architecture d'un système de monitoring audio	12
2.6 Méthodes proposées dans la littérature pour la détection et la localisation des sons impulsifs	16
2.7 Conclusion	17

CHAPITRE 3 : METHODES PROPOSEES POUR LA DETECTION ET LA LOCALISATION DES EVENEMENTS SONORES18

3.1 Introduction	18
3.2 Architecture globale du système proposé	18
3.3 Acquisition	20
3.4 Détection	20
3.5 Localisation	25
3.6 Évaluation des performances de la détection des sons	29
3.7 Mesure de la précision de localisation	30
3.8 Conclusion	30

CHAPITRE 4 : RESULTATS ET EXPERIMENTATIONS	31
4.1 Introduction	31
4.2 Logiciels et matériel de développement	31
4.3 Conception du prototype de détection et localisation des sons impulsifs	36
4.4 Scénarios de test.....	40
4.5 Résultats expérimentaux	43
4.5 Conclusion.....	47
CHAPITRE 5 : CONCLUSIONS ET TRAVAUX FUTURES.....	48
5.1 Conclusion.....	48
5.2 Travaux futurs	49
BIBLIOGRAPHIE.....	50

LISTE DES FIGURES

Figure 2.1: Champ de vision limité d'une caméra de surveillance.	10
Figure 2.2: Trajectoires sonores directes et réfléchies dans une pièce.	11
Figure 2.3: Système Boomerang I installé sur un véhicule militaire [9].....	12
Figure 2.4: Architecture des systèmes de monitoring audio.....	13
Figure 2.5: Localisation des sources sonores avec deux microphones	15
Figure 3.1: Détection et localisation des sources sonores avec deux microphones...	19
Figure 3.2: Étapes de l'algorithme de détection	20
Figure 3.3: Bruit blanc ajouté a un signal audio	21
Figure 3.4: Densité de probabilité du bruit ajouté.	21
Figure 3.5: Énergie d'un signal audio impulsif.....	22
Figure 3.6: Création de la séquence d'énergie dans le régime transitoire.	23
Figure 3.7: Création de la séquence d'énergie dans l'état permanent.	23
Figure 3.8: Variance d'un signal audio lors d'un son impulsif.	24
Figure 3.9: Seuil de détection.	24
Figure 3.10: Propagation des ondes dans l'environnement.....	26
Figure 3.11: Calcul de TDOA.	27
Figure 3.12: Localisation d'une source sonore en utilisant deux microphones [34]....	28
Figure 4.1: Interface du logiciel Jupyter Notebook	31
Figure 4.2: Interface du service Google Colab.....	32
Figure 4.3: Interface du logiciel PyCharm.....	32
Figure 4.4: Interface du logiciel Android Studio	33
Figure 4.5: Microsoft Xbox Kinect v1.....	34
Figure 4.6: Taille d'un microphone MEMS en millimètre.....	34
Figure 4.7: Emplacement des microphones MEMS dans un smartphone	35
Figure 4.8: Sonomètre	35
Figure 4.9: Fonctionnement de l'application développée sur Android studio	36
Figure 4.10 : Interface de l'application mobile.....	37
Figure 4.11 : Fenêtre principale de l'interface graphique.....	38
Figure 4.12 : Fenêtre de visualisation de l'angle	38
Figure 4.13 : Fenêtre pour l'ajustement des paramètres	39
Figure 4.14 : Distance entre les microphones d'une Kinect.....	39
Figure 4.15 : Environnement de test	42
Figure 4.16 : Sonomètre utilisé.....	42
Figure 4.17 : Haut-parleur utilisé pour l'émission de sons impulsifs	42
Figure 4.18 : F1-score pour une taille de bloc de 20 ms	43
Figure 4.19 : F1-score pour une taille de bloc de 30 ms	44
Figure 4.20 : F1-score pour une taille de bloc de 60 ms	44
Figure 4.21 : F1-score pour une taille de bloc de 120 ms	44
Figure 4.22 : F1-score pour une taille de bloc de 240 ms	45
Figure 4.23 : Erreur absolue de l'écart entre l'angle mesuré et sa valeur exacte de tous les scénarios de tests	45
Figure 4.24 : Erreur absolue de l'écart entre l'angle mesuré et sa valeur exacte. L'acquisition des signaux a été faite par les microphones 1 & 4 de la Kinect	46

LISTE DES TABLEAUX

Tableau 3.1 : Variation des indices i et j	22
Tableau 3.2 : Matrice de confusion	29

Chapitre 1 : Introduction

1.1 Motivations

Le besoin croissant de la protection des citoyens et des biens a entraîné une forte demande à la surveillance des lieux. Les aéroports, les stations de trains, les parkings, les autoroutes, et les centres commerciaux ont besoins d'être équipés par des systèmes de surveillance efficaces.

L'incorporation de l'intelligence artificielle et la vision par ordinateur a permis aux systèmes de surveillance conventionnels de reconnaître les événements dangereux en analysant les scènes vidéo. Cependant, ces systèmes sont incapables de signaler les événements dangereux qui se produisent en dehors du champ de vision des caméras. L'incorporation de l'information audio peut offrir une capacité d'écoute aux systèmes conventionnels. Plus particulièrement, les méthodes avancées de traitement audio peuvent être utilisées pour pivoter les caméras vers l'évènement dangereux. La seule information requise est l'acquisition permanente des sons de l'environnement par plusieurs capteurs. La majorité des événements acoustiques dangereux, tel que les coups de feu et les bris de glace, ont la particularité d'être des ondes acoustiques à caractère impulsif, d'intensité forte et de durée courte (quelques millisecondes seulement). La détection et la localisation de ce type d'évènement sont les tâches cruciales pour la réalisation de tels systèmes de surveillance.

Certains travaux de recherche sur la surveillance audio ont été proposées dans la littérature. Cependant, le problème de la localisation en temps réel des sons impulsifs n'a pas été encore étudié.

1.2 Contributions

Nous nous intéressons dans le cadre de ce projet à la détection et la localisation des sons impulsifs pour la surveillance audio. Notre contribution consiste à concevoir un logiciel de détection et de localisation des sons impulsifs en temps réel nécessitant peu de ressources matérielles.

La méthode de détection que nous avons implémentée est basée sur la détection par seuillage. Elle repose sur le calcul de la variance de l'énergie du signal. Nous avons proposé une approche d'optimisation des paramètres de la méthode de détection pour réduire la complexité algorithmique. L'acquisition du signal audio est accomplie en utilisant le réseau de microphones incorporé dans une Microsoft Xbox Kinect v1.

D'autre part, nous avons utilisé le generalized cross-correlation phase transform (GCC-PHAT) pour estimer l'angle d'incidence de la source.

1.3 Impact du projet

Le travail mené dans le cadre de ce projet de fin d'études à un impact socio-économique important pour notre pays. La majorité des systèmes de surveillance disponible dans le marché Algérien sont basés sur les systèmes conventionnels de surveillance.

A notre connaissance, les systèmes de surveillance audio ne sont pas assez répandus dans notre pays. De ce fait, une bonne maîtrise de ces systèmes pourra faire l'objet d'une idée de startup. Cette idée sera sans doute intéressante pour plusieurs institutions telles que les banques, les sociétés, et les ministères.

1.4 Organisation du mémoire

Le mémoire est organisé comme suit :

Le deuxième chapitre introduit des généralités sur les technologies de surveillance et se focalise sur la modalité audio. Un état de l'art sur les méthodes de détection et de localisation des événements acoustiques est aussi fourni dans ce chapitre.

Le troisième chapitre présente l'architecture générale du système proposé ainsi que les différentes méthodes utilisées pour sa conception et son évaluation.

Le quatrième chapitre s'articule autour de la présentation des résultats expérimentaux.

Le cinquième chapitre comprend une conclusion générale et des perspectives (travaux futurs).

Chapitre 2 : Systèmes de surveillance audio

2.1 Introduction

Nous introduisons dans ce chapitre les systèmes de surveillance audio. Nous présentons des généralités sur les technologies de surveillance et nous exposons l'avantage de la modalité audio par rapport à la modalité vidéo. Par la suite, nous citons quelques applications de la surveillance audio et nous décrivons l'architecture globale d'un système de monitoring audio. Nous présentons un état de l'art succinct sur les méthodes de surveillance à base de la modalité audio. Nous clôturerons ce chapitre par une conclusion.

2.2 Généralités sur les technologies de surveillance

La surveillance, secrète ou évidente, a toujours été présente dans l'histoire. De nos jours, elle est devenue un moyen primordial pour assurer la sécurité des citoyens et des biens. Afin de répondre à ces besoins de sécurité, les systèmes de surveillance ont connu plusieurs évolutions. La première génération de ces systèmes, parue dans les années 1960 [1] était basée sur des enregistrements vidéo et impliquait la surveillance de l'activité par un opérateur humain afin de détecter des situations ou des événements anormaux.

Plus récemment, des systèmes automatisés sont développés pour tenter d'effectuer cette tâche en utilisant des techniques de vision par ordinateur et de reconnaissance de formes. L'avantage de cette approche réside essentiellement dans la réduction des coûts [2]. Cependant les images obtenues par les systèmes de surveillance vidéo ont un champ de vision limité et sont généralement de faible résolution en raison des contraintes de stockage.

Pour surmonter ces contraintes, une troisième génération de systèmes de surveillance a vu le jour. Ces systèmes incorporant des capteurs pour compléter ou même remplacer la surveillance vidéo [3]. Plusieurs capteurs peuvent être utilisés. Nous citons : les capteurs de mouvements, de température et même les microphones. Les données audio représentent une source très importante à partir de laquelle nous pouvons extraire de nombreuses informations précieuses pour la surveillance et la détection d'événements potentiellement dangereux. Dans certaines situations, il est impossible d'extraire de l'information utile à partir d'une image ou d'une vidéo. Nous citons l'exemple des cris ou des coups de feu.

La quatrième génération de monitoring englobe toute la chaîne de surveillance (vidéo + capteurs + traitement) dans un système embarqué [4].

2.3 Avantages de l'audio pour le monitoring

Par rapport à l'utilisation de la vidéo pour la surveillance, l'audio présente plusieurs avantages intéressants, on peut citer les points suivants :

- Un flux audio est généralement beaucoup moins onéreux qu'un flux vidéo en termes de bande passante, stockage de mémoire et d'exigences de calcul. La nature unidimensionnelle du flux audio par opposition à la nature tridimensionnelle du flux vidéo (largeur × hauteur × temps) est la raison principale de cette réduction [5].
- La luminosité et les conditions météorologiques n'affectent pas l'acquisition des événements audio [5], [6].
- Les microphones peuvent être omnidirectionnels, ce qui signifie qu'ils sont sensibles aux sons provenant de toutes les directions de manière égale quelle que soit l'orientation du microphone. **(voir la figure 2.1).**
- Plusieurs événements audio importants pour la surveillance, tels que des cris ou des coups de feu, ne peuvent pas être capturés avec une surveillance vidéo.
- La réflexion des ondes acoustiques sur les surfaces nous permet de capturer des événements audio même lorsque des obstacles sont présents le long du chemin direct [5]. **(voir la figure 2.2).**

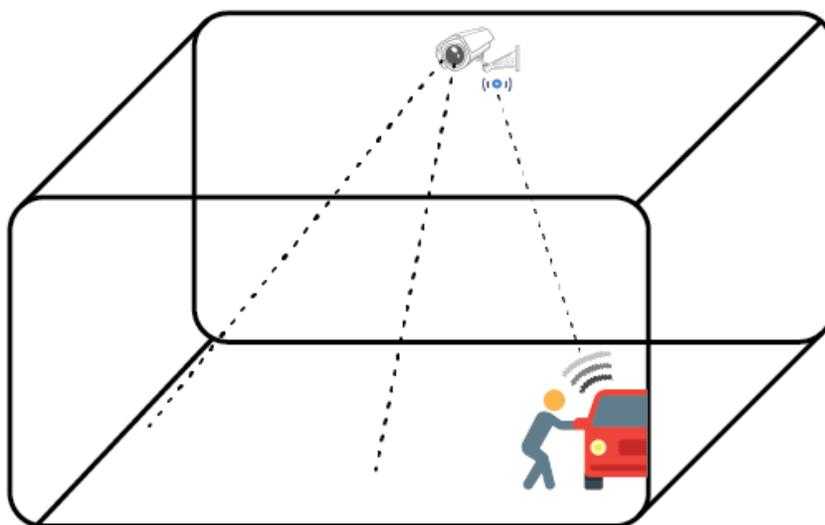


Figure 2.1: Champ de vision limité d'une caméra de surveillance.

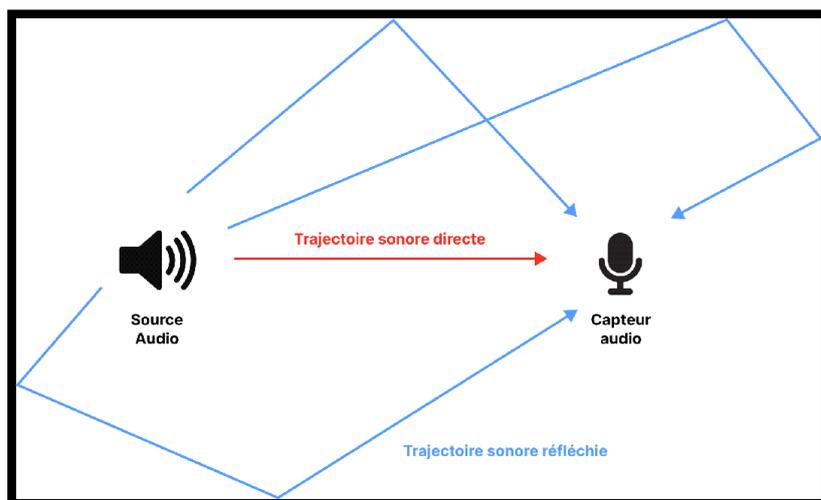


Figure 2.2: Trajectoires sonores directes et réfléchies dans une pièce.

2.4 Applications de la surveillance audio

Aujourd'hui, la sécurité des lieux publics et privés est une tâche très importante. Le développement de nouvelles techniques de détection et de reconnaissance d'événements suspects est primordial. L'introduction de la surveillance audio dans certains domaines, comme la sécurité des lieux publics, est devenue indispensable. En outre, dans de nombreuses situations dangereuses, tels que les incendies, les tremblements de terre, les inondations et les menaces d'agressions physiques, l'audio transmet une information plus importante que la vidéo [7].

La surveillance audio est également incorporée dans de nombreuses applications telles que les armes militaires, les parcs de stationnement, les baby phones et les ascenseurs [8].

Le Boomerang est une technologie militaire qui incorpore des techniques avancées de traitement audio pour la détection et la localisation des sources sonores [9]. Il s'agit d'un système acoustique installé sur des véhicules qui visent à déterminer la position exacte des tireurs (**voir la figure 2.3**).

La modalité audio est aussi importante pour la surveillance des parcs de stationnement. Selon les analystes de l'industrie automobile d'Hedges & Company, il y avait plus de 275 millions de véhicules immatriculés aux États-Unis en 2018. Accueillir ce nombre de véhicules nécessite un énorme réseau de parkings équipés par des technologies de pointe pour la sécurité des personnes et des biens [10].

La surveillance de tels espaces est une tâche complexe qui demande beaucoup de ressources humaines et matérielles. La modalité audio, dans ce cas, complète la surveillance

vidéo et assure une détection efficace d'événements acoustiques tels que les alarmes de voitures, les cris de personnes et les bris de glace [11].

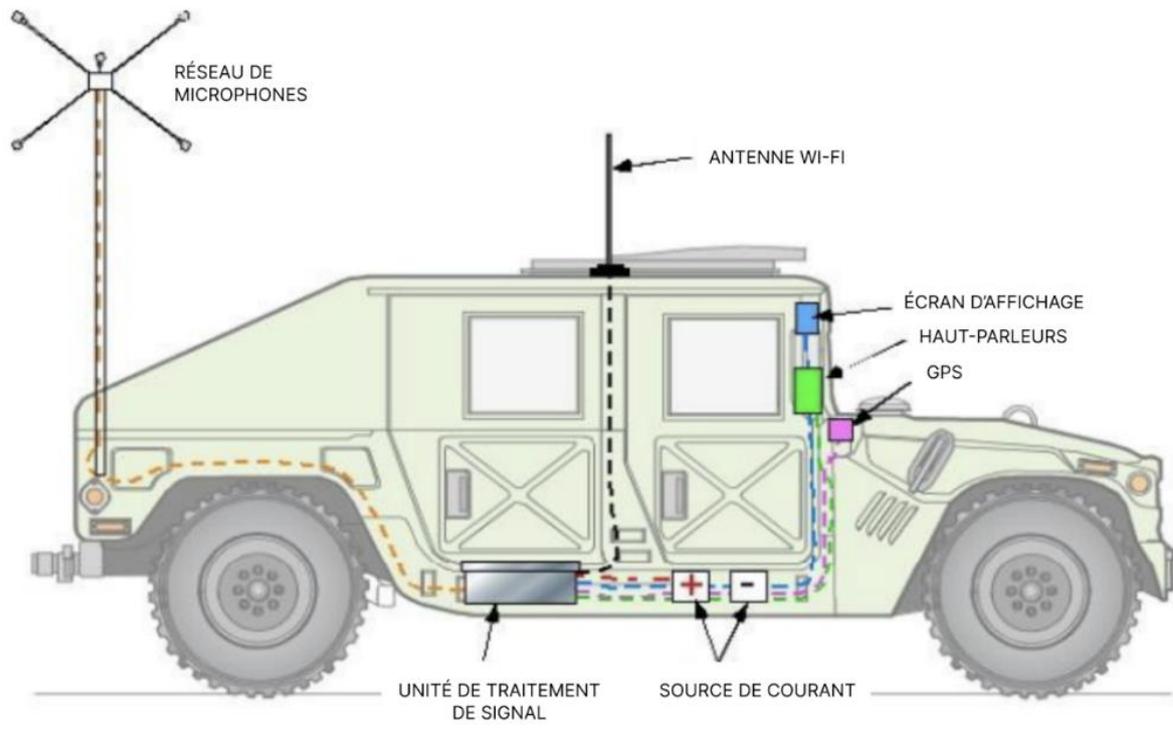


Figure 2.3: Système Boomerang I installé sur un véhicule militaire [9]

2.5 Architecture d'un système de monitoring audio

Il existe dans la littérature de nombreuses propositions d'architectures de systèmes de monitoring audio. Chacune d'elles essaie de répondre à un cas d'utilisation particulier, mais en général, il y a beaucoup de ressemblances entre elles. Ce qui fait réellement la différence c'est les méthodes utilisées pour l'implémentation de cette architecture (Méthode de détection, Méthode de localisation etc.). Ci-dessous une représentation d'une architecture qui résume les tâches les plus fréquentes dans un système de monitoring audio.

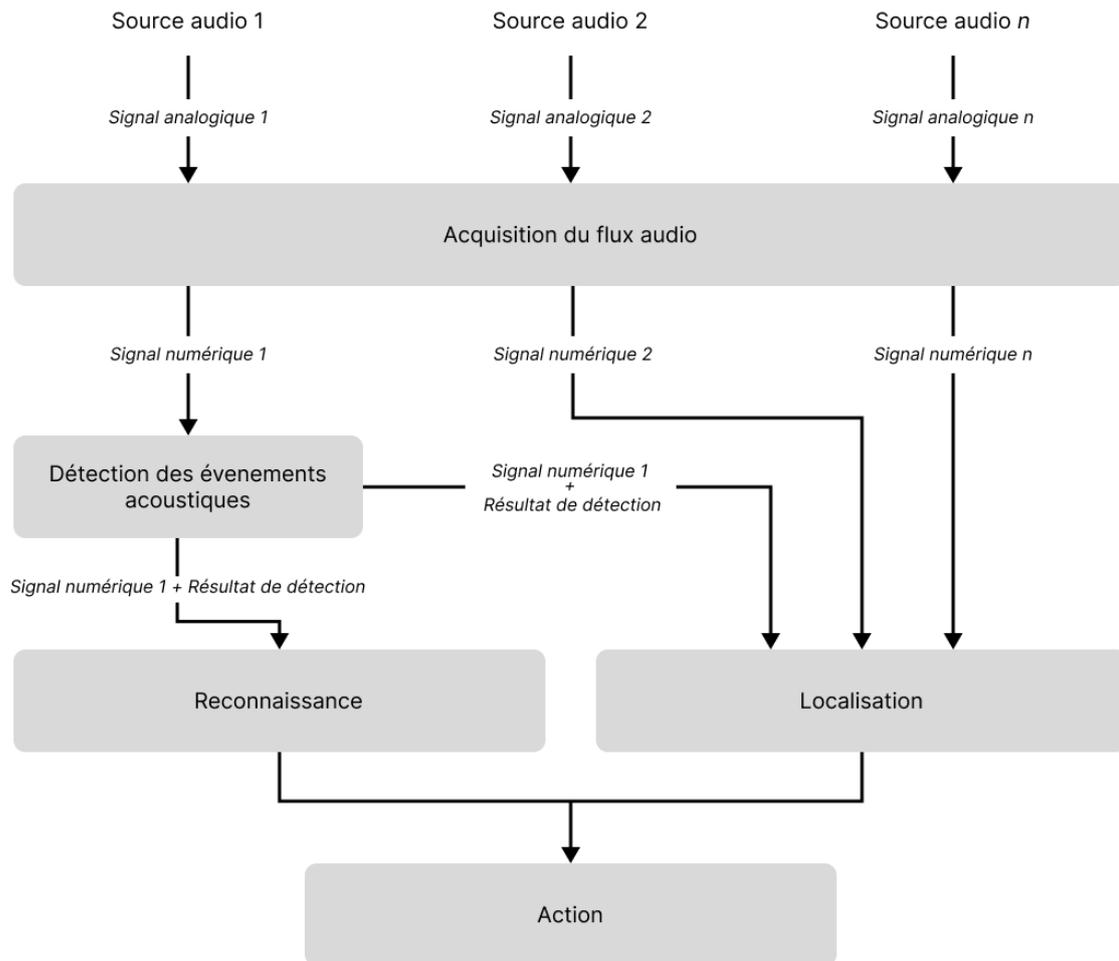


Figure 2.4: Architecture des systèmes de monitoring audio

2.5.1 Acquisition du flux audio

Le son est produit par la vibration d'un corps solide, liquide ou gazeux qui constitue la source sonore. Cette vibration produit des ondes de pression qui peuvent être captées par des microphones. Lors d'un enregistrement, le système d'acquisition, muni d'un ou plusieurs microphones (capteurs audio), convertit ces ondes en oscillations électriques. Le signal résultant est transmis à un convertisseur analogique-numérique qui échantillonne le signal analogique avec un taux d'échantillonnage spécifique [12]. Ce système d'acquisition peut être composé d'un seul capteur audio ou un réseau de capteurs allant de 2 jusqu'à plusieurs centaines, ce qui est le cas dans le projet LOUD (Large acOUSTic Data array project), un projet du Laboratoire d'Informatique et d'Intelligence Artificielle du MIT (Massachusetts Institute of Technology). Ce dernier contient 1020 capteurs et il génère près de 50 Mo de données chaque seconde [13].

2.5.2 Détection d'un événement acoustique

La détection d'événements acoustiques vise à identifier les types d'événements dans un flux audio. La détection de tels événements est une solution efficace pour de nombreux problèmes [14],[15] Cette action est généralement accomplie en utilisant des techniques à base de machine learning ou des méthodes de détection par seuillage. Le choix entre ces deux approches dépend du domaine, du cas d'utilisation ainsi que des limitations matérielles à disposition (performances du processeur, espace RAM disponible etc.). Généralement les scènes auditives sont assez complexes en matière d'événements sonores, ce qui rend la détection plus difficile [16], et impacte la performance des algorithmes.

2.5.3 Reconnaissance des sons

La reconnaissance des sons est l'un des domaines les plus marquants au cours de la dernière décennie. Le nombre de recherches effectuées a augmenté et les chercheurs se concentrent de plus en plus sur des tâches bien particulières telles que les problèmes de la classification et de l'identification des sons environnementaux [17].

La reconnaissance de l'audio est exploitée dans plusieurs domaines tels que la recherche audio [18], le balisage automatique des fichiers audio [19], l'amélioration de la navigation d'un robot dans un environnement [20], et l'analyse de vidéos pour les applications de surveillance [21]. La majorité des méthodes proposées dans la littérature pour reconnaître ces événements sont basées sur l'intelligence artificielle. On peut citer les modèles de mélange Gaussiens (GMMs), les machines à vecteurs de support (SVMs) et les réseaux de neurones convolutifs (CNNs) [17].

2.5.4 Localisation de la source sonore

La localisation des sources sonores est la capacité à déterminer la direction par laquelle provient un son dans un espace [16]. L'opération de positionnement peut être accomplie en 2 ou 3 dimensions. La localisation en 2 dimensions nécessite au minimum deux capteurs audio. Cependant, la localisation en 3 dimensions nécessitera au minimum trois capteurs [22].

La localisation des sources sonores est une tâche complexe, soumise à plusieurs facteurs qui peuvent la rendre encore plus difficile comme le bruit, la réflexion des ondes acoustiques sur les surfaces, l'amplitude du son, et le chevauchement des fréquences [23].

Il existe principalement trois catégories d'algorithmes pour la localisation des sons :

- **Beamforming** : Cette technique utilise un réseau de microphones répartis dans un espace pour localiser la source d'un son. La précision de cette technique est très élevée. Cependant, elle est lourde en termes de calculs. Par conséquent, le Beamforming est un mauvais choix pour la conception d'applications en temps réel. En outre, elle nécessite plus de deux microphones pour l'acquisition de sons [24].
- **Estimation Spectrale à Haute Résolution** : Cette deuxième catégorie de techniques comprend les méthodes modernes de Beamforming adaptées du domaine de l'analyse spectrale à haute résolution. En termes de complexité, elles sont moins exigeantes que les méthodes de beamforming classique ; mais elles ont aussi tendance à être nettement moins robustes [24], [25]
- **Différence entre les Temps d'Arrivée (TDOA)** : Cette méthode consiste à mesurer la différence du délai de propagation d'un signal sur deux capteurs audio de réception ou plus à des positions différentes. Cette méthode est efficace en termes de complexité, ce qui en fait un choix très intéressant pour les applications qui requièrent un traitement en temps réel [24], [26], [27].

Les algorithmes qui se basent sur cette approche comportent deux étapes principales :

1. Estimation de la différence de temps d'arrivée entre une paire de signaux audio.
2. Calcul de l'angle duquel la source sonore vient (**voir la figure 2.5**).

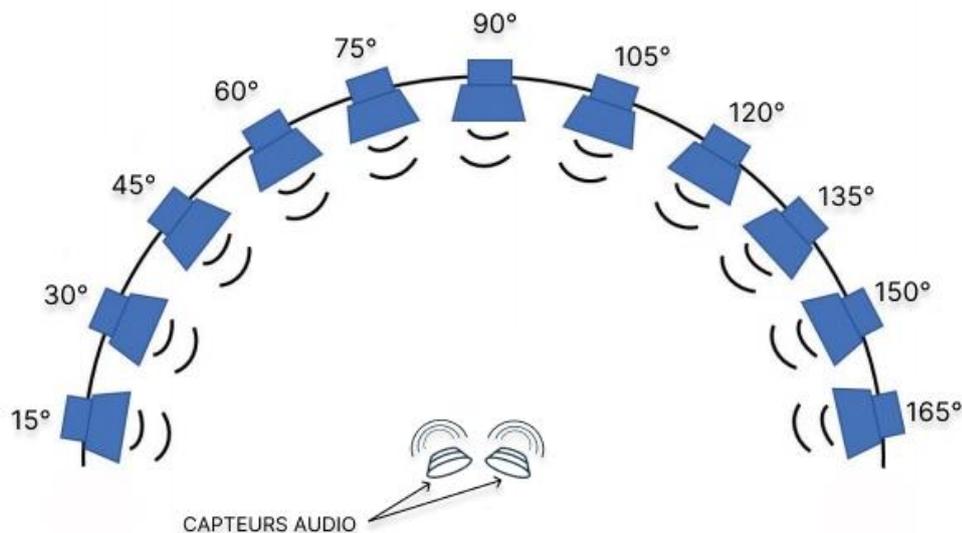


Figure 2.5: Localisation des sources sonores avec deux microphones

2.5.5 Action

Dépendant des cas d'utilisation du système de surveillance, le système devra faire une action lors de la détection, la reconnaissance ou la localisation d'un évènement dangereux. Cette action peut être un déclenchement d'alarme, un appel aux agents de sécurités ou bien aux autorités, un enregistrement vidéo ou audio de la scène, et la rotation du champ de vision de la caméra vers la scène concernée.

2.6 Méthodes proposées dans la littérature pour la détection et la localisation des sons impulsifs

Dans cette partie, nous présentons quelques travaux sur la surveillance audio. Nous examinerons les principales solutions proposées dans la littérature pour résoudre ce problème.

G. Valenzise et al. [28] ont proposé un système de surveillance permettant la détection et la localisation des coups de feu et les cris. Pour la détection, ces auteurs implémentent deux classifieurs à base de GMMs, fonctionnant en parallèle, pour distinguer, respectivement, les cris vs le bruit, et les coups de feu vs le bruit. Chacun des classifieurs est entraîné en utilisant différentes caractéristiques. La localisation de ces événements sonores est ensuite effectuée en calculant la différence entre le temps d'arrivée du son aux 4 microphones du système. De très bons résultats de détection et de localisation ont été reportés.

C. Zieger et al. [29] ont proposé un système de surveillance pour la détection d'intrusion. Ce dernier se base sur un réseau de microphones réparti sur plusieurs pièces d'une maison. Le signal retourné par ces microphones passe par 3 modules principaux. (i) Le premier module vise à identifier les événements acoustiques dont l'énergie excède un certain seuil. (ii) Le deuxième module identifie les sons dont le contenu spectral change de manière significative. (iii) Le troisième et dernier module localise la direction de la source sonore en se basant sur la méthode GCC-PHAT pour l'estimation de la différence de temps d'arrivée. Le système a montré une grande efficacité pour la détection et la localisation d'intrusions. Cependant, le taux de fausses alarmes est assez important.

2.7 Conclusion

Dans ce chapitre, nous avons présenté les technologies de surveillance. Nous nous sommes focalisés sur la surveillance audio, ses avantages et ses usages. Nous avons aussi constaté un manque de recherches menées dans ce domaine, surtout lorsqu'il s'agit de surveillance en temps réel, car la majorité des systèmes proposés utilisent des algorithmes très complexes qui ne le permettent pas, du moins en utilisant des machines plein public. Notre objectif est de proposer un système de surveillance permettant la détection et la localisation des sons impulsifs en temps réel. Les méthodes utilisées pour l'implémentation de ce système doivent être efficaces afin de permettre un traitement en temps réel avec des ressources minimales. Dans le chapitre suivant, nous verrons plus en détail l'architecture du système proposé.

Chapitre 3 : Méthodes proposées pour la détection et la localisation des événements sonores

3.1 Introduction

Au cours du chapitre précédent, nous avons présenté les travaux de recherche menés dans le domaine de la surveillance audio. Nous avons constaté l'importance et l'apport de l'audio pour la surveillance. Dans ce chapitre nous présentons l'architecture globale du système proposé. Nous décrivons les différentes méthodes utilisées pour la réalisation et l'évaluation de ce système. Enfin, nous terminerons ce chapitre par une conclusion.

3.2 Architecture globale du système proposé

Le choix d'une bonne architecture pour la localisation et la détection des sons impulsifs est crucial. La complexité d'une architecture peut affecter le fonctionnement et les performances globales du système de localisation. Nous avons proposé une architecture qui répond à nos besoins et qui assure de bonnes performances avec peu de ressources matérielles (**voir la figure 3.1**). En plus, le système que nous proposons peut-être installer dans des machines à capacités limitées.

Notre architecture est basée sur trois phases :

- L'acquisition : Elle consiste à utiliser deux capteurs audio pour capter les sons environnementaux.
- La détection : Après l'acquisition, nous allons analyser le signal audio pour décider si le son est impulsif ou non en le faisant passer par un algorithme de détection par seuillage. Cette étape est très importante vu qu'elle doit fonctionner en temps réel.
- La localisation : Dès qu'un événement acoustique impulsif est détecté, la phase de localisation est déclenchée. La méthode utilisée exploite l'onde acoustique venant des deux capteurs audio. En se basant sur l'écart entre les temps d'arrivés du signal aux deux capteurs audio, l'algorithme de localisation estime l'angle d'arrivé de la source sonore.

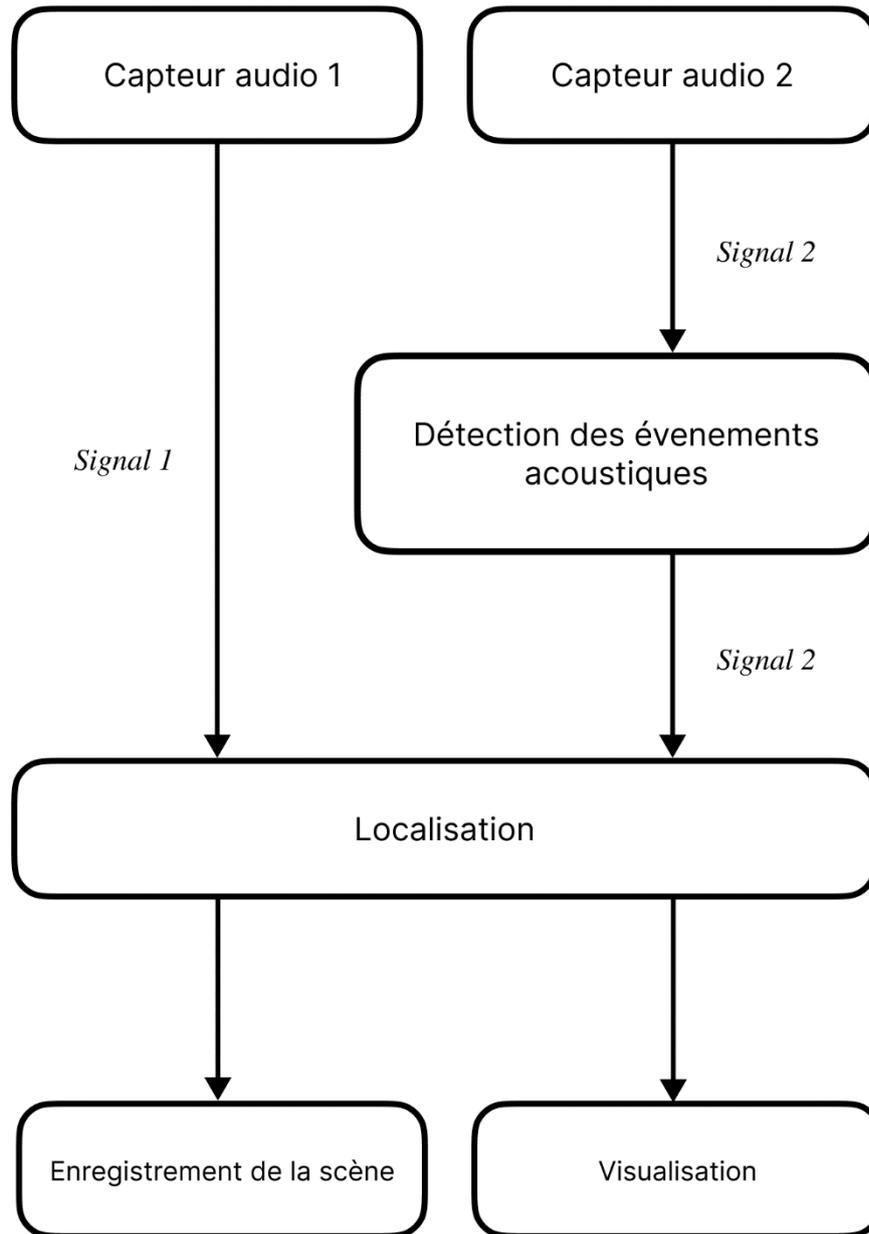


Figure 3.1: Détection et localisation des sources sonores avec deux microphones

Chacune de ces phases doit fonctionner en temps réel et donner de bonnes performances. Dans les sections suivantes, nous détaillerons les phases principales de notre système. Nous introduisons aussi les métriques d'évaluation que nous avons utilisées pour mesurer les performances globales de la localisation et la détection des sons dangereux.

3.3 Acquisition

L'acquisition du son se fait avec un ou plusieurs capteurs audio qui sont connectés à une carte son. Les cartes intégrées dans les PCs ne répondent pas à nos besoins. Le plus souvent, ces cartes ne sont pas capables de reproduire et d'enregistrer des signaux audio multicanaux. La stabilité du signal issu de ces cartes est également discutable car la fonctionnalité audio n'est pas la priorité principale du système [30]. Pour ces raisons, en pratique, une carte son externe possède de meilleures caractéristiques par rapport à une carte intégrée.

La carte son requise pour la localisation des sources sonores doit avoir au moins deux entrées audio (microphones). Certains appareils électroniques incorporent un système d'acquisition de son en mode stéréo. Nous citons le cas des smartphones. La majorité des modèles récents sont dotés de deux microphones MEMS ou plus. Le son généré par ces microphones est de bonne qualité. Ils sont principalement utilisés pour isoler la voix et réduire le bruit lors des appels téléphoniques. D'autres appareils, comme la Microsoft Xbox Kinect, comportent 4 microphones et plusieurs caméras. Ces appareils sont destinés pour le développement des jeux vidéo de la Xbox [31]. Les quatre microphones dans la Kinect peuvent être exploités pour localiser les sources sonores.

3.4 Détection

L'un des principaux objectifs de notre système est de pouvoir détecter les événements audio impulsifs en temps réel. Cette étape impose des contraintes de complexité algorithmique. Pour cette raison, nous avons choisi d'utiliser un algorithme de détection par seuillage au lieu d'un algorithme de détection par classification. L'algorithme de détection par seuillage est composé de six étapes principales [32] (voir la figure 3.2).

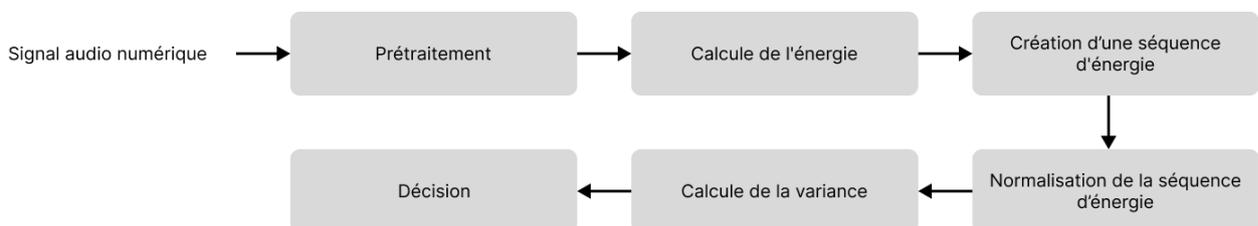


Figure 3.2: Étapes de l'algorithme de détection

Étape 1 : Prétraitement : Cette étape consiste à ajouter un bruit blanc $y(n)$ à chaque nouveau bloc du signal audio afin de masquer les sons du background qu'on ne veut pas détecter (**voir la figure 3.3**). Ce bruit est défini selon une distribution normale de moyenne μ et de variance σ^2 (**voir la figure 3.4**). La distribution de probabilité de y est donnée comme suit :

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (3.1)$$

Le taux de bruit ajouté est défini par l'utilisateur selon l'environnement de test [32].

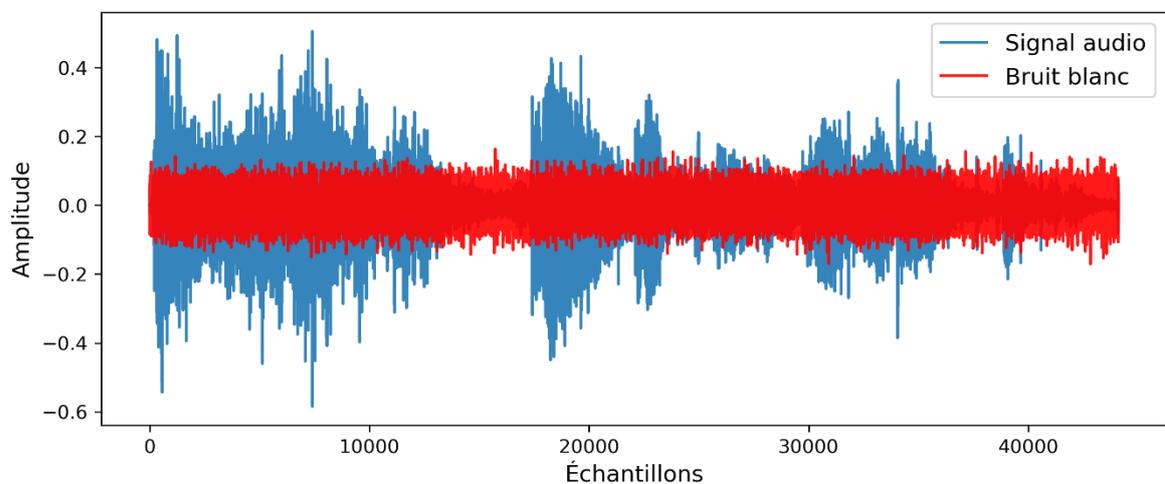


Figure 3.3: Bruit blanc ajouté a un signal audio

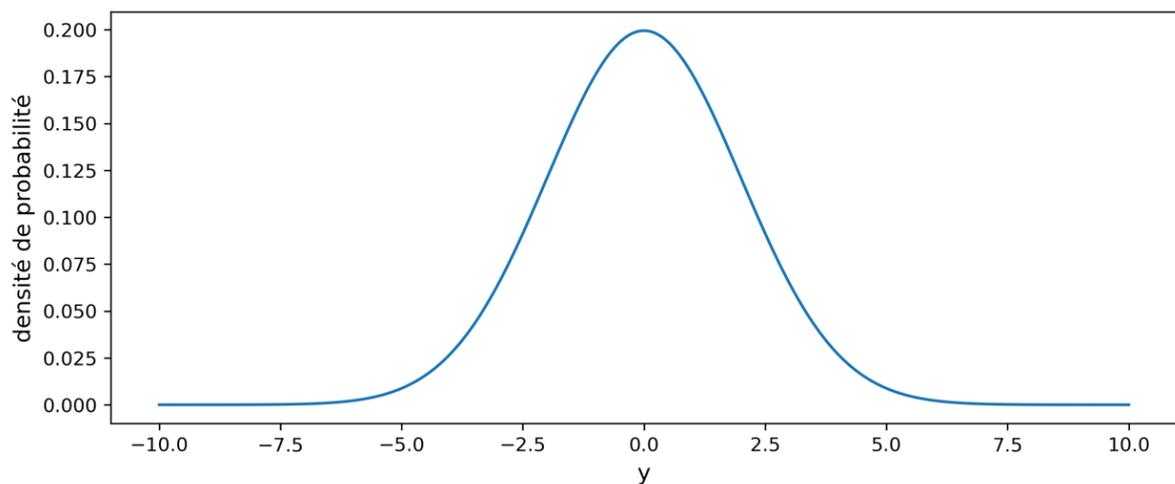


Figure 3.4: Densité de probabilité du bruit ajouté.

Étape 2 : Calcul de l'énergie du signal. Théoriquement, tous les sons impulsifs provoquent une augmentation soudaine de l'énergie (**voir la figure 3.5**).

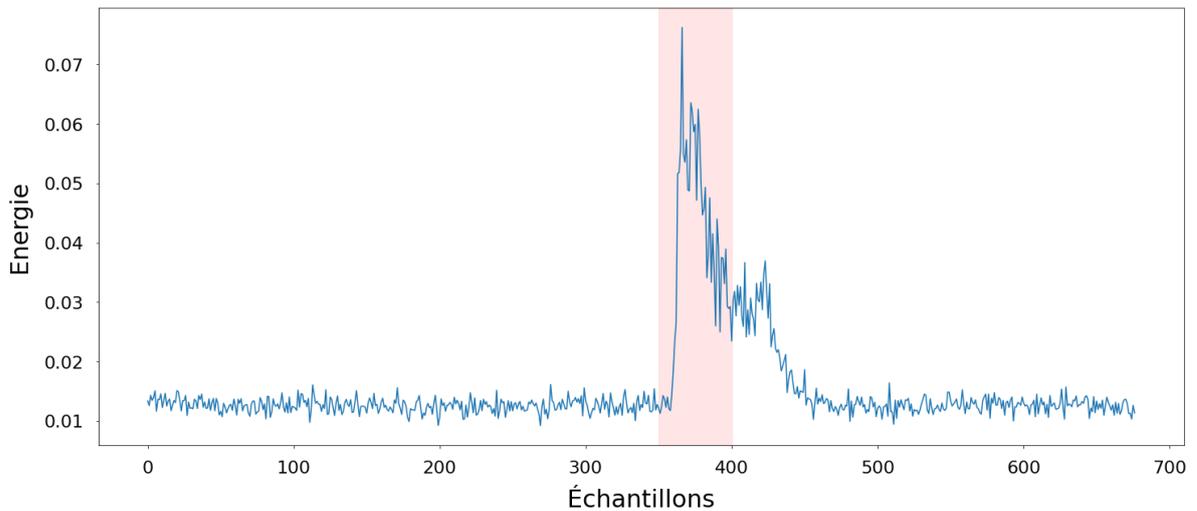


Figure 3.5: Énergie d'un signal audio impulsif.

Le signal audio est échantillonné à une fréquence d'échantillonnage F_e . Cette fréquence est le nombre d'échantillons audio enregistrés chaque seconde. Afin de trouver l'énergie du signal audio, ces échantillons seront divisés en plusieurs blocs, chaque bloc ayant N échantillons. L'énergie du k -ième bloc d'énergie est exprimée comme suit :

$$e(k) = \frac{1}{N} \sum_0^{N-1} x^2(n + kN) \quad (3.2)$$

$x(n)$: n -ième bloc du signal audio échantillonné à F_e

k : L'indice du bloc d'énergie. $k \in [0, +\infty]$

N : La taille des blocs d'énergie

Étape 3 : Création d'une séquence d'énergie $e_{win}(j/k)$ de taille L tel que :

$$e_{win}(j) = e(i)$$

La variation des indices i et j est liée à k . On peut distinguer deux régimes :

Tableau 3.1 : Variation des indices i et j

	Variation de i	Variation de j
Régime transitoire ($k < L$) :	$i \in [0, k - 1]$	$j \in [0, k - 1]$
Régime permanent ($k \geq L$) :	$i \in [k - L + 1, k]$	$j \in [0, k]$

Les figures 3.6 et 3.7 montrent respectivement les deux régimes de traitement.

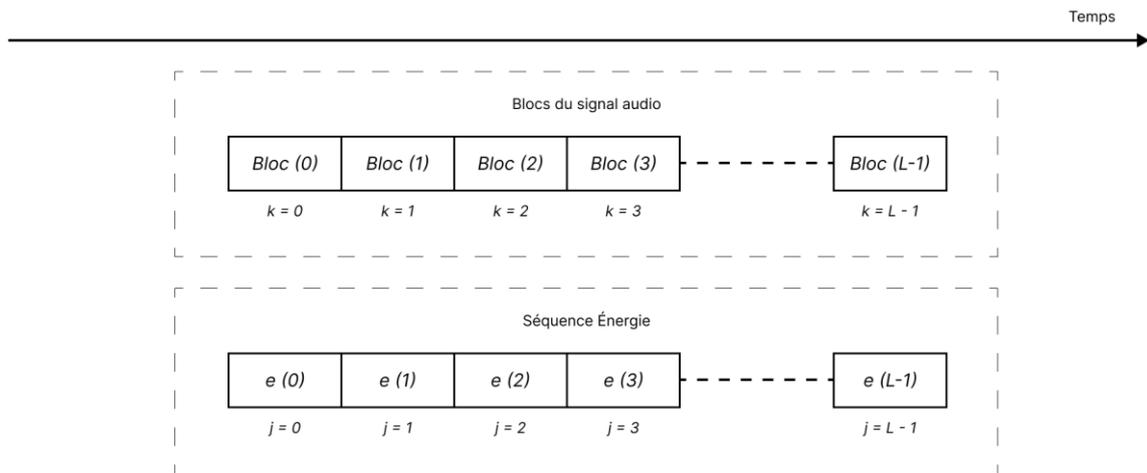


Figure 3.6: Création de la séquence d'énergie dans le régime transitoire.

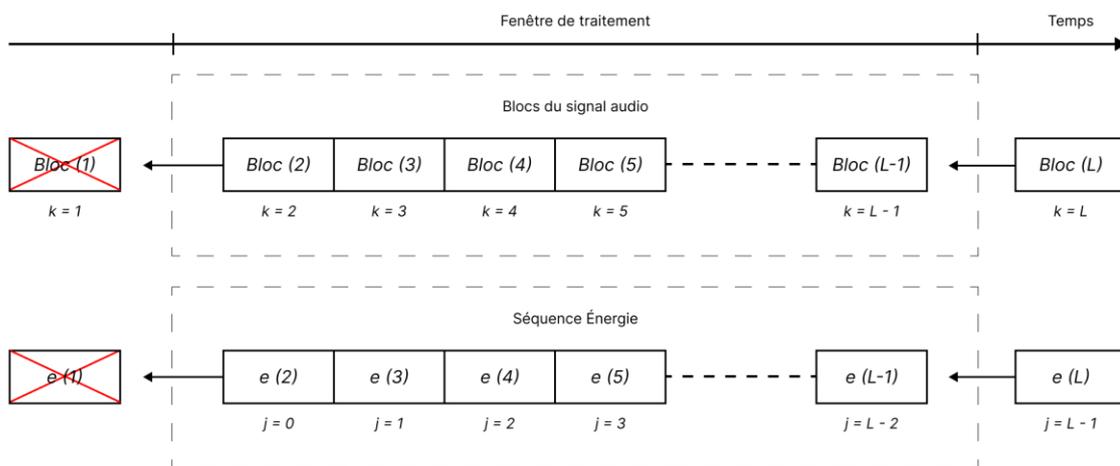


Figure 3.7: Création de la séquence d'énergie dans l'état permanent.

Étape 4 : Normalisation de la séquence d'énergie :

$$e_{norm}(j) = \frac{e_{win}(j) - \min(e_{win}(j=0:L-1))}{\max(e_{win}(j) - \min(e_{win}(j=0:L-1)))} \quad (3.3)$$

Tel que $j \in [0, L - 1]$ et $e_{norm}(j) \in [0, 1]$

Étape 5 : Calcul de la variance de la séquence d'énergie normalisée. La variance se définit comme suit :

$$var(k) = \frac{1}{L-1} \sum_{j=0}^{L-1} \{ e_{norm}(j) - \underline{e}_{norm}(k) \}^2 \quad (3.4)$$

$\underline{e}_{norm}(k)$ représente la moyenne des premières L-1 valeurs de $e_{norm}(j)$

La valeur de la variance baisse considérablement lors d'un son impulsif (**voir la figure 3.8**).

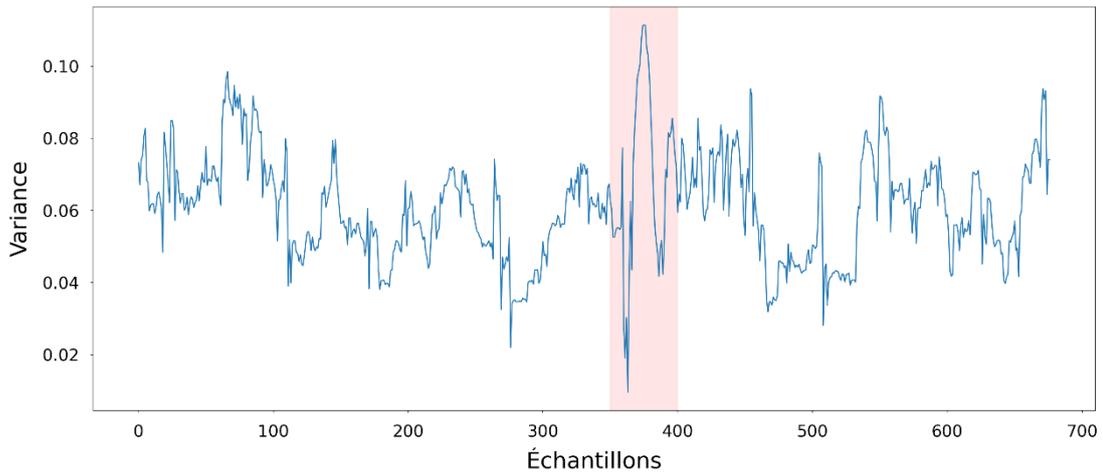


Figure 3.8: Variance d'un signal audio lors d'un son impulsif.

Étape 6 : La dernière étape est de comparer cette variance avec un seuil (Th) défini à l'avance. Si la variance est en dessous du seuil Th , l'événement audio est considéré comme étant un son impulsif (**voir la figure 3.9**).

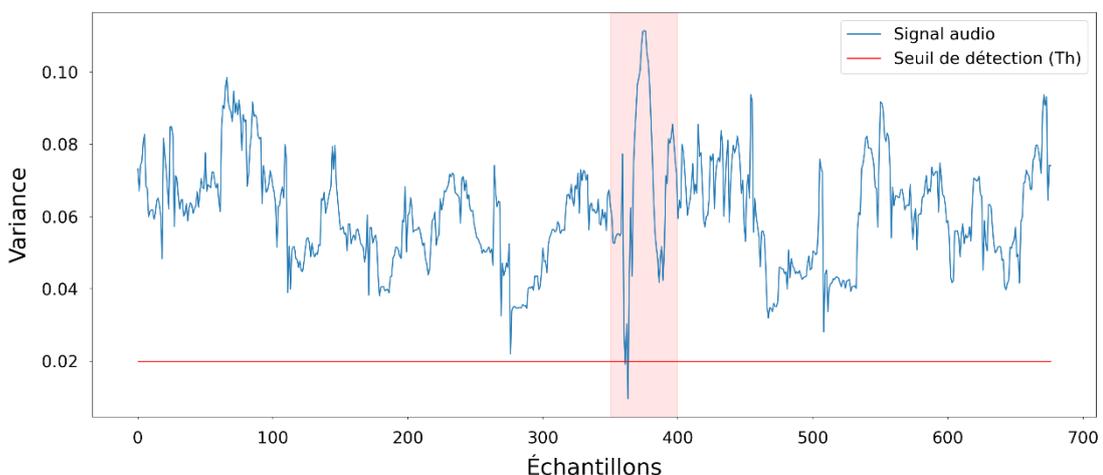


Figure 3.9: Seuil de détection.

3.5 Localisation

Dès qu'un son impulsif est détecté par la technique décrite précédemment, l'algorithme de localisation est déclenché. Notre solution repose sur le calcul de la différence entre les temps d'arrivée des deux signaux captés par les deux microphones. Pour cela, nous avons utilisé l'algorithme de la Corrélation Croisée Généralisée (en anglais Generalized Cross Correlation (GCC)). Cet algorithme a été proposé par Knapp et Carter [33]. La différence entre les temps d'arrivée (TDOA) des deux signaux est estimée en calculant le décalage temporel qui maximise la corrélation croisée entre ces signaux [33]. Cet algorithme nécessite l'utilisation d'une fonction de pondération pour améliorer la précision de localisation. Nous citons le Smoothed Coherence Transform (SCOT), le ROTH processor, et la Phase Transform (PHAT). Il a été reporté dans la littérature que la PHAT est plus adéquate pour la localisation dans les environnements bruités [34].

La GCC-PHAT est composée de deux phases de traitement : (i) l'estimation de TDOA entre les deux signaux, (ii) et la prédiction de l'angle d'incidence de la source sonore.

Le calcul du TDOA passe par 4 étapes :

Étape 1 : Supposons un signal $s(n)$ qui se propage dans un environnement bruité à 2 dimensions. $x_1(n)$ et $x_2(n)$ sont les deux signaux acquis par deux capteurs (**voir la figure 3.10**).

Les signaux sont représentés comme suit :

$$x_1(n) = s(n) + n_1(n) \quad (3.5)$$

$$x_2(n) = s(n - \tau) + n_2(n) \quad (3.6)$$

$n_1(n)$ et $n_2(n)$ sont des bruits gaussiens blancs additifs.

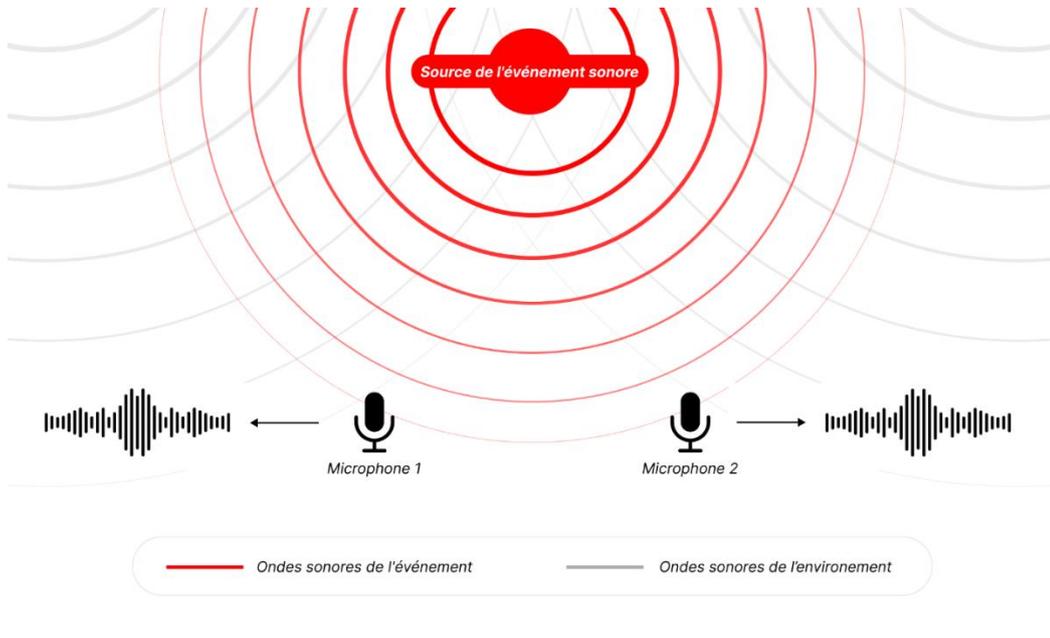


Figure 3.10: Propagation des ondes dans l'environnement.

Étape 2 : La fonction de pondération (PHAT) est définie comme suit :

$$\psi_{PHAT}(\omega) = |X_1(\omega)X_2^*(\omega)| \quad (3.7)$$

$X_1(\omega)$ et $X_2(\omega)$ représentent respectivement la transformation de Fourier de $x_1(n)$ et $x_2(n)$.

Étape 3 : La corrélation croisée généralisée $R_{x_1x_2}(\tau)$ peut s'exprimer sous la forme suivante :

$$R_{x_1x_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{X_1(\omega)X_2^*(\omega)}{\psi_{PHAT}(\omega)} e^{j\omega\tau} d\omega \quad (3.8)$$

[.]* représente la matrice transconjuguée.

ω est la vitesse angulaire.

Étape 4 : Le TDOA entre $x_1(n)$ et $x_2(n)$ qui maximise l'équation 3 est défini comme suit :

$$\tau_{x_1,x_2} = \text{ArgMax}(R_{x_1,x_2}(\tau)) \quad (3.9)$$

τ_{x_1,x_2} est l'instant de temps qui correspond au pic le plus élevé de $R_{x_1x_2}(\tau)$ (l'instant où les deux signaux corrèlent au maximum).

On peut résumer les étapes de calcul de TDOA dans la **figure 3.11**

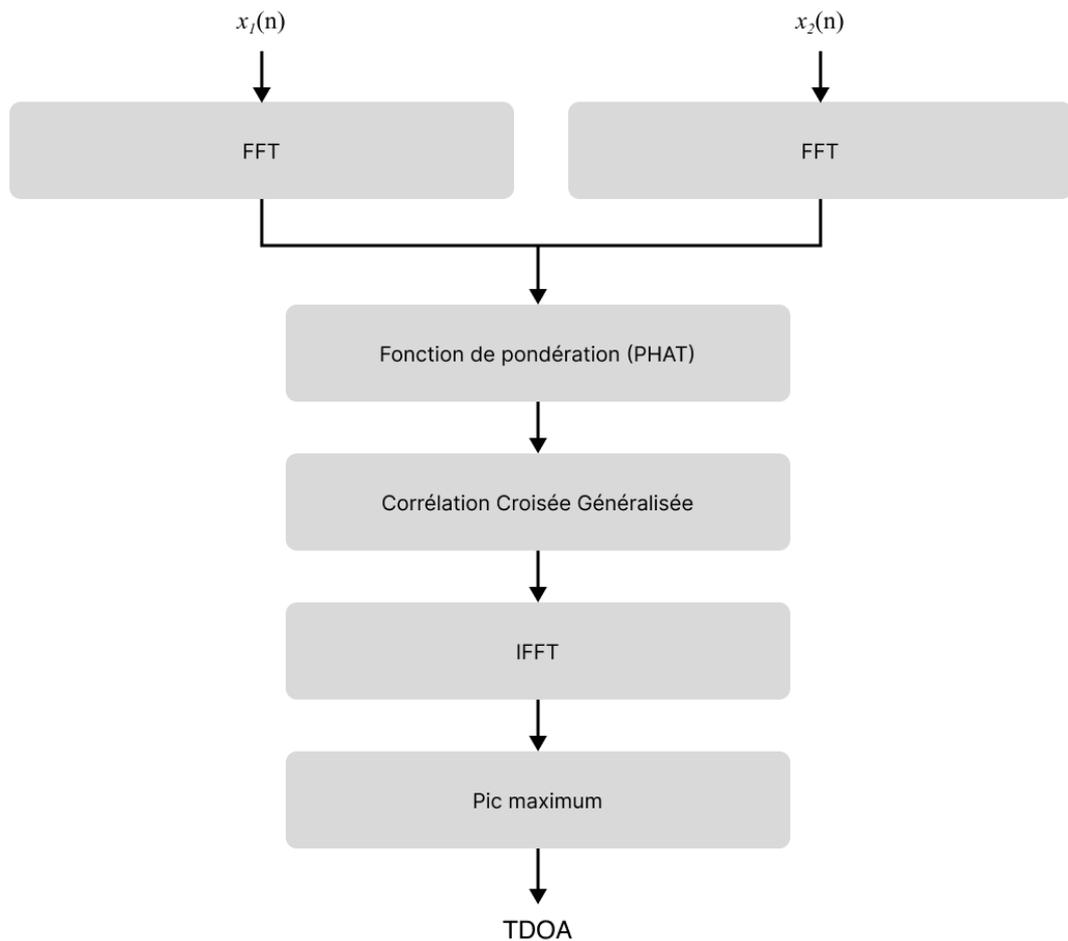


Figure 3.11: Calcul de TDOA.

Une fois qu'on obtient une valeur de τ_{x_1,x_2} on passe à la deuxième phase de l'algorithme qui est l'estimation de l'angle. **La figure 3.12** représente une paire de capteurs audio A et B avec un source sonore S.

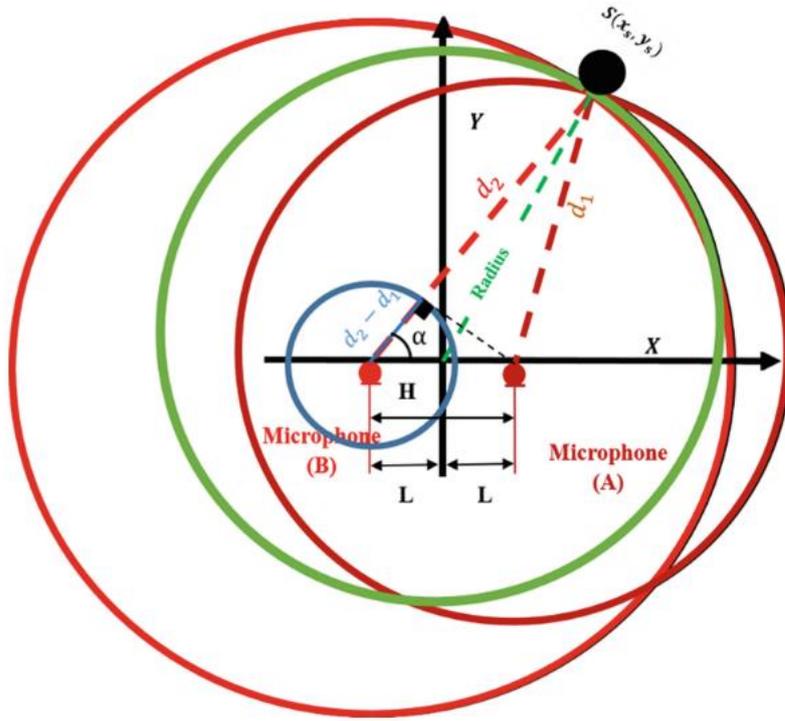


Figure 3.12: Localisation d'une source sonore en utilisant deux microphones [34].

Les deux capteurs peuvent être représentés en fonction de L comme suit :

$$L = \frac{H}{2} \quad (3.10)$$

Les coordonnées des capteurs A et B dans un plan à 2 dimensions sont respectivement $(L, 0)$ et $(-L, 0)$. Si l'on suppose que la source sonore est émise à partir des coordonnées (x_s, y_s) , donc les expressions suivantes peuvent être établies :

$$\{d1 = \sqrt{(x_s - L)^2 + y_s^2}\} \quad (3.11)$$

$$\{d2 = \sqrt{(x_s + L)^2 + y_s^2}\} \quad (3.12)$$

$d1$ et $d2$ sont les distances entre la source sonore et les capteurs audio A et B respectivement.

On se basant sur les équations du mouvement, on peut représenter la distance Δd entre les deux capteurs avec l'équation suivante :

$$\Delta d = c * \tau \quad (3.13)$$

c représente la vitesse de propagation du son dans l'air qui est égale à 343 m/s. Et $\tau = \tau_{x1,x2}$ est la TDOA entre $x_1(n)$ et $x_2(n)$.

$$\tau_{x1,x2} = \tau_{micA,micB} = \frac{\sqrt{(x_s+L)^2 + y_s^2} - \sqrt{(x_s-L)^2 + y_s^2}}{C} \quad (3.14)$$

On peut estimer $\cos(\alpha)$ en utilisant l'équation suivante :

$$\cos(\alpha) = \frac{d_2 - d_1}{|H|} \quad (3.15)$$

La dernière étape est d'utiliser l'équation précédente pour obtenir l'angle α :

$$\alpha = \cos^{-1}\left(\frac{C * \tau_{x1,x2}}{|H|}\right) \quad (3.16)$$

3.6 Évaluation des performances de la détection des sons

Afin d'évaluer la précision de notre algorithme de détection nous avons utilisé le F1-Score. Cette métrique est le rapport entre les prédictions positives correctes (TP) et les erreurs (FN+FP) commises par l'algorithme. Sa formule se présente comme suit :

$$F1 \text{ Score} = \frac{TP}{TP + \frac{1}{2}(FN+FP)} \quad (3.17)$$

TP représente le nombre de fois où le modèle prédit correctement la classe positive, FP représente le nombre de fois où le modèle prédit de manière incorrecte la classe positive, FN représente le nombre de fois où le modèle prédit de manière incorrecte la classe négative (**voir le tableau 3.2**).

Tableau 3.2 : Matrice de confusion

		Classe réelle	
		Positive	Négative
Classe prédite	Positive	True Positive (Vrais positifs)	False Positive (Faux positifs)
	Négative	False Negative (Faux négatifs)	True Negative (Vrai négatifs)

3.7 Mesure de la précision de localisation

Afin d'évaluer la précision de l'algorithme de localisation, nous avons utilisé l'erreur absolue (EA):

$$EA = |\alpha_m - \alpha_R| \quad (3.18)$$

α_R représente l'angle réel de la source sonore

α_m représente l'angle estimé par l'algorithme de localisation

3.8 Conclusion

Dans ce chapitre, nous avons présenté l'architecture globale du système de détection et localisation des sons impulsifs. Nous avons expliqué les différentes méthodes utilisées pour concevoir notre système. Dans le chapitre suivant, nous allons évaluer les performances de notre solution et présenter le logiciel développé.

Chapitre 4 : Résultats et Expérimentations

4.1 Introduction

Dans ce chapitre, nous présentons trois sections importantes : (i) les logiciels et le matériel de développement, (ii) la conception des prototypes de détection et localisation simultanées des sons impulsifs (iii) et les résultats obtenus. Nous terminons ce chapitre par une conclusion.

4.2 Logiciels et matériel de développement

Dans cette partie nous présentons les logiciels et les équipements que nous avons utilisés afin de mettre en œuvre notre prototype.

4.2.1 Jupyter Notebook

Ce projet a été partiellement réalisé à l'aide du logiciel open-source Jupyter Notebook (voir la figure 4.1). Une application qui permet d'éditer et d'exécuter du code (Python, R et autres) via un navigateur Web. Il permet l'intégration de textes, d'images et de code dans un même bloc-notes ce qui facilite la compréhension des différentes parties du code entre les collaborateurs. Il permet aussi l'exécution de parties spécifiques du code, très utile pour le débogage et le test des algorithmes [35].

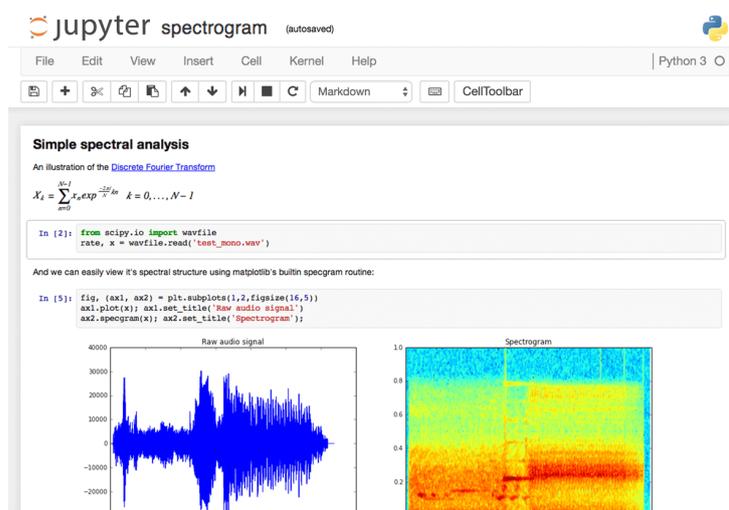


Figure 4.1: Interface du logiciel Jupyter Notebook

4.2.2 Google Colab

Google Colaboratory, ou Colab, est un service hébergé de notebooks Jupyter proposé par Google (**voir la figure 4.2**). Il ne nécessite aucune configuration et permet d'accéder rapidement, sans frais, à des ressources informatiques, dont des GPU [36]. Dans la version gratuite, l'exécution est limitée à 12 heures par session sur des serveurs avec les ressources suivantes :

- 2x vCPU
- GPU K80
- 12 GB de RAM
- 5 GB de stockage interne.

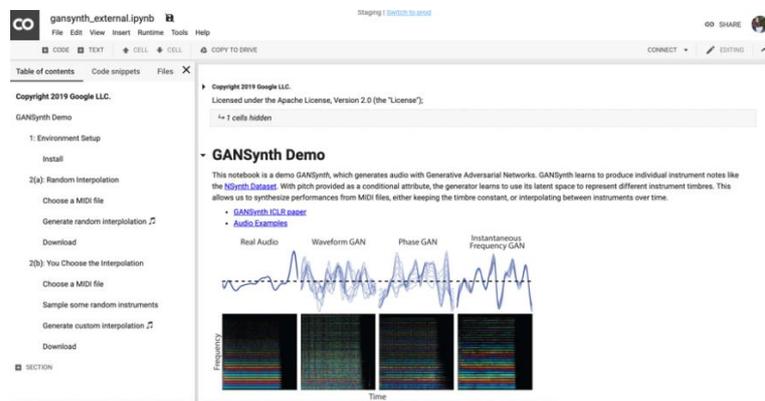


Figure 4.2: Interface du service Google Colab

4.2.3 PyCharm

Un environnement de développement intégré (IDE) développé par JetBrains pour la programmation en Python (**voir la figure 4.3**). Il offre une saisie de code intelligente, des inspections de code, la mise en évidence des erreurs à la volée et des correctifs rapides [37].

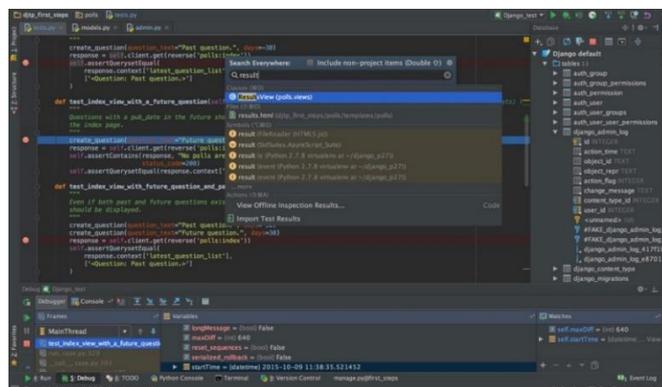


Figure 4.3: Interface du logiciel PyCharm.

4.2.4 Android Studio

Un environnement de développement d'applications mobiles pour Android basé sur IntelliJ Idea utilisant le moteur de production Gradle. Il permet principalement d'éditer les fichiers Java/Kotlin et les fichiers de configuration XML (Extensible Markup Language) d'une application Android (**voir la figure 4.4**) [38].

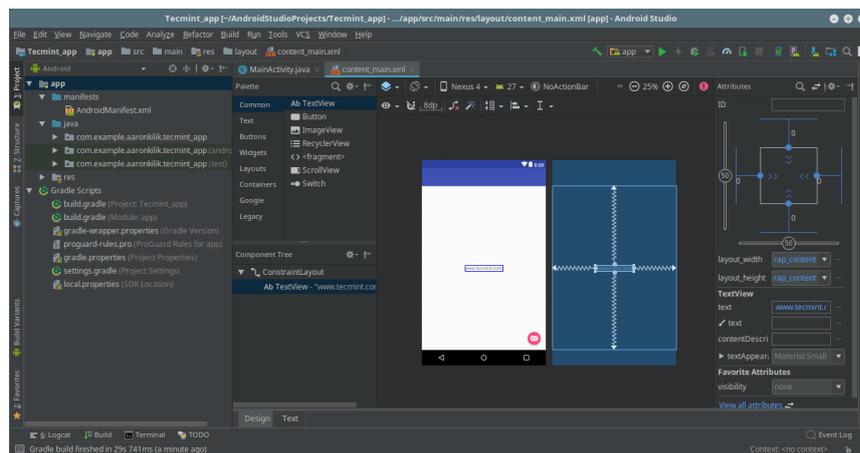


Figure 4.4: Interface du logiciel Android Studio

4.2.5 Microsoft Xbox Kinect V1

Un périphérique destiné au matériel Microsoft, la Xbox 360. Conçue par Microsoft, la Kinect permet de contrôler une interface sans utiliser de manette. Elle est destinée principalement aux jeux vidéo. Elle est équipée d'une caméra RGB qui peut capturer jusqu'à 30 images par seconde, un émetteur, un récepteur infrarouge ainsi qu'un réseau de 4 microphones (**voir la figure 4.5**). Chacun de ces microphones traite le signal audio à une fréquence maximale de 16 KHz [39]. Ces microphones sont principalement utilisés pour isoler les voix des joueurs du bruit de l'environnement pour la discussion audio lors des sessions de jeu. Ce périphérique a un coût beaucoup plus accessible par rapport à des cartes son externes et des microphones.

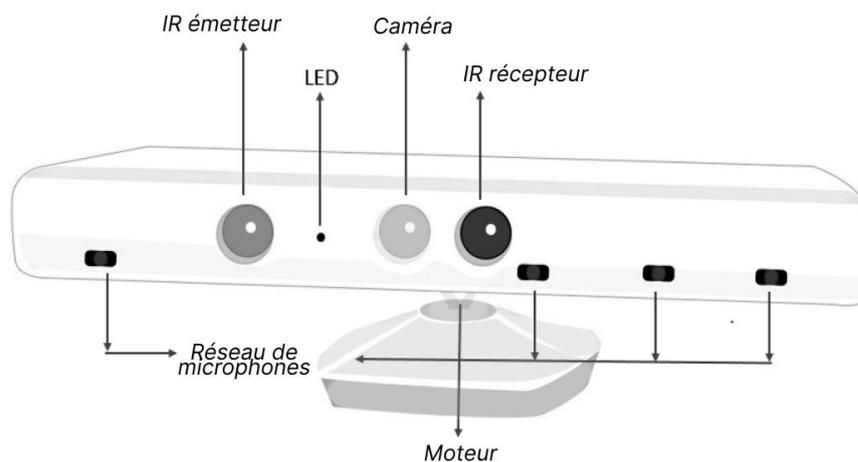


Figure 4.5: Microsoft Xbox Kinect v1

4.2.6 Microphones MEMS / Smartphone

Les microphones MEMS (en anglais Micro-Electro-Mechanical Systems) sont des microphones miniatures réalisés par des technologies MEMS (**voir la figure 4.6**). Ils sont généralement utilisés dans les smartphones. Ils offrent de bonnes performances, particulièrement dans les environnements bruyants. Les microphones MEMS sont des microphones omnidirectionnels ce qui veut dire qu'ils sont capables de détecter les sons venant de toutes les directions ce qui fait un très bon choix pour notre projet de localisation de son [40].

La majorité des smartphones modernes incorporent au minimum 2 microphones MEMS placés sur les deux côtés (**Voir la figure 4.7**). Ces microphones sont principalement utilisés pour isoler la voix lors des appels et réduire le bruit de l'environnement.

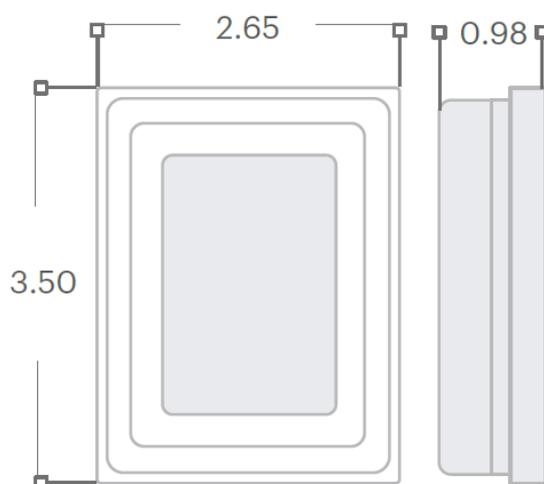


Figure 4.6: Taille d'un microphone MEMS en millimètre.

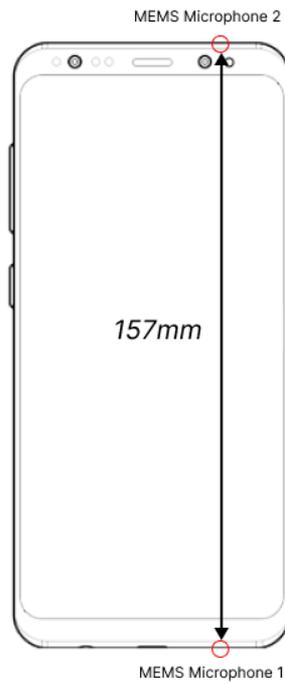


Figure 4.7: Emplacement des microphones MEMS dans un smartphone

4.2.7 Sonomètre

Un sonomètre est un appareil utilisé pour mesurer le niveau de la pression acoustique. C'est une grandeur physique liée au volume sonore. Elle est utilisée pour calculer le bruit sonore. Nous avons utilisé le sonomètre de marque kk-moon pour mesurer le niveau de pression acoustique dans notre environnement de test (**voir la figure suivante**) [41].



Figure 4.8: Sonomètre

4.3 Conception du prototype de détection et localisation des sons impulsifs

Afin de tester les algorithmes proposés dans le chapitre 3, nous avons utilisé le langage Python pour les implémenter. Nous avons utilisé la Kinect et le Smartphone pour l'acquisition du son.

4.3.1 Prototype 1 : acquisition des sons par le smartphone

Le premier prototype réalisé utilise les microphones MEMS d'un smartphone. Afin de transférer le signal audio venant de ces microphones vers l'ordinateur pour le traitement, nous avons essayé plusieurs applications mobiles (WoMic, Iruin, etc.). Le seul inconvénient de ces applications est qu'elles ne permettent pas l'acquisition du signal en mode stéréo ce qui est crucial pour la tâche de localisation comme nous l'avons expliqué dans le chapitre 3.

Cela nous a poussé à développer une application qui permet l'acquisition et le transfert du signal stéréo vers un serveur python. Le transfert des données se fait à travers une librairie de communication nommée Socket.io. C'est une bibliothèque de communication bidirectionnelle qui permet le transfert des données en temps réel. Elle utilise le protocole TCP (**voir la figure 4.9**).

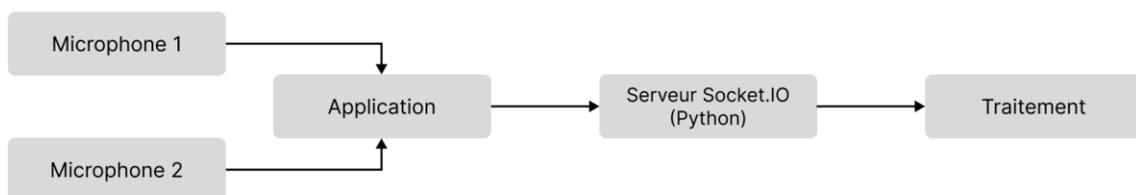


Figure 4.9: Fonctionnement de l'application développée sur Android studio

Cette application offre une interface simple et permet à l'utilisateur d'introduire l'adresse IP et le port de serveur afin d'établir la connexion (**voir la figure 4.10**). Elle permet aussi de modifier la taille de bloc et d'ajuster la sensibilité des microphones. Dès que l'utilisateur saisit ces paramètres, il clique sur le bouton « Start » qui va directement lancer l'envoi des signaux audio au serveur. Une fois que le serveur reçoit les signaux, il exploite l'un des deux pour la détection. Cependant, l'algorithme de localisation utilise les deux signaux en même temps.

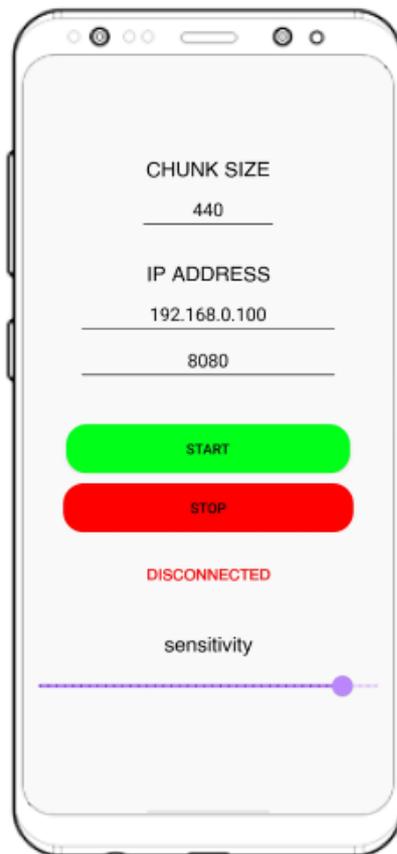


Figure 4.10 : Interface de l'application mobile

4.3.2 Prototype 2 : acquisition des sons par la Kinect

Le deuxième prototype que nous avons réalisé utilise l'un des microphones de la Kinect pour détecter les sons impulsifs et une paire de microphones pour localiser la source sonore.

Nous avons développé une interface graphique en utilisant PyCharm qui permet à l'utilisateur de contrôler le système sans aucune interaction avec le code source en lui offrant la possibilité d'introduire les paramètres qui conviennent à son environnement (**voir les figures 4.11, 4.12 et 4.13**). L'utilisateur peut changer les paramètres N, Th, Fe et d'autres.

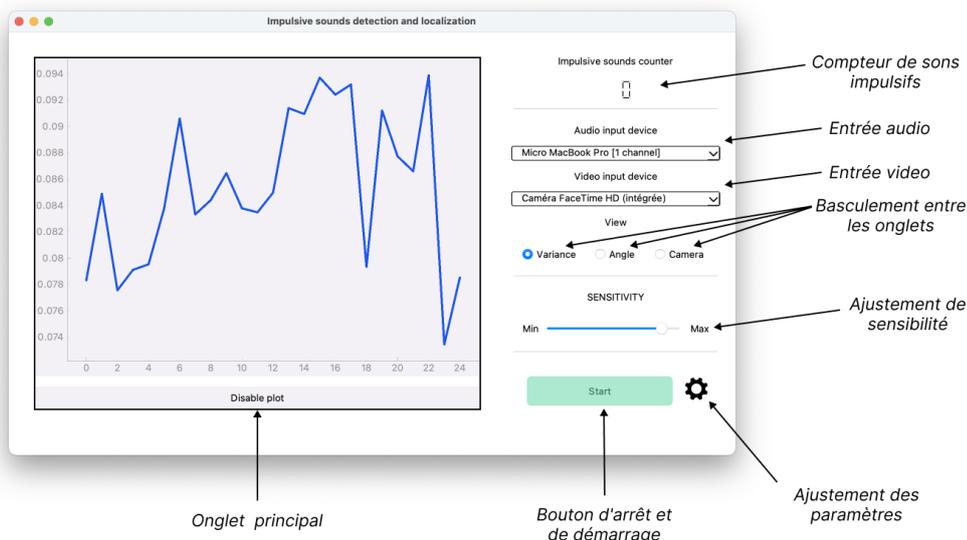


Figure 4.11 : Fenêtre principale de l'interface graphique

L'interface affiche également un compteur qui calcule le nombre de sons impulsifs détecté. De plus, elle contient trois volets qui offrent à l'utilisateur la possibilité de basculer entre eux. Le premier volet est réservé au graphe de la variance, là où l'utilisateur aura la possibilité d'analyser ce graphe et peut visualiser ses variations. Le deuxième volet contient une interface où nous affichons le retour de la caméra. En cas de détection d'un son impulsif, cette caméra enregistre, dans un fichier vidéo, la scène en cours, et la nommera avec la date et l'heure exactes où le son impulsif s'est produit. Il permet aussi de la stocker dans la mémoire de l'ordinateur afin que l'utilisateur puisse la consulter plus tard. Le dernier volet contient une interface pour afficher l'angle exact d'incidence du son impulsif (**voir la figure 4.12**).

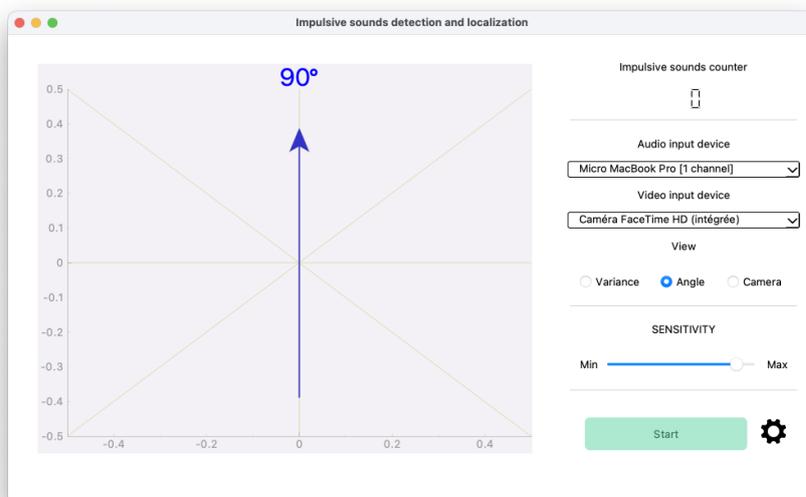


Figure 4.12 : Fenêtre de visualisation de l'angle

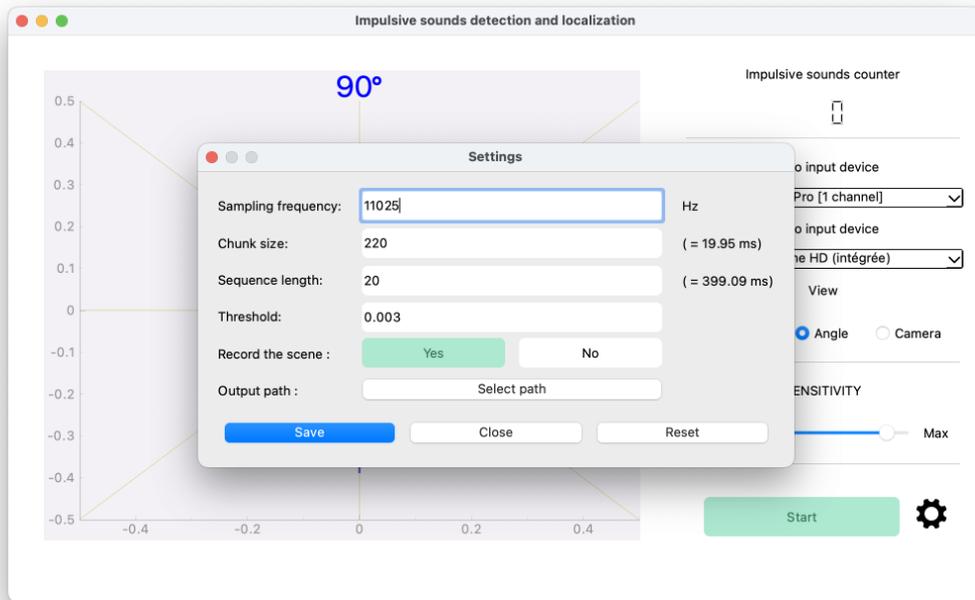


Figure 4.13 : Fenêtre pour l’ajustement des paramètres

Pour l’acquisition du son, nous avons utilisé la librairie Python PyAudio, qui permet de choisir les paramètres d’acquisition qu’on veut (F_s , N , et le nombre de canaux). Ce signal est ensuite transmis vers l’algorithme de détection. Une fois un son impulsif est détecté, nous nous intéressons à l’estimation de son angle d’incidence.

B. Douaer et al. [34] ont prouvé que la distance entre les deux microphones influence les performances de l’algorithme de localisation. De ce fait, nous avons testé plusieurs paires de microphones de la Kinect pour trouver la paire qui donne les meilleures performances de localisation. La distance entre ces capteurs est définie par le fabricant (**voir la figure 4.14**).

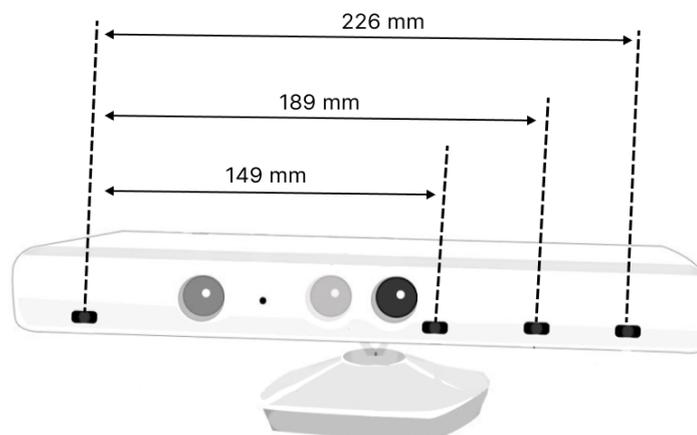


Figure 4.14 : Distance entre les microphones d’une Kinect

4.4 Scénarios de test

4.4.1 Évaluation et optimisation des paramètres de l'algorithme détection

Nous avons testé la détection des sons impulsifs en variant les paramètres N, Th et L de l'algorithme décrit dans la section 3.4 comme suit.

N prend les valeurs 20 ms, 30 ms, 60 ms, 120 ms, 240 ms

L variant entre 20 et 65 avec un pas de 5

Th variant entre 0 et 0.03 avec un pas de 0.001

Une combinaison de 1500 valeurs a été expérimentée sur un corpus de 200 sons impulsifs. La Fe de ce corpus est de 11025 Hz. Les signaux sont enregistrés à 16 bits. Nous avons marqué l'instant du début et de la fin des sons impulsifs qui composent le fichier audio. Ces instants ont été sauvegardés dans un fichier Excel.

On a adapté notre code pour fonctionner hors ligne sur ce fichier audio. Par la suite, nous l'avons exécuté sur la plateforme Google Colab.

Le programme va d'abord détecter les sons impulsifs et enregistrer l'instant de sa production. L'évaluation se fait en comparons les valeurs prédites avec celles enregistrés dans le fichier Excel. Le but de ce test est de trouver la meilleure combinaison de paramètres en évaluant la performance de chacune d'elles.

4.4.2 Détection et localisation simultanées des sons impulsifs en temps réel

Nous avons fait les tests finaux de notre application dans deux endroits différents en utilisant le matériel suivant :

- Haut-parleur : source des sons impulsifs relié à un ordinateur.
- PC Portable : Exécute m'application de notre projet et possède les ressources suivantes :
 - o Processeur Intel i3 5005U
 - o 4 Go de RAM
- Smartphone (Oppo F9) : joue le rôle d'un microphone sans fil.
- Kinect : joue le rôle de deux microphones.

Expérimentation n°1 :

La première expérimentation a été effectuée dans une salle du pavillon 1 à l'université Saad Dahlab de Blida. Nous avons testé les prototypes 1 et 2 citées ci-dessus. L'objectif de cette expérimentation est d'investiguer l'influence de la distance entre les microphones sur les performances globales de notre système.

Nous avons expérimenté 5 scénarios différents. La procédure expérimentale consiste à varier l'angle du son impulsifs de 10° à 170° par un pas de 20°.

Scénario 1 : Les microphones d'acquisition sont ceux d'un Smartphone. La distance entre ces derniers est de 157 millimètres **(voir la figure 4.7)**.

Scénario 2 : Les microphones d'acquisition sont les microphones 1 et 2 de la Kinect. La distance entre ces microphones est de 149 millimètres **(voir la figure 4.14)**.

Scénario 3 : Les microphones d'acquisition sont les microphones 1 et 3 de la Kinect. La distance entre ces microphones est de 189 millimètres **(voir la figure 4.14)**.

Scénario 4 : Les microphones d'acquisition sont les microphones 1 et 4 de la Kinect. La distance entre ces microphones est de 226 millimètres **(voir la figure 4.14)**.

Scénario 5 : Les microphones d'acquisition sont les microphones 2 et 3 de la Kinect. La distance entre ces microphones est de 40 millimètres **(voir la figure 4.14)**.

Expérimentation n°2 :

La deuxième expérimentation a été effectuée dans la salle de conférences du CDTA. La surface de cette salle est de 250 m². L'objectif de cette expérimentation est d'investiguer l'influence de la distance entre la source sonore et les capteurs sur la performance de la méthode de détection et localisation des sons impulsifs. La pression acoustique des sons impulsifs émis par le haut-parleur a été fixée à 81 dB. Nous avons utilisé un sonomètre pour l'ajustement des pressions acoustiques. La pression acoustique de cet environnement en l'absence de sons est de 50 dB **(voir les figures 4.15, 4.17)**.

Nous avons expérimenté le scénario suivant :

- Les microphones d'acquisition sont les microphones 1 et 4 de la Kinect. La distance entre ces microphones est de 226 millimètres **(voir la figure 4.14)**.

La procédure expérimentale consisté à varier :

- (i) L'angle du son impulsifs de 10° à 170° par un pas de 20°.
- (ii) La distance entre la source sonore et les microphones de 2m à 10m avec un pas de 2m.

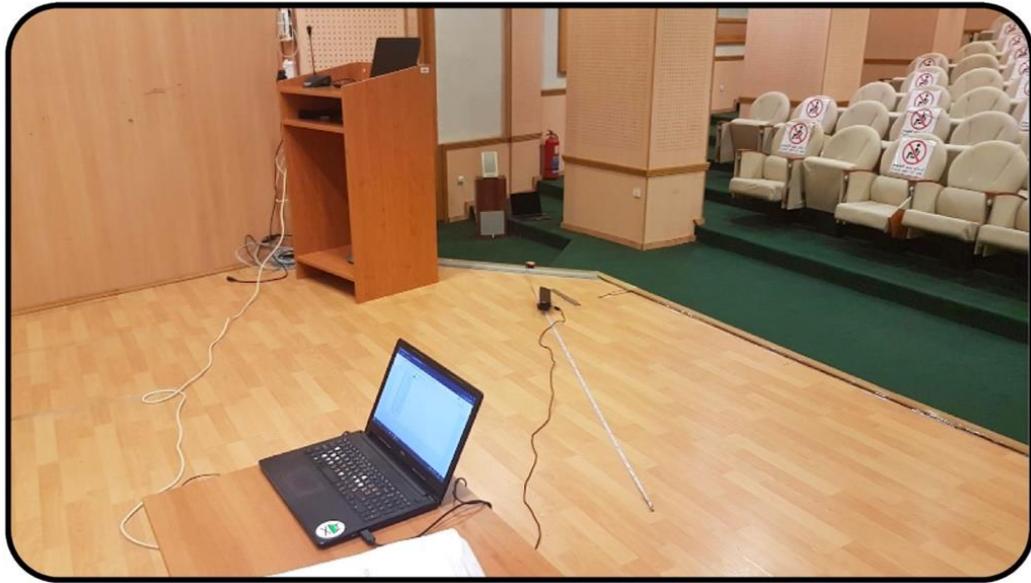


Figure 4.15 : Environnement de test



Figure 4.16 : Sonomètre utilisé



Figure 4.17 : Haut-parleur utilisé pour l'émission de sons impulsifs

4.5 Résultats expérimentaux

4.5.1 Résultats de la détection des sons impulsifs

Les paramètres qui doivent être ajustés pour assurer l'efficacité de l'algorithme de détection sont:

- La taille du bloc N
- La longueur de la séquence d'énergie L
- Le seuil de décision T_h

La F_e a été fixée à 11025 Hz.

La taille du bloc N : La valeur de ce paramètre doit être suffisamment faible afin de réduire la complexité de l'algorithme. Si la valeur est trop faible, beaucoup de détails inutiles dans la séquence d'énergie peuvent apparaître, ce qui va générer des fausses détections. Le meilleur compromis a été trouvé pour $N = 20$ ms.

La taille de la séquence L : Comme la taille du bloc N, ce paramètre doit aussi avoir une taille acceptable pour réduire la complexité algorithmique, tout en garantissant de bons résultats de détection. Une valeur de $L=20$ produira un signal impulsif d'une durée de 400ms. Cette durée est suffisante pour le fonctionnement en temps réel sur une machine avec des ressources limitées.

Le seuil T_h : ce seuil doit être suffisamment élevé pour détecter tous les sons impulsifs, mais il résulte en beaucoup de fausses détections s'il est trop élevé. Une valeur $T_h = 0.003$ a donné les meilleurs résultats pour les paramètres $N = 20$ ms et $L = 20$.

Les résultats trouvés sont représentés par les figures suivantes :

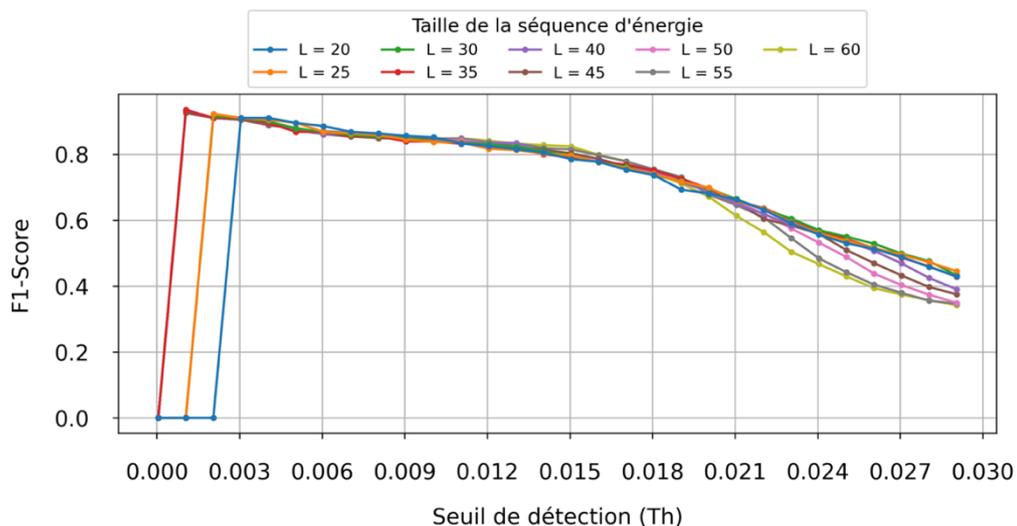


Figure 4.18 : F1-score pour une taille de bloc de 20 ms

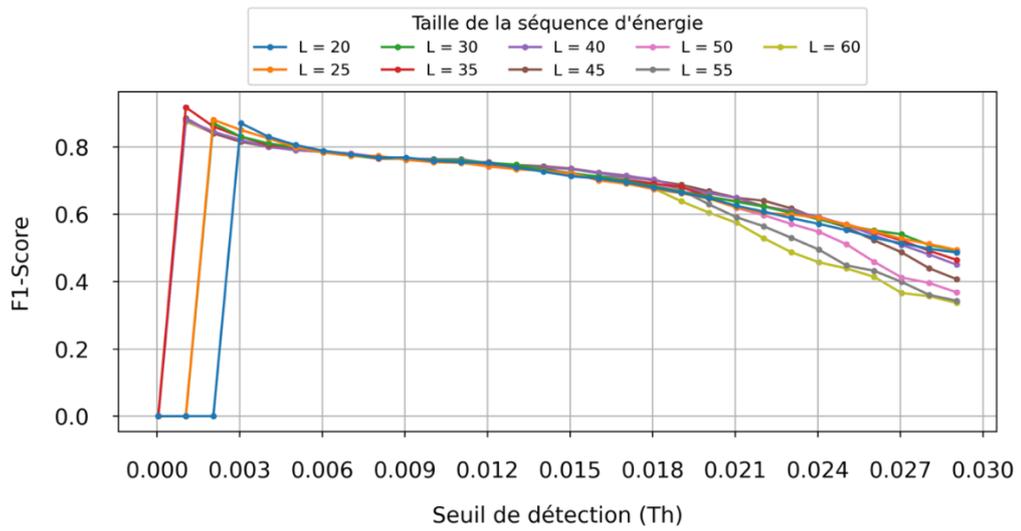


Figure 4.19 : F1-score pour une taille de bloc de 30 ms

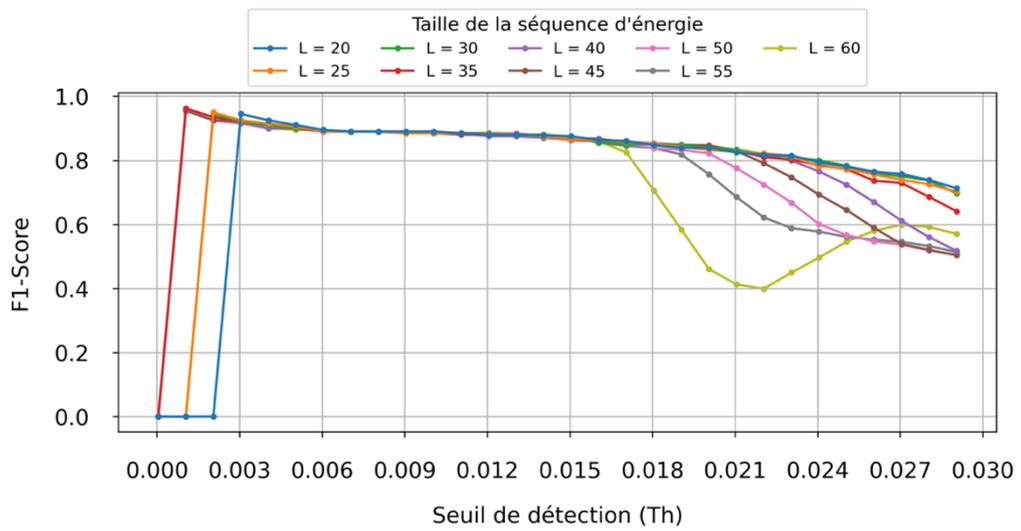


Figure 4.20 : F1-score pour une taille de bloc de 60 ms

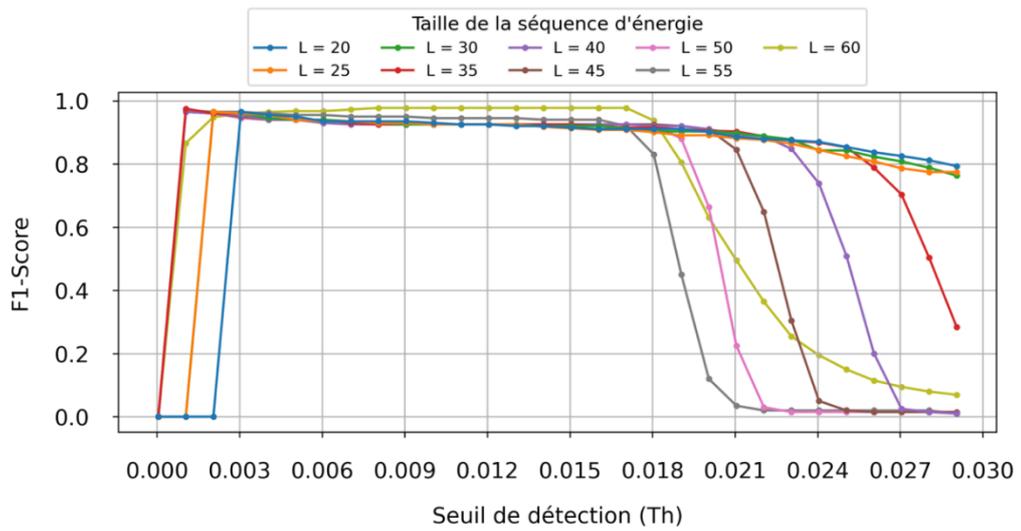


Figure 4.21 : F1-score pour une taille de bloc de 120 ms

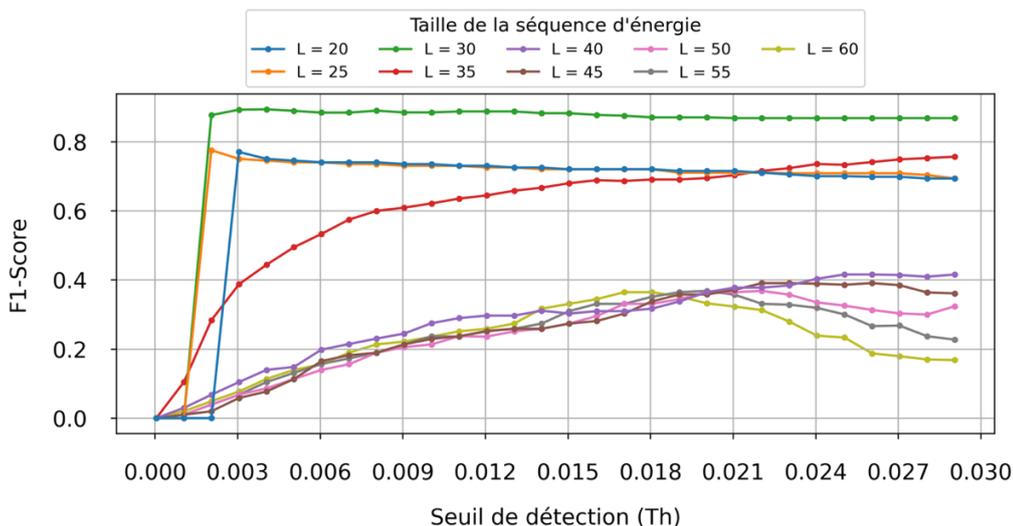


Figure 4.22 : F1-score pour une taille de bloc de 240 ms

4.5.2 Résultats de la détection et la localisation des sons impulsifs en temps réel

Pour les tests finaux de notre application, nous avons gardé les paramètres de détection choisis dans la section 4.4.1, en modifiant seulement la sensibilité qui dépend du bruit de l'environnement. Si l'environnement est calme, nous pouvons augmenter la sensibilité. Dans le cas où l'environnement est bruyant, nous diminuons la sensibilité, ce qui va ajouter plus de bruit blanc.

a) Résultats de l'expérimentation 1 :

La sensibilité du microphone été fixée à une valeur = 0.98. La courbe ci-dessous (figure 4.23) représente EA de tous les scénarios. La distance entre la source sonore et les microphones est égale à 2 mètres.

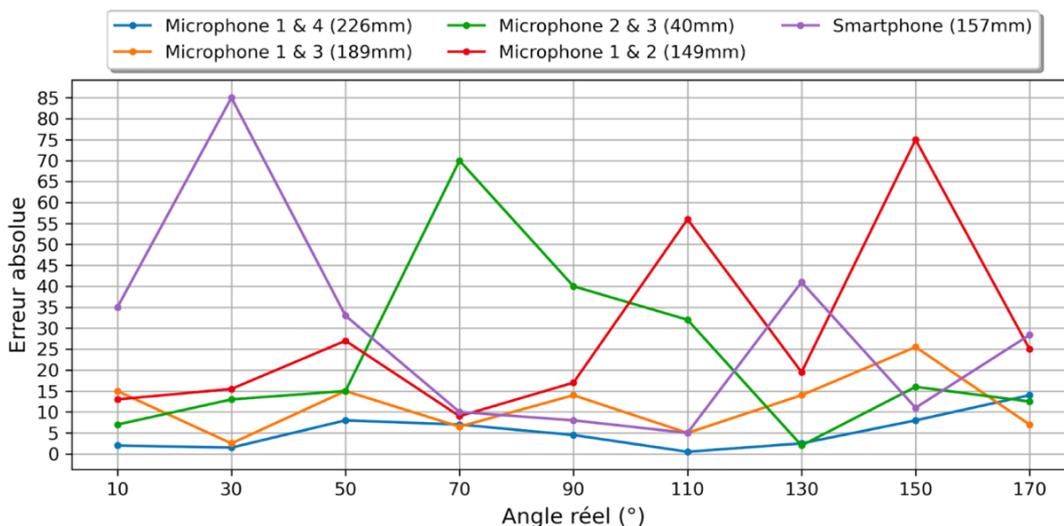


Figure 4.23 : Erreur absolue de l'écart entre l'angle mesuré et sa valeur exacte de tous les scénarios de tests

Nous remarquons que quand nous augmentons la distance entre les microphones, EA diminue. Ces résultats ne sont pas toujours parfaits, il existe toujours des valeurs aberrantes résultantes de plusieurs facteurs tels que l'écho et la réverbération à l'intérieur de la salle. Le quatrième scénario, lorsque nous avons utilisé la plus grande distance entre les microphones de la Kinect (226mm), les résultats obtenus sont les meilleurs.

b) Résultats de l'expérimentation 2 :

Parmi les scénarios de l'expérimentation 1, nous avons choisi le cas qui a donné les meilleurs résultats. La sensibilité du microphone est fixée à une valeur de 0.94. La figure 4.24 représente EA de cette expérimentation.

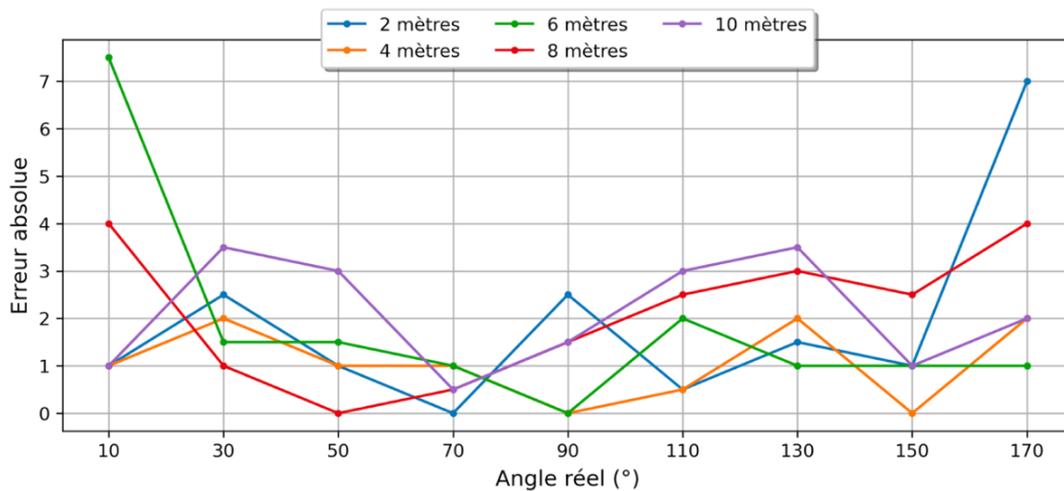


Figure 4.24 : Erreur absolue de l'écart entre l'angle mesuré et sa valeur exacte. L'acquisition des signaux a été faite par les microphones 1 & 4 de la Kinect

Nous remarquons que jusqu'à une distance de 10 mètres, l'algorithme donne de bons résultats avec une moyenne d'erreur absolue de 2.17° seulement. Nous remarquons aussi que quand le son provient d'un angle inférieur à 10° ou supérieur à 170°, EA augmente plus que la moyenne. Cette augmentation peut être expliquée par la réverbération des ondes acoustiques sur les murs dans l'environnement de test. Nous avons aussi remarqué que cela arrive quand les microphones sont près des murs.

4.5 Conclusion

Dans ce chapitre, nous avons présenté les résultats expérimentaux de notre système de détection et de localisation des sons impulsifs en temps réel. Nous avons défini l'ensemble des logiciels et matériels utilisés pour la conception et l'évaluation du prototype. Les différentes expérimentations effectuées en vue de détecter les sons impulsifs ont été décrites en détail. Le meilleur jeu de paramètres a été identifié en utilisant une approche expérimentale. Le prototype de détection et localisation simultanées des sons impulsifs a été présenté dans ce chapitre.

D'après les résultats obtenus, nous pouvons conclure que le système donne de bons résultats en utilisant des microphones distants d'une valeur de 226mm.

Les avantages de notre système sont :

- Une faible complexité qui n'exige pas beaucoup de ressources matérielles
- Facilement adaptable aux conditions de l'environnement
- Fonctionne en temps réel

Néanmoins, les principaux inconvénients de notre solution sont :

- Le système ne peut pas détecter deux événements impulsifs consécutifs si le décalage temporel entre ces événements est inférieur à la taille de la séquence d'énergie L .
- La dimension de l'environnement de test était assez limitée et pas très représentative d'un cas réel tel que les parcs de stationnement.
- La distance entre les microphones utilisés pour la localisation est assez petite.

Chapitre 5 : Conclusions et travaux futures

5.1 Conclusion

Ce travail consiste en la conception et l'implémentation d'une méthode de détection et de localisation en temps réel des événements sonores en vue de la conception d'un système de surveillance audio. La méthode que nous avons implémentée pour la détection des sons impulsifs est une approche par seuillage. Quant à la localisation, nous avons utilisé la GCC-PHAT. Ces deux méthodes sont très efficaces en termes de complexité algorithmique et ressources nécessaires. Notre système combine ces deux méthodes pour détecter et localiser les sons impulsifs en temps réel.

Dans le deuxième chapitre du mémoire, nous avons présenté des généralités sur les technologies de surveillance. Nous nous sommes principalement focalisés sur la modalité audio et ses avantages par rapport à la modalité vidéo.

Dans le troisième chapitre, nous avons présenté l'architecture globale du système proposé ainsi que les différentes méthodes utilisées pour la conception et l'évaluation du système.

Dans le quatrième chapitre, nous avons présenté les résultats expérimentaux, Nous avons comparé les performances de la détection en utilisant 1500 combinaisons de paramètres afin de trouver la combinaison qui soit la plus adaptée pour un système de détection des ondes impulsives en temps réel. Pour mesurer la performance de l'algorithme de détection, nous avons utilisé le F1 score. Nous avons trouvé le meilleur compromis entre fiabilité et performances avec une taille de bloc N égale à 20 ms, une taille de séquence d'énergie L égale à 20 ms et un seuil de détection Th de 0.003. Le F1 score de cette combinaison est de 0.91.

Afin d'implémenter le système, nous avons aussi conçu deux prototypes,

- (i) Le premier en utilisant les microphones MEMS d'un smartphone,
- (ii) Le deuxième en utilisant les microphones 1 et 4 d'une Xbox Kinect v1.

Pour évaluer l'algorithme de localisation, nous avons utilisé l'erreur absolue.

Les résultats de l'évaluation avec le prototype 1 n'étaient pas très satisfaisants. La moyenne d'erreur absolue est de 28.5° . Cela est dû à la faible distance entre les microphones du smartphone (157 mm).

Afin de vérifier ce résultat, nous avons testé plusieurs combinaisons de microphones de la Kinect. Nous avons trouvé que plus la distance entre les microphones est élevée, plus l'algorithme est précis. Nous avons aussi trouvé que d'autres travaux de recherche ont aussi confirmés ce résultat [34].

Nous avons exploité ce résultat pour monter une deuxième expérimentation en utilisant les microphones 1 et 4 de la Kinect (la distance entre les microphones est de 226 mm). Ce deuxième prototype a été testé en variant la distance entre la source sonore et la Kinect de 2 à 10 mètres par un pas de 2 mètres. Nous avons trouvé que, jusqu'à une distance de 10 mètres, les résultats obtenus sont satisfaisants. La moyenne d'erreur est de 2.17° seulement. Nous avons aussi remarqué que l'erreur absolue augmente quand le son provient d'un angle inférieur à 10° ou supérieur à 170° à cause des réverbérations des ondes acoustiques sur les murs de la salle de test. Donc pour assurer une meilleure précision, nous recommandons de placer les microphones loin des murs. Nous concluons que la détection par seuillage et la localisation à base de la GCC-PHAT sont des méthodes appropriées pour un système de surveillance en temps réel. Ces méthodes possèdent l'avantage d'être :

- i. Rapides et efficaces en termes de complexité algorithmique.
- ii. Précises dans la localisation.
- iii. Fonctionnelles en temps réel.
- iv. Facilement adaptable aux conditions de l'environnement.

5.2 Travaux futurs

Les travaux futurs consistent à :

- Implémenter les modules de détection et de localisation avec un système de surveillance vidéo afin de faire pivoter la caméra vers la source des événements.
- Implémenter un prototype de détection et localisation avec des microphones externes et une carte son de meilleure qualité.
- Mesurer les performances de la détection et la localisation dans un environnement réel tel que les parcs de stationnement ou la source sonore est distante de plus de 10 mètres.

Bibliographie

- [1] C. A. Williams, "Police Surveillance and the Emergence of CCTV in the 1960s," *Crime Prevention and Community Safety* 2003 5:3, vol. 5, no. 3, pp. 27–37, Jul. 2003, doi: 10.1057/PALGRAVE.CPCS.8140153.
- [2] R. Surette, "The thinking eye: Pros and cons of second generation CCTV surveillance systems," *Policing*, vol. 28, no. 1. pp. 152–173, 2005. doi: 10.1108/13639510510581039.
- [3] T. D. Rätty, "Survey on contemporary remote surveillance systems for public safety," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 40, no. 5, pp. 493–515, Sep. 2010, doi: 10.1109/TSMCC.2010.2042446.
- [4] S. AHMED HAMADA, "Détection des sons impulsifs en vue de la mise en œuvre d'un système de surveillance audio," Saad Dahleb Blida University , Blida.
- [5] M. Crocco, M. Cristani, A. Trucco, and V. Murino, "Audio surveillance: A systematic review," *ACM Computing Surveys*, vol. 48, no. 4. Association for Computing Machinery, Mar. 18, 2016. doi: 10.1145/2871183.
- [6] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, and M. Vento, "Audio surveillance of roads: A system for detecting anomalous sounds," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 279–288, Jan. 2016, doi: 10.1109/TITS.2015.2470216.
- [7] C. Clavel, T. Ehrette, and G. Richard, "Events Detection for an Audio-Based Surveillance System," in *2005 IEEE International Conference on Multimedia and Expo*, 2005, pp. 1306–1309. doi: 10.1109/ICME.2005.1521669.
- [8] S. Chandrakala and S. L. Jayalakshmi, "Environmental Audio Scene and Sound Event Recognition for Autonomous Surveillance," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–34, May 2020, doi: 10.1145/3322240.
- [9] J. A. Mazurek *et al.*, "Boomerang mobile counter shooter detection system," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV*, May 2005, vol. 5778, p. 264. doi: 10.1117/12.607616.
- [10] "USGS estimates impervious parking lot coverage for all 3,109 U.S. counties - Stormwater Report." <https://stormwater.wef.org/2019/06/usgs-estimates-impervious-parking-lot-coverage-for-all-3109-u-s-counties/> (accessed May 21, 2022).
- [11] G. Ciaburro, "Sound event detection in underground parking garage using convolutional neural network," *Big Data and Cognitive Computing*, vol. 4, no. 3, pp. 1–14, Sep. 2020, doi: 10.3390/bdcc4030020.
- [12] Vázquez y del Árbol and Esther, "Modern Journal of Language Teaching Methods (MJLTM)," *Modern Journal of Language Teaching Methods*, 2018, doi: 10.26655/mjltm.2018.12.1.
- [13] E. Weinstein, K. Steele, A. Agarwal, and J. Glass, LOUD: A 1020-Node Microphone Array and Acoustic Beamformer. International Congress on Sound and Vibration (ICSV), July 2007, Cairns, Australia.

- [14] Y.-T. Peng, C.-Y. Lin, M.-T. Sun, and K.-C. Tsai, "Healthcare audio event classification using Hidden Markov Models and Hierarchical Hidden Markov Models," *2009 IEEE International Conference on Multimedia and Expo*, pp. 1218–1221, 2009.
- [15] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "On acoustic surveillance of hazardous situations," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 165–168, 2009.
- [16] J. CLARKSON, "6 - HUMAN CAPABILITY AND PRODUCT DESIGN," in *Product Experience*, H. N. J. Schifferstein and P. Hekkert, Eds. San Diego: Elsevier, 2008, pp. 165–198. doi: <https://doi.org/10.1016/B978-008045089-6.50009-5>.
- [17] S. Sigtia, A. M. Stark, S. Krstulović, and M. D. Plumbley, "Automatic Environmental Sound Recognition: Performance Versus Computational Cost," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 11, pp. 2096–2107, Nov. 2016, doi: 10.1109/TASLP.2016.2592698.
- [18] T. Virtanen and M. Helen, "Probabilistic Model Based Similarity Measures for Audio Query-by-Example," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007, pp. 82–85. doi: 10.1109/ASPAA.2007.4393031.
- [19] S. Duan, J. Zhang, P. Roe, and M. W. Towsey, "A survey of tagging techniques for music, speech and environmental sound," *Artificial Intelligence Review*, vol. 42, pp. 637–661, 2012.
- [20] S. Chu, S. Narayanan, C. -c. J. Kuo, and M. J. Mataric, "Where am I? Scene Recognition for Mobile Robots using Audio Features," in *2006 IEEE International Conference on Multimedia and Expo*, 2006, pp. 885–888. doi: 10.1109/ICME.2006.262661.
- [21] M. Cristani, M. Bicego, and V. Murino, "Audio-Visual Event Recognition in Surveillance Video Sequences," *Multimedia, IEEE Transactions on*, vol. 9, pp. 257–267, May 2007, doi: 10.1109/TMM.2006.886263.
- [22] A. Pourmohammad and S. M. Ahadi, "N-dimensional N-microphone sound source localization," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 1, p. 27, 2013, doi: 10.1186/1687-4722-2013-27.
- [23] E. L. Ferguson, S. B. Williams, and C. T. Jin, "Sound Source Localization in a Multipath Environment Using Convolutional Neural Networks," *CoRR*, vol. abs/1710.10948, 2017, [Online]. Available: <http://arxiv.org/abs/1710.10948>
- [24] S. Hoseini, A. Rezaie, and Y. Zanjireh, "Time difference of arrival estimation of sound source using Cross Correlation and modified maximum likelihood weighting function," *Scientia Iranica*, vol. 24, May 2017, doi: 10.24200/SCI.2017.4355.
- [25] S. Aleinik and M. Stolbov, "SNR Estimation Based on Adaptive Signal Decomposition for Quality Evaluation of Speech Enhancement Algorithms," Jun. 2015, pp. 364–371. doi: 10.1007/978-3-319-23132-7_45.
- [26] A. K. Zad Tehrani, B. Makkiabadi, A. Parsayan, and S. H. Hozhabr, "Sound Source Localization Using Time Differences of Arrival; Euclidean Distance Matrices Based Approach," in *2018 9th International Symposium on Telecommunications (IST)*, 2018, pp. 91–95. doi: 10.1109/ISTEL.2018.8661037.

- [27] J. Wang, X. Qian, Z. Pan, M. Zhang, and H. Li, "GCC-PHAT with Speech-oriented Attention for Robotic Sound Source Localization," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 5876–5883. doi: 10.1109/ICRA48506.2021.9561885.
- [28] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," in *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, 2007, pp. 21–26. doi: 10.1109/AVSS.2007.4425280.
- [29] C. Zieger, A. Brutti, and P. Svaizer, "Acoustic Based Surveillance System for Intrusion Detection," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2009, pp. 314–319. doi: 10.1109/AVSS.2009.49.
- [30] N. Pekez, N. Kaprocki, and J. Kovacevic, "Implementation of PC Application for Controlling RT-AG External Sound Card," *2018 Zooming Innovation in Consumer Technologies Conference, ZINC 2018*, pp. 135–139, Aug. 2018, doi: 10.1109/ZINC.2018.8448593.
- [31] L. Pei *et al.*, "Sound positioning using a small-scale linear microphone array," in *2013 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2013*, Jun. 2013, pp. 1–7. doi: 10.1109/IPIN.2013.6817869.
- [32] F. Ykhlef, S. Ahmed Hamada, F. Ykhlef, A. Derbal and D. Bouchaffra, "Real-time detection of impulsive sounds for audio surveillance systems," 3rd edition of the National Study Day on Research on Computer Sciences (JERI 2019), volume 2351, paper 49, 2019.
- [33] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976, doi: 10.1109/TASSP.1976.1162830.
- [34] F. and Y. F. Douaer Belgacem and Ykhlef, "Experimental Investigation into the Influence of the Distance Between Microphones for 2D Real-Time Sound Source Localization Using GCC-PHAT Technique," in *Advances in Computing Systems and Applications*, 2021, pp. 354–362.
- [35] K. M. Mendez, L. Pritchard, S. N. Reinke, and D. I. Broadhurst, "Toward collaborative open data science in metabolomics using Jupyter Notebooks and cloud computing," *Metabolomics*, vol. 15, no. 10, p. 125, 2019, doi: 10.1007/s11306-019-1588-0.
- [36] T. Carneiro, R. V. Medeiros Da Nóbrega, T. Nepomuceno, G.-B. Bian, V. H. C. de Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.
- [37] "PyCharm : l'IDE Python pour développeurs professionnels par JetBrains." <https://www.jetbrains.com/fr-fr/pycharm/> (accessed Jun. 22, 2022).
- [38] "Android Studio and SDK tools | Android Developers." <https://developer.android.com/studio> (accessed Jun. 22, 2022).
- [39] "Microsoft Kinect – Hardware | Geospatial Modeling & Visualization." <https://gmvc.cast.uark.edu/scanning/hardware/microsoft-kinect-resourceshardware/?fbclid=IwAR2srj-8QQdeXQhRangvfDrJDHIJj0UfHzYMsqQlUiRfRMskhCX4BFFSKUQ> (accessed Jun. 19, 2022).

- [40] F. Khoshnoud and C. W. de Silva, "Recent advances in MEMS sensor technology-mechanical applications," *IEEE Instrumentation & Measurement Magazine*, vol. 15, no. 2, pp. 14–24, 2012, doi: 10.1109/MIM.2012.6174574.
- [41] "What Is A Sound Level Meter? | Brüel & Kjær."
<https://www.bksv.com/en/knowledge/blog/sound/what-is-a-sound-level-meter> (accessed Jun. 22, 2022).