

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

Présenté par

Sarah AHMED HAMADA

Pour l'obtention du diplôme de Master en Électronique

Spécialité : Traitement de l'Information et Systèmes Électroniques (TISE)

Thème

Détection des sons impulsifs en vue de la mise en œuvre d'un système de surveillance audio

Proposé par :

Fayçal YKHLEF, Maître de Recherche B, CDTA, Alger.

Farid YKHLEF, Maître de Conférences A, Université SAAD DAHLEB, BLIDA.

Année Universitaire 2016-2017

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 B 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 remercions également notre co-encadreur Dr. Farid YKHLEF, *Maitre de Conférence A* au niveau de l'*Université SAAD DAHLEB*, pour sa disponibilité et ses conseils précieux.

Nous devons un remerciement à tous les enseignants de l'*université SAAD DAHLEB* pour leurs qualités scientifiques et pédagogiques.

DEDICACE

Je dédie ce mémoire à :

Mes parents :

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui trouve ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit. Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Mes frères et sœurs qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

A tout ma famille chacun son nom et spécialement ma grand-mère qui a été toujours présente pour m'encourager

A mes meilleures amies qui ont été toujours à mes côtés surtout dans les moments difficiles.

A tous ceux qui, par un mot, m'ont donné la force de continuer.

Et tous ceux qui j'aime et qui m'aiment.

A mes collègues de promotion TISE 2017.

Sarah

RESUME

ملخص:

الهدف الاساسي من هذا المشروع لنهاية المشوار الدراسي هو تحقيق كاشف للأحداث الصوتية المندفعة في الوقت الحقيقي. التقنية المستخدمة لهذا المشروع تستند إلى تحليل تباين وتغير في الطاقة على المدى القصير لإشارة صوتية. تم استقبال البيانات الصوتية بواسطة ميكروفون لاسلكي ميمس (ميكروفون من الهاتف الذكي). عملية الكشف عن الاحداث الصوتية تمت عن طريق ضبط حساسية الميكروفون. قمنا بتنفيذ نسخة نموذجية في برنامج ماتلاب. واستخدمنا مقاييس للتقييم وهي المعدل الإيجابي الحقيقي (بفت) والمعدل الإيجابي الكاذب (بت) ومنحنى روك. تم اختبار البرمجيات المنفذة في غرفة مغلقة (داخلية). والنتائج التي تم الحصول عليها كانت مشجعة جدا.

الكلمات الدالة: الأصوات المندفعة، الكشف عن الحدث الصوتي، معالجة الإشارات الصوتية

Résumé :

L'objectif principal de ce projet de fin d'étude est la mise en œuvre d'un détecteur d'évènements acoustiques impulsifs en temps réel. La méthode utilisée est basée sur l'analyse de la variance des énergies à courts termes d'un signal audio. L'acquisition des données acoustiques a été accomplie par un microphone MEMS sans fils (microphone d'un smartphone). Le détecteur d'évènements acoustiques est doté d'un module pour l'ajustement de la sensibilité du microphone. Nous avons implémenté une version prototype sous MATLAB. Les métriques d'évaluation que nous avons utilisées sont : le taux de vrais positifs (TVP), le taux de faux positifs (TFP) ainsi que la courbe ROC. Le logiciel implémenté a été testé dans une salle fermée (indoor). Les résultats obtenus sont très encourageants.

Mots clés : Sons impulsifs, détection des évènements acoustique, traitement du signal audio.

Abstract:

The main purpose of this project is to implement a real time impulsive events detector. The method we have used is based on the variance-analysis of short-term energies extracted from a given audio waveform. Data acquisition was accomplished using a wireless MEMS microphone. The impulsive sound detector is equipped with a specific module for tuning the sensitivity of the microphone. We have implemented a prototype version using MATLAB software. The evaluation metrics we have used are: the true positive rate (TPR), the false positive rate (FPR), and the ROC curve. The implemented software has been tested within an indoor environment. The obtained results are very promising.

Keywords: Impulsive sounds, acoustic events detector, audio processing.

TABLE DES MATIERES

Résumé	(i)
Remerciements	(ii)
Table des matières	(iii)
Liste des acronymes et abréviations	(iv)
Liste des figures	(v)

Chapitre 1 : Introduction

1.1	Motivations.....	9
1.2	Contributions.....	9
1.3	Impact du projet.....	10
1.4	Organisation du mémoire.....	10

Chapitre 2 : Généralités sur les systèmes de surveillance

2.1	Introduction.....	12
2.2	Systèmes de surveillance.....	12
2.3	Modalité audio pour la surveillance.....	14
2.4	Sons impulsifs.....	14
2.5	Détection et reconnaissance des sons impulsifs.....	16
2.6	Conclusion.....	17

Chapitre 3 : Détection des sons impulsifs

3.1	Introduction.....	18
3.2	Etat de l'art.....	18
3.3	Critères de choix des méthodes de détection des sons impulsifs.....	20
3.4	Détection des sons impulsifs à base de la variance des séquences d'énergies.....	21
3.5	Métriques d'évaluation.....	28
3.6	Conclusion.....	29

Chapitre 4 : Résultats expérimentaux

4.1	Introduction.....	30
4.2	Logiciels et matériel de développement.....	30
4.3	Corpus de tests.....	34

4.4	Résultats expérimentaux.....	35
4.5	Application développée.....	44
4.6	Conclusion	45

Chapitre 5 : Conclusions et travaux futures

5.1	Conclusions.....	47
5.2	Travaux futurs.....	48

6	Bibliographie	49
----------	----------------------------	----

LISTE DES ACRONYMES ET ABREVIATIONS

CCTV : *Closed-Circuit TeleVision*

DB SPL: Decibel Sound Pressure Level

EDI : Environnement de Développement Intégré

Fe : Fréquence d'échantillonnage

GMM : *Gaussian Mixture Models*

HMMs : *Hidden Markov Models*

MEMS: *Micro-Electrical-Mechanical Systems*

RSB: Rapport Signal sur Bruit

SI : Sons impulsifs

SS : Systèmes de surveillance

SVMs : *Support Vector Machines*

TEO : *Teager Energy Operator*

TFP : Taux de faux positifs

TFN : Taux de faux négatifs

TVP : Taux de vrais positifs

Wi-Fi : *Wireless Fidelity*

LISTE DES FIGURES

Figure 2.1: Système de surveillance exploitant la modalité audio	15
Figure 2.2: Onde de Friedlander	16
Figure 2.3: Système de détection et de reconnaissance des sons impulsifs	17
Figure 3.1: Densité de probabilité gaussienne	28
Figure 4.1 : MATLAB R2015b.....	30
Figure 4.2 : GoldWave.....	31
Figure 4.3 : Wo Mic Client et serveur (PC et Smartphone).....	32
Figure 4.4 : MyPublicWiFi	33
Figure 4.5 : Sonomètre KMOON GM1352	33
Figure 4.6 : Microphone MEMS	34
Figure 4.7 : Connexion entre le Pc et le Smartphone à l'aide de MyPubliWiFi et Wo Mic	36
Figure 4.8 : Disposition du matériel de test.....	37
Figure 4.9 : Instants de début des sons impulsifs	39
Figure 4.10 : Valeurs expérimentale du seuil de décision	42
Figure 4.10 : Courbe ROC pour un niveau d'intensité de sons impulsifs de 71 dB SPL.....	42
Figure 4.12 : Courbe ROC pour un niveau d'intensité de sons impulsifs de 82 dB SPL.....	43
Figure 4.13 : Courbe ROC pour un niveau d'intensité de sons impulsifs de 90 dB SPL.....	43
Figure 4.14 : Interface MATLAB pour la détection en temps réel des sons impulsifs.....	45

Chapitre 1 : Introduction

1.1 Motivations

La sécurité des citoyens est une problématique de grande importance dans tous les pays du monde. L'installation des systèmes de surveillance est devenue incontournable dans les environnements urbains. La troisième génération des systèmes de surveillance intègre des capteurs complémentaires aux caméras pour assurer une vigilance automatisée. Nous citons comme exemples, (i) les capteurs thermiques, (ii) les capteurs infra-rouges, (iii) les détecteurs de mouvements, (iv) les capteurs sismiques et (v) les microphones acoustiques.

En particulier, l'information audio, captée par les microphones de surveillance, complète efficacement les séquences vidéo car elle permet de capturer des événements dangereux qui se déroulent en dehors du champ de vision des caméras. En plus, la donnée audio est très nécessaire lorsque les conditions climatiques sont défavorables (pluies, brouillarde et autres). Pour assurer une surveillance vigilante de la région, l'information audio peut être utilisée pour pivoter les caméras de surveillance vers les sources d'évènements acoustiques dangereuses (cris humains, coups de feu, bris de verre, aboiement de chiens, accident de voitures et autres). La technologie utilisée pour pivoter les caméras est la localisation des sources sonores.

Théoriquement, la particularité acoustique commune à ces évènements est le caractère impulsif. On parle dans ce cas des **évènements audio impulsifs**. La **détection** et la **reconnaissance** de ces évènements est une phase clé pour la mise en œuvre d'un système de surveillance de troisième génération.

1.2 Contributions

Nous nous focalisons dans le cadre de ce projet à la **détection** des évènements impulsifs. La reconnaissance de ces évènements n'est pas abordée dans le cadre de ce mémoire.

Notre contribution consiste à mettre en œuvre un logiciel de détection des sons impulsifs en temps réel et cela dans un environnement fermé (Indoor). Le logiciel fonctionne sur un Pc portable sous Windows. Quatre volets essentiels ont été abordés dans ce projet : (i) l'étude d'une méthode de détection des sons impulsifs à base de l'analyse de la variance des énergies à courts termes, (ii) l'optimisation des performances de cette méthode en se basant sur plusieurs

paramètres algorithmiques (iii) l'exploration d'un capteur de type MEMS sans fils pour l'acquisition en temps réel des données acoustiques (exploitation du microphone d'un smartphone) et (iv) la proposition d'un module de réglage de la sensibilité du capteur en vue de l'amélioration des performances de détection.

Les paramètres algorithmiques de cette méthode ont été optimisés par une approche pratique à base du calcul d'erreurs et une représentation graphique à l'aide des courbes ROC.

1.3 Impact du projet

Le travail présenté dans ce projet de fin d'étude est très utile pour le développement des technologies industrielles de notre pays. Il s'agit d'un module primordial pour la mise en œuvre des systèmes de surveillance audio. En Algérie, ce genre de systèmes est actuellement à fort impact socioéconomique.

Une fois nous maîtrisons la conception et l'évaluation des systèmes de surveillance (détection et reconnaissance en temps réel des sons impulsifs), nous pouvons sans aucun doute intéresser plusieurs clients en Algérie. Nous citons : la DGSN (direction générale de la sûreté interne), la MDN (ministère de la défense nationale), les supermarchés, les banques, les sociétés, les universités, les écoles, les lycées et les ambassades. De plus, ce système pourra facilement être intégré dans un projet de cité intelligente (*Smart City*).

Finalement, nous tenons à souligner que le système développé dans ce mémoire cadre exactement avec la spécialité de notre Master : Traitement de l'Information et Systèmes Électroniques (TISE).

1.4 Organisation du mémoire

Le mémoire est organisé comme suit :

Le deuxième chapitre : présent des généralités sur les systèmes de surveillance et se focalise principalement sur la modalité audio. Des définitions succinctes sur la détection et reconnaissance des sons impulsifs sont aussi données dans ce chapitre.

Le troisième chapitre : expose, en premier lieu, un état de l'art sur les méthodes de détection des événements impulsifs. Par la suite, une étude détaillée sur la détection des événements impulsifs à base de la variance des séquences d'énergies est présentée. Les

métriques d'évaluation utilisées pour mesurer les performances de cette méthode sont aussi données dans le même chapitre.

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

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

Chapitre 2 : Généralités sur les systèmes de surveillance

2.1 Introduction

Dans ce chapitre, nous décrivons brièvement les systèmes de surveillance. Nous nous focalisons sur l'apport de la modalité audio à ce genre de système. Nous donnons aussi des définitions sur les sons impulsifs suivis par une description des processus de détection et reconnaissance des événements impulsifs. Le chapitre est clôturé par une conclusion.

2.2 Systèmes de surveillance

2.2.1 Historiques

Dans les années soixante-dix, le Royaume-Uni installa pour la première fois des caméras de surveillance dans des lieux publics afin de visualiser à distance et d'enregistrer l'activité qui s'y déroule. L'objectif étant de lutter contre les attentats terroristes [1].

Depuis les années quatre-vingt-dix, des millions de caméras ont été déployés dans la majorité des villes, sur les routes et dans les transports dans tous les pays du monde. La surveillance vidéo s'est aussi développée dans les immeubles, les parkings, les magasins, et même dans les propriétés privées [1].

2.2.2 Evolution des systèmes de surveillance

L'évolution technologique des systèmes de surveillance (SS) a commencé avec les systèmes analogiques d'enregistrement vidéo.

Ces systèmes se composent d'un certain nombre de caméras situées dans des emplacements distants. Les caméras sont reliées à un ensemble de moniteurs, habituellement placés dans une seule salle de commande par l'intermédiaire de commutateurs (une matrice vidéo) [1].

2.2.2.1 Première génération des systèmes de surveillance

Les SS de première génération (1960-1980) sont basés sur des techniques analogiques pour l'acquisition des séquences vidéo (Caméras CCTV : Closed-Circuit TeleVision). Les séquences acquises sont affichées en permanence sur plusieurs moniteurs analogiques [2].

Les SS de première génération souffrent de plusieurs problèmes, nous citons : (i) la contrainte d'archivage de données et (ii) la contrainte de la surveillance permanente des événements risqués.

Les SS à cette époque souffrent du problème de stockage de données issues des systèmes d'acquisition analogiques.

L'enregistrement des séquences vidéo nécessite l'utilisation de bandes magnétiques de longue durée ce qui complique par la suite les procédures de récupération des séquences vidéo sauvegardées.

L'utilisation de plusieurs moniteurs d'affichage embrouille la procédure de surveillance. En fait, l'identification des événements dangereux par la visualisation permanente des écrans est une tâche difficile, particulièrement si le nombre de moniteurs est élevé [2].

2.2.2.2 Deuxième génération des systèmes de surveillance

Afin de faire face au problème de surveillance rencontrés par les SS de première génération, les SS de deuxième génération (1980-2000) utilisent des ordinateurs dotés d'algorithmes intelligents pour l'identification automatique et en temps réel des événements dangereux [1], [2]. Ces algorithmes sont basés sur des techniques de traitement d'image et reconnaissance de formes.

2.2.2.3 Troisième génération des systèmes de surveillance

La troisième génération des SS se concentrent sur l'intégration de capteurs complémentaires aux caméras afin de développer une vigilance automatisée au sein des systèmes de surveillance [3]. Ces systèmes ont vu le jour à partir des années 2000.

Les capteurs intégrés peuvent être : (i) des caméras thermiques ou infra rouges, (ii) des détecteurs de mouvements, (iii) des microphones acoustiques et des (iv) capteurs sismiques [2].

L'utilisation de plusieurs capteurs permet de diversifier l'information disponible et de renforcer la qualité de la détection des situations anormales, y compris dans de mauvaises conditions d'observation.

2.3 Modalité audio pour la surveillance

La modalité audio complète efficacement l'information vidéo car elle permet de capturer des événements qui se déroulent en dehors du champ des caméras, ou lorsque les conditions sont défavorables (par exemple, une visibilité trop limitée due à des conditions climatiques) [3].

L'analyse auditive peut fournir une prise de conscience automatique afin de concentrer l'attention des opérateurs de surveillance sur des situations réellement dangereuses tels que les coups de feu, les explosives, les cris et les bris de vitres. Les sons issues suite à ces évènements acoustiques sont appelés **sons impulsifs** (SI).

Les opérations de détection et reconnaissance de ces sons sont utilisées pour déclencher des alarmes et enregistrer les séquences vidéo acquises via des caméras réseaux [4].

La Figure 2.1 schématise un système de vidéo-surveillance exploitant la modalité audio.

2.4 Sons impulsifs

Les sons impulsifs (SI) sont des bruits transitoires qui résultent d'une libération soudaine d'énergie. Les sources typiques de ces sons transitoires sont les décharges explosives [5].

Un SI est caractérisé par une courte durée, d'un début brutal et d'une désintégration rapide [6].

La signature acoustique des SI est souvent modélisée par l'onde de Friedlander P(t) (Figure 2) [5] :

$$P(t) = \left(1 - \frac{t}{c}\right) e^{-t/c}, \quad 0 \leq t < \infty \quad (2.1)$$

C : est la durée d'impulsion (C=0.65s) ;

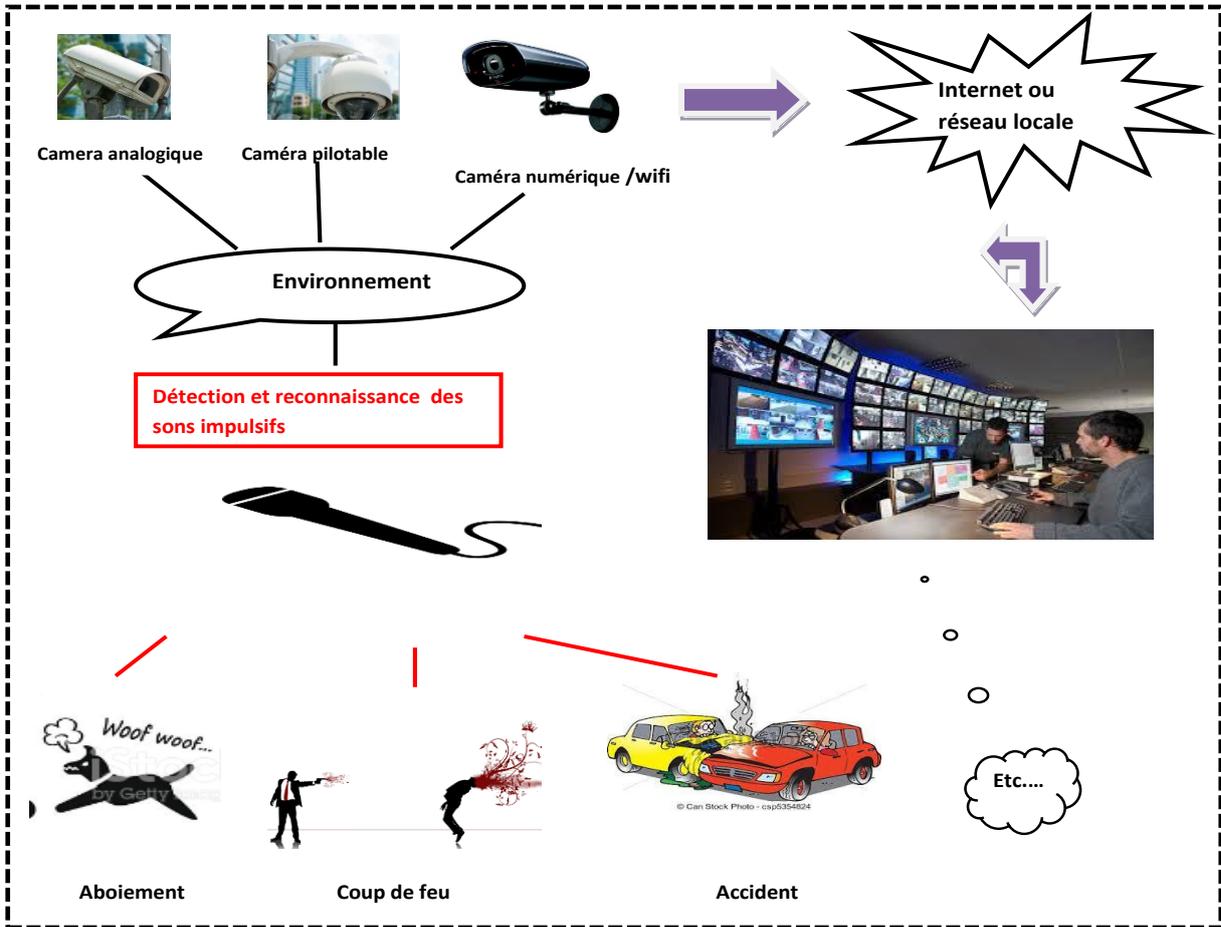


Figure 2.1: Système de surveillance exploitant la modalité audio

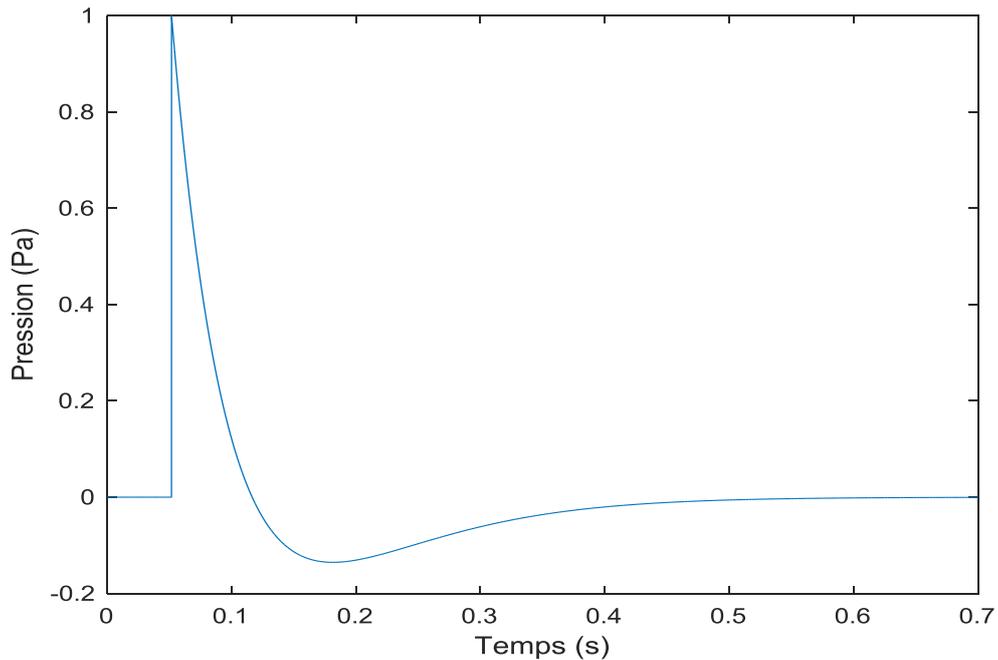


Figure 2.2: Onde de Friedlander

Nous citons les exemples suivants des sons impulsifs :

- Coups de feu ;
- Bris de verre ;
- Applaudissements ;
- Aboiements des chiens ;
- Cris des humains ;
- Alarmes.

2.5 Détection et reconnaissance des sons impulsifs

La détection des évènements impulsifs est l'opération de discrimination entre les SI et les autres sons de l'environnement. A l'issue de cette opération, la seule information fournie est qu'un évènement spécial (souvent dangereux) s'est produit suite à une augmentation soudaine de l'énergie [7].

Aucun autre détail, sur le type de son produit, n'est fourni. L'opération de reconnaissance de l'évènement acoustique consiste à identifier le type exact du son produit [7]. La Figure 3 donne un aperçu sur la détection et la reconnaissance des SI.

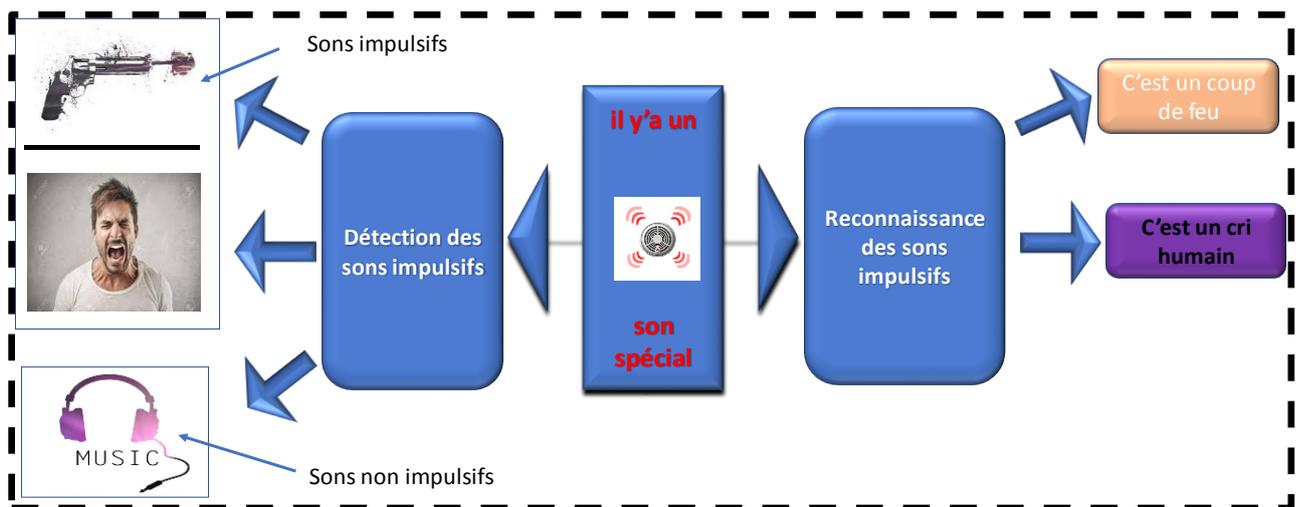


Figure 2.3: Système de détection et de reconnaissance des sons impulsifs

L'opération de détection des SI peut être considérée comme une étape de prétraitement avant de reconnaître le type exact du son produit. Il opère une présélection des signaux candidats à identifier (reconnaître) et empêche le processus de reconnaissance d'être constamment activé. Cette étape est très importante en raison de la complexité des algorithmes de reconnaissance utilisés dans la pratique [7].

2.6 Conclusion

Dans ce chapitre, nous nous sommes focalisés sur les systèmes de surveillance. Nous avons présenté au cœur de ce chapitre leur évolution. Nous avons trouvé que la troisième génération des systèmes de surveillance utilise plusieurs modalités physiques. Nous nous sommes focalisés sur la modalité audio vu que nous voulons concevoir un système de détection des sons impulsifs à base de l'audio.

Dans le prochain chapitre, nous nous intéressons particulièrement aux méthodes de détection des sons impulsifs ainsi qu'à la solution adoptée dans notre projet de fin d'étude.

Chapitre 3 : Détection des sons impulsifs

3.1 Introduction

La détection des SI est une étape très importante pour la reconnaissance des sons de l'environnement.

Dans ce chapitre, nous présentons trois sections importantes : (i) un état de l'art sur les méthodes de détection, (ii) une description de la solution que nous avons adoptée et (iii) les critères de son évaluation. Nous terminons ce chapitre par une conclusion.

3.2 Etat de l'art

La détection des SI est un domaine de recherche actif en traitement du signal et vise plusieurs champs d'applications.

Pour certaines problématiques, il est nécessaire de détecter les événements impulsifs en temps réel. Nous citons les exemples suivants :

- i) Système de surveillance de troisième génération [8] ;
- ii) Systèmes de détection des coups de feu [9] ;
- iii) Systèmes d'alertes pour les personnes âgées ayant des troubles d'audition [7] ;

Dans ce genre d'applications, les algorithmes et les méthodes de détection employées doivent être simples et non-complexes.

Pour d'autres problématiques, la détection en temps réel des événements impulsifs n'est pas une nécessité primordiale. A cet effet, la complexité des méthodes adoptées n'est pas un critère de choix contrairement à la première catégorie de méthodes [10].

Le problème de détection peut être abordé de deux manières différentes [11] :

- Détection par seuillage ;
- Détection par classification.

3.2.1 Détection par seuillage

La détection par seuillage consiste à comparer une caractéristique du signal audio avec un seuil. Cette approche est moins exigeante en termes de complexité de calcul que celle par

classification. Les méthodes par seuillage sont inspirées des approches de détection de l'activité de la voix utilisées en codage de la parole. Une simple approche à base du niveau d'intensité du signal pourrait être exploitée pour détecter les événements impulsifs. L'inconvénient majeur de cette mesure est qu'elle est dépendante du rapport signal sur bruit (RSB). Il est impossible dans ce cas de détecter les événements acoustiques dont le RSB est inférieur au niveau du bruit de fond [11].

Nous citons par la suite d'autres approches plus sophistiquées :

- R. Levorato [8] a présenté une méthode de détection et reconnaissance des SI basée sur plusieurs opérateurs et modèles statistiques de type GMMs (*Gaussian Mixture Models*). En fait, la détection proprement dite des événements impulsifs est basée sur un opérateur énergétique nommé *Teager Energy Operator* (TEO). L'inconvénient majeur de cette méthode est qu'elle souffre d'un taux élevé de faux positifs. D'après les auteurs de cet article, ces erreurs ont été corrigées dans l'étape de reconnaissance des SI.
- A. Defaux [7] a proposé une méthode de détection des SI basée sur la variance des séquences d'énergie. Cette méthode a été conçue pour fonctionner en temps réel et dans des conditions bruitées.
- A. Talal et al. [9] ont proposé une approche d'identification d'un type bien spécifique de sons impulsifs, il s'agit des coups de feu. La méthode proposée est basée sur deux étapes de traitement, la détection des événements acoustiques et la reconnaissance. L'étape de détection utilise un formalisme probabilistique à base d'une distribution Chi-square. La reconnaissance est basée sur les modèles SVMs (Support Vector Machines). Les résultats ont été satisfaisantes.

3.2.2 Détection par classification

De nombreuses solutions dans l'état de l'art n'utilisent pas d'algorithmes de détection en soi, mais exploitent des méthodes de reconnaissance de formes pour détecter les SI.

Les méthodes de reconnaissance analysent les données audio en temps réel et indiquent si un SI existe ou non.

Cette approche est basée sur un vecteur de caractéristiques dont le nombre peut atteindre une centaine de valeurs. Un nombre élevé d'attributs assure théoriquement une bonne

précision, cependant, les méthodes de reconnaissance utilisées peuvent être susceptibles de fausses alertes. Un autre inconvénient de cette approche est sa complexité algorithmique élevée. Plusieurs méthodes de reconnaissance peuvent être exploitées, nous citons les GMMs, les HMMs (Hidden Markov Models) et les SVMs.

Nous présentons par la suite une solution de détection des SI à bases des méthodes de reconnaissance.

- A. Rabaoui et al. [11] ont exploité les modèles SVMs à une seule classe pour détecter et classifier des évènements audio. Plusieurs catégories de sons ont été utilisé. Nous citons : (i) les coups de feu, (ii) les bris de verres, (iii) les explosions et autres. L'étape de détection des sons est basée sur une approche combinée entre une famille de fonctions dites « exponentielles » avec un modèle SVM à une seule classe. Cette solution vise à approximer le rapport de vrais semblance généralisé. L'étape de classification des sons est basée aussi sur un formalisme à base des SVMs.

3.3 Critères de choix des méthodes de détection des sons impulsifs

Notre objectif dans le cadre de ce projet de fin d'étude est de mettre en œuvre un détecteur des évènements acoustiques impulsifs en temps réel. La solution que nous utilisons devra être implémentée sur un PC ordinaire ayant de bonnes performances. Alors, nous nous sommes focalisées sur les méthodes de détection par seuillage.

Les critères de choix d'une méthode de détection efficace sont :

- La complexité algorithmique ;
- L'efficacité de détection (taux d'erreurs) ;
- La robustesse aux bruits de l'environnement ;
- Le fonctionnement en temps réel.

Parmi les méthodes étudiées dans le cadre de ce projet, nous avons trouvé que la détection des SI à base de la variance des séquences d'énergie, proposée initialement par A. Dufaux [7], est celle qui répond mieux aux critères ciblés ci-dessus.

En fait, la complexité algorithmique de cette solution est beaucoup plus réduite par rapport aux autres méthodes décrites auparavant. De plus, elle est conçue pour détecter les évènements acoustiques en temps réel. D'après les résultats reportés par A. Dufaux [7], cette méthode fournie de bons résultats dans des conditions bruitées.

Néanmoins, nous avons trouvé que cette méthode souffre de plusieurs types d'erreurs liés à la sensibilité du capteur acoustique (microphone).

Nous décrivons dans ce qui suit, la méthode de détection ainsi que la solution que nous proposons pour faire face aux problèmes de sensibilité souligné ci-dessus.

3.4 Détection des sons impulsifs à base de la variance des séquences d'énergies

Cette méthode a été initialement proposée par A. Dufaux [7]. Elle utilise un processus de détection basé sur l'évolution de la variance des séquences des énergies à courts termes du signal audio.

Elle est conçue pour détecter des événements acoustiques impulsifs en temps réel.

Néanmoins, les performances de cette méthode dépendent de deux contraintes importantes :

- (i) La sélection des paramètres de la méthode ;
- (ii) Le niveau de la pression acoustique capté par le dispositif d'acquisition (Microphone).

Pour défaire la première contrainte, on doit choisir des paramètres algorithmiques assurant une complexité réduite et une bonne exactitude de détection (un nombre d'opération réduit et un bon taux de détection). Selon la plateforme d'implémentation de la méthode (Sur Pc, Téléphone portable, FPGA ou carte DSP), on cherche à trouver le jeu de paramètres qui garantira le bon fonctionnement de la méthode.

Le niveau de la pression acoustique des événements sonores est un paramètre très important. La sensibilité du capteur est le seul facteur qui détermine la qualité et l'allure de l'onde acoustique capté par le microphone. L'efficacité de détection des événements impulsifs est principalement liée à la qualité de l'onde acoustique du signal audio.

Dans cette section, nous donnons une description détaillée de la méthode de détection. Nous présentons les différents paramètres de cette méthode ainsi que les principales étapes de détection.

Par la suite, nous discutons les critères pratiques d'optimisation de ces paramètres. Nous proposons aussi une approche de réglage de la sensibilité du capteur en vue de l'obtention de bonnes performances de détection.

3.4.1 Description de la méthode

L'estimation de l'énergie $e(k)$ d'un signal numérique est donnée comme suit :

$$e(k) = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n + kN) \quad k = 0, 1, \dots, \dots, +\infty \quad (3.1)$$

$x(n)$: est un signal audio échantillonné à une fréquence F_e ;

k : est l'indice des blocs ;

n : est l'indice des échantillons ;

$e(k)$: est l'énergie du k^{eme} bloc calculée à partir du signal $x(n)$;

N : représente la longueur du k^{eme} bloc.

Il est nécessaire de tenir compte des propriétés temporelles des sons impulsifs lors du choix de la valeur de N .

Une faible valeur de N engendre trop de détails inutiles dans la séquence d'énergie $e(k)$, perturbant parfois le processus de détection.

D'autre part, si la valeur de N est trop large, certaines difficultés pourraient survenir lors de la détection d'impulsions fines, surtout en cas de bruit de fond important.

La méthode de détection décrite dans ce chapitre est subdivisée en deux phases importantes :

- a) Fenêtrage et normalisation de la séquence d'énergie,
- b) Calcul de la variance des fenêtres normalisées.

3.4.1.1 Fenêtrage et normalisation de la séquence d'énergie

A chaque nouveau bloc k , une nouvelle valeur de l'énergie $e(k)$ est obtenue. Afin de conserver les L valeurs récentes de $e(k)$, la séquence d'énergie est sauvegardée dans une fenêtre dynamique $e_{win}(j/k)$ tel que : $j = 0 \dots L - 1$, et $k = 0 \dots +\infty$.

Selon les valeurs de « k », on distingue deux régimes :

- (i) Régime transitoire ;
- (ii) Régime permanent.

i) Régime transitoire

Le régime transitoire a pour but d'initialiser les composantes de la séquence $e_{win}(j/k)$ pour les valeurs de « $k < L$ ».

$$e_{win}(j) = e(i) \tag{3.2}$$

tel que:
 $i, j = 0, \dots, k - 1$

Durant ce régime, la variation des indices « i et j » est effectuée d'une manière simultanée. L'exemple suivant donne un aperçu sur le calcul des énergies.

Exemple :

$$\begin{aligned} e_{win}(0) &= e(0) \\ e_{win}(1) &= e(1) \\ e_{win}(2) &= e(2) \\ &\cdot \\ &\cdot \\ &\cdot \\ e_{win}(L - 1) &= e(L - 1) \end{aligned}$$

ii) Régime permanent

Le régime permanent a pour but de mettre à jours la séquence $e_{win}(j/k)$ pour les valeurs de $k \geq L$ et de décider par la suite sur la nature de la séquence captée (Impulsif ou non). C'est le régime dynamique.

Pour chaque nouvelle valeur de k , les L composantes récentes de la séquence $e_{win}(j/k)$ sont calculés comme suit :

$$e_{win}(j)^1 = e(i) \tag{3.3}$$

tel que:
 $i = k - L + 1, \dots, k$
 $j = 0, \dots, L - 1$

¹ L'indice k a été omis dans cette équation pour une simplicité de représentation.

L'indice « j » est un entier statique, il prend toujours des valeurs entre 0 et L-1. Cependant, l'indice « i » est un entier dynamique vu qu'il dépend principalement de la variation de « k ». La variation des indices « i et j » est effectuée d'une manière simultanée.

Nous donnons par la suite des exemples illustratifs de l'évolution de la séquence d'énergie en fonction des valeurs de « k » :

❖ **k = L :**

$$\begin{aligned} i &= 1, \dots, L \\ j &= 0, \dots, L - 1 \end{aligned}$$

$$\begin{aligned} e_{win}(0) &= e(1) \\ e_{win}(1) &= e(2) \\ &\vdots \\ e_{win}(L - 1) &= e(L) \end{aligned}$$

❖ **k = L + 1 :**

$$\begin{aligned} i &= 2, \dots, L + 1 \\ j &= 0, \dots, L - 1 \end{aligned}$$

$$\begin{aligned} e_{win}(0) &= e(2) \\ e_{win}(1) &= e(3) \\ &\vdots \\ e_{win}(L - 1) &= e(L + 1) \end{aligned}$$

❖ **k = L + 5 :**

$$\begin{aligned} i &= 6, \dots, L + 5 \\ j &= 0, \dots, L - 1 \end{aligned}$$

$$\begin{aligned} e_{win}(0) &= e(5) \\ e_{win}(1) &= e(6) \\ &\vdots \\ e_{win}(L - 1) &= e(L + 5) \end{aligned}$$

Afin de standardiser la dynamique des séquences d'énergie, le vecteur $e_{win}(j)$ doit être normalisé et transformé dans une plage de valeurs comprises entre 0 et 1 :

$$e_{norm}(j) = \frac{e_{win}(j) - \min_j(e_{win}(j))}{\max_j(e_{win}(j)) - \min_j(e_{win}(j))} \quad j = 0 \dots L - 1 \quad (3.4)$$

3.4.1.2 Calcul de la variance des fenêtres normalisées

Une fois la normalisation de la séquence d'énergie est accomplie, l'étape suivante consiste à calculer la variance de $e_{norm}(j)$ en excluant la dernière composante :

$$\text{var}(k) = \frac{1}{L-1} \sum_{j=0}^{L-2} [e_{\text{norm}}(j) - \bar{e}_{\text{norm}}(k)]^2 \quad (3.5)$$

$\bar{e}_{\text{norm}}(k)$: est la valeur moyenne du vecteur $e_{\text{norm}}(j)$;

En fait, la normalisation de la séquence d'énergie est la clé de la méthode de détection. Lorsque le signal est plutôt stable, avec un niveau constant ou légèrement variable, toutes les composantes de $e_{\text{norm}}(j)$ sont réparties entre 0 et 1, ce qui génère une variance assez importante. Cependant, une fois qu'une impulsion significative survient, la dernière valeur de la fenêtre atteint 1, tandis que les valeurs précédentes varient aux alentours de 0. La valeur de la variance chute dans ce cas.

L'analyse de la variance de la séquence d'énergie est un critère très efficace pour détecter une augmentation soudaine du niveau du signal. Lorsque la variance est inférieure à un seuil donné « Th », un événement impulsif est détecté.

3.4.2 Adaptation de la méthode aux cas pratiques

3.4.2.1 Positionnement du problème

Comme nous l'avons souligné ci-dessus, deux types de problèmes doivent être résolus pour assurer un bon fonctionnement de la méthode de A. Dufaux [7] :

- L'ajustement des paramètres ;
- Le niveau de la pression acoustique capté par le microphone.

Les paramètres de la méthode qui doivent être ajustés pour une efficacité de détection sont :

- ✓ La fréquence d'échantillonnage « Fe » ;
- ✓ La durée (taille) de la fenêtre d'analyse « N » ;
- ✓ La longueur de la séquence d'énergie « L » ;
- ✓ Le seuil de décision « Th ».

A. Dufaux [7] a présenté une approche de minimisation des erreurs produites en utilisant plusieurs classes de sons impulsifs. Les catégories d'erreurs qui ont été utilisées dans son étude sont :

- (i) Le taux de faux positifs (TFP) ;
- (ii) Le taux de faux négatifs (TFN).

Pour une F_e de 44100Hz, il a trouvé que le meilleur jeu de paramètres assurant une bonne détection est donné comme suit : $N= 3969$ (une durée de 90ms), $L= 30$ et $Th=0.1$. Le choix d'une valeur élevée de F_e est nécessaire si on veut reconnaître le type exact du son impulsif (voir le chapitre 2). Dans le cas où la reconnaissance n'est pas envisagée, une valeur inférieure à 44.1kHz est souhaitée.

Une fois l'optimisation de ces paramètres est effectuée, la méthode doit théoriquement fonctionner dans des situations réelles.

Les questions importantes qui se sont posées lors de l'implémentation de la méthode sont les suivantes :

- Es ce que le choix d'un seuil de décision « Th » adéquat est suffisant pour détecter tous les types d'évènements impulsifs en cas de variation du bruit de l'environnement ?
- Es ce que la détection des évènements impulsifs est évidente quel que soit le type et la qualité du capteur utilisé ?

Les deux réponses à ces questions sont négatives.

Nous avons trouvé que la sensibilité du capteur (microphone) est un paramètre important et très influant sur l'efficacité de la détection des évènements impulsifs indépendamment de l'étape de sélection ou d'ajustement des paramètres. En fait, la méthode décrite ci-dessus se base sur la détection des évènements acoustiques au moment où un changement critique de la variance de $e_{norm}(j)$ est signalé.

Pratiquement, les sons impulsifs peuvent se produire à cause de sources et actions divers.

Nous citons les exemples suivants :

- (i) Les applaudissements ;
- (ii) Un objet qui tombe sur terre ;
- (iii) Les coups de feu ;
- (iv) Une Fermeture brusque d'une porte ;
- (v) Une alarme de voiture ;
- (vi) Une articulation correcte des consonnes plosives lors d'un discours continu ;
- (vii) Les claquements de doigts ;
- (viii) Autres.

Ces actions génèrent des ondes acoustiques de différentes formes et niveaux d'intensités acoustiques. La seule factrice commune entre ces évènements est le caractère impulsif.

De plus, le niveau de la pression acoustique capté par le microphone varie selon la distance entre le microphone et le lieu où l'évènement acoustique s'est produit.

Par exemple, la pression acoustique d'un son impulsif généré suite à une fermeture d'une porte située à une distance proche de l'emplacement du microphone est généralement plus intense que celle d'un coup de feu lointain. De plus, le rapport signal sur bruit peut influencer la détection de ces évènements acoustiques.

L'estimation du niveau de la pression acoustique réel peut jouer un rôle important dans le réglage de la sensibilité du capteur si on cherche à détecter des évènements acoustiques dont le niveau d'intensité sonore dépasse un certain seuil.

A cause de ces perturbations, la détection des évènements impulsifs en utilisant cette méthode souffre de problème de faux positifs même si l'optimisation des paramètres de détection est effectuée préalablement.

3.4.2.2 Solutions envisagées

a) Paramètres de la méthode

Vu que nous ciblons l'implémentation d'un détecteur de sons impulsifs sur un Pc portable, nous allons déduire dans ce projet un autre jeu de paramètres qui assure de bonnes performances de détection tout en imposant une fréquence d'échantillonnage réduite. Nous cherchons aussi à trouver les valeurs minimales de N et L qui assurent une détection correcte ainsi que de bonnes performances de reconnaissance (nécessaire pour l'étage suivant). La procédure expérimentale adoptée dans notre étude sera présentée dans le chapitre suivant.

b) Niveau de la pression acoustique

Nous proposons d'ajouter un étage pour le réglage de la sensibilité du microphone en vue de résoudre les problèmes cités ci-dessus.

La solution proposée consiste à ajouter au signal d'entrée un bruit blanc $b(n)$ d'une moyenne μ et d'une variance σ^2 .

La densité probabilité du bruit est donnée comme suit :

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.6)$$

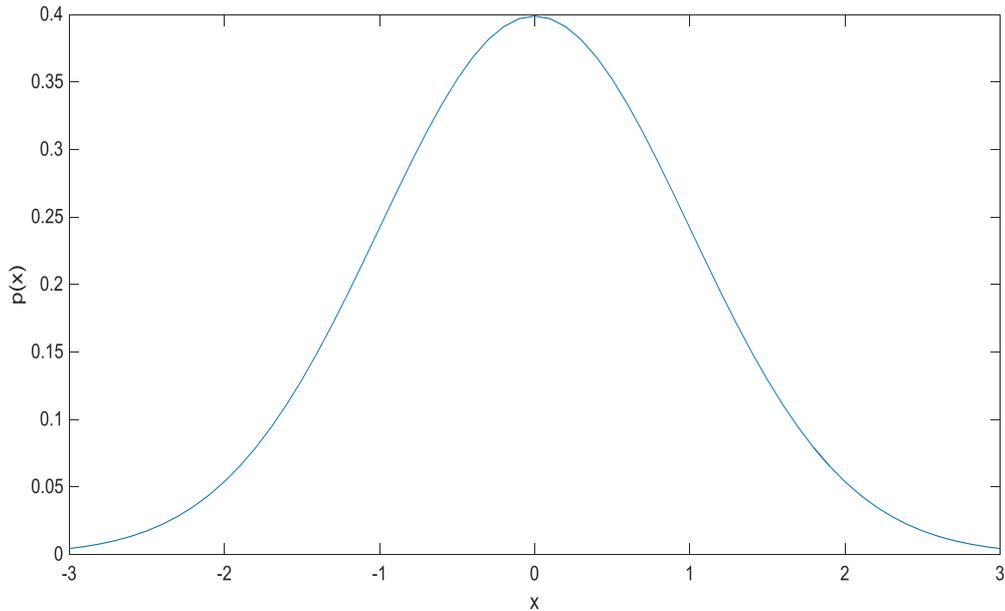


Figure 3.1: Densité de probabilité gaussienne

L'équation (3.1) doit être remplacée par l'équation suivante :

$$e(k) = \frac{1}{N} \sum_{n=0}^{N-1} (x^2(n + kN) + b(n)) \quad k = 0, 1, \dots, \dots, \dots, +\infty \quad (3.7)$$

La reste des équations est inchangée. De cette manière, la sensibilité de la détection est ajustée en manipulant la variance du signal $b(n)$. La moyenne du signal est fixée à zéro car elle n'influe pas sur l'efficacité de détection.

3.5 Métriques d'évaluation

Les performances des algorithmes de détection des SI sont évaluées en utilisant les métriques suivantes [11] :

- (i) Le taux de vrais positifs (TVP) ;
- (ii) Le taux de faux positifs (TFP) ;
- (iii) Le taux de faux négatifs (TFN) ;

- Taux de vrais positifs (TVP) : représente le rapport entre le nombre des événements correctement détectés (ECD) et le nombre de tous les événements présents²(TEP). Un

² Dans notre cas, un évènement acoustique correspond à un son impulsif.

événement est compté comme vrais positif, s'il est présent dans le signal de test et il a été détecté au bon moment par la méthode employée.

$$TVP = \frac{ECD}{TEP} \quad (3.8)$$

- Taux de faux positifs (TFP) : également appelé le taux de faux alertes ; il représente le rapport entre le nombre des événements détectés d'une manière incorrecte (EDI) et le nombre de tous les événements détectés (TED). D'une autre manière, cette mesure correspond aux événements qui ne sont pas présents dans le signal de test mais qui sont détectés par la méthode employée.

$$TFP = \frac{EDI}{TED} \quad (3.9)$$

- Taux de faux négatifs (TFN) : également appelé le taux de faux rejet ; il représente le rapport entre le nombre des événements manqués (EDI) et le nombre de tous les événements présents (TEP).

$$TFN = 1 - TVP = \frac{EDI}{TEP} \quad (3.10)$$

Ces mesures sont mutuellement reliées par l'une des deux représentations [11]:

- Detection Error Tradeoff (DET) (TFN vs TFP);
- Receiver Operating Characteristics (ROC) (TVP vs TFP);

Ces représentations servent à définir le point de fonctionnement de la méthode de détection pour lequel la valeur du seuil de décision est optimale.

3.6 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art concis sur le problème de la détection des sons impulsifs. Par la suite nous avons exposé une méthode de détection basée sur l'évolution de la variance des séquences d'énergies.

La dernière partie a été consacrée aux métriques d'évaluation des méthodes de détections des événements acoustiques. Dans le chapitre suivant, nous présentons les résultats expérimentaux ainsi que leurs interprétations.

Chapitre 4 : Résultats expérimentaux

4.1 Introduction

Dans ce chapitre, nous exposerons les résultats obtenus par l'application de la solution proposée pour détecter les sons impulsifs en temps réel.

4.2 Logiciels et matériel de développement

4.2.1 Environnement MATLAB

C'est en exploitant les fonctionnalités de l'Environnement de Développement Intégré (EDI) MATLAB que les principaux travaux de ce projet ont été réalisés (Figure 4.1).

C'est la version 2015 qui a été utilisée (Matlab R2015b). Plusieurs fonctions dédiées au traitement de signal et aux statistiques sont disponibles [13] .

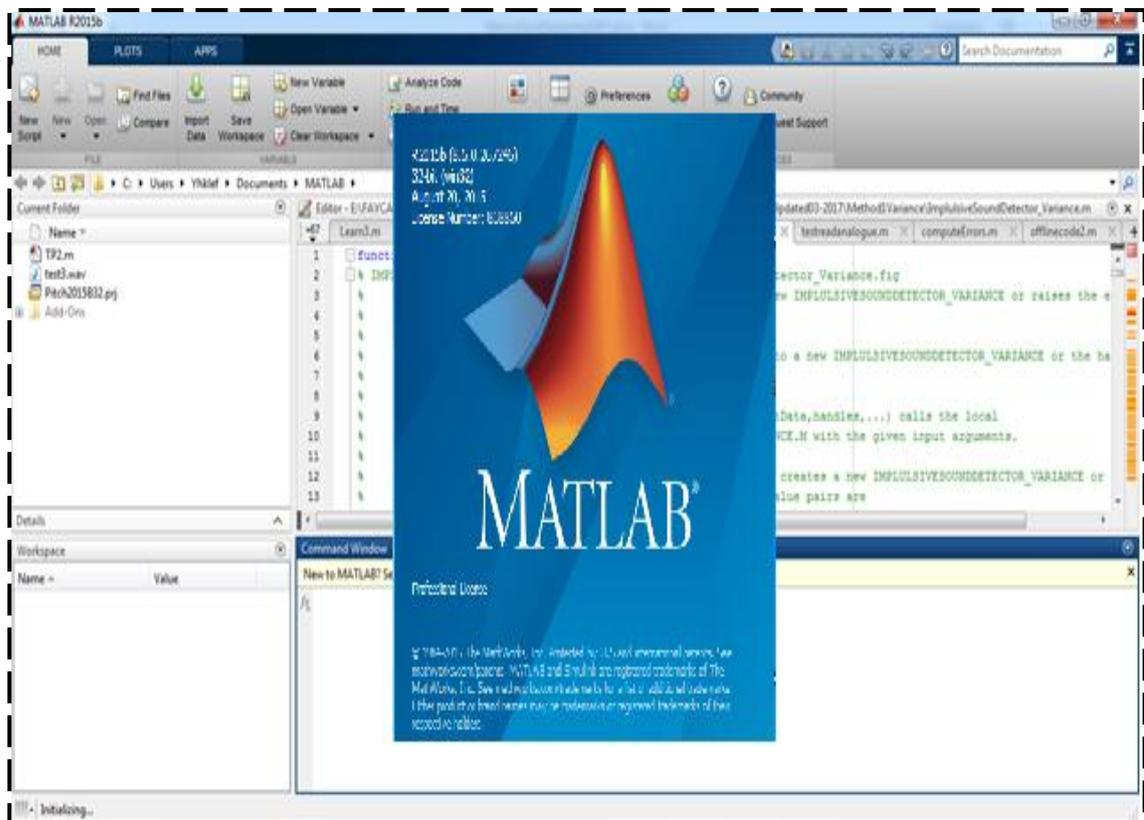


Figure 4.1 : MATLAB R2015b

4.2.2 GoldWave

GoldWave est un logiciel sous Windows qui permet la lecture et l'enregistrement des sons. L'enregistrement peut être effectué à partir du microphone de l'ordinateur ou bien l'entrée de la carte son (Line-in). GoldWave permet la lecture et l'écriture de fichiers audio sous formats divers (.WAV, .MP3, .Wma, .Aiff, etc...).

Il est doté de plusieurs effets acoustiques comme Doppler, Dynamics, Echo, Filter, et autres. Il permet aussi d'éditer, copier, coller et convertir les fichiers audio [14].



Figure 4.2 : GoldWave

4.2.3 Wo Mic

WO Mic est une application informatique composée de deux plateformes différentes. La première plateforme est une application sous Windows. Elle est utilisée sur ordinateur comme client (Wo Mic Client). La deuxième plateforme est une application Android. Elle est utilisée sur un smartphone comme serveur (Wo Mic server).

Le rôle principal de Wo Mic est de transformer un smartphone à un microphone pour n'importe quel ordinateur.

Il existe trois modes de connexion entre le smartphone et l'ordinateur : (i) Wi-Fi (Wireless Fidelity), (ii) Bluetooth et (iii) USB. Dans notre cas, nous avons utilisé le mode Wi-Fi [15].

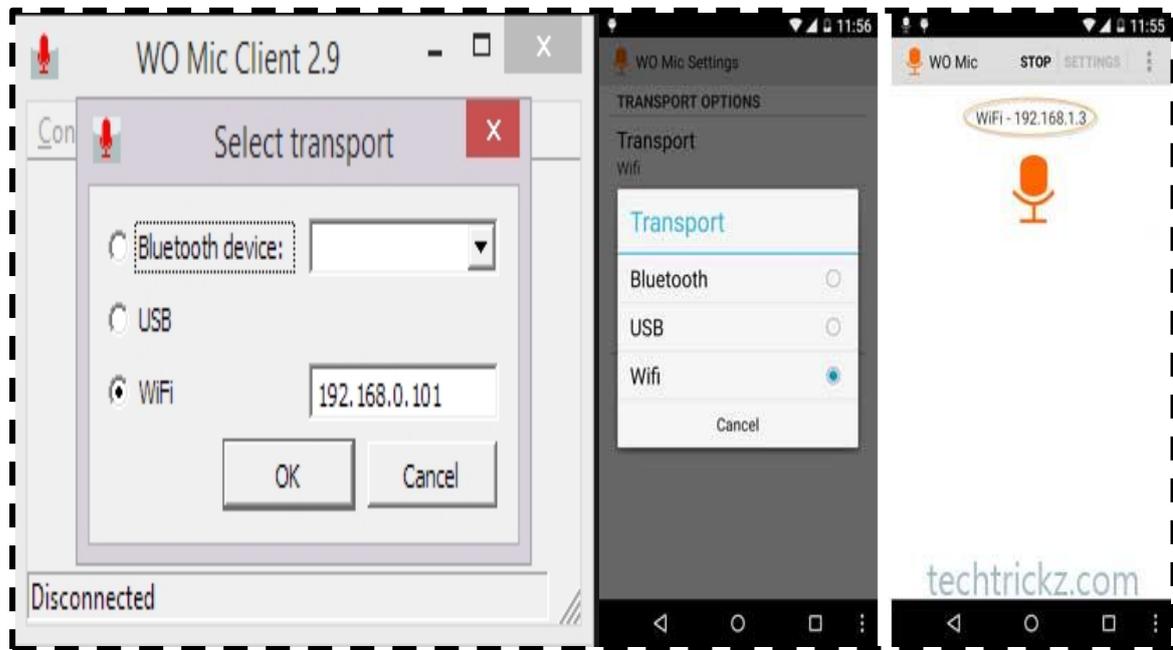


Figure 4.3 : Wo Mic Client et serveur (PC et Smartphone)

4.2.4 MyPublicWiFi

MyPublicWiFi est une application informatique sous Windows qui a pour objective de changer un ordinateur en point d'accès WI-FI. Plusieurs appareils intelligents (smartphones, tablettes, smartwatches, phabletes, laptops, desktop) peuvent être connectés au même point grâce à cette application informatique. MyPublicWiFi possède aussi d'autres fonctionnalités plus évoluées (partage de fichiers). Ces fonctionnalités n'ont pas été utilisées dans notre projet [16].



Figure 4.4 : MyPublicWiFi

4.2.5 Sonomètre

Le sonomètre est un dispositif de mesure du niveau de bruit acoustique de l'environnement. L'unité de mesure est le dB SPL (Decibel Sound Pressure Level). Il est nommé aussi décibel mètre [17]. Dans notre projet, nous avons utilisé le KMOON GM1352.



Figure 4.5 : Sonomètre KMOON GM1352

4.2.6 Microphone MEMS

Les microphones MEMS (Micro-Electrical-Mechanical Systems) apparaissent dans les produits de téléphones intelligents (smartphones) et de tablettes. Il y a 2 types de microphones MEMS : (i) analogique et (ii) digital. Le niveau des décibels maximaux captés par ces dispositifs est environ 94 dB SPL [18]. Nous avons utilisé dans notre projet un smartphone de type SAMSUNG Galaxy pour l'acquisition des données acoustiques. La procédure d'acquisition est expliquée dans les sections suivantes.

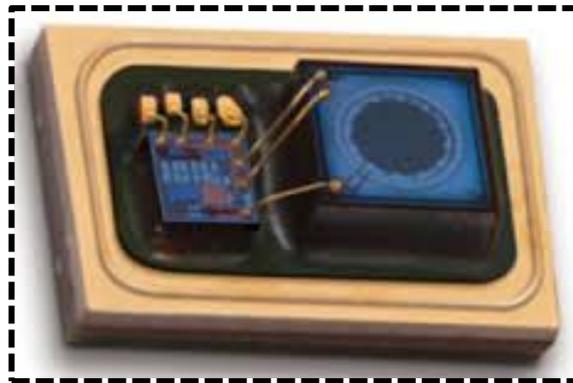


Figure 4.6 : Microphone MEMS

4.3 Corpus de tests

Le système de détection des sons impulsifs a été évalué sur un corpus sonore de 200 sons impulsifs. Nous avons choisi les armes à feu de type pistoles comme signaux de test. Cet corpus a été téléchargé à partir du site sounddogs [19].

Sounddogs est une base de données audio contenant environ 694 395 fichiers englobant toutes les catégories de sons (sons des guitares, la nature, fermeture des portes, explosions, armes à feu, les animaux, les cris humains et autres). Les sons peuvent être mono ou stéréo. Leur fréquence d'échantillonnage est de 11025 Hz. Les signaux sont enregistrés à 16 bits. Le téléchargement des sons est gratuit.

La seule contrainte que nous avons trouvée avec les sons de cette base de données est liée à leur niveau de volume variable (certains sons possèdent un niveau de volume faible, d'autres moyen et certains d'entre eux possèdent un volume fort).

Pour cela, nous avons ajusté la dynamique des 200 sons de notre corpus en utilisant l'option Dynamic Range de Goldwave pour obtenir des amplitudes comparables.

4.4 Résultats expérimentaux

4.4.1 Lieu de test

Nous avons effectué nos expérimentations dans une salle fermée d'une superficie de 22 m². La salle est dotée d'un (i) climatiseur en marche, (ii) deux bureaux (immobiliers), (iii) cinq chaises et (iv) deux armoires. Le bruit de fond est à environ 46 dB. Il a été mesuré par le sonomètre KMOON.

4.4.2 Matériel utilisé

Nous avons utilisé dans notre expérimentation le matériel suivant :

- Haut-parleurs : source des sons impulsifs,
- Pc de Bureau : pour la génération des sons émis via les haut-parleurs,
- Smartphone : joue le rôle d'un microphone sans fil,
- Pc-portable : contient l'application de notre projet,

4.4.3 Installation et connexion des software

Tout d'abord, à l'aide des logiciels Wo Mic et MyPublicWifi, nous avons converti notre smartphone en un microphone sans fil connecté à un ordinateur portable comme le montre la Figure 4.7.

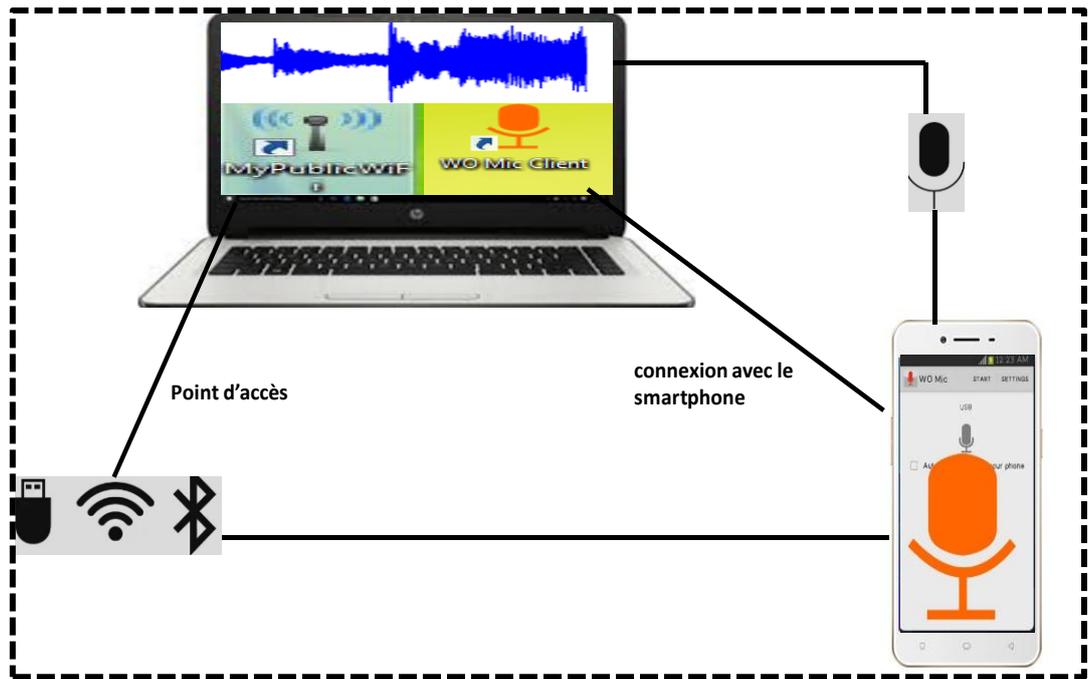


Figure 4.7 : Connexion entre le Pc portable et le Smartphone à l'aide de MyPubliWiFi et Wo Mic

- MyPubliWiFi : Etablissement du point d'accès (Pc Portable) + connexion WI-FI du smartphone avec le Pc portable,
- Wo Mic (client & server) : Conversion du smartphone à un microphone sans fil,

La disposition du matériel de test est schématisée sur la Figure 4.8. L'emplacement du smartphone est fixé sur l'une des deux armoires. Les haut-parleurs (source des sons impulsifs) sont posés sur le bureau à une distance approximative de quatre mètres par rapport au smartphone. Les sons sont générés en utilisant un ordinateur de bureau. La détection automatique des sons impulsifs (en utilisant la méthode reportée dans le chapitre précédent) a été implémentée sous MATLAB et cela en utilisant un laptop (Pc Portable).



Figure 4.8 : Disposition du matériel de test

4.4.4 Enregistrement du corpus : Déroulement de l'expérience

L'objectif de cette partie expérimentale est de créer un scénario d'évènements impulsifs à l'intérieur d'une salle fermée (In door). Ces événements seront utilisés par la suite pour la sélection des paramètres algorithmiques. Pour cela, nous avons utilisé l'ensemble de sons reportés dans la section 4.3 comme signaux de test (200 sons impulsifs). Les étapes de cette expérience sont données comme suit :

- Lecture des événements impulsifs : Cette étape a pour objectif de lire séquentiellement les fichiers audio de notre base de données en utilisant un ordinateur de bureau équipé de haut-parleurs de très bonne qualité. La lecture des sons est effectuée par une routine sous MATLAB. Un intervalle de silence d'une durée de trois secondes a été imposé entre chaque deux sons consécutifs. La disposition des sources sonores (haut-parleurs) est faite comme indiqué sur la figure 4.8.
- Enregistrement des événements impulsifs : L'enregistrement des événements impulsifs est achevé en se basant sur le microphone du smartphone. Goldwave est utilisé pour

l'enregistrement. La disposition du smartphone est faite comme indiqué sur la figure 4.8.

- Répétition : Les deux étapes précédentes ont été répétées pour trois niveaux acoustiques : (i) faible 71 dB(+/-4dB), (ii) moyen 82dB(+/-4dB), et (iii) fort 90 dB (+/-4dB). Les mesures moyennes ont été faite par un sonomètre.
- Prétraitement : La procédure expérimentale décrite précédemment est effectuée dans une salle au niveau de la division ASM du CDTA. Les événements acoustiques qui peuvent surgir à l'extérieur de la salle (des personnes qui marches dans le couloir, l'ouverture et la fermeture des portes des autres salles voisines, les discours, etc.) peuvent influencer sur la qualité et le contenu du corpus enregistré. La moindre irrégularité de la pression acoustique peut générer un évènement impulsif non attendu dans le scénario souhaité. Pour corriger des éventuelles situations, nous avons été obligés de faire une étape de prétraitement de notre corpus à l'aide du logiciel Goldwave.
- Extraction des instants exactes d'événements impulsifs : Cette étape a pour objectif d'identifier les marques (instants) de début de chaque son impulsif en secondes et cela pour chaque niveau d'intensité de notre expérience. Ces instants sont utilisés par la suite pour le calcul des erreurs de détection (voir le chapitre 3, section 3.5) (Figure 4.9).

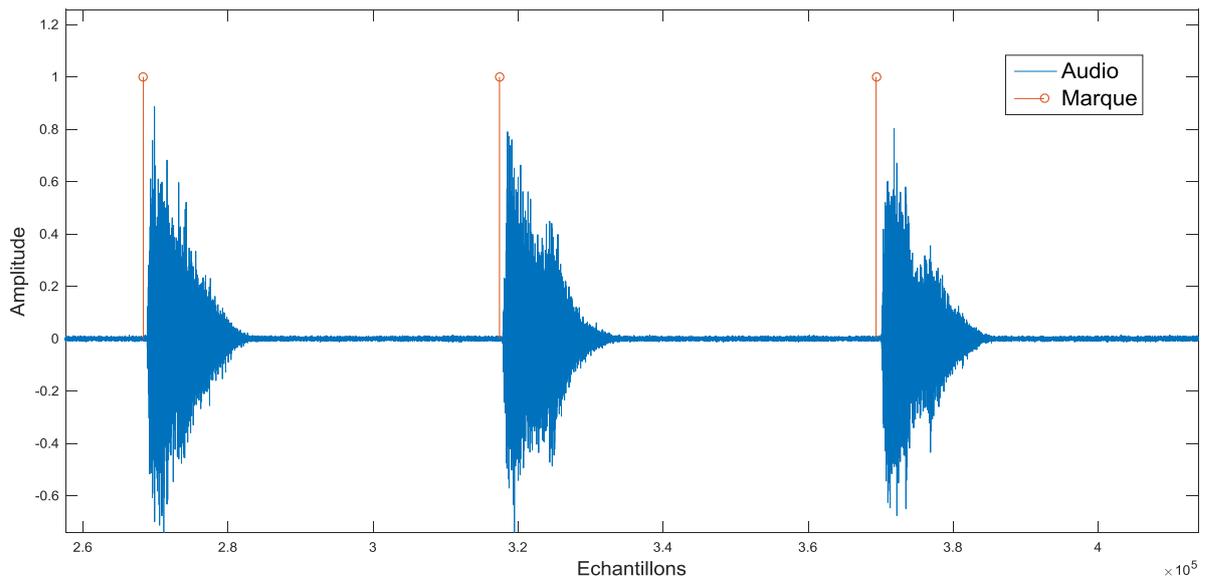


Figure 4.9 : Instants de début des sons impulsifs

4.4.5 Paramètres algorithmiques de la méthode de détection

Les paramètres de la méthode qui doivent être ajustés pour une efficacité de détection sont :

- ✓ La fréquence d'échantillonnage « F_e »,
- ✓ La durée (taille) de la fenêtre d'analyse « N »,
- ✓ La longueur de la séquence d'énergie « L »,
- ✓ Le seuil de décision « Th ».

- Fe : La fréquence d'échantillonnage de l'algorithme est fixée à 11025 Hz vu que les sons de la base de données téléchargée du site Sounddogs sont prédéfinis à cette valeur.
- N : Le choix de ce paramètre doit satisfaire deux contraintes essentielles : (i) sa valeur doit être assez faible pour réduire la complexité de la méthode et permettre une exécution rapide (Real Time) (ii) Dans le ou cette valeur est trop faible, trop de détails inutiles dans la séquence d'énergie peuvent surgir ; ainsi perturbant le processus de détection. Pour cela, et pour une Fe de 11025 Hz, le meilleur compromis a été trouvé pour N= 20ms.
- L : La longueur de la séquence d'énergie doit être choisie de telle sorte que la durée de l'évènement détecté soit assez large pour permettre l'obtention de bonnes performances de reconnaissance (l'étape qui vient en suite du processus de détection : voir le chapitre 2, section 2.5). Cette valeur est dépendante aussi de la durée de la fenêtre d'analyse. Une valeur de L=30 produira un signal impulsif d'une durée de 0.6 secondes (pour les paramètres fixés préalablement). Cette valeur est jugée suffisante pour une reconnaissance correcte.
- Le seuil « Th » : le choix du seuil de décision est très important pour un fonctionnement optimal de l'algorithme de détection des évènements impulsifs. D'après nos expérimentations, nous avons trouvé que la plage de variation de ce paramètre est comprise entre 10^{-5} et **0.25**. La valeur maximale représente la variance normalisée de la séquence d'énergie instantanée dans le cas où aucun évènement impulsif ne se produit. La valeur minimale représente le seuil de blocage pour lequel aucun son impulsif n'est détecté quel que soit son intensité. Le choix de « Th » dépend de deux critères importants :
 - (i) Le niveau d'intensité acoustique des évènements : sons impulsifs à niveau élevé, moyen ou faible (en dB SPL),
 - (ii) Les bruits de l'environnement : les sons perturbateurs de l'environnement (les discours, les sons de la nature, bruit des machines, et autres). Ces perturbations sont de nature différente. Selon la sensibilité du capteur utilisé, certains d'entre eux peuvent ressembler aux signaux impulsifs à faible variation d'énergie.

Le choix d'un seuil de décision optimal nécessite l'exploitation de données acoustiques de nature différente. Il est très difficile d'imaginer des scénarios qui englobent tous les évènements acoustiques décrits ci-dessus (sons impulsifs à des intensités variables avec des bruits d'environnement variables aussi).

Pour cela, nous avons adopté dans notre étude la solution suivante :

- (i) Trouver les seuils de décisions optimaux dans des conditions acoustiques à faible bruits d'environnement, et cela en utilisant des évènements impulsifs à trois niveaux d'intensités différents : faible, moyen et fort (voir la section 4.4.4),
- (ii) Chercher un seuil global de décision,
- (iii) Exploiter le facteur de sensibilité du capteur (chapitre 3, section 3.4.2.2) pour le réglage des performances de la méthode de détection des sons impulsifs par rapport aux conditions d'environnement de salle.

Les valeurs du seuil de décision utilisées dans notre expérience sont représentées dans la Figure 4.10 (67 niveaux différents). La méthode de détection est très sensible aux variations du seuil lorsque les valeurs sont inférieures ou égales à 0.1. Pour cela, une échelle logarithmique a été utilisée. Au-delà de 0.1, la précision n'est pas obligatoire. Une échelle linéaire est alors jugée suffisante.

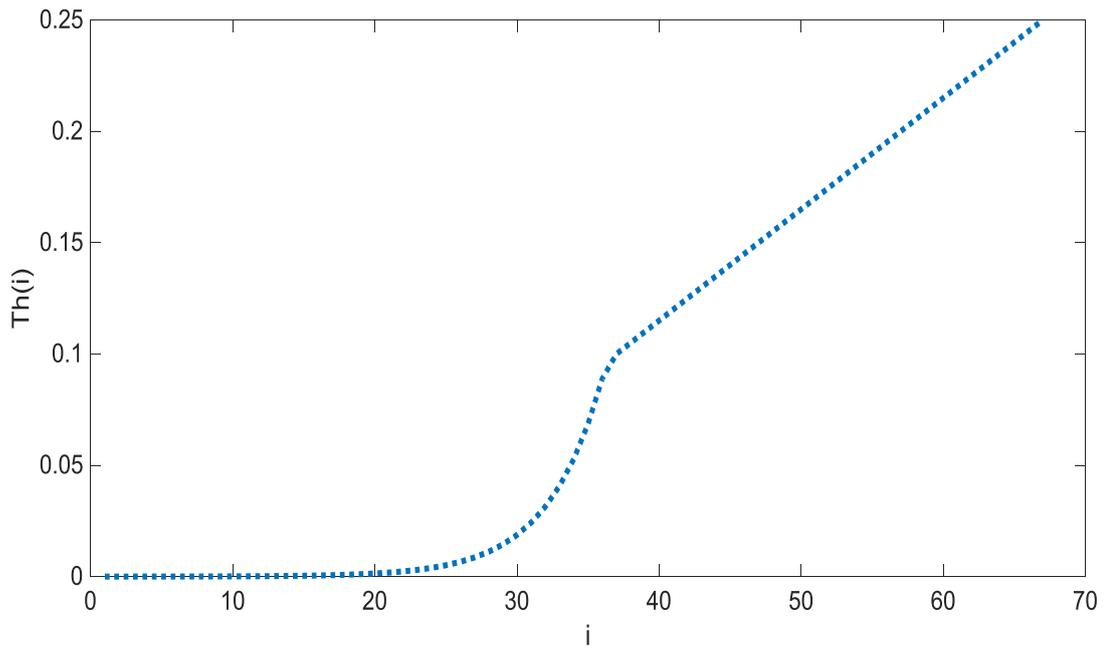


Figure 4.10 : Valeurs expérimentale du seuil de décision

Le choix de la valeur optimale du seuil est fait à l'aide des courbes ROC obtenues après le calcul des TVP et TFP pour toutes les valeurs de $Th(i) : i=1 :67$. Les résultats obtenus pour les trois niveaux de pressions (71dB, 82 dB et 90 dB) sont représentés respectivement dans les figures 4.11, 4.12 et 4.13.

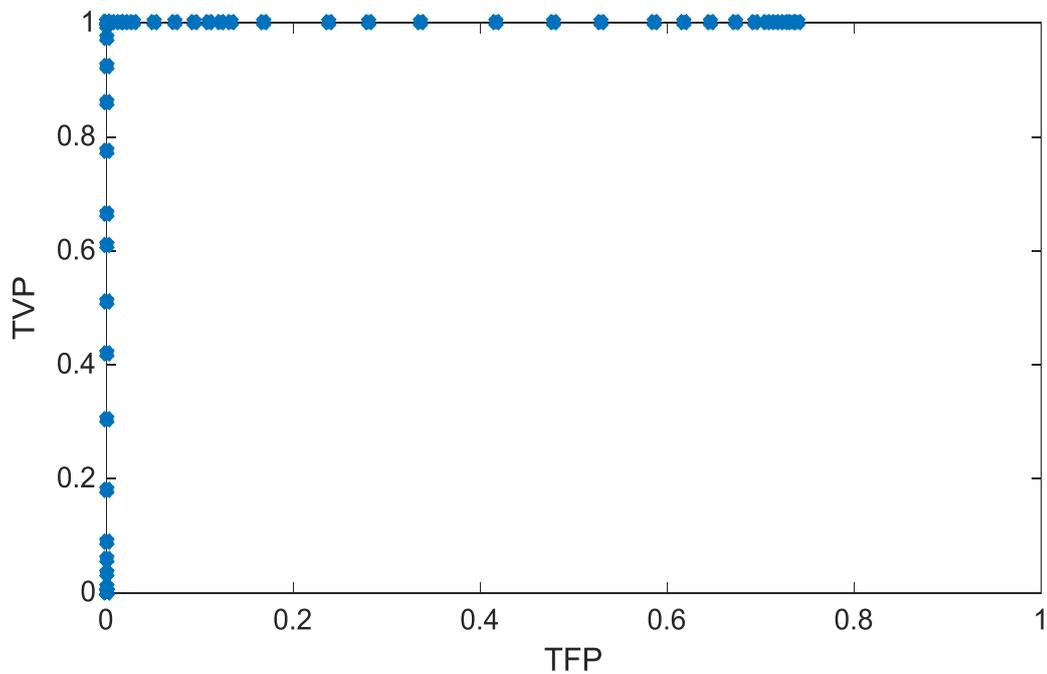


Figure 4.11 : Courbe ROC pour un niveau d'intensité de sons impulsifs de 71 dB SPL

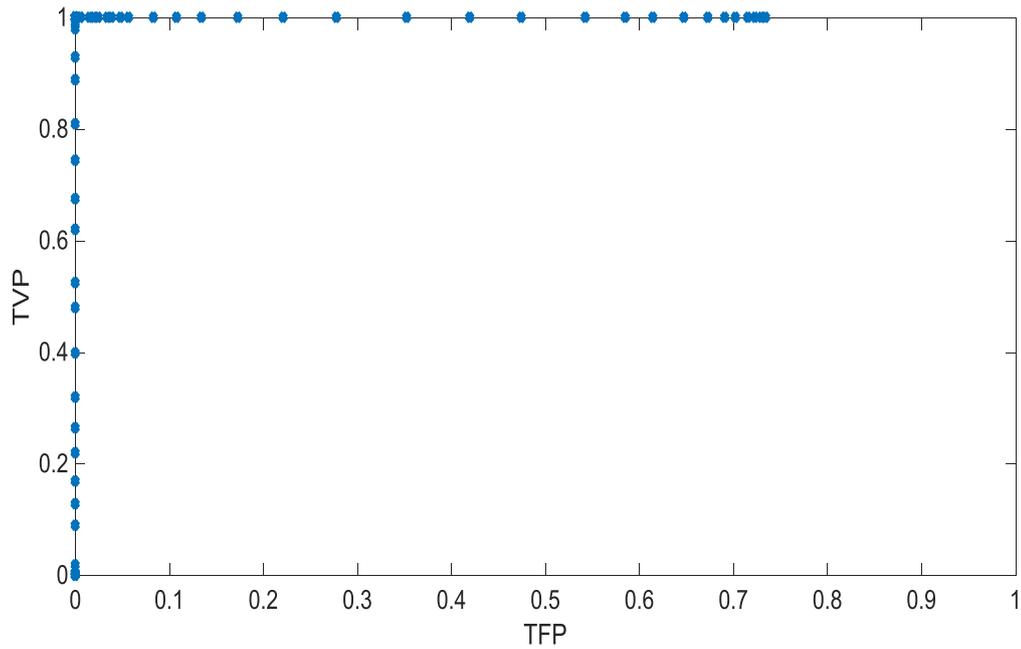


Figure 4.12 : Courbe ROC pour un niveau d'intensité de sons impulsifs de 82 dB SPL

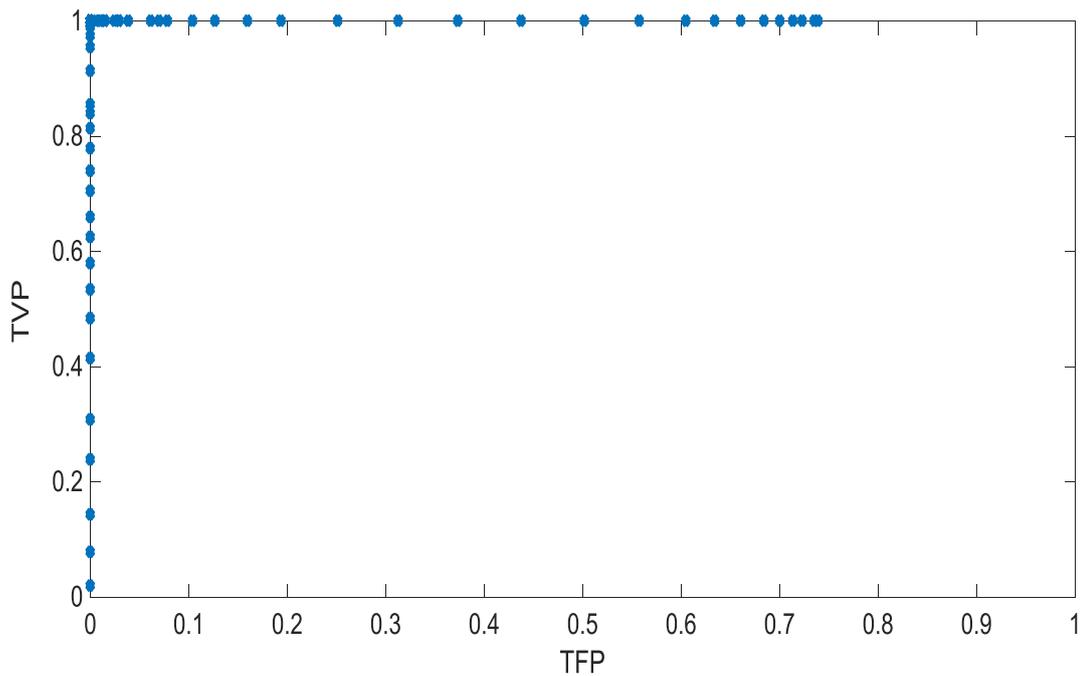


Figure 4.13 : Courbe ROC pour un niveau d'intensité de sons impulsifs de 90 dB SPL

Nous nous sommes basés sur une procédure de sélection des seuils pour lesquels les TVP sont maximisés et les TFP sont minimisés. Nous avons utilisé les seuils dont les TFP sont inférieurs ou égaux à 0.1. Les seuils qui assurent ce compromis sont :

- Niveau 71dB : les seuils varies de Th(29) à Th(40),
- Niveau 82dB : les seuils varies de Th(25) à Th(38),

- Niveau 90dB : les seuils varient de Th(25) à Th(36),

L'intersection des trois suites de seuils est représentée par l'ensemble S :

$$S = [\text{Th}(29) \text{ Th}(30) \text{ Th}(31) \text{ Th}(32) \text{ Th}(33) \text{ Th}(34) \text{ Th}(35) \text{ Th}(36)].$$

Le seuil optimal peut être obtenu en cherchant la médiane du vecteur S. Vu que S est pair, le seuil optimal est donné par :

$$\text{Th}_{\text{optimal}} = \frac{\text{Th}(32) + \text{Th}(33)}{2} \quad (\text{Dans notre expérimentation, cette valeur est de } 0.0361).$$

Il s'agit du seuil utilisé dans des situations réelles. Ce choix assure de bonnes performances de détection avec une grande sensibilité (les événements impulsifs de faibles niveaux d'intensités peuvent être détectés en utilisant cette combinaison de paramètres).

- **Réglage de la sensibilité de détection**

La sensibilité de détection dans des situations réelles est ajustée en choisissant un écart type adéquat du bruit additif de stabilisation (voir le chapitre 3, section 3.4.2.2). Certains événements acoustiques, à faible niveau d'intensité, peuvent être acoustiquement considérés comme sons impulsifs (sons de claquement de doigts dans des conditions de silence, son de fermeture d'une porte lointaine, etc.).

Néanmoins, et pour des applications de surveillance audio, ce type d'événements ne doit pas nécessairement être détecté par le système développé. Le rôle du bruit de masquage est de masquer ce genre d'événements impulsifs. Un exemple explicatif est donné dans les sections suivantes.

4.5 Application développée

Nous avons implémenté notre application en utilisant Matlab R2015b. Il s'agit d'un logiciel de surveillance audio en temps réel (online). Une fois un événement impulsif surgit, le logiciel de surveillance lance une signalisation visuelle et enregistre l'instant d'apparition de l'événement. Le système enregistre aussi le nombre des événements impulsifs détectés.

Il offre la possibilité d'ajuster les paramètres algorithmiques de la méthode développée. La figure suivante représente un exemple illustratif de l'application développée.

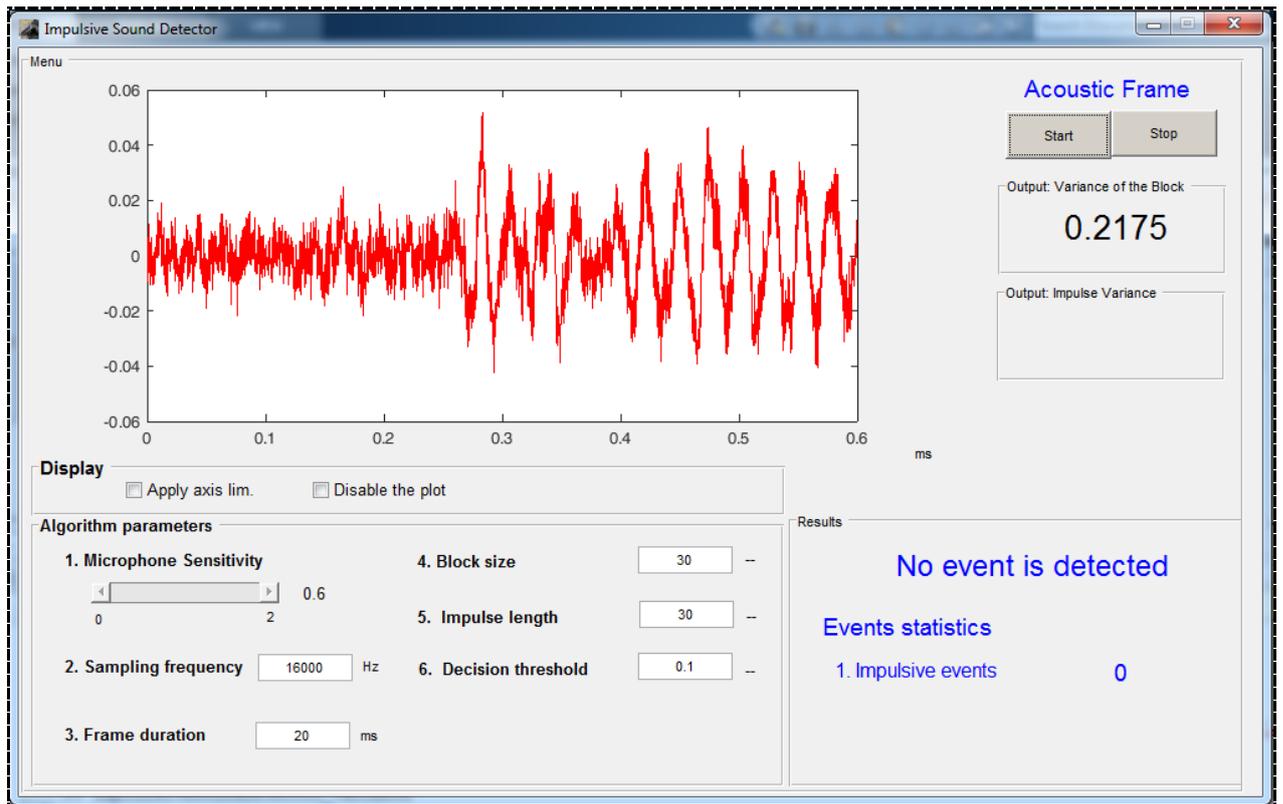


Figure 4.14 : Interface MATLAB pour la détection en temps réel des sons impulsifs

4.6 Conclusion

Dans ce chapitre, nous avons présenté les résultats expérimentaux de la détection des sons impulsifs en temps réel. Nous avons tout d'abord présenté les logiciels de développements nécessaires à la réussite de la solution proposée. Par la suite, nous avons présenté le corpus sonore utilisé dans nos expérimentations.

Les détails concernant : (i) le lieu de test, (ii) le matériel nécessaire, (iii) la connexion des softwares, (iv) l'enregistrement des sons de test, (v) le choix et l'optimisation des paramètres algorithmiques sont fournis dans une section à part. La dernière partie de ce chapitre est consacrée à la présentation de l'application développée.

D'après les résultats obtenus, nous pouvons conclure que la méthode développée a donné de bons résultats dans des environnements fermés (Indoor). Elle possède l'avantage d'être :

- (i) Peu complexe (soft et hard),
- (ii) Facilement adaptable aux conditions de l'environnement,

- (iii) Facile à implémenter en temps réel,
- (iv) Formalismes mathématiques non complexes.

Néanmoins, les principaux inconvénients de notre solution sont :

- (i) La configuration décrite ci-dessus ne peut pas détecter deux évènements impulsifs consécutifs si le décalage temporel entre ces évènements est inférieur à 0.6 secondes (la taille du bloc d'analyse L). La valeur de « L » doit être optimisée pour plus de précision,
- (ii) Notre système de surveillance audio est seulement capable de contrôler une seule zone fermée,
- (iii) La dimension de la zone surveillée dans nos expérimentations est très petite,
- (iv) La portée du capteur est un paramètre qui doit être étudié pour une solution complète. Une comparaison des performances de la méthode en exploitant d'autres types de microphones est souhaitée.

Chapitre 5 : Conclusions et travaux futures

5.1 Conclusions

L'objectif de cette étude est de mettre en œuvre un système de détection d'évènements impulsifs en temps réel. Il s'agit d'un module spécifique aux systèmes de surveillance de la troisième génération. La chaîne complète de ce genre de systèmes a été présentée dans le deuxième chapitre de ce mémoire.

Pour réaliser ce module, nous avons fait une étude détaillée sur les méthodes de détection des sons impulsifs ainsi que les critères de choix d'une solution adéquate. Le chapitre trois couvre les notions et les détails nécessaires à l'accomplissement de cette tâche.

Nous avons opté pour une méthode basée sur l'analyse de la variance des énergies à courts termes d'une forme d'onde acoustique. La mission principale de notre projet est de mettre en œuvre un logiciel de détection des sons impulsifs en temps réel et cela dans un environnement fermé (Indoor). Le logiciel doit fonctionner sur un Pc portable muni d'un système d'exploitation Windows. Un capteur MEMS (microphone d'un smartphone) est utilisé dans notre étude. L'optimisation des paramètres algorithmiques est faite en se basant sur une séquence audio de durée fixe. De plus, un module de réglage de la sensibilité du capteur est proposé.

Le quatrième chapitre a été consacré à la présentation des expérimentations. Plusieurs logiciels et outils ont été utilisés pour réaliser le logiciel de détection des évènements impulsifs. Les paramètres algorithmiques de notre méthode ont été optimisés en calculant les erreurs de prédiction. Nous nous sommes basés principalement sur les métriques suivantes : le taux de vrais positifs (TVP), le taux de faux positifs (TFP) et la courbe ROC.

Nous avons trouvé que notre solution donne de bons résultats dans un environnement fermé (Indoor).

Les avantages de notre solution sont donnés comme suit :

- Solution peu complexe (soft et hard),
- Facilement adaptable aux conditions de l'environnement,
- Facile à implémenter en temps réel,
- Formalisme mathématique non complexe,

Les inconvénients de notre solution sont donnés comme suit :

- La détection des évènements impulsifs consécutifs est seulement possible si le décalage temporel entre ces deux évènements est inférieur à 0.6 secondes.
- La surveillance audio n'est assurée qu'à l'intérieur d'une zone fermée.
- La dimension de la zone surveillée (salle) dans nos expérimentations est très petite.
- La portée du capteur est un paramètre qui doit être étudié pour une solution complète.

5.2 Travaux futurs

Les travaux futurs seront focalisés sur deux volets principaux :

- (i) L'amélioration des performances du système, à savoir :
 - L'optimisation de la taille des blocs « L » en vue d'obtenir de bons résultats en termes de détection et reconnaissances des sons impulsifs,
 - L'utilisation des réseaux de capteurs sans fils (Microphones) connectés (Wireless Sensor Network) pour la démonstration de l'efficacité de la solution proposée,
 - Le test de la solution proposée à l'intérieur des salles à grandes dimensions (en tenant compte des effets de réverbération),
- (ii) L'implémentation d'un module de reconnaissance de sons impulsifs : Le choix d'un formalisme mathématique peu complexe en vue de la détection et la reconnaissance des évènements impulsifs en temps-réel.

Bibliographie

- [1] M. Valera and S. A. Valastin, "Intelligent Distributed Surveillance Systems: a review," *IEE Proceedings, Vision, Image and Signal Processing*, vol. 152, no. 2, 2005.
- [2] Y. Benabbas, "Analyse du Comportement Humain à Partir de la Vidéo en Etudiant l'Orientation du Mouvement," Thèse de Doctorat, Université des Sciences et Technologies de Lille, Lille, France, 2012.
- [3] S. Lecomte, "Classification Partiellement Supervisée par SVM Application à la Détection d'Evènements en Surveillance Audio," Thèse de Doctorat, Université de Technologie de Troyes, France, 2013.
- [4] A. Guide, "Guide technique de la vidéo sur IP," 2006-2015. [Online]. Available: <https://www.axis.com/global/fr/learning/web-articles/technical-guide-to-network-video>. [Accessed 04 02 2017].
- [5] P. Roger and D. Hamernik and Keng, "Impulse Noise: Some Definitions, Physical Acoustics and Other Considerations," *The Journal of the Acoustical Society of America*, vol. 90, no. 1, 1991.
- [6] A. Michael and W. M. DRao, "Characterization of Sound Quality of Impulsive Sounds," in *Proceedings of 20th International Congress on Acoustics*, Sydney, 2010.
- [7] A. Dufaux, "Detection and Recognition of Impulsive Sound Signals," Phd Thesis, Neuchatel University, 2001.
- [8] R. Levorato, "GMM Classification of Environmental Sounds for Surveillance Applications," Master Thesis, University Of Padova, Italie, 2010.
- [9] A. Talal, U. Momin and A. Muhammad, "Improving Efficiency and Reliability of Gunshot Detection Systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'13)*, Vancouver, BC, Canada, 2013.
- [10] P. Chetan , J. Nyoupane, J. D. Shrestha et S. Jha, «Nepali Speech Recognition,» Technical report of the (flipkarma) <http://flipkarma.com>, 2008.
- [11] K. Lopatka, "Adaptive System for Recognition of Sounds Indicating Threats to Security of People and Property Employing Parallel Processing of Audio Data Streams," Phd thesis, GDANSK University of Technology, Poland, 2015.
- [12] A. Rabaoui, H. Kadri, Z. Lachiri and N. Ellouze, "One-Class SVMs Challenges in Audio Detection and Classification Applications," *EURASIP Journal on Advances in Signal Processing*, 2008.

- [13] "MATLAB," [Online]. Available: <https://fr.mathworks.com/>. [Accessed 2017].
- [14] "Goldwave," 2015. [Online]. Available: <https://www.goldwave.com/>. [Accessed 2017].
- [15] «Wo Mic,» [En ligne]. Available: <http://www.wirelessorange.com/womic/>. [Accès le 2017].
- [16] "True Software," [Online]. Available: <http://www.mypublicwifi.com/publicwifi/fr/index.html>. [Accessed 2017].
- [17] L. L. Beranek, Acoustics, Acoustical Society of America, 1993.
- [18] "Chipworks," [Online]. Available: <http://www.chipworks.com/about-chipworks/overview/blog/survey-of-mems-microphone-technology>. [Accessed 2017].
- [19] «Sounddogs,» [En ligne]. Available: <https://www.sounddogs.com/>. [Accès le 2017].